3-30-2004

The United States Marine Corps Data Collaboration Requirements: Retrieving and Integrating Data From Multiple Databases

Pamela J. Cole

THE UNITED STATES MARINE CORPS

DATA COLLABORATION REQUIREMENTS:

RETRIEVING AND INTEGRATING DATA FROM MULTIPLE DATABASES

THESIS

Pamela J. Cole, Gunnery Sergeant, USMC

AFIT/GIR/ENV/04M-04

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Marine Corps, United States Air Force, Department of Defense, or the United States Government.

AFIT/GIR/ENV/04M-04

THE UNITED STATES MARINE CORPS
DATA COLLABORATION REQUIREMENTS:
RETRIEVING AND INTEGRATING DATA FROM MULTIPLE DATABASES

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Information Resource Management

Pamela J. Cole, B.A.

Gunnery Sergeant, USMC

March 2004

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT/GIR/ENV/04M-04

THE UNITED STATES MARINE CORPS

DATA COLLABORATION REQUIREMENTS:

RETRIEVING AND INTEGRATING DATA FROM MULTIPLE DATABASES

Pamela J. Cole,
Gunnery Sergeant, USMC

Approved:

________//SIGNED//___________________ _30 Mar 04_____
Alan R. Heminger, PhD (Chairman) date

_______//SIGNED//____________________ _17 Mar 04_____
Lt Col Stephen M. Swartz, PhD (Member) date

______//SIGNED//_____________________ _17 Mar 04_____
Lt Col Stephan P. Brady, PhD (Member) date

Abstract

The goal of this research was to identify those strategies used in the private sector that may help the Marine Corps to better share information across its many different databases. This research is exploratory; it focuses on only one initiative: the IT-21 initiative. The IT-21 initiative dictates *The Technology for the United States Navy and Marine Corps, 2000-2035: Becoming a 21st Century Force*. The IT-21 initiative states that Navy and Marine Corps information infrastructure will be based largely on commercial systems and services, and the Department of the Navy must ensure that these systems are seamlessly integrated and that information transported over the infrastructure is protected and secure. The Delphi Technique was used to identify strategies, and to assess their value for helping organizations to share information better. Data was primarily collected from mid-level to senior information officers, with a focus on Chief Information Officers. The participants were able to identify measures used in the civilian sector to enhance information sharing strategies that helped them to successfully share information across different databases, in a secure, cost effective, efficient, and flexible manner. It is hoped that the Armed Forces and the Department of Defense will benefit from future development of the information sharing and database integration Holistic Model.

AFIT/GIR/ENV/04M-04

Dedication Page

In Memory Of
Major William D. Wood, Jr. (USAF)

Acknowledgements

Many people have contributed to the creation of thesis, and it is impossible to acknowledge them all explicitly. There were some, however, who directly participated in this endeavor, whom I would like to acknowledge.

I would like to thank Jesus for blessing me with the opportunity to excel, and for all the people that he provided to support and teach me a valuable lesson in "FAITH."

Specifically, I would like to make special thanks to my thesis advisor, Dr. Alan R. Heminger for his support and guidance. I would like to thank the other members of my committee, Dr. Stephen M. Swartz (Lt Col, USAF) for spending his valuable time to assist me with the formulation of my research and investigative questions; and Dr. Stephan P. Brady (Lt Col, USAF) for suggesting that I look at generalized information sharing.

In addition, I would like to thank the following people for their support, review and efforts in assisting me to complete this study: Major Timothy Webb (USAF), Professor Daniel E. Reynolds, Major William D. Wood, Jr. (USAF), Captain Scott Angerman (USAF), First Lieutenant Mary M. King (USAF), Mrs. Jeannette Gwin, Mr. Peter A. Dreher, Mr. William G. Queerer, Miss Teresa E. Cunningham, Mrs. Valerie I. McAlister, Mr. Dwight Wilson, Mr. Terence G. Sampson, and the AFIT library staff.

Additionally, special thank you to Colonel Rita Jordan (USAF), Lt Col Summer E. Bartczak (USAF), Captain David A. Weinstein (USMC), USAF Captains Grace M. Beck, Richard O. Day, and Tia A. Jordan, Mr. Bao Nguyen, and Larry B. Rainey, Ph.D.

Pamela J. Cole

Table of Contents

Page

Page

List of Figures

List of Tables

THE UNITED STATES MARINE CORPS

DATA COLLABORATION REQUIREMENTS:

RETRIEVING AND INTEGRATING DATA FROM MULTIPLE DATABASES

I. Introduction

Overview

This research explores the United States Marine Corps (USMC) efforts to enhance collaboration through retrieval and integration of information from multiple databases by differentiating the various approaches of information sharing as it relates to database integration.

This study looked at the commercial world for new systems developments. The *Technology for the United States Navy and Marine Corps, 2000-2035: Becoming a 21st Century Force* dictates that Navy and Marine Corps information infrastructure will be based largely on commercial systems and services. Moreover, that the Department of the Navy (DON) must ensure that these systems are seamlessly integrated and that the information transported over the infrastructure is protected and secure (NSB, 1997). Specifically,

> ...The Department of the Navy must establish an integrated organizational structure with the responsibility for planning, programming, and budgeting for all information systems not unique to individual platforms or weapons. ...Information superiority will be achieved only when a robust, seamless, and secure information infrastructure is established to support naval forces and provide them with the necessary information content in a timely and interpretable manner. The information infrastructure will be based largely on commercial systems and services, and the Department of the Navy

> must ensure that these systems are seamlessly integrated and that the information transported over the infrastructure is protected and secure.
> ...A mechanism must be found to coordinate all aspects of information superiority across both Navy and Marine Corps C^4ISR [Command, Control, Communications, Computing, Intelligence, Surveillance, and Reconnaissance] endeavors, giving due consideration to the evolving missions for naval forces and to current and future capabilities for ISR performed by other Services and agencies. [Where feasible, DON should] establish a clear policy designating responsibility in the [DON] for identifying, organizing, classifying, and assuring all relevant information sources that permit information extraction and communication from multiple remote locations. Invest in research and development tools and techniques to facilitate this shared information environment. Ensure timely and convenient access to all relevant information sources by naval assets...

Technology for the United States Navy and Marine Corps, 2000-2035: Becoming a 21st Century Force

If the information and resources our forces require are not readily available in a time of crisis, they cannot effectively accomplish their mission. This research explored different strategies that may provide improved operational secure information sharing, cost effectiveness, efficiency, and flexible application interfaces for heterogeneous databases. An extensive literature review and Delphi interviews were used to gather information pertaining to the different strategies.

With regards to database security and privacy, information sharing depends on authorized or proper disclosure of information to outside organizations or individuals. As Sushil (1996) states, information should be disclosed only when specifically authorized and solely for the limited use specified (Sushil, 1996).

Purpose of the study

Transformation restructuring and fiscal controls have focused attention on developing integrated database systems to foster improved operational secure information sharing within and outside the Marine Corps (NSB, 1997). According to the *C^4ISR Architecture Framework Version 2.0*, recent government legislation is placing more emphasis on the need to pursue interoperable, integrated, and cost-effective business practices and capabilities within each organization and across the Department of Defense (DoD). According to the National Research Council (NSB, 1997), the complexity of the DON massive business enterprise is growing. The DON business enterprise manages multiple large-scale processes including platform and weapons acquisitions; supply management; logistics management; resource planning; and personnel management and training. "Powerful new information technologies are becoming available that can be applied to these enterprise processes to significantly improve overall efficiency and effectiveness" (NSB, 1997). The DON is being forced to do more with less: funding constraints and a confluence of complexities are driving the DON to be or become more efficient (NSB, 1997). Every phase of the DON's acquisition process and operations is directed to acquire resources which are *faster, better, and cheaper*. The combination of increasing complexity and constrained resources will drive the development of new operational processes in order to achieve mission-required capabilities. Becoming a 21st Century force requires building technological capabilities necessary to meet Navy and Marine corps requirements. The requirement of the new technological capabilities demands extremely specialized analyses, comprehension, and "the integration of multiple disciplines into a set of enterprise processes that will extend across the entire spectrum of naval activities (NSB, 1997).

The Marine Corps' current information infrastructure operates under too many restrictions (Brady, 2003a). The information infrastructure is stovepiped. This system impedes information sharing (Brady, 2003b). The Marine Corps' information infrastructure is faced with a confluence of impediments: (1) lack of trained and skilled personnel, (2) increasing user demands, and (3) increased system sophistication (Brady, 2003b). Cross-functional platforms are needed to ensure transparent information flow (Brady, 2003b). Currently, needed similar critical and non-critical information are located in separate databases. For one reason or another, these databases operate in stovepiped environments. Independent database systems often cannot handle applications that cross organizational boundaries. Since databases exist on various platforms and applications, measures must be taken to ensure connectivity and secure interfaces. Thus, the USMC has an additional need to create a shared data environment across database systems so that such critical and non-critical information may be shared.

According to the National Research Council, the Cold War has left the United States with two parallel industrial infrastructures: defense and commerce (NSB, 1997). Each sector has its own business practices and distinct manufacturing technologies. Ultimately, however, success in an environment dominated by continuous and unpredictable changes requires the ability to quickly respond. This quick response drives the need for a unified, dual-purpose industrial base that can cater to both defense and commercial needs (NSB, 1997). Kutler (2003) stated that, the defense community no longer has the monopoly on technology development it once enjoyed. It will, over time, be forced to further marry commercial research & development into its unique knowledge of system integration, customer requirements and funding.

Accordingly, the purpose of this research was to explore those choices that hinder or facilitate sharing of information among multiple databases. Therefore, this research attempted to gather information from experts in the commerce sector. They identified options that may provide for improved secure information sharing, cost effectiveness, efficiency, and flexible application interfaces for heterogeneous databases.

The results of this research may assist the Marine Corps in its efforts to better share its information. Utilization of a successful data integration strategy will enable the USMC to outfit itself with the capability to retrieve and integrate data via a more secure and responsive database application platform, while forecasting asset requirements to support quick response in any crisis. This research will place emphasis on identifying successful private sector database integration strategies that could be appropriate for USMC use.

Statement of Problem

In order to accomplish a mission, a commander synthesizes information provided by staff and uses it as a significant decision-making tool. If information is not available or presented accurately and on time, the commander and personnel may be unable to make timely and accurate decisions.

Background

Over the past decade, many private companies became aware of the need to manage data as an asset. They realized that the requirement for flexibility is a significant component in managing data in order to compete in a very challenging world. With this in mind, organizations are faced with the challenge to ensure that organizational growth and survival are not at risk. As Dyck (2002) stated, ensuring that all parts of the business are working with correct as well as current information and keeping IT system maintenance costs down all depend on a consistent, universal, and integrated information management strategy.

In the past, database systems played a big part in the Armed Forces' mission accomplishment. For example, United States Code, Title 10, and joint doctrine, the Marine Corps, in coordination and cooperation with the Navy, have made logistical self-sufficiency an essential element of Marine Air-Ground Task Force (MAGTF) expeditionary warfighting capabilities. This means that the Marine Corps' logistics mission, at all command and support levels, is to generate MAGTFs that are rapidly deployable, self-reliant, self-sustaining, flexible, and capable of rapid reconstitution.

The Marine Corps logistics core capabilities are essential to the expeditionary character that distinguishes MAGTFs from other military organizations. Fundamental to all logistics operating systems are distribution systems that consist of functional resources and procedures. Logistics operating systems joined with command and control (C2) address all logistics functions at every level of war (MCWP 20:1-2). C2 of logistics enables a commander to recognize requirements and provide the required resources. C2 must provide visibility of both capabilities and requirements. This visibility allows a commander to make decisions regarding the effective allocation of scarce, high-demand resources. Accordingly, the C2 organization uses comprehensive data from a variety of sources, which are accessible by communications and information systems architecture, to provide a common logistics picture.

The effectiveness of C2 is based largely on the data retrieved from the various logistics database systems. There are three functional levels of logistics: strategic, operational, and tactical. The three levels of logistics functions interact and interconnect like sections of a pipeline, tying together logistics support (MCDP 20:1-3). Only when C2 effectively supports the logistics effort can logistics effectively and efficiently support the mission, manage distribution of capabilities, provide a shared real-time picture of the battlespace, anticipate requirements, allocate resources, and effect the timely distribution of resources (MCWP 20:1-3).

There are six functional areas in Marine Corps logistics: supply, maintenance, transportation, general engineering, health services, and services. Each logistics functional area is critical; all functions must be integrated into the overarching logistics support operation to ensure total support of MAGTF operations (MCWP 20: 4-1).

Today, as the Armed Forces right size, personnel are forced to do more with less, while operations in a joint service environment are becoming commonplace and demand change. Fiscal control and transformation restructuring have focused attention toward developing integrated database systems (NSB, 1997).

Private organizations and the government, as integral parts of the management support system, heavily use databases in their decision-making efforts. In the military, timely and accurate access to information can drastically change a commanders' course of action (Hamilton, 1993).

With this in mind, is there a commercial solution capable of meeting the Marine Corps' information sharing and database integration needs?

In order to answer the research problem, the following questions will be investigated:

(1) What are the characteristics of the private sector's incompatible database?

(2) What data sharing strategies were developed in private industries?

(3) What data sharing strategies were successful in private industries?

(4) Under what circumstances was database integration successful in private organizations?

Method of Inquiry

This research will explore strategies employed in the civilian organizations to see if they may be useful for implementation by USMC. A literature review of works from academic and industry points of view will be explored. Input will also be elicited from private sector's Chief Information Officers (CIOs) and other corporate Information Technology (IT) subject matter experts (SME). The Delphi Technique will be used to gather information relevant to the research questions.

Research Objectives

The objective of this research is to explore strategies, which offer integrated database systems, and enterprise information systems alternatives that will identify and provide options for improved operational secure information sharing, cost effectiveness, efficiency, and flexible application interfaces for USMC heterogeneous databases. Of interest are the design and attributes of established database integration strategies and enterprise information systems that will afford the United States Marine Corps flexibility during a crisis; reliability, security, efficiency, and responsiveness will also be explored.

Thesis Overview

Chapter I provides the introduction and background for the research, as well as the problem statement and investigative questions, which are keystones for the research. Chapter II will provide a review of current academic and practitioner relevant literature to identify what is already known about integrating different database systems and enterprise information systems. Chapter III will present the methodology selected to gather and analyze data for this research. Chapter IV will provide the results of the study and the analysis derived. Chapter V will discuss the results mentioned in Chapter IV, as well as the implications, lessons learned, limitations and suggestions for future research.

II. Literature Review

Introduction

The purpose of this literature review is to identify what is already known about integrating different database systems and enterprise information systems. Transformation restructuring and fiscal control have focused attention on developing integrated database systems to foster more improved operational secure information sharing within and outside the Marine Corps. Given the current state of today's stove-piped information infrastructure, the Marine Corps' lack of trained and skilled personnel, and ever-increasing user demands and increased system sophistication, cross-functional platforms are needed to ensure transparent information flow (Brady, 2003b). Currently, badly needed similar critical and non-critical information are located in separate databases. For one reason or another, these databases operate in stove-piped environments. Independent database systems often cannot handle applications that cross organizational boundaries. Since databases exist on various platforms and applications, measures must be taken to ensure connectivity and secure interfaces. The USMC has additional need for the sharing of information to create a shared data environment across database systems. According to Lt Col F. X. Brady, HQMC C4 CP CIO:

> Formerly, Legacy applications developed to satisfy specific business or operational objectives included procurement of independent and diverse hardware as well as accompanying databases and were hosted in variety of locations. As a result, we have a proliferation of application servers and databases through the Marine Corps that are excessively expensive to purchase, deploy, manage, and maintain.

Lieutenant Colonel F. X. Brady, USMC
Shared Data Environment, Chief Information Officer
MARADMIN Number 568/03 dtd 12/09/2003

The literature review is divided into two parts: (1) strategic information management as it relates to different information integration and sharing which offer database systems and enterprise information systems alternatives, and (2) discussion of attributes which contribute to database integration and information sharing strategies which will identify options that will provide for improved information sharing, cost effectiveness, efficiency, and flexible application interfaces for heterogeneous databases.

The strategic information management review is a chronological explanation of information from its initial inclusion in information systems, through the clarifications and modification brought about by various forms of database systems and enterprise information systems. The chronology illustrates the maturation of information from its initial lack of definition to its current definition, and from its lack of ownership to information stewardship. The second part of the literature review discusses obstacles to information sharing, cost effectiveness, efficiency, and flexible application interfaces for heterogeneous databases.

Data and Information

According to *The American Heritage® Dictionary of the English Language, Fourth Edition,* in computer science, data are numerical or other information represented in a form suitable for processing by computer. Data is defined as known facts or objects that have meaning in the user environment. The application of architecture transforms data into information (Evernden and Evernden, 2003). Data are observations of the environment while information is that which affects ongoing decisions (Maconachy, 2001). Along this same line of thought, in computer science, information is defined as processed, stored, or transmitted data. Information is found in one or more of three states: stored, processed, or transmitted (Maconachy, 2001). Information is seen as a resource; it is understood that it can be assessed, valued, and used (Myburgh, 1998). Information is data that has been replaced in a context or processed and presented in a form suitable for human interpretation. Information can be produced in digital or analog form. There are four forms of business information: voice, data, image, and video, as well as the implication of distributed requirements (Stallings, 2001). In a broader sense, from an architectural perspective, there are three main interrelated categories of information (Evernden and Evernden, 2003). Table 1, provides a description of each category.

Table 1. Three Main Interrelated Categories of Information

CATEGORIES	DESCRIPTIONS
Organizational or Management Information	This category is used to understand and make decisions about the organization itself. It is information that assists in the administration of the organization, in strategic planning and direction setting, and in managing personnel and their skills.
Business or Operational Information	This category describes the business that an organization manages. This information assists in meeting customer needs and in providing products and services.
Information about Supporting Technologies	This category is describes information about the technical infrastructure that supports business operations and management decisions, for instance, information about software applications and interfaces, communication networks, and system platforms. This helps develop and manage information systems

Information is the glue that binds everything together (Evernden and Evernden, 2003). To capitalize on the benefits of information as an asset, “an organization can either take a proactive approach by developing new skills, expertise, experience and capability in architecture, or it can choose to make do, in a random and ad hoc way, without these skills” (Evernden and Evernden, 2003). In view of that, now more than ever, there is a need to improve the use of the accrued data (Sokol, 2002). Information “needs to be managed through a strategically planned combination of organizational structures" (Langemo, 1988). Currently, badly needed similar critical and non-critical information is located in separate databases; thus, a process that can change different types of explicit and structured data into liable information is required (Sokol, 2002).

Strategic Information Management

All organizations need quality Information Resource Management (IRM) programs and systems (Langemo, 1988). IRM is used to assess a firm's strategic positions (Crook, 2003). IRM involves management of information as a corporate asset, (Kerr, 1991). Langemo (1988) states that information is a critical resource that:

> needs to be managed through a strategically planned combination of organizational structures, systems, technologies, work methods, and people all combined to receive, create, process, communicate, use, store, retrieve, and eliminate or archive information to do that work more effectively and efficiently.
>
> Dr. Mark Langemo, CRM
> Record Management Quarterly, 1988

IRM involves all functions and systems necessary to efficiently and effectively manage information through the entire life cycle of the information in an office and throughout an organization (Langemo, 1988). The convergence of computers, communication technologies, and demographics is transforming the way enterprises conduct themselves and carries out their organizational directive. Information is at the center of the transformation. An organization that ignores its information will fall by the wayside and are left behind in the global race for a competitive edge (Crook, 2003). In order to achieve success in the marketplace, organizations must rely on the right combinations of organizational resources working together in a dedicated effort to penetrate and achieve leadership. Information is the resource that organizations must use to achieve successful transformation (Myburgh, 1998).

In order to assess a firm's strategic positions, managers must gain insight internal and external to the organization. "Managers must collect and interpret data regarding the firm itself, its corporation, its stakeholders, and the industry" (Crook, 2003). To be more precise, managers must manage information through a confluence of "strategically planned" organizational structures, systems, technologies, work methods, and people skills all of which are combined to receive, create, process, communicate, use, store, retrieve, and eliminate or archive information to do that work more effectively and efficiently (Langemo, 1988). Manager's identification and use of information plays a large role in an organization's achievement of competitive advantage (Myburgh, 1998). The focus of information work has changed from the achievement of the effective and efficient managing of documents and technologies to the strategic use and application of information itself (Myburgh, 1998).

According to Myburgh, established writers on information management, describe "five revealing stages" in the history of information management:

Stage 1: Paper management

Stage 2: Management of corporate automated technologies

Stage 3: Management of corporate information resources

Stage 4: Business competitor analysis and intelligence

Stage 5: Strategic information management (SIM)

Of the five stages, the SIM paradigm is directly relevant to this study. SIM is on the rise. SIM focuses on corporate strategy and direction (Myburgh, 1998). SIM also focuses on the management of strategic information to achieve organizational objectives (Myburgh, 1998). SIM draws upon internal and external resources – explicit and tacit – that are recorded in documents and imbedded in people (Myburgh, 1998) by matching internal resources with external opportunities (Crook, 2003). SIM is a strategy that turns an organization's intellectual assets, recorded information, corporate memory, and employee expertise, into greater productivity, increased competitiveness, and increased collaborative efficiency and effectiveness (Myburgh, 1998). Managers can use the information to develop market and non-market strategies. Crook (2003) stated that market-based strategies seek to provide an advantage for the firm over its competitors by appealing to specific customer attributes. Non-market strategies take into account aspects of the environment not directly related to customers, including the actions of government, shareholders, and special interest groups.

Data Management (Lack of Data Ownership)

Government and private sector leaders are realizing that as organizations grow more dependent upon the availability of quality up-to-the-second information to survive and be competitive, the need for an all-embracing and effective IRM programs intensifies (Langemo, 1988). Organizational leaders are also realizing that to make changes, improve products and services, speed deliveries of products and services, reduce cost, improve profits, and be more productive necessitates managing information better than before (Langemo, 1988). Most businesses and corporations have 'evolved' to their present state without benefit of strategic planning or control (Langemo, 1988).

Case study of one organization attempted to see if the presence of problems with information management and behaviors related to ownership could be identified (Plant, 1996). The result of the study indicated that the presence of both problems exists. Plant (1996) states that Information Resource Management (IRM) has as its goal the management of information as a resource, but has not been implemented with the level of success expected. Problems with the implementation of IRM are indicated by the presence of redundant or inconsistent data, inability to share information across systems, and difficulty finding the information on systems (Plant, 1996).

The need for a unified infrastructure that will support a single query across all data sources would allow the integration of different collection stove pipes, including text, structured data, images, faxes, audio, and video (Sokol, 2002). Creation of a unified structure would transform the collection stove pipes into sets of derived data that are integrated with structured data (Sokol, 2002).

Data Integration

In *Data integration in a Bandwidth-Rich World*, it is stated that the key to deriving insight and knowledge is often the correlation of data from multiple sources (Foster and Grossman, 2003). Furthermore, in a world of more and more data, storage systems, computers, and networks, it is both necessary and feasible for system architects to think in terms of a new paradigm based on data integration - the flexible and managed federation, exploration, and processing of data from many sources (Foster and Grossman, 2003). Vital for effective data integration is the distributed system middleware that is beginning to allow distributed communities, or virtual organizations, to access and share data, networks, and other resources in a controlled and secure manner. Applications impossible in the past are achievable today over optical networks with the help of data services (Foster and Grossman, 2003). For example, virtual data warehouses allow users to extract the most important parts of data from disparate legacy applications, without the time, expense, and risk to data required by traditional data warehousing. The term "virtual enterprise" has been used in articulating the strategy for the 21st century global manufacturing enterprises. One of the key requirements is to develop an information system infrastructure to integrate and control the interoperability of the distributed, heterogeneous and concurrent systems in the participating organizations (Park and Favrel, 1999). The virtual enterprise concept focuses on three types of technology: data warehouse, process warehouse, and intranet/extranet (Park and Favrel, 1999).

For example, in terms of database integration and the supply chain, the virtual enterprise concept is extremely beneficial. Humby (1996) states that the information from a data warehouse is advantageous if the data in it is converted into information and then used to create knowledge for key decision makers. The front end of the data warehouse must be easy to use and intuitive for data exploration. If the data and its integrity are vital, the interface and the ease of use will determine how much it becomes central to the decision-making process of the organization. By using process warehousing, more technology can be introduced into the database integration and data-warehousing concept. The Process Warehousing is a growing information source that is accessible through a standard web browser on the corporate intranet (Nishiyama, 1999).

The degree to which information technology is used in organizations depends on the nature of the business and the operational needs in terms of fast and continuous access to information vital to the organization's continued success. Data replication can be used to improve the availability of data in a distributed database system (Her-Kun and Shvan-Ming, 2000).

In terms of data integration and availability of data in a distributed database system, the concept of using Java sensors to stream-based distributed processing of sensor data is made possible by utilizing Java virtual machines. This process "provides a virtual, homogeneous platform for distributed and parallel computing on a global scale" (Zhou and Zhou, 1995). The sensors are attached to applications to monitor some events about the application. Information gathered about the events is sent to the decision-making managers (Zhou and Zhou, 1995).

Distributed data sources can be diverse in their formats, schema, quality, access mechanisms, ownership, access policies, and capabilities (Foster and Grossman, 2003): For example, data discovery involves the utilization of a computer system that automates information retrieval from many data sources. In the article, *Warehousing Wherewithal*, Mattison (1996) illustrate the concept of *data discovery and access* in that "a private data warehouse is made up of 3 very different functional areas, each of which must be customized to meet the needs of a business. One component handles acquisition of data from legacy systems and outside sources. Another component of the warehouse is the storage area, which is managed by relational databases;" including specialized hardware, or software. An access area is the third component of the warehouse (Mattison, 1996). "Three different end-user PCs and workstations are used to draw data from the warehouse with the help of multidimensional analysis products, neural networks, data discovery tools, or analysis tools" (Mattison, 1996). In this case, the widest range of unique products can be found in the area of user access, sometimes referred to as data mining (Mattison, 1996).

During *data discovery and access*, data exploration and analysis are performed. The primary attributes required from these tools are: intuitive interface, transparent access to data, support for a catalog of information, multiple query and analysis methods and multiple presentation styles for information.

Information is the resource that organizations must use to achieve successful transformation (Myburgh, 1998). Organization's practice of data discovery and access as well as data exploration and analysis requires managing information resources as key enterprise assets in addition to the development of security and policy practices(Foster and Grossman, 2003). Organizations must start realizing that data is only half of the systems integration picture (Ambler, 1998).

Secure Information Sharing

Increasingly, government and corporate databases have become vulnerable to threats from internal and external sources. The efficacy of secure information sharing depends on a secure computing platform. A secure computing platform is designed so that individuals and devices who should not be able to perform certain actions cannot while those who should be able to perform certain actions can. The actions in question can be reduced to operations of access, modification and deletion (Wikipedia, 2002).

Secure information sharing is a primary concern to everyone. Trusted Computer System Evaluation Criteria (TCSEC) commonly called the 'Orange Book', is a United States government standard for computer security. Elsewhere, Canada used their own Canadian Trusted Computer Product Evaluation Criteria (CTCPEC) and Europe and several other parts of the world used the competing Information Technology Security Evaluation Criteria (ITSEC) standard (Wikipedia, 2002).

TCSEC was issued by the United States Government National Computer Security Council (an arm of the U.S. National Security Agency) as 'Trusted Computer System Evaluation Criteria, DOD standard 5200.28-STD, December 1985' (Wikipedia, 2002).. The National Computer Security Center (NCSC), through its Trusted Product Evaluation Program, analyzes the security features of commercially produced and supported computer systems (Gallagher, Jr., 1991).

TCSEC defines criteria for trusted computer products. There are four levels, A, B, C, and D. Each level adds more features and requirements as shown in Tables 2 and 3 below:

Table 2. TCSEC Level C and D Criteria for Commercial Computer

CRITERIA	DESCRIPTIONS
D	For non-secure system.
C1	Requires user log-on, but allows group ID
C2	Requires individual log-on with password and an audit mechanism. (Most Unix implementations are roughly C1, and can be upgraded to about C2 without excessive pain).

Levels B and A necessitate mandatory control. Access is based on standard DoD clearances:

Table 3. TCSEC Level A and B Criteria for Commercial Computers

CRITERIA	DESCRIPTIONS
B1	Requires DoD clearance levels.
B2	Guarantees the path between the user and the security system and provides assurances that the system can be tested and clearances cannot be downgraded.
B3	Requires that the system is characterised by a mathematical model that must be viable.
A1	Requires a system characterized by a mathematical model that can be proven.

These standards have now been superseded by the Common Criteria (CC). The CC are an international standards (ISO 15408) for computer security. Their purpose is to allow users to specify their security requirements, allow developers to specify the security attributes of their products, and allow evaluators to determine if products actually meet their claims.

As part of the NCSC Technical Guidelines Program, the Trusted Database Management System Interpretation (TDI) was also issued. The TDI extends the evaluation classes of the Trusted Computer System Evaluation Criteria to trusted applications in general and database management systems in particular. The TDI serves as an adjunct to TCSEC by providing a technical context for the consolidation of entire systems constructed of parts and by presenting database-specific interpretation of topics that require direct comment (NCSC, 1991). Thus the TDI is relevant to applications which support sharing of computer services and devices and which enforce access control policies. More specifically, it provides insight into the design, implementation, evaluation, and accreditation of database management systems" (Gallagher, 1991).

Databases

A database is a collection of organized and structured data (Ranade, 2002). Organized and related information is stored in a database. A database is a model of the real world. Databases are used to store, manipulate, and retrieve data in every type of organization. The major purpose of a database is not to store information but to retrieve it. The term organized means that the data are structured. Structured data can be easily stored, manipulated, and retrieved by users. The term related means that the data describe a domain of interest to a group of users, and that those users can use the data to answer questions concerning the domain (Hoffler, 2002). All over the world, databases are incorporated into companies' computing networks (Harris, 2002). "Databases are an important class of applications, and because of their complexity, effort is required to tune them for best performance" (Ranade, 2002). "Databases not only permit easy retrieval of data, they also provide for the creation of new knowledge from the same data" (Thede, 1991). Over the past two decades, the notion of database functionality has remained unchanged; however, the way in which they are used has changed significantly (Harris, 2002). The Internet explosion and its multiple technologies can be credited for changes that allow users so many different ways to interact with databases (Harris, 2002). Company databases are accessed from various sources: customers, vendors, and employees through Web sites, different applications, extranets, and Virtual Private Networks (Harris, 2002). Interactions with the underlying storage layers can help or hinder a database (Ranade, 2002). A database is a self-describing collection of integrated records. It is self-describing because it contains a description of itself in a data dictionary. The information about the meaning of the data stored in a database can be stored in a data dictionary. The data dictionary contains all relevant information about data items in a database. The computer must know where data is stored in a database. In order to access and provide accurate and correct data to the requestor, when a user or computer application requests data, the computer must know what data are stored in the database, how they are organized, and how to access them from the database (Narayan, 1988). "These rules may be contained in a file layout, and 'include' file, a copybook, a subschema, a relational table, or a data dictionary" (Narayan, 1988). Schema is the term used to describe the complete database logical design. Sub-schemas are subsets of the schema. Each user has a subschema. The subschema can be thought of as the user's view of the database. The data dictionary is a table that defines all of the names that are used in a system model. The data dictionary contains information from the names in all of the documents describing a software product. Thus, the data dictionary can be analyzed, checked for inconsistencies and edited. The data dictionary is also known as *metadata*.

Metadata

The term *metadata* has been used often since the early 1980s to describe the properties or characteristics of other data contained within databases. Metadata is descriptions of databases: data that describe other data. The data are organized in the form of entities, attributes, and relationships, and are generally stored in a data dictionary (Narayan, 1988).

In order to facilitate the storage and retrieval of information, a data dictionary is divided into entities. An entity can be a subject, or an object, depending upon its activity (Harris, 2002). "An entity is a person place or thing, or event about which data may be recorded" (Narayan, 1988). "For example, a program can be an object when a user requests information from it or requests it to process data. The program can also be a subject if it initiates communication with another program" (Harris, 2002). Entities represent a class of similar things. The differences among the instances of a class are denoted by the values of the entity's attributes. Entities never exist inside the computer; they belong to the problem being modeled. Entities have descriptors. These descriptors are the attributes of the entity.

Attributes are the type of data that is stored about an entity. Attributes can be used to "contain a pointer or an address to another entity in the data dictionary, a file in the dictionary structure, or another physical object" (Narayan, 1988). The values of the attributes are machine-readable. This necessary to facilitate those applications "that have to extract the pointer or address of that entity from the dictionary and then have to go to the address in a mechanical manner to operate on whatever exists at that address" (Narayan, 1988).

The data dictionary system is driven by parameters that are stored in the control region. These parameters consist of the entities and relationships between them. The database administrator establishes valid attributes classes, access keys, and entity-to attribute combinations. The user, however, assigns the actual characteristics to the attributes. "Many archivists and records managers are now adopting metadata concepts to describe electronically recorded information and then elaborating on these issues to incorporate a more robust set of descriptive data elements that suits their needs" (Phillips, 1995).

To that end, a database is a collection of integrated records because the relationships among the records are stored in the database. To be more specific, a database is a collection of interrelated data stored in a meaningful way to allow multiple users and applications to access, view, retrieve, use, and store information.

Database Models.

The term *database model* is used to describe two related but different ideas (Kay, 2003). One aspect describes the function of the system, and the other aspect describes its behavior: a conceptual model and a physical model. The fundamental structure of a database is a data model (Silberschatz et al, 1996). The structure of a database should reflect what it is designed to model. A data model is described as a collection of conceptual tools for describing the real-world entities to be modeled in the database and the relationships among these entities (Silberschatz et al, 1996). "Conceptual models are the means by which complex software-intensive systems, as well as systems in the more general sense, are conceived, architected, designed, and built" (Dori, 2003). Conceptual models are not system specific. Physical data models are used to describe data at the lowest level. The amount of semantic detailed representation and the primitives available to describe the data influence the difference in physical data model (Silberschatz et al, 1996). Various data models exist; however, they fall into three different categories: object-based logical models, record-based logical models, and physical data models.

Flat Files

The features of a flat file structure database are simple; however, it is not convenient for 21st century business applications. "A flat database can access only one source or table at a time" (Thede, 1991). Flat databases consist of a single file and can be though of as rows and columns. A row is a record, and a column is a field. The physical structures of flat files are sequential and direct access. Flat file feature is similar to spreadsheets: features such as sorting, counting and aggregating are included in many spreadsheets.

Hierarchical Database Model

The hierarchical database model is made up of a collection of tree structures, to form a directory structure. Each tree consists of records. Within each record, there are two modules or fields: a root and a subordinate field. The root is the master key that identifies the field responsible for the ordering of the records. Data is represented as a series of parent-child relationships. The constraint for a one-to-many relationship between parent and child can result in redundant data. To avoid data redundancy, data is stored in one place and referenced by links or physical pointers in other places. However, the links are hard-coded into the data structure. This has a negative impact on the hierarchical model; it is inflexible. Users are restricted to work on data in only one way. Furthermore, due to the physical links, hierarchical database modifications tend to require major rewriting. For the most part, with the exception of the Data Language I (DL/I), the hierarchical database model is no longer a practical commercial option.

Network Database Model

The network database model is almost extinct. This database model is an expansion on the hierarchical model. It has more than one parent –child relationship. It provides multiple paths among segments; nevertheless, it is not very practical. By connecting individual records, it only supports a plain network relationship. Since there are no restriction no restrictions on how many relations the network database can contain, it can get complicated.

Relational Database Model

The relational model provides the same flexibility offered by the network model but is much easier to work with. In a relational database, the logical design is independent of the physical design.

The features of relational databases are complex. Relational databases consist of multiple flat files or tables, with relationships among files. New tables are created from joining, filtering or splitting existing tables.

“The true power of a relational database resides in its ability to break the link between data access and the underlying data itself” (MSND). The relational database provides flexibility that allows changes to the database structure. Because the data reside in tables, the structure of the database can be modified without having to change any applications that were based on that structure. Users can access all of their organization’s data dynamically without any knowledge of how the underlying data is actually stored by using a high-level access language such as SQL (structured query language). Query optimizer is used to input queries and convert them to a format that efficiently accesses the stored data. This is done to maintain both system performance and throughput, so that the relational database system can accept a variety of user queries and convert them to a format that efficiently accesses the stored data (MSND). Unlike the hierarchal database model, in terms of retrieval, the structure of the relational database model allows flexibility.

Object-Oriented Database Model

The object-oriented database model is very sophisticated. It holds actions as well as data. It is an improvement on the relational database model. The object-oriented database model is different from the traditional database models in that it deals with Binary Large Objects (BLOBs). BLOBs are complicated data types such as images, documents, e-male messages, dictionary structure, and Computer Aided Designs. BLOBs are not easily represented in relational database. Relational database use pointers to reference BLOBs and storage of BLOBs are outside the database. Unlike relational database model, the object-oriented approach provides a natural way to represent the hierarchies that occur in complex data. Object-oriented database treats everything as an object that can be manipulated.

Database Management System (DBMS)

A DBMS allows a user to store data and retrieve information. Using a DBMS is more efficient that file processing. A DBMS is more advantageous in that data is located in one location. There are many advantages to having data in one location: reduced data redundancy; improved security; improved data integrity, and improved consistency. Data is entered into a database from existing files from other databases; input is made from keyboard or as output from other applications. DBMS allows skillful data manipulation and quick and easy development of custom applications. A user can use the DBMS to query the database for specific information and can create output in the form of reports. Based on queries created by a user, the data is processed into information by the DBMS. Using a DBMS, a user can perform many functions on data; such as record; store; retrieve; view; select; modify; sort; merge/join; compute; and display.

Semantic and Information Retrieval

During the design of the physical model, knowledge of semantic is crucial. Physical database design is made from the perspective of the programmer or the computer. Physical database is concerned with things such as media type and file type. Semantic refers to the science of meanings of words. Frequently users of information retrieval systems and document authors use different terms to refer to the same concept (Fabio, 2003). Concepts are abstractions and they are defined in terms of properties, individuals, and instances of properties (Flater, 2003). "For any given abstraction, it is possible to construct an integration scenario in which a failure will occur because of some property that was not explicitly modeled" (Flater, 2003). When different terms or properties are used to reference the same concept, the conceptual integrity of the system is said to be compromised.

Bad conceptual design can affect and compromise the conceptual integrity of a system, thus resulting in 'semantic faults, which are commonly blamed for hidden integration bugs (Flater, 2003). A 'semantic faults' is a violation of conceptual integrity (Flater, 2003). Conceptual integrity in system architecture "allows the system to become a cohesive and sensible whole" (Flater, 2003). Accordingly, if the conceptual integrity of a system is compromised, information retrieval is inconsistent. Information retrieval is shaped by the term mismatch problem: incompatible terms (Fabio, 2003). Retrieving and integrating information from various sources is a serious problem (Arens et al, 1993). The incompatibility of terms "does not only have the effect of hindering the retrieval of relevant documents, it also produces bad rankings of relevant documents (Fabio, 2003). One factor that contributes to incompatible terms is database design.

Customarily, "database design activities are partitioned into distinct phases in which a logical design phase precedes physical database design" (Ling, 1996). The logical design phase illustrates the compulsory business functions and products of the system without any sign of the technology used to achieve it. Using the concept of data dependencies, the purpose of the logical design step is to get rid of redundancies and updating anomalies, "while leaving the physical design step to consider how the database schema may be restructured to provide more efficient access" (Ling, 1996). The separation of the logical and the physical design steps repeatedly results in the physical database design not being able to benefit from knowledge of the semantics of data captured in the earlier phases of the database design life cycle (Ling, 1996). One way to resolve this situation is to introduce strong and weak functional dependencies. They can be "extended to capture data semantics relevant to the design of the database schema; which are more desirable from the efficiency point of view" (Ling, 1996).

In one sense, a data model "is somewhat abstract in nature and refers to a database's overall structure, or type. The best-known example is the relational model" (Kay, 2003). The other description includes flat-file, hierarchical, network, and object, semantic and dimensional models (Kay, 2003). The second description of "data model, or schema, takes the overall structure of one of the standard database models and tailors it to a specific application, company, project or task (Kay, 2003). Mostly, the schema type description of data model illustrates the overall structure and details "to specific data items, including their names, values, and granularity and how they relate to one another (Kay, 2003).

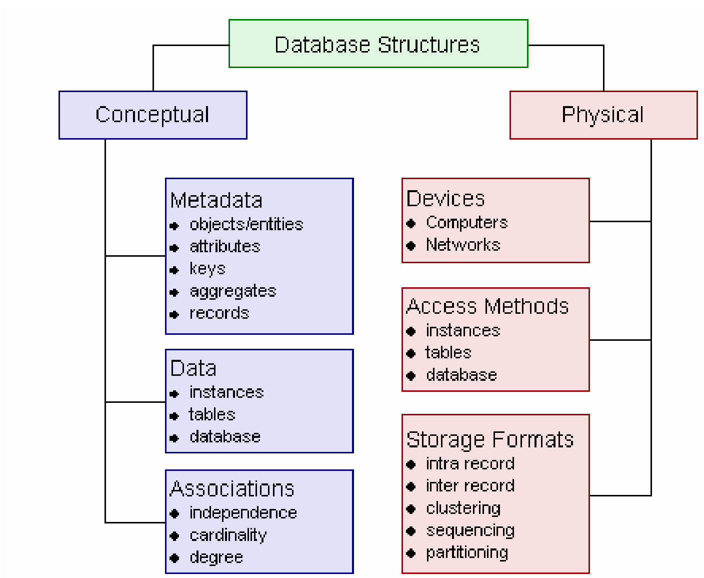

Figure 1. A Database Structure Model (Source: http://famed.ufrgs.br/pdf/csih/mod3/Mod_3_2.htm)

Integrity of Stored Information

Retrieval inconsistencies can occur if the integrity of stored information is ignored. Factors such as logical and physical integrity may contribute to the integrity of stored information.

Logical database design is constructed and displayed from the perspective of the user. There is a logical connection among the fields of each record. Logical integrity pertains to conceptual integrity, as well the requirements of applications integrity. "Conceptual integrity is required for the result of data integration to be cohesive and sensible" (Flater, 2003). Applications integrity requires that combination of related database changes work correctly. This type of application integrity is called logical integrity. Logical integrity of data involves making sure that the content in a physical database or file matches the organization's logical definition for those data elements. For example, "Today, the risks incurred by users of relational database management systems are largely with respect to the local integrity of their databases. The earlier the logical integrity of data is enforced and maintained the less traumatic it will be to introduce support for this concept (Codd, E. F., 1990).

Physical integrity involves many things. Primarily database physical integrity is concerned with safeguarding and managing the physical systems that maintain, store, and deliver data. With this in mind, physical integrity involves a confluence of measures:

> (1) Physical integrity is concerned with the preservation of stored data "from physical degradation of the storage media and technological changes to data formats and storage methods" (Hunt, 1999). Physical integrity is concerned with the selection and refresh of storage options that will assure the physical integrity of information; as well as to the longevity of useful information in terms of preservation and restoration (Heminger and Robertson, 1998).

(2) Database physical integrity is also concerned with the threats to internal and physical integrity hardware failures caused by malfunctions or external events such as fire or loss of electric power. There are two solutions designed to assist in the database internal and physical integrity recovery process: write image journaling, and establishing a constant journaling and backup strategy. The write image journaling is a two-phase write protocol safeguard structural database integrity solution. However, if hardware failures occur before a complete write of an update to the write image journal, the recovery process does not prevent data loss. To recover databases from a loss of structural integrity, one must restore the backup and then apply the ensuing database changes from the journal. Each organization must choose a database backup and recovery strategy. Database backup strategy may entail backing up the database or backing up the database and the transaction log. The database backup strategy selected will determine the extent to which committed transactions are recovered. The simplest backup strategy is backing up the database only. However, selecting this process will not save all committed transactions that occurred after the last backup. To ensure recovery at the point of failure, performing database and transaction log recovery is a better option. Transaction log backups ensure that the information required redoing changes made after a database backup was performed. However, extra precaution, such as placing the transaction log files on a mirrored disk, must be taken to protect the active transaction log. This will ensure that during backup, only uncommitted transactions will be lost.

(3) Physical integrity also affects the organization's disaster recovery plan. Physical integrity is also contingent on who can reach the system and how easily they can do so (Beaty, 1988). Along with general security-oriented procedures, an organization may want to consider badges, exterior lighting, and physical barriers. "Further, unauthorized use of software should be controlled, and redundancy needs to be built into system design. Complete backup does not have to be provided on-site. Planning should cover mainframes as well as personal computers" (Beaty, 1988).

The data stored in a database management system (DBMS) is often crucial to the organization's survival and is regarded as an asset. Organizations must consider ways to ensure privacy and control access to data that must not be revealed to certain groups of users.

Access Control and Data Security

Access Control and Data Security are directly related to the integrity of stored information. There are distinct approaches to database access control mechanism to support a security policy: views and authentication (Groff and Weinberg, 2002). Views are a valuable tool in enforcing security policies, in that they can be used to create a window on a collection of data that is appropriate for some group of users. Views allow limited access to sensitive data by providing access to restricted version of that data, rather than the data itself. In addition, before access is granted, it is crucial to authenticate a user to the database system.

Access control measures, as well as data security precautions must be in place to determine how and who will have access to stored information. Access control measures use tools such as security policies, access rights, passwords, authorization certificates, and firewalls to control access and enforce data security. To be more specific, these tools are used to ensure that only authorized individuals and resources can access information, database, computer, networks and other devices. There are two distinct DBMS approaches to access control: discretionary and mandatory access control.

Discretionary access control is an access policy, not a security principle. It allows users to make access decisions about their files (Anderson, 2001). The notion of an authorization or access control is a discretionary access control model. A two-dimensional matrix with columns corresponding to the data items and the rows corresponding to the users may be used to specify access rights. Each element in the matrix specifies the access rights a user has to a data item. Access rights represent operations performed by users on data items and may include retrieve, insert, delete and update. The data item may be a relation, a tuple or an attribute. Data items may also be defined using views.

Discretionary access control gives users access rights, or privileges, and mechanisms for privileges. A privilege allows a user to access some data object privileges to access specific data files, records, or fields in a specific way, such as read, insert, delete, or update. A user who creates a database object such as a table or a view automatically gets all applicable privileges on that object. The owner of a relation can create a view then grant the view to other users. While very effective, discretionary access control is also considered vulnerable to viruses and malicious attacks, it offers no protection against malicious or buggy codes. However, mandatory access control offers restrictive access; it is based on system-wide policies that cannot be changed by individual users.

Mandatory access control is an access policy, not a security principle. It is built to enforce a security policy independently of user actions (Anderson, 2001). Mandatory access control classifies users and data into multiple levels of security, and then imposes appropriate rules. Each database object is assigned a security class; each user, subject, or program is assigned clearance for a security class. Security classes are organized according to a partial order, with a most secure class and a least secure class. Security classes could be Top Secret (TS), Secret (S), Confidential (C), and Unclassified (U).

In this system, TS > S > C > U, where A > B means that class A data is more sensitive than class B data. Rules are imposed on reading and writing of database objects (e.g., tables, views, rows, and columns), subjects (e.g., users, programs), security classes, and clearances.

Granularity of Context.

Access control and data security primary concern is the context of multilevel security of relations and polyinstantiation of information flow control. For example, in order to apply mandatory access control policies in a relational DBMS, a security class must be assigned to each database object. The objects can be at the granularity of tables, rows, or even individual column values. This granularity of context leads to the concept of a multilevel table, which is a table that users with different security clearances see at different collection of rows when they access the same table.

Threat Model-Security Policy-Security Mechanism.

When a top-down approach to security implementation is possible, it takes the form of *threat model-security policy-security mechanism* (Anderson, 2001). In the form of a security policy, factors such as logical and physical integrity measures contribute to access control and data security measures that will influence the design of the security protection mechanism. "The best-known example of a security policy model was proposed by David Bell and Len LaPadula in 1973, in response to U. S. Air Force concerns over the security of time-sharing mainframe systems" (Anderson, 2001).

The context of classification of military and intelligence information flow control influenced the construction of the Bell LaPadula (BLP) model of computer security. The BLP is also known as *multilevel security;* "systems that implement it are often called *multilevel secure*, or MLS, systems" (Anderson, 2001). The basic principle is that information can flow downward.

The BLP model was formulated to protect against vulnerabilities such as malicious and buggy codes. Its purpose is to enforce *mandatory access control*, which entails a security policy independent of user actions. The BLP model enforces two axioms (Anderson, 2001):

(1) The *simple security property:* no process may read data at a higher level. This is also known as *no read up* (NRU).

(2) The (star)*-*property*: no process may write data to a lower level. This is also known as *no write down* (NWD).

Reliability of Data.

The goals of database security are availability, secrecy (or confidentiality), and integrity of stored information. Integrity pertains to only authorized users should be allowed to modify data. Secrecy is the protection of data from unauthorized disclosure. Availability is the assurance that, when needed, information and services can be accessed reliably. If security fails then the database will not be available. Physical integrity is directly related to availability of information and services; in that information and IT resources must be physically protected from malicious attacks and disasters. Malicious attacks and disaster are various; however, the most common threats are viruses, power outages, system failures, and overloads. Physical integrity ensures that necessary precautions, such as database backups, transaction log backup, anti-virus, firewalls and other access control measures are in performed and in place to assure availability.

Availability of information and services is not the only focus of access control and data security. In order to assure the integrity of stored information, several other access control and data security defense-in-depth measures are accessible: authenticity, confidentiality, integrity, and non-repudiation.

These defense-in-depth, Information Systems Security (INFOSEC) measures are now classified as the five attributes of information assurance (Maconachy et al, 2001).

Information assurance (IA) "includes the products, procedures, and policies that allow the timely transfer of information in an accurate and secure way among all parties involved" (McKnight, 2002). Central to IA is the provisioning of the five security services (Maconachy et al, 2001): availability, integrity, authentication, confidentiality, and non-Repudiation. Accordingly, McKnight (2002) describes the five attributes as follows: (1) Availability means access; it is the "state where information is in the place needed by the user, at the time the user needs it, and in the form needed by the user." The issues that most directly affect availability are information system reliability (is it up and running?), the information level of importance (some information is more critical than others are), and timely information delivery (delay of some information has a greater impact than other information). (2) Integrity is "sound, unimpaired, or perfect condition." Here we are looking at system integrity instead of data integrity (although both can be considered). (3) Confidentiality is "the concept of holding sensitive data in confidence, limited to an appropriate set of individuals or organizations". Confidentiality is often referred to as information security. Here we deal with two issues clearances and data security. Access to data is based on two criteria: a security clearance and a need to know. Building private networks, encrypting the data that travel across unprotected sections of the network, providing protective distribution systems, or building secure enclosures where data can be processed, can provide data security. (4) Non-Repudiation is "a service that provides proof of integrity and origin of data, both in an unforgeable relationship, which can be verified by any third party at any time; or, an authentication that with high assurance can be asserted to be genuine, and that cannot subsequently be refuted." (5) Authentication is defined by the National Computer Security Center as: "to verify the identity of the user, device, or other entity in a computer system, often as a prerequisite to allowing access to resources in a system, and 2) to verify the integrity of data that have been stored, transmitted, or otherwise exposed to possible unauthorized modification."

Authentication ensures that you have the right to see the information, and that you are who you say you are. Not only do people need to be authenticated, so do devices. They must go through a network protocol authentication process validation. Authenticity is a major concern; especially with increase usage of Internet databases. Identity theft is rampant. To prevent against fraud, identity verification is necessary to make sure that individuals are able to prove that they are who they should be. Authenticity access control measures include the use of logins and passwords, biometrics, digital certificates, and digital signatures or shared secrets.

Business Rules in Corporate Database

Chen and et al. (1992) state that as a means to guarantee the integrity of information stored in a shared database environment, business rules in an organization are set up by top managements to reflect their knowledge and wishes about how the organization should operate. Business rules provide a direct and indirect framework of regulations and constraints on the organization's database. These rules are vital to the organization's survival; these rules can be used to communicate to other levels of management the organization's data policies. Policy is a set of guiding principles designed to influence or determine future decisions and actions. Business rules are used to set the organization's expectations. The organization's expectations in the form of business rules are also available to employees as guidelines to assess their own functions (Chen, 1992). In general, business rules are described by natural language sentences and it is difficult to incorporate the semantic constraints of the business rules into the corporate model. In order to clarify the interrelationships among data entities and business rules, entity-relationship diagrams, production rules, and a reference able for attributes are often used to provide a pictorial representation of the organization's business rules (Chen, 1992).

Incompatible Data Format.

Many impediments that cause incompatible data format, thus preventing disparate databases from sharing information, necessitate the use of policies to correct information sharing hindrances. Enforcing business rules would be the most effective means to fix some of these incompatible data formats which are caused by impediments such as lack of hierarchical structure; insufficient self description; lack of pointers or referencing mechanism; restricted number of records; difficulty in extensibility or interoperability; restricted record length; restricted number of characters for variable names; and inability to directly query multiple files.

Database Interface Language.

Database Interface language is necessary for data integration. "Data integration can be difficult, expensive and error-prone" (Anthes, 2002). Great care must be taken to build interfaces between applications and databases that ensure accuracy and timeliness of information and that answer the needs between applications and databases that ensure accuracy and timeliness of information and that answer the needs of disparate communities of end users (Anthes, 2002).

Concurrency Issues.

The goal of the database system is to simplify and facilitate access to data. Performance is important. Views provide simplification. Concurrency issues occur when one operator's access interferes with another human operator's access (Papazoglou, and Valder, 1989). Concurrency controls in the form of transactions are implemented to enable several operators to access a database simultaneously. "In order to be able to organize the control of shared access to the database the DBMS must provide appropriate mechanisms to identify and define sequences of actions, including reads and writes to the database, that are logically linked to one another" (Papazoglou and Valder, 1989).

Middleware.

Middleware is used to enable collaboration. It provides standard interface communication services and acts as an integration broker by focusing on creating universal interfaces. "A middleware service is a general-purpose service that sits between platforms and applications" (Bernstein, 1996). The Application Program Interface (API) and protocols it supports define application middleware service. In addition, it meets the needs of a wide variety of applications across many industries, and it must have implementations that run on multiple platforms.

Ontology.

Ontology is defined as a taxonomized set of terms, ranging from very general terms at the top down to very specialized ones at the bottom (Hovy, 2003). Ontology systems are designed to decompose data requests into database queries according to the content and nature of the data sources, retrieves data from data them, and reassembles the results appropriately (Hovy, 2003). A query in the context of information retrieval is a request for information from the database (Papazoglou and Valder, 1989). "The data requested is stated in a query language which conveys to the database management system the request of the user in an understandable form" (Papazoglou and Valder, 1989). A query is stated in a query language and is processed by a query processor (Papazoglou and Valder, 1989).

SEMEDA

Kohler states that SEMEDA (Semantic Meta Database) is an application for semantic database integration. One of its key components is its ontology editor. Like many other tools, SEMEDA uses a relational database to store ontologies. Most ontology editors load the data-structure of the ontology to memory, where it is edited and finally written back to the database (Kohler et al). "However, this approach has some disadvantages: i) When several users edit ontologies at the same time, conflicts between different versions have to be resolved when the ontology is written back to the database. ii) Users cannot immediately see changes that other users apply to the ontology. iii) The size of the ontology is limited by the size of the data-structure that the application, which loads the ontology from the database to memory, can handle." (Kohler et al). "SEMEDA avoids these disadvantages by editing the ontologies directly in the database" (Kohler et al).

Integrating Databases.

Organizations gain benefits from higher levels of integration (Herman, 2002). Nevertheless, due to the following impediments, few companies have gotten to the high levels of integration (Herman, 2002):

(1) Overcoming distrust in sharing proprietary information

(2) Overcoming functional stovepipes within one's own company

(3) Gaining executive sponsorship

(3) Creating business and technology architectures that lay out the process and technology design the organization wish to receive.

Business Integration

"Business change is occurring at more than seven times the rate of IT change, forcing organizations to realize they cannot code their way out of the backlog. Instead, organizations must consider alternative application development approaches that focus on reuse of existing assets" (Addington, 2004). Business integration is the technology that simplifies the integration of internal systems, makes it easier to connect with internal and external customers, and removes cost. "Automating and optimizing business processes is a priority that touches all areas of the enterprise" (Addington, 2004). According to CIOs, integration is the biggest single problem they face. However, integration covers everything from low-level data integration through to Business Process Management.

System Integration Strategies

According to Ambler (1998), there are five systems integration strategies used in organizations:

Stovepipes Strategy

Level 0: The Stovepipe strategy in which applications do not interact with one another. The Stovepipe strategy does not involve any systems' integration. (Ambler, 1998). Stovepiped systems involve related data stored in unconnected databases. For various reasons, data is stored in separate databases: technological, personal, or security reasons. Stovepiping affects retrieval of information from separate, unconnected databases. In some instances, stovepiping is deliberate when it is necessary to protect against unauthorized access or hacking.

Data Exchange Strategy

Level 1: The Data Exchange strategy in which data is extracted from stand-alone applications and copied into the databases of other applications.

Shared Data Store

Level 2: The Shared Data Store strategy where applications access common, shared databases. Shared Data Store is the norm within the industry today (CITE, 2003). Shared data normally simplifies the way business is done. Ranade (2002) states that resources may be shared between nodes of a cluster. "A cluster is built from general-purpose computers that are interconnected using a sufficiently powerful network. The computers of a cluster are called *nodes*" (Ranade, 2002). Shared data storage is an important resource, in that it is generally made accessible to all nodes over a Storage Area Network (SAN). According to Ranade (2002), shared data is beneficial in a number of ways:

(1) Computational resources can be deployed without constraints arising from inaccessibility of data. An application can be deployed on any node that has the required computational resources because data storage required by the application is available anywhere through a SAN.

(2) It avoids needless replication of data. Before SANs, the only way to get many computers running one application independently was to duplicate the required data on local disks. Replicated copies need to be kept synchronized and they consume more storage. Shared storage saves on disks.

(3) It avoids fragmentation of storage space. All available storage is in a single large pool that can be used to carve out storage volumes as needed. Shared storage saves on disks.

(4) It allows better management of data, such as backups, remote replication for disaster recovery, and archival storage.

(5) Data availability can be managed independently of data use. That is, storage is not only a fluid resource; its properties can be fluid too. For example, a particular logical storage volume can be made more reliable by adding mirrors.

Objects strategy

Level 3: The Objects strategy in which object-oriented (OO) technology is used to build new applications and to wrap legacy applications. The Objects strategy is effective from a systems integration point of view only when you choose to make it so.

The Brokered Objects strategy

Level 4: The Brokered Objects strategy in which brokered, distributed object technology is used to integrate systems. The Brokered Objects strategy is best represented by the use of CORBA (Common Object Request Broker Architect) compliant object request brokers (ORBs) to both develop new applications and to wrap access to legacy applications. The Brokered Objects strategy supports the integration of both data and behavior within a distributed environment often comprised of different hardware and software platforms.

Client/Server

Client/server architecture is very successful and popular. Client/server balances the processing load between the client machine and the server machine. If a LAN /WAN connection exists, multiple users or client workstations can work simultaneously on a centralized database system using the Client/Server configuration, or the Intranet configuration. The increase use of Intranet and Internet applications has refocused attention on centralized databases.

From a security point of view, centralized databases are easy to maintain, manage and control. However, due to its configuration, utilization of centralized database is both beneficial and unfavorable. Since most of the processing exists on the Application Server and the database server, it is beneficial. The client machine is freed from excessive processing. On the other hand, the Client/Server reliance on a centralized database configuration not advantageous. Reliance on the centralized database can result in a single point of failure: Should the database fail, work for all those who relies on the centralized database is interrupted.

Enterprise Database

An enterprise is an organization that operates on an extremely large scale. For example, a multinational company that has interconnected computer users located around the world could be considered an enterprise (Chapple, 2003). Applications designed for huge organizations are called enterprise applications. Enterprises database are designed to operate at the utmost echelon of database involvedness. Enterprise databases are exceptional in the database community; they power industry and are used for managing tremendously large volumes of data. Enterprise databases are most often integrated into business processes such as supply chain management, e-commerce and transaction functions. Enterprise database platforms are the organization's primary platform; they are obtained from third-party vendors; they are customized or pre-packaged applications. For example, the database that stores an organization's global sales information is both an enterprise application and an enterprise database. Enterprise database differs from traditional flat file or relational database in that, the data is separate from the program that uses it: an application program is used to separate the data from the physical storage.

While very effective, enterprise databases are not very specific in their form, and they are more difficult to define or codify.

Enterprise Information Systems

According to CompInfo, the Computer Information Center, an Enterprise Information Systems (EIS) collect data from across the entire enterprise. Data may be taken from many different databases and presented in a common way. Report writers will print reports based on a template specifying the fields to be included in the report and which fields are to be totaled. The user can produce reports without recourse to programming. Data mining can be used to look for hidden trends and other previously unknown information within the data (CompInfo, 2003).

Flexible Application Interface

A Flexible Application Interface is easily supported, and adaptable to changes in the internal and external organizational environment. Until recently, Flexible Application Interfaces were inefficient. However, now Flexible Application Interfaces are enabled by Web services functionality and processes (Boodro, 2003). Flexible Application Interfaces are now much more flexible, cost-effective and attainable. "Web services provide the technology infrastructures for different software applications to work together. These services can now bring together technology solutions that were not originally designed by a single vendor. Web services allow disparate applications to work seamlessly through a standard interface, thus delivering new or enhanced business functionality" (Boodro, 2003).

Another aspect of Flexible Application Interface is attributed to user interface (UI). Applications do not consist of a database alone, but also a UI through which the user can access and manipulate data in a controlled environment (http://www.arepo.com/platform/theory/theory-rebuild2.asp).

Data Mediation

Data heterogeneity is a huge problem; it necessitates data standardization. However, since each database user has unique requirements, data interoperability will continue to be a challenge. One solution to data interoperability is data mediation. A data mediator is a computer program, which translates data between two systems with different data schemas. The mediator handles an information exchange between a source and receiver system in two steps. It beginning with a query from the receiver's schema, then it translates it into the equivalent query against the source schema. Next, it executes the source query and translates the retrieved source data into the receiver's format. The result is that the mediator acts as a *semantic gateway* between the systems, permitting the receiver to view the source as an extension of its own database, without concern for the differences in names and representations of data (Lin et al, 2001).

Internet Databases

An Internet database is designed to strengthen and deploy applications for the Internet, to manage Internet content and to facilitate the capability of businesses to conduct business online (Pravica, 1999). The concept of Internet databases is to partition data across multiple geographically or administratively distributed sites where each site runs an almost autonomous database system (Khan et al, 2001). The Internet databases are appropriate for organizations consisting of a number of almost independent sub-organizations such as a University with many departments or a bank with many branches (Khan et al, 2001).

In terms of cost effectiveness, Khan (2001) conducted a study that focused on join queries, and how a static query optimizer might choose an expensive plan by mistake. The study addressed several shortcomings, such as lack of a priori knowledge of the run-time environment, inaccurate statistical assumptions in size estimation, and neglecting the cost of remote method invocation.

Summary

This chapter presented literature reviews that identify what is already known about integrating different database systems and enterprise information systems. The next chapter will present the methodology and data collection procedures for the research.

III. Methodology

Overview

This chapter presents the methodology used to meet the research objectives. This chapter will seek to accomplish the following: restatement of the research objective, presentation of the research methodology, and the reason this was an appropriate methodology for this research.

Research Objective

The objective of this research is to explore information integration and sharing strategies that offer database systems and enterprise information systems alternatives to improve secure operational secure information sharing, cost effectiveness, efficiency, and flexible application interfaces for USMC heterogeneous databases.

Delphi Technique

The Delphi Technique (Dalkey and Helmer, 1963) is most widely known for its application to technology forecasting. Forecasting is a critical factor in policy formulation and planning. It helps determine the direction of future actions (Ono and Wedemeyer, 1994). According to Ono and Wedemeyer (1994), the Delphi Technique has been called the 'cornerstone of futures research.' The Delphi approach builds on the principle that multiple heads are better than one (Wiegers, 2000). The Delphi Technique is a data producing method of generating ideas and facilitating consensus among and from subject matter experts (SME) who have special knowledge to share (Campbell, 1966; Linstone and Turoff, 1975; Basu and Schroeder, 1977).

In a broad sense, the Delphi Technique originated when one of airpower's greatest proponents, General Henry H. "Hap" Arnold (Commanding General of U.S. Army Air Forces) began planning the post-war Air Force many months before the end of World War II. General Arnold felt that the Air Force ought to employ ‘all the scientific minds’ it could find and turn their ‘wondrous’ theories into useful tools. The future, he believed, was tied to new technology -- without it, aviation science would stagnate (Cleary, 1991). In 1944, General Arnold asked the prominent aerodynamicist, Dr. Theodore von Karman, to develop a prospectus for future Air Force research (Cleary, 1991), and a forecast of future technological capabilities that might be of interest to the military. According to Cleary (1991), Von Karman organized a group of his fellow scientists into the Scientific Advisory Group (later known as the Scientific Advisory Board), and this group produced its initial report, *Where We Stand*, in August 1945. The Scientific Advisory Group presented General Arnold with a 33-volume series, Toward New Horizons, four months later. The comprehensive survey of research and development options -- with applications to the Air Force of the future -- accentuated the Group's belief that the Air Force would have to ‘draw on the technological potential of the entire nation’ to acquire and maintain technological superiority over any potential enemy (Cleary, 1991). In order to study “the broad subject of inter-continental warfare other than surface,” in 1946, General Arnold persuaded the Douglas Aircraft company to establish a Project RAND (an acronym for Research and Development).

In 1953, Olaf Helmer and Norman Dalkey of the RAND Corporation invented the Delphi Technique for the purpose of addressing a specific military problem (Helmer, 1983). RAND Corporation's very first application of the Delphi technique was to assess the direction of long-range trends, with special emphasis on science and technology, and their probable effects on society (Gordon and Helmer, 1964). The study addressed six areas: scientific breakthroughs, population control, automation, space progress, war prevention, and weapon systems (Gordon and Helmer, 1968).

Fowles (1978) cites that, in *The Epistemology of the Inexact Science*, Helmer and Rescher (1959) provided a philosophical base for forecasting: the Delphi technique recognizes human judgment as legitimate and useful inputs in generating forecasts.

The Delphi technique is a method for the systematic solicitation and collation of judgments on a particular topic through a set of carefully designed sequential questionnaires interspersed with aggregated information and response of opinions derived from earlier responses (Delbecq, Van de Ven, & Gustafson, 1975). Helmer (1977) points out that Delphi technique represents a useful communication device among a group of experts and thus facilitates the formation of a group judgment. Wissema (1982) emphasizes the significance of the Delphi Method as a 'monovariable' exploration technique for technology forecasting. Baldwin (1975) asserts that lacking full scientific knowledge, decision makers have to rely on their own intuition or on expert opinion.

Opponents of the Delphi Technique conducted studies to show that face-to-face interaction is superior to the Delphi Technique's private and dispersed non-face-to-face opinion gathering (Sackman, 1974). Another concern is "the credibility of Delphi results" Jones, 1975). Some opponents' concerns are that "individual experts may bias their responses so that they are overly favorable toward areas of personal interest" (Jones, 1975). Studies later proved that experts "were able to rise above the desire" to protect self interest (Jones, 1975). To that end, the Delphi Technique has been widely used to generate forecasts in technology, education, and other fields (Dalkey, 1972). Dalkey (1972), co-creator of the Delphi Technique (Dalkey and Helmer, 1963), states that in its simplest form, the method has three features:

Table 4. Delphi Technique Features

FEATURES	DESCRIPTION
Anonymity	Each member of the panel submits his own independent answer(s) to the relevant question(s) by questionnaire or computer query.
Controlled Feedback and Iteration	The results of a given round of responses are summarized and reported to the group, who are then asked to reassess their replies in light of the feedback.
Formal Group Judgment	Given the final set of individual answers, the group answer is expressed as a formal aggregation; e.g., if the questions involve numerical answers, the group judgment may be formulated as the mean, median, or other measure of central tendency.

Debecq (1975) asserts that the Delphi procedure is appropriate for data gathering, because the usual Delphi procedure is used to obtain ideas in writing. The act of writing forces participants to contemplate the subject thoughtfully and tends to produce a high volume of ideas (Debecq, 1975). Finally, the Delphi technique has been shown to be a successful approach to conduct research when the responses being sought are value judgments rather than factual information. Although it is more difficult to assess the appropriateness of value judgments, it is generally agreed upon that value judgments are not all equal, but can, in fact, be more 'right' or more 'wrong.' Dalkey and Rourke (1972) conducted an experiment to test the value of using Delphi procedures in obtaining non-factual data. In the experiment, university students were asked about the objectives of higher education. They determined from the outcome of these experiments that Delphi procedures are "appropriate for generating and assessing value material" (Dalkey and Rourke, 1972). Linstone and Turoff (1975) agree that Delphi is particularly useful for studies that call for subjective judgment rather than precise statistical analysis."

The goal of the Delphi technique is to build consensus within the group by first eliciting their opinions and developing common themes. After that, each panel member is presented with the common topic and asked to assess it. Each member's reasoning may also be presented to the panel so that individuals can begin to 'think around' the problem. As the process progresses, the group tends to move toward consensus. Linstone and Turoff (1975) suggest that the Delphi procedure should be considered for research problems when one of several conditions exist (see Table 5).

Table 5. Conditions Under which the Delphi Technique is Appropriate (Linstone and Turoff, 1975)

DELPHI TECHNIQUE CONDITIONS
The problem does not lend itself to precise analytical techniques, but can benefit from subjective judgments on a collective basis.
The individuals needed to contribute to the examination of a broad or complex problem have no history of adequate communication and may represent diverse backgrounds with respect to experience or expertise.
More individuals are needed than can effectively interact in a face-to-face exchange.
Time and cost make frequent group meetings infeasible
The efficiency of face-to-face meetings can be increased by a supplemental group communication process
Disagreements among individuals are so severe or politically unpalatable that the communication process must be refereed and/or anonymity assured
The heterogeneity of the participants must be preserved to assure validity of the results, i.e., avoidance of domination by quantity or by strength of personality

In order to have a successful Delphi study, careful consideration must be given to the selection and quality of the participants. According to Delbecq et al (1975), potential respondents should meet four criteria. It is unrealistic to expect effective participation unless respondents:

(1) Feel personally involved in the problem of concern to the decision makers

(2) Have pertinent information to share

(3) Are motivated to include the Delphi task in their schedule of competing tasks

(4) Feel that the aggregation of judgments of a respondent panel will include information which they too value and to which they would not otherwise have access

Dalkey (1975) suggest that the full Delphi attributes should consist of the following conditions (see Table 6).

Table 6. Full Delphi Techniques/Attributes

ATTRIBUTES	DESCRIPTION
Size of group	The average error of the group responses declined monotonically with the size of the group, with decreasing returns with increasing size. Roughly, one half of the individual error was observed with groups of 7 members. An additional 20 members reduced the average group error by an additional 10%. The reduction of error with size of group is analogous to, but not identical with, the rule for the dispersion of the sample mean in random sampling.
Iteration with feedback.	There was monotonic reduction in the dispersion of individual responses (convergence) with iteration, again with decreasing effect with additional iterations. However, the accuracy of the group answer improved with the first iteration and fluctuated with additional iterations. It is my present belief that a single iteration furnishes the major benefit obtainable with iteration.
Dispersion.	There is a generally held belief that greater agreement (smaller dispersion is associated with a greater likelihood of the group being correct). This was born out in the experiments. Roughly, the average group error was about 2/3 of the observed dispersion.
Individual and group self-ratings	In many of the exercises, individuals were asked to rate their confidence in their answer to each question on a scale of 1-5 where 5 meant 'I know the answer' and 1 meant ' I'm just guessing'. A group self-rating could then be computed for each question by taking the average of the individual ratings. Between a group self-rating of 1.2 and 4, average error dropped by a factor of 5.

The Delphi Technique can be modified in many ways. In terms of procedures (i.e., a series of rounds with selected experts) and intent (i.e., to predict future events and to arrive at consensus), the modified Delphi technique is similar to full Delphi technique (Custer et al, 1999). “Advantages related to the use of the modified Delphi technique include reducing the effects of bias due to group interaction, assuming anonymity, and providing controlled feedback to participants" (Custer et al, 1999).

The panel size for a Delphi study varies. "Helmer and Dalkey used a panel of seven experts in their original Delphi experiment in 1953" (Helmer, 1983). Turoff (1975) suggests a panel size of anywhere from ten to fifty participants (p.86). A Delphi panel that consists of a homogeneous group, such as a group of experts from the same general discipline area, need only involve ten to fifteen people (Delbecq et al, 1975). Wicklein (2000) used a panel of 25 experts in his study. Dalkey et al (1969) found that error decreased rapidly as the group size increased from one to about thirteen; further small decreases in error continued to a size of about 25 people, at which point the error rate stabilized (1969). Based on these findings, they continued their experiments using groups of fifteen to twenty people" (Dalkey et al, 2000).

Research Approach

Given its ability to obtain expert input from individuals who are geographically dispersed, the Delphi method was selected for this research. The Delphi method typically consists of three or four rounds of questionnaires among the experts. However, since 'the accuracy of the group answer improved with the first iteration and fluctuated with additional iterations (Dalkey, 1972), a modified Delphi process of two iterations will be used to first educe and then rate the importance of characteristics of integrated database systems and enterprise information systems alternatives that will identify and provide options for improved operational secure information sharing, cost effectiveness, efficiency, and flexible application interfaces. In order to improve the round response rate, and provide a solid grounding in previously developed work, additional modification of the technique consists of (1) beginning the process with a set of carefully chosen items. These pre-selected items were drawn from various sources including synthesized reviews of the literature, and interviews (Rubin and Rubin, 1995) with selected content experts. (2) The survey interview will be primarily web-based vice paper-based. The primary participants will be mid-level to senior information officers, with a focus on Chief Information Officers (CIOs). "The CIO's role is primarily one of architecture and the process of integrating" (Iansiti, 2003). CIOs need strong architectural expertise to figure out how the different integration alternatives fit (Iansiti, 2003). Additional participants will be solicited from government contractors, as well as military and civilian IT officials. Participant's anonymity and privacy will be assured. Extra precaution will be taken by ensuring that participants register, generate a Password and User ID to access the interview.

The Delphi methodology will be employed as follows:

(1) During the first round of this study, the researcher will select fifty of the top Fortune 500 organizations and request via email and web survey that the CIO/IT official complete and return the interview questionnaire cited in Appendix A, Data Integration/Information sharing, 1st Round, Delphi Interview to Private Sector's Chief Information Officer.

(2) After all the response from the first wave of the Delphi study are received, the researcher will compile the results and email a second iteration of the interview instrument to the panelists, the same CIOs. "For most event statements the final-round interquarantile range is smaller than the initial round range (Linstone and Turoff, 2002). Appendix B, 2nd Phase of Private Sector's CIO Interview (Open-ended questions). In addition, the CIOs will be asked to rank order their top 3 items.

(3) After their responses are collected, all identifiers will be removed from the data before analysis commence. To keep coherent data strings, the researcher will identify data in such a manner as Data from respondent 001A, 001B, 001C, 002A, 002B, 002C, 0003A, 003B, and 003C, etc.

(4) Using the results of the Delphi study, the researcher will construct a model.

Statistical Analysis Approach

For the most part, *data* is the term used to describe sets of factual information collected as part of some study. However, in statistics, *data* refers to sets of measurements (Bernstein and Bernstein, 1999). The four levels of measurement are four types of measurement scales: nominal, ordinal, interval, and ratio. Nominal scales produce qualitative measurement variables. Ordinal, interval, and ratio scales produce quantitative measurement variables.

This research is a qualitative study. Thus, nominal-level measurement will be used on the data gathered in this study. Nominal-level measurement measures things by classifying them in categories. Categories on nominal scales are not ordered in any way (e.g., from small to large), and numbers are used only as labels for categories (Bernstein and Bernstein, 1999). There can be as many categories as needed (Bernstein and Bernstein, 1999). Because there is no intrinsic numerical order to the data collected for this research, the data collected for this study is classified as qualitative data. Qualitative data is not measured on a natural numerical scale (McClave, 2001). Qualitative data is measured and classified categorically. Qualitative data can be sub-classified as either nominal data or ordinal data. The categories of an ordinal data set can be ranked or meaningfully ordered, but the categories of a nominal data set cannot be ordered

(McClave, 2001). For this research, the Delphi Technique will be used to gather data; the Pareto Analysis (80 – 20 Rule), and Group Work decision making methodology (Ngwenyama and Bryson, 1999). Qualitative analysis will be performed on the data gathered.

Pareto Analysis

Data analysis for this research will be conducted, based on the *Pareto principle – The 80-20 Rule*. The Pareto principle illustrates the idea of 'the vital few and the trivial many" (McClave et al, 2001). The *Pareto analysis* is based on the concept of categorization. The Pareto analysis focuses on the categorization of items and the determination of which categories contain the most observations (McClave et al, 2001). Data that consist of small number of values, each corresponding to a specific category value or label is classified as categorical data.

For this research, Paerto concepts will be used in round two of the Delphi study, where each participant will be asked to select and rank the top 3 items in the relevant lists that were developed from round one. Additionally, Group work decision making methodology will be used to aggregate rank ordering across multiple participants (Ngwenyama and Bryson, 2003).

Group Work Decision Making

The purpose of Group Work Decision Making is to evaluate and decide upon various decision alternatives. According to Ngwenyama and Bryson (1999), Group Work decision making is one area of team support that is often desired in the scoring and ranking of decision alternatives on qualitative/subjective domains, and the aggregation of individual preferences into group preferences. Group Work decision analysis techniques advances the state of the art of group decision making by addressing four common limitations: (1) the inability to deal with vagueness of human decision makers in articulating preferences; (2) difficulties in mapping qualitative evaluation to numeric estimates; (3) problems in aggregating individual preferences into meaningful group preference; and (4) the lack of simple user friendly techniques for dealing with a large number of decision alternatives.

Group Work decision making offers some basic techniques for: (1) eliciting preferences from users of diverse backgrounds; (2) mapping qualitative evaluations to numeric estimates; (3) analyzing data relevant to evaluating consensus formation; (4) easy implementation in manual and computer supported group activities (Ngwenyama and Bryson, 1999).

Concept Analysis and Affinity Diagram

In addition to the development of and selection from lists, participants' opinions will be obtained via open-ended questions, as well. Content analysis for this research will be conducted on the subject matter experts' (SMEs) open-ended questions, by using the Affinity Diagram process (Brassard, 1997). Concept analysis can be applied to many forms of inquiry and contexts (Neuendorf, 2002). For this research, test analysis is the context of interest. Thus, the Affinity Diagram process will be used to gather ideas and opinions form the SMEs' response provided in their open-ended questions. By definition, an Affinity Diagram is a tool. "This tool gathers large amount of language data (ideas, opinions, issues, etc.), organizes it into groupings based on the natural relationship between each item, and defines groups of items. It is largely a creative rather than a logical process" (Brassard, 1997).

Phase I Interview and Instrument Development

Phase I of the Delphi interview required data aggregation of the participants' responses. The aim of interview Questions 1 and 2, Appendix A is to focus on the first objective of this research: to *explore information integration and sharing strategies that offer database systems and enterprise information systems alternatives.* Numerous steps are required to identify successful information sharing strategies that the private sector's organizations have tried.

Interview Questions 1 and 2 are designed to discover a pervasive successful sharing strategy. Interview Question 1, ask participants to characterize both past and present impediments which contributed to data integration and information sharing incompatibilities. Steps were taken to identify all the impediments:

(1) Each category of *Past Impediment* and *Present Impediment* was separately compiled. A new list was generated which served as input for the second iteration of questions.

(2) An aggregate of each category was generated. Additionally, using an Excel spreadsheet, a summation of each category's past and present impediments was calculated.

(3) The total number of participants who responded was divided into the total of each category. The quotient was then used to generate tables and graphs of each category, for Pareto analysis.

Interview Question 2 was designed to identify all the private sector's current successful information sharing strategies. The question, ask participants to characterize their organization's *Successful* and *Unsuccessful* sharing strategies; as well as information sharing strategies *Not Used* and *Currently Used*. Each category of *Successful*, *Unsuccessful*, *Not Used*, and *Currently Used* information sharing strategies was compiled, and tallied. Tables and categorical Pareto charts were generated from the results.

Interview Questions 3, in Appendix A concentrates on the second objective of this research: to identify *which database integration strategies and enterprise information systems will provide options for improved operational secure information sharing, cost effectiveness, efficiency and flexible application interfaces. As a result, this will afford the United States Marine Corps flexibility, reliability, security, efficiency, and responsiveness during a crisis.*

The data analysis procedures used in the previous questions are also performed on each category of interview Question 3: *Secure Information sharing*, *Cost Effectiveness*, *Efficiency*, and *Flexible Application Interfaces*; all of which identifies attributes that characterize the strengths of the private sector's information sharing strategies.

Finally interview Questions 4 in Appendix A attends to the third objective of this research: Systems Integration Strategies – specifically, *which database integration strategies and enterprise information systems will provide options for improved operational secure information sharing, cost effectiveness, efficiency, and flexible application interfaces for USMC heterogeneous databases*.

A comprehensive list was produced from the responses provided in interview Question 4 citing the various database models utilized by the private sector; as well as the ubiquitous model used in today's organization. The number of occurrences of each database model was tallied and the quotient was then used to generate tables, for analysis.

Phase II Interview and Instrument Development

A compilation of the product from the first round of the Delphi interview is used as source data to generate questions for the second round of the Delphi study. Participants who responded and provided an email address are emailed the second phase of the interview. Essentially, they are asked to identify and rank order, the top three items from the initial set of responses.

Questions 1, 2, 3, 4 and 5 of the second interview, Appendix B focused on the first objective of this research: *exploration of information integration and sharing strategies that offer database systems and enterprise information systems alternatives.*

The responses provided from the Phase I interview are compiled into separate lists. The compiled lists are regenerated; along with a variant of the questions asked in Phase I and are repeated in Phase II of the Delphi Interview. The aim of Phase II is to narrow down the myriad of categories characterized in each question of Phase I.

In Questions 1 thru 6, from each consolidated list, participants are asked to select their top three choices in order of priority. After selections are made, they are asked, via open-ended question to explain from what standpoint the selection is made. Questions 1 thru 6 follow:

> (1) In terms of data integration and information sharing, out of all the impediments, which of the following incompatibilities would you classify to be the most hindrance? **Please rank your <u>top three</u> (3) in order of priority** (1 – the most hindrance)
>
> (2) What is your reason for selection of the number one information sharing impediment? (Briefly explain.)
>
> (3) Which of the following information sharing strategies would you characterize as most successful? **Please rank your <u>top three</u> (3) in order of priority** (1 – the most successful)

(4) What characterize your selection as a success? (Please explain)

(5) Which of the following information sharing strategies would you characterize as most unsuccessful? **Please rank your top three (3) in order of priority** (1 – the most unsuccessful).

(6) What characterize your selection as an unsuccessful? (Please explain)

After the all the responses of the second wave are returned, group work decision making methodology is will be used to aggregate answers in Questions 1, 3, and 5 (Ngwenyama and Bryson, 1999).

Each category is separately aggregated and weighted. A new list is generated which provided a prioritized aggregation of the participant's choices. From this new prioritized list, the preeminent impediment and information sharing strategies are identified.

Questions 2, 4 and 6, are open-ended questions; and necessitate interpretation of content (Scheele, 1975). Concerning desirable and undesirable attributes of information sharing strategies, Questions 1 and 2 are designed to stimulate responses indicating what concepts and terms support their opinions. Most of the participants' "thinking process cannot directly shared," so [Questions 1 and 2] attempted to depict for the group some typical points of view (Scheele, 1975).

Questions 7 of the second interview, is designed to answer the second and third objective of this research.

A study was conducted on the efficiency of interval-scale effectiveness in ranking opinions obtained in a Delphi study; and found that the most common methods of scaling which could be used in a Delphi study are simple ranking, the rating-scale method (likerttype scale) and the pair comparison methods (Scheibe et al, 1975). One of the results found that the simple ranking method was considered by the participants as the most comfortable to perform. The ranking method is fairly easy for a small number of goals (Scheibe et al, 1975).

Final Phase

Using the results of the Delphi study, a model will be constructed that will be compared to the diagrams derived during the Pareto analysis. Prior to presenting a model to the USMC, comparisons of the models will be necessary to build on what was already known, recognize similarities and recognize differences.

Data Collection Procedure

To gather qualitative data, individual email was sent to fifty IT officials inquiring if they would take part in a Delphi interview. The interview consisted two sequential phases. The interviews are shown in Appendix A. This chapter explains how the research question and hypotheses were answered. First, the research design is addressed. Next, the variables used for the study are discussed. The population, sample selection, pilot survey, and survey administration are then explained. Finally, a method for data analysis is proposed.

Summary

This chapter presented the methodology used to meet the research objectives. In order to answer the research question, a Delphi Technique was used to gather data; in addition, Pareto Analysis, Group Work decision making, and content analysis are used for data analysis. The next chapter will present the data analysis and results. Chapter V will present the discussion.

IV. Results and Analysis

Overview

Chapter IV consists of five sections, which outlines the result of the Delphi research process. Section I outlines the demographics of Phase I participants. Section II outlines Phase I interview results. Section III outlines the demographic of Phase II participants. Section IV outlines Phase II interview results. And Section V presents a final model of the results from the Delphi research in its entirety.

Section I: Phase I Participants Demographic

Participants were geographically dispersed across the United States. The initial contact of fifty potential participants obtained nineteen initial participants. Consequently, the results cited in Section II of this chapter, illustrates the results from a panel size of nineteen Delphi participants.

Section II: Phase I Interview Results

Phase I interview questions 1 and 2 were designed to focus on the first objective of this research: to *explore information integration and sharing strategies, which offer, database systems and enterprise information systems alternatives.*

In phase I of the research, participants were asked to characterize their organization's *Past* and *Present impediments* to information sharing. Next, each participant was asked to characterize their organization's *successful* and *unsuccessful*; as well as *Not Used* and *Currently Used* information sharing strategies. Phase I of the research ended with the participants being asked to characterize the strength of the strategies presently used within their organization.

Table 7, shows the first part of Question 1's interview results: a list which the participants characterize as past impediments.

Table 7. Respondents Identification of Impediments to Information Sharing

(N = 19)		
CATEGORIES OF PAST IMPEDIMENTS	**FREQUENCY OF PAST IMPEDIMENTS**	**FREQUENCY OF PRESENT IMPEDIMENTS**
Access Control (Data Security)	13	8
Integrity of Stored Information	13	11
Business Rules in Corporate Database	11	13
Incompatible Data Format	11	10
Incompatible Databases	11	8
Database Interface Language	9	5
Granularity of Context	9	8
Client/Server	8	2
Concurrency Issues	8	9
Incompatible Data Definition Format	2	1
Data Stored Across Multiple Systems	1	1
Reliability of Data	1	1
Data Management (Lack of Data ownership)	1	0
Exploiting Semantic Constraints	0	1
Not Exploiting Semantic Constraints	1	0
Reliability of Custom Interface Code	1	0

Table 7 shows the respondent identification of both past and present impediments. From this information, a new list was generated for phase two of the Delphi study. During the Delphi Phase II interview, using the Pareto Principle, the participants were asked to select their top 3 hindrances from the new list.

Table 8 consists of an aggregation of Question 2's interview results: a general list of characteristics that the participants characterize as successful and unsuccessful information sharing strategies within their organization.

Table 8. Respondents Identification of Successful Strategies

(N = 19)		
CATEGORY OF INFORMATION SHARING STRATEGIES	# OF RESPONSE SUCCESSFUL	FREQUENCY # OF RESPONSE UNSUCCESSFUL
Enterprise databases	8	0
Enterprise Information Systems	7	1
Middleware	6	1
Collaborative Planning Forecasting and Replenishment	5	5
Internet databases	5	1
Persistency	4	3
Federated databases	3	3
APIs	1	0
Data Warehouse	1	0
Enterprise Data Management (Owners of Data)	1	0
Exploiting Semantic Constraints	1	3
Internet Delivery	1	0
Mediation	1	2
Quality Validations	1	0
Web Access	1	0
Brokered Objects	0	0
Data Exchange	0	0
Objects (O-O)	0	0
Shared Data Store	0	0

Table 8 shows the respondents' identification of both successful and unsuccessful information sharing strategies. From this information, a new list was generated for phase two of the Delphi study. During the Delphi Phase II interview, using the Pareto Principle, the participants are asked to select their top 3 successful and unsuccessful information sharing strategies.

The reported *Successful* and *Currently Used* characteristics of information sharing strategies are shown in Table 9.

Table 9. Comparison of Successful and Currently Used Sharing Strategies

(N = 19)		
CATEGORY OF CHARACTERISTICS	**# OF RESPONSE**	**FREQUENCY # OF RESPONSE**
INFORMATION SHARING STRATEGIES	**SUCCESSFUL**	**CURRENTLY USED**
Enterprise databases	8	15
Enterprise Information Systems	7	12
Middleware	6	12
Collaborative Planning Forecasting and Replenishment	5	5
Internet databases	5	7
Persistency	4	5
Federated databases	3	3
APIs	1	1
Data Warehouse	1	1
Enterprise Data Management (Owners of Data)	1	1
Exploiting Semantic Constraints	1	5
Internet Delivery	1	1
Mediation	1	1
Quality Validations	1	1
Web Access	1	1
Brokered Objects	0	0
Data Exchange	0	0
Objects (O-O)	0	0
Shared Data Store	0	0

It is important to note that in some cases, the number of respondents that reported these strategies as currently used is almost double the amount of successful characteristics reported.

Question 2 analyses conclude with a comparison between the reported Successful and Not Used categories of information sharing strategies. Table 10 presents the results of this comparison.

Table 10. Comparison of Successful Information Sharing Strategies Not Used

(N = 19) CATEGORY OF CHARACTERISTICS INFORMATION SHARING STRATEGIES	# OF RESPONSE SUCCESSFUL	FREQUENCY OF RESPONSE CURRENTLY NOT USED
Enterprise databases	8	1
Enterprise Information Systems	7	2
Middleware	6	2
Collaborative Planning Forecasting and Replenishment	5	5
Internet databases	5	5
Persistency	4	5
Federated databases	3	10
APIs	1	0
Data Warehouse	1	0
Enterprise Data Management (Owners of Data)	1	0
Exploiting Semantic Constraints	1	9
Internet Delivery	1	0
Mediation	1	13
Quality Validations	1	0
Web Access	1	0
Brokered Objects	0	0
Data Exchange	0	0
Objects (O-O)	0	0
Shared Data Store	0	0

In the case of the primary successful and not used information sharing strategies, the observed differences between successful strategies as compared to not used strategies are higher. Phase I analysis continues with, an aggregation of interview Question 3 dealing with characterization of the strength of the information sharing strategies. In order to answer the thesis research question, Phase I interview question 3 was designed to focus on the second objective of this research: *to identify which database integration strategies and enterprise information systems will provide options for improved operational secure information sharing, cost effectiveness, efficiency and flexible application interfaces.*

Table 11. Strength of Current Strategy: Secure Information sharing

(N = 19)	
STRENGTH OF STRATEGIES	**FREQUENCY OF SECURE INFORMATION SHARING**
Enterprise databases	12
Enterprise Information Systems	8
Middleware	7
Collaborative Planning Forecasting and Replenishment	6
Federated databases	6
Exploiting Semantic Constraints	4
Internet databases	3
APIs	1
Internet Delivery	1
Persistency	1
Brokered Objects	0
Data Exchange	0
Data Warehouse	0
Enterprise Data Management (Owners of Data)	0
Mediation	0
Objects (O-O)	0
Quality Validations	0
Shared Data Store	0
Web Access	0

Table 11 shows respondent identification of how they characterize the strength of their organization's currently used information sharing strategies.

The Phase I analysis continues with evaluation of *Cost Effectiveness*, an aggregation of interview question 3 regarding characterization of the strength of the information sharing strategies.

Table 12. Strength of Current Sharing Strategies: Cost Effectiveness

(N = 19)	
STRENGTH OF STRATEGIES	**FREQUENCY OF COST EFFECTIVENESS**
Middleware	9
Enterprise databases	8
Enterprise Information Systems	8
Internet databases	5
Collaborative Planning Forecasting and Replenishment	4
Exploiting Semantic Constraints	4
Persistency	3
Federated databases	2
Mediation	2
APIs	1
Brokered Objects	0
Data Exchange	0
Data Warehouse	0
Enterprise Data Management (Owners of Data)	0
Internet Delivery	0
Objects (O-O)	0
Quality Validations	0
Shared Data Store	0
Web Access	0

Table 12 shows respondent identification of how they characterize the strength of their organization's currently used information sharing strategies. In this case, cost effectiveness is relevant. Several unpredictable factors can influence cost effectiveness: lack of a priori knowledge of the run-time environment, inaccurate statistical assumptions in size estimation, and neglecting the cost of remote method invocation. In addition, a static query optimizer that does not consider the characteristics of the environment or only considers the a priori knowledge on the run-time parameters might end up choosing expensive plans (Khan et al, 2001).

A new list was generated from the information on Table 12 and used in phase two of the Delphi study. During the Delphi Phase II interview, the participants were asked to rate how they feel that "Cost Effectiveness" contributes towards their organization's successful information sharing strategy.

The Phase I analysis continues with evaluation of *Efficiency*, an aggregation of interview question 3 regarding characterization of the strength of the information sharing strategies.

Table 13. Strength of Current Strategy: Efficiency

	(N = 19)
STRENGTH OF STRATEGIES	**FREQUENCY OF EFFICIENCY**
Enterprise Information Systems	12
Collaborative Planning Forecasting and Replenishment	9
Enterprise databases	9
Middleware	8
Persistency	7
Internet databases	4
Exploiting Semantic Constraints	3
Federated databases	3
Mediation	3
APIs	1
Brokered Objects	0
Data Exchange	0
Data Warehouse	0
Enterprise Data Management (Owners of Data)	0
Internet Delivery	0
Objects (O-O)	0
Quality Validations	0
Shared Data Store	0
Web Access	0

Table 13 shows respondent identification of how they characterize the strength of their organization's currently used information sharing strategies Information from Table 13 will be used to form a new list, which is used in phase two of the Delphi study. During the Delphi Phase II interview, the participants are asked to rate how they feel that "Efficiency" contributes towards their organization's successful information sharing strategy.

The Phase I interview analysis continues with evaluation of *Flexible Application Interface*, an aggregation of interview question 3regarding characterization of the strength of the information sharing strategies.

Table 14. Strength of Current Strategy: Flexible Application Interface

	(N = 19)
STRENGTH OF STRATEGIES	FERQUENCY OF FLEXIBLE APPLICATION INTERFACE
Internet databases	8
Enterprise databases	7
Enterprise Information Systems	6
Middleware	7
Collaborative Planning Forecasting and Replenishment	3
Federated databases	3
Exploiting Semantic Constraints	3
APIs	1
Mediation	1
Persistency	0
Brokered Objects	0
Data Exchange	0
Data Warehouse	0
Enterprise Data Management (Owners of Data)	0
Internet Delivery	0
Objects (O-O)	0
Quality Validations	0
Shared Data Store	0
Web Access	0

Table 14 shows the respondents' identification of how they characterize the strength of their organization's currently used information sharing strategies. In this case, Flexible Application Interface is relevant. A Flexible Application Interface is easily supported, and adaptable to changes in the internal and external organizational environment. Until recently, Flexible Application Interfaces were inefficient. However, now Flexible Application Interfaces are enabled by Web services functionality and processes (Boodro, 2003).

Information from Table 14 will be used to form a new list, which is used in phase two of the Delphi study. During the Delphi Phase II interview, the participants are asked to rate how they feel that "Efficiency" contributes towards their organization's successful information sharing strategy.

The Phase I analysis concludes with evaluation of the final question, an aggregation of interview Question 4 regarding database models currently used in the private sector.

Table 15. Database Models Currently Used in the Private Sector

(N = 19)	
PHASE I DATA ANALYSIS DATABASE MODEL USED	
TYPE	FREQUENCY OF RESPONSE
RELATIONAL	19
HIERARCHICAL	12
FLAT FILE	11
OBJECT-ORIENTED	9
NETWORK	7
OBJECT-RELATIONAL	0

Table shows the entire database models utilized in today's industry. "The true power of a relational database resides in its ability to break the link between data access and the underlying data itself" (MSND). Users can access all of their organization's data dynamically without any knowledge of how the underlying data is actually stored by using a high-level access language such as SQL (structured query language). Query optimizer is used to input queries and convert them to a format that efficiently accesses the stored data. This is done to maintain both system performance and throughput, so that the relational database system can accept a variety of user queries and convert them to a format that efficiently accesses the stored data (MSND).

Section III: Phase II Participants Panel Size

The initial contact of participants obtained a panel size of nineteen Delphi participants. These participants were e-mailed the final Delphi interview. Fifteen participants responded. Consequently, the results cited in Section IV of this chapter, illustrates the results from fifteen Delphi participants.

Section IV: Phase II Interview Results

The aim of Phase II is to narrow down the myriad categories characterized in each question of Phase I. Phase II of the Delphi interview required data aggregation of the participants' responses. The final Delphi interview questions were designed to focus on the research question, specifically the third objective of this research: *which database integration strategies and enterprise information systems will provide options for improved operational secure information sharing, cost effectiveness, efficiency, and flexible application interfaces for USMC heterogeneous databases.*

In questions 1 thru 6, in each consolidated list, participants are asked to select their top three choices in order of priority. After selections are made, they are asked an open-ended question to explain from what standpoint their selection is made.

The participants' ranking of questions 1, 3, and 5 are aggregated. Questions 2, 4 and 6, are open-ended questions; which necessitates interpretation of content. The Affinity Diagram process is used to extract relevant ideas and opinions. Question 7 is designed to rate the participant's selection and obtain their opinion via open-ended questions. This question also necessitates interpretation of content. The results are provided in a series of Tables.

Phase II of the research began with asking the participants to classify their organization's top three incompatibilities and hindrance (in order of priority where 1 is the biggest hindrance) to information sharing. Next, in order to narrow down which strategies are successful, it was necessary to identify all of the generalized information sharing strategies individual organizations have tried. Again, each participant was asked to characterize their organization's most *successful* and *unsuccessful information sharing strategies.* Phase II of the research ended with the participants being asked to characterize and rate on a scale of 1 – 5, the strength of the strategies presently used within their organization. In addition, they were asked to specify how each feature impact their organization's selection of its successful information sharing selection (1 – no impact). These questions were asked, in order to determine the overall successful information sharing strategies.

A compiled list from Phase I results of past and present impediments to information sharing was provided to the participants. The results from Phase II are shown in Table 16: a list that the respondents characterize as hindrances of information sharing. Using the Pareto Principle, each participant was asked to rank the top three hindrances faced by their organization, in order of priority, with one being the most hindrance. The Group Systems methodology was used to aggregate the rank ordering of hindrances across multiple participants.

Table 16. Respondents Identification of Hindrances to Information Sharing

(N = 15)	
CATEGORY OF HINDERANCES TO INFORMATION SHARING	FREQUENCY OF TOTAL WEIGHT
Data Source Across Multiple Systems	20
Business Rules in Corporate Database	10
Access Control (Data Security)	9
Incompatible Data Definition Format	9
Reliability of Data	7
Incompatible data format	6
Data Management (Lack of Data ownership)	5
Granularity of context	5
Integrity of stored information	5
Database Interface Language	4
Incompatible databases	4
Client/Server	2
Concurrency Issues	2
Not Exploiting Semantic Constraints	1
Reliability of Custom Interface Code	1
Exploiting Semantic Constraints	0
Other	0

Total Weight = 3 * 1st Priority + 2 * 2nd Priority + 1 * 3rd Priority

Table 16 shows the respondents' identification of how they characterize hindrances to information sharing strategies in their organization. Participants' responses as to what characterizes their selection as a hindrance vary. Some of the open-ended responses include: (1) "Stovepipes remain the number one hindrance to information sharing. The technology does exist to overcome many of the interoperability issues. Client Server remains a significant problem due to the burden of having to have the client available on every desktop for specific applications"; (2) "Lack of cohesive/holistic approach"; (3) "First I think access to data is hardest – especially with physically disparate users. Second (and third) is the ability to aggregate data and perform meaningful analysis"; (4) "If concrete business rules are established up front a lot of the other impediments may never come into play"; (5) Information sharing is a challenging issue for a decentralized organization."

The Phase II interview analysis continues with evaluation of *Successful* strategies, an aggregation of interview question 3 regarding characterization of successful information sharing strategies.

Table 17. Identification of Successful Strategies to Information Sharing

(N = 15)	
CATEGORY OF SUCCESSFUL TO INFORMATION SHARING	**TOTAL WEIGHT**
Enterprise Databases	14
Middleware	13
Enterprise Information Systems	11
Data Warehouse	10
Web Access	10
Internet Delivery	9
Shared Data Stored	8
APIs	4
Exploiting Semantic Constraints	3
Internet Databases	3
Federated Databases	2
Brokered Objects	1
Data Exchange	1
Objects-Oriented (O-O)	1
Collaborative Planning Forecasting	0
Enterprise Data Management (Owners of Data)	0
Mediation	0
Other	0
Persistency	0
Quality Validations	0

Total Weight = 3 * 1st Priority + 2 * 2nd Priority + 1 * 3rd Priority

Table 17 shows the respondents' identification of how they characterize successful information sharing strategies in their organization. Participants' responses as to what characterizes their selection as a success vary. Few of the open-ended responses include (1) "Most reliable, accessible and fit for purpose"; (2) "Impact of results gained from technical solution"; (3) "Utilizing the same database for multiple applications means no replication delay or inconsistent data concerns."

The Phase II interview analysis continues with evaluation of *Unsuccessful* strategies, an aggregation of interview question 5: characterization of unsuccessful information sharing strategies.

Table 18. Identification of Unsuccessful Strategies to Information Sharing

(N = 15)	
CATEGORY OF UNSUCCESSFUL TO INFORMATION SHARING	**FREQUENCY OF TOTAL WEIGHT**
Middleware	18
Federated Databases	12
Data Warehouse	10
Exploiting Semantic Constraints	10
Internet Delivery	5
Enterprise Databases	4
APIs	2
Brokered Objects	0
Collaborative Planning Forecasting and Replenishment (CPFR)	0
Data Exchange	0
Enterprise Data Management (Owners of Data)	0
Enterprise Information Systems	0
Internet Databases	0
Mediation	0
Objects-Oriented (O-O)	0
Other	0
Persistency	0
Quality Validations	0
Shared Data Stored	0
Web Access	0

Total Weight = 3 * 1st Priority + 2 * 2nd Priority + 1 * 3rd Priority

Table 18 shows the respondents' identification of how they characterize unsuccessful information sharing strategies in their organization. Participants' responses as to what characterizes their selection as an unsuccessful vary. It is important to note that some participants did not answer question 5. A few of the open-ended responses include: (1) "While middleware can work well for smaller databases, it is not a good solution for the large databases that are typically involved at our company"; (2) "Owners of legacy systems strongly resist transition to the data warehouse"; (3) "Data warehouses usually wind up with too much information. It makes it difficult to extract what is needed, unless you know exactly what to look for."

The Phase II interview analysis continues with evaluation of *Secure Information Sharing* as it relates to the organization's number one selection of information sharing strategies, an aggregation of interview question 7 information sharing strategies whose selection was based on secure information sharing are presented in Table 19.

Table 19. Secure Information Sharing's Impact on Selection of Strategies

(N = 15)	
INFORMATION SHARING SELECTION BASED ON OF SECURE INFORMATION SHARING	**FREQUENCY OF TOTAL WEIGHT**
Web Access	12
Data Warehouse	8
Enterprise Databases	8
Exploiting Semantic Constraints	5
Internet Delivery	5
Middleware	5
Enterprise Information Systems	4
Shared Data Stored	4
APIs	0
Brokered Objects	0
Collaborative Planning Forecasting	0
Data Exchange	0
Enterprise Data Management (Owners of Data)	0
Federated Databases	0
Internet Databases	0
Mediation	0
Not Exploiting Semantic Constraints	0
Objects-Oriented (O-O)	0
Other	0
Persistency	0
Quality Validations	0

Total Weight = 5 * Rate #1 + 4 * Rate #2 + 3 * Rate #3 + 2 * Rate #4 + 1 * Rate #5

Table 19 shows the respondents' identification of which information sharing strategies were selected based on the strength of secure information sharing. Participants' responses as to their number one selection base on secure information sharing vary. Some of the open-ended responses include: (1) "It is entirely possible to adopt secure information sharing within an architecture that includes our identified and selected technologies. This is important in an enterprise application; (2) "Secure means meeting the security specs of the enterprise. Usually this means that only the authorized users can have access to the information, and no one else;" (3) "Secure Information Sharing is very important. However, too much unnecessary control can be a roadblock to share information and prevent an organization to maximize the ROI due to information sharing;" (4) "Data in the wrong hands can be detrimental, and thus should be secured to ensure only those folks that have a need to know has access to data that is pertinent to their department or their needs;" (5) "Meeting the requirements (mandatory) with the least resources necessary."

The Phase II interview analysis continues with evaluation of *Cost Effectiveness* as it relates to the organization's number one selection of information sharing strategies. An aggregation of interview question 7 information sharing strategies whose selection was based on cost effectiveness, are presented in Table 20.

Table 20. Cost Effectiveness' Impact on Selection of Strategies

(N = 15)	
INFORMATION SHARING SELECTION BASED ON COST EFFECTIVENESS	**FREQUENCY OF TOTAL WEIGHT**
Enterprise Databases	22
Enterprise Information Systems	7
Data Warehouse	6
Shared Data Stored	5
Internet Delivery	4
Middleware	4
Exploiting Semantic Constraints	2
APIs	0
Brokered Objects	0
Collaborative Planning Forecasting	0
Data Exchange	0
Enterprise Data Management (Owners of Data)	0
Federated Databases	0
Internet Databases	0
Mediation	0
Not Exploiting Semantic Constraints	0
Objects-Oriented (O-O)	0
Other	0
Persistency	0
Quality Validations	0
Web Access	0

Total Weight = 5* Rate #1 + 4 * Rate #2 + 3 * Rate #3 + 2 * Rate #4 + 1 * Rate #5

Table 20 shows the respondents' identification of which information sharing strategies were selected based on the strength of cost effectiveness. Participants' responses as to their number one selection base on cost effectiveness vary. Some of the open-ended responses included: (1) "In the private sector, IT decisions are mostly driven by ROI and Cost." (2) "Anytime you can do more with less, its beneficial, however it's my opinion that you get what you pay for. Cost should not be spared when you are dealing with data and information that in the lifeline or your organization." (3) "Cost effectiveness is the ratio of the cost to produce the information over the value it has to the user." (4) "We try to look for solutions that are the most cost effective in the longer term.

Sometimes, project requirements dictate a faster, more costly solution but that should be justified by the cost benefit from delivering that system earlier." (5) "Pricing strategies generally reflect an organization's product maturity and supportability, and prices generally are set to be market competitive."

The Phase II interview analysis continues with evaluation of *Efficiency* as it relates to the organization's number one selection of information sharing strategies. An aggregation of interview question 7 information sharing strategies whose selection was based on cost effectiveness, are presented in Table 21.

Table 21. Efficiency's Impact on Selection of Strategies

(N = 15)	
INFORMATION SHARING SELECTION BASED ON EFFICIENCY	**FREQUENCY OF TOTAL WEIGHT**
Enterprise Databases	14
Enterprise Information Systems	8
Web Access	8
Data Warehouse	6
Exploiting Semantic Constraints	5
Shared Data Stored	5
Internet Delivery	3
Middleware	3
APIs	0
Brokered Objects	0
Collaborative Planning Forecasting	0
Data Exchange	0
Enterprise Data Management (Owners of Data)	0
Federated Databases	0
Internet Databases	0
Mediation	0
Not Exploiting Semantic Constraints	0
Objects-Oriented (O-O)	0
Other	0
Persistency	0
Quality Validations	0

Total Weight = 5 * Rate #1 + 4 * Rate #2 + 3 * Rate #3 + 2 * Rate#4 + 1* Rate #5

Table 21 shows the respondents' identification of which information sharing strategies were selected based on the strength of efficiency. Participants' responses as to their number one selection base on efficiency vary. A few of the open-ended responses include: (1) "Through efficiency company can save money, achieve high ROI, stay competitive and stay in business." (2) "We have previously identified that concept-based searching is important to overall efficiency; the value comes from the algorithms used to implement this capability along with its use within an architectural framework that exploits its capabilities. (3) Cost efficiency is the ration of the actual cost to produce the information over the potential lowest cost." (4) "While efficiency is not the most important, it ranks pretty high on the scale. You should store and display data in the most efficient means possible. If information and data are not efficient, it won't get used, and just takes up valuable storage space on your appliances."

The Phase II interview analysis concludes with evaluation of *Flexible Application Interfaces* as it relates to the organization's number one selection of information sharing strategies, an aggregation of interview question 7 regarding information sharing strategies whose selection was based on flexible application interfaces are presented in Table 22.

Table 22. Flexible Application Interfaces' Impact on Selection of Strategies

INFORMATION SHARING SELECTION BASED ON FLEXIBLE APPLICATION INTERFACES	FREQUENCY OF TOTAL WEIGHT
Web Access	11
Enterprise Databases	9
Data Warehouse	7
Enterprise Information Systems	7
Exploiting Semantic Constraints	5
Internet Delivery	4
Shared Data Stored	4
Middleware	2
APIs	0
Brokered Objects	0
Collaborative Planning Forecasting	0
Data Exchange	0
Enterprise Data Management (Owners of Data)	0
Federated Databases	0
Internet Databases	0
Mediation	0
Not Exploiting Semantic Constraints	0
Objects-Oriented (O-O)	0
Other	0
Persistency	0
Quality Validations	0

Total Weight = 5 * Rate #1 + 4 * Rate #2 + 3 * Rate #3 + 2 * Rate #4 + 1 * Rate #5

Table 22 shows the respondents' identification of which information sharing strategies were selected based on the strength of flexible application interface. Participants' responses as to their number one selection base on flexible application interface vary. A few of the open-ended responses include: (1) "If possible organizations should try to achieve standard application interfaces and if possible re-use. Many vendors today are serving ties with other software partners and this puts the burden of APIs to either the customer or third party integration firms. Non standard and inflexible interfaces create issues with efficiencies and effectiveness;" (2) "To me, this can be useful to a point, but if the interface is made properly the first time, it is not a critical factor. I would give it a 4 or 5, except that the interface does have an effect on the efficiency of the info system. A good interface makes it more efficient, a poor one makes less so;" (3) "Our newer applications are taking advantage of API's to help insulate the applications from database changes. Continuing to utilize large, central databases works well with this strategy" (4) "API's are important because if not generated and used effectively can render the data or information useless, there again taking up valuable storage space, and driving up costs.

Content Analysis and Affinity Diagram

Question 1 of the final Delphi interview asked the participants to list their top three information sharing hindrance, with one being the biggest. Question 2 of the final Delphi interview asked the participants to explain their reason for their number one selection.

The participants' open-ended questions provided a wealth of information. Thus, content analysis on the open-ended question yields applicable aspect of information sharing relative to organizational and functional processes. Processes were looked at because insight gained form the literature review and the participants indicated that both organizational and functional processes are the drivers for information sharing and the need for data integration.

For this research, the Affinity Diagram process is the tool used to organize the respondents' opinions into groups, based on a particular aspect of information sharing.

The first Affinity Diagram, Figure 2 captured the respondents' opinions about their organization's biggest hindrance to information sharing.

Impediment Aspects

Organizational	Functional
The more systems you must deal with, the more difficult it becomes to share the data due to individual policies on data sharing.	Joining large tables across multiple servers, especially when they are located in different data centers, is very inefficient.
Data Source across Multiple Systems	The ability to aggregate data and perform meaningful analysis
I think access to data is the hardest, especially with physically disparate users	Old Development Culture Lack of cohesive/holistic
Information sharing is a challenging issue for a decentralized organization	Too many times, there is data I need, but I do not have the right level of access.
Stovepipes remain the number one hindrance to information sharing.	The definition of "market" varied by application and in some cases by business units using the definition.
If concrete business rules are established up front, many of the other impediments may never	Management of Data Lack of Data Ownership
Client Server remains a significant problem due to the burden of having to have the client available on every desktop for specific applications.	…Users are often unable to explain business rules clearly or even to agree among themselves what the business rules are.

Figure 2. Affinity Diagram of Respondents' Aspects of Hindrances to Information Sharing

Questions 3 and 5 of the final Delphi interview asked the participants to list their top three successful and unsuccessful information sharing strategies, respectively, with one being the biggest. Questions 4 and 6 asked the participants to explain their reason for their number one selection. The next Affinity Diagram, Figure 3 captured the respondents' opinions about their organization's selection of its most successful and unsuccessful information sharing strategy.

Successful Aspects		**Unsuccessful Aspects**	
Organizational	**Functional**	**Organizational**	**Functional**
The consolidation of information to a standard database, that enforces business rules across the board, has been the most successful.	Web access because you are not restricted to internal network You can access and manipulate your database based on your permissions, anytime	We are currently setting the stage for collaborative work through organizational transformation and re-focusing	Middleware is most unsuccessful because it is usually developed by specialist, and may not be consistent across the board
Exploitation of semantic constraints… They define the concepts used by the organization.	Most reliable, accessible and fit for purpose	We need to improve our collaborative work.	The inability to share information across different tools
As the CIO, I am the data steward. Enterprise data management and quality validation	Impact of results gain from technical solutions	Middleware is not a good solution for the large databases that are at our company.	Owners of legacy systems strongly resist transition to the data warehouse.
If customers are able to get the data transparently, no matter the data is stored.	Utilizing the same database for multiple applications means no replication delay.	Mediation is unsuccessful; takes more time to implement,	Mediation – I am not a believer in organizations releasing control of their data.
Middleware tools have made it easier to connect disparate data sources.	Utilizing the same database for multiple applications means no inconsistent data	We cannot get to the right view of data because of excessive and lumpy	Collaborative planning has not worked – we cannot get to the right views of data…
Data warehouse by definition will allow information sharing when well designed.	The use of information systems over a database makes all the difference.	Security considerations make Internet delivery a very doubtful	Technology personnel come and go – change middleware or train the new person…
We exploit an enterprise data repository environment that can be evolved to remove redundancies and create source record data repositories…	Rather than just giving raw data, information systems compile it into a format that is useful information, rather than data that still needs to be processed.	Data warehouse and enterprise databases, consisting of structured data repositories, now represent only 20% of an enterprise's core stored and	Data warehouses usually wind up with too much information. It makes it difficult to extract what is needed, unless you know exactly what to look for.

Figure 3. Affinity Diagram of Respondents' Aspects of Successful and Unsuccessful Strategies

This research also looked at four desirable attributes relative to the desired information sharing infrastructure. Question 7 of the final Delphi interview asked the participants to rate each feature as it pertains to their organization's number one successful information sharing strategy, with five being the biggest. Accordingly, the next two Affinity Diagrams, Figures 4 and 5 captured the respondents' opinions about their organization's selection and the impact each of these attributes made in the selection of the organization's number one information sharing strategy.

Secure Information Sharing Aspects		Cost Effectiveness Aspects	
Organizational	**Functional**	**Organizational**	**Functional**
Secure means meeting the security specs of the enterprise. Usually this means that only the authorized users can have access to the info…	Too much unnecessary control can be a roadblock to share information and prevent an organization to maximize the ROI …	No one, except DoD, can afford to neglect the cost of doing business.	Cost should not be spared when you deal with data and information that is the lifeline of your organization
Security must be an integral part of any systems life cycle management.	If you cannot depend on data you get being reliable, you have no business getting it.	Cost is a major consideration in sharing data.	One must balance prioritization of work and limited resources.
Security must focus on data sensitivity and level of	Data integrity, reliability	In the private sector, IT decisions are mostly driven by ROI and Cost.	Cost effectiveness is the ratio of the cost to produce the information over the value it has to the user.
It is entirely possible to adopt secure information sharing within an architecture that includes our identified and selected technologies. This is important in enterprise application.	EIS allow access to multiple sources of data through the system, with security guaranteed by restricted access (user accounts).	By integrating multiple databases into one info system, the cost of maintaining user interfaces into each separate database is reduced/eliminated	Efficiency and effectiveness both play a role in capability delivery through focus mandatory requirements.

Figure 4. Affinity Diagram of Respondents' Aspects of Secure Information Sharing and Cost Effectiveness Qualities Which Impacts Information Sharing Strategy Selection

Efficiency Aspects		Flexible Application Interface Aspects	
Organizational	**Functional**	**Organizational**	**Functional**
Through efficiency company can save money, achieve high ROI, stay competitive and in business.	Cost efficiency is the ratio of the actual cost to produce the information over the potential lowest cost.	You must be able to adapt to your customer's needs through a flexible interface.	You must have flexibility when accessing data from potentially numerous sources.
Using centralized database on large database servers- data manageable	Timely, reliable, complete, accurate	API's insulate the applications from database changes.	The ability of using the interface in more than one setting.
Concept-based searching is important to overall efficiency.	The job can still be accomplished while doing it less efficient.	Easily supported, adaptable to business environment changes	Interface does not have an effect on the efficiency of the info system.
Efficiency is important, but not as much so as the previous two categories.	Less time spent looking for the right data, and info system format that immediately useful	This is essential to the evolution of capabilities for the Global Info Grid	If the interface is made properly the first time, it is not a critical factor.
Companies that are sharp will do things more efficiently than those that do not have the same talent or resources.	You should store and display data in the most efficient means possible. If information and data are not efficient, it will not get used, and just takes up valuable storage space on your appliances.	Two major components of overall IT cost are: Development Cost and Maintenance and Enhancement cost. Flexible API can reduce dev and M&E cost.	API's are important because if not generated and used effectively can render the data or information useless, thus taking up valuable storage space and driving up cost.

Figure 5. Affinity Diagram of Efficiency and Flexible Application Interfaces Qualities

Section V: Holistic Model of Successful Information Sharing Strategies

Figure 6, the Holistic Model, the final model derived from the research, illustrates a holistic model of successful information sharing strategies. The model was developed from the knowledge gained from the information obtained from the literature review, as well as the panel's responses. In addition to the participants' responses, the literature review provided insight of what was already known about recognized similarities and differences in strategic information management, information sharing strategies, and database integration strategies, information sharing and database integration.

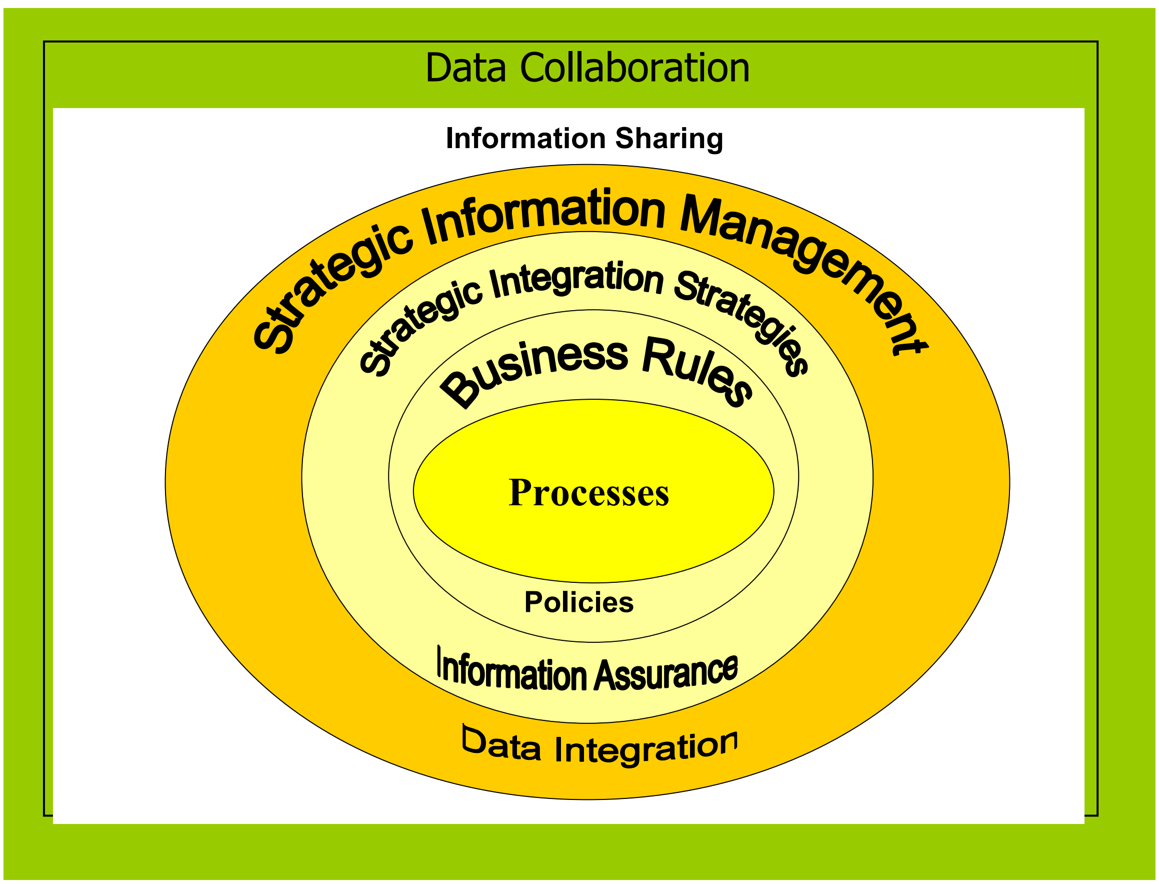

Figure 6. Holistic Model, Final Model Derived From the Research

The model illustrates the life cycle of information sharing, from cradle to grave. Data exist everywhere. Data is fact. Data collaboration is valuable when data is gathered and used to information. Data collaboration is the foundation of information sharing. Since information is a vital asset to an organization's survival, effective information sharing involves careful strategic information management. By its nature strategic implies long term future broad planning. Strategic information management includes a confluence of processes and resources, all of which relies on data integration. Processes and resources are only as strong as the foundation that supports them. Thus, strategic information management is beneficial to organizations that utilize sound information architecture and policies to support its processes and resources. Strategic integration strategies are information architecture used by successful organizations to support a variety of data integration needs. All the attributes of information assurance is available to assure the integrity of the information and ensure its availability.

In addition, as part of a successful organization's strategic information management, business rules are established to effect policies to support the organization's data, its information sharing practices, its processes, its resources, as well as its information architecture.

Summary

Phase I of the Delphi study included 19 participants. Their responses resulted in data gathering and data generation. Basically, during the first interview, the participants were asked to identify, their organization's information sharing impediments; successful information sharing strategies, unsuccessful information sharing strategies, currently used information sharing strategies; as well as the desirable features of their information sharing strategies: secure information sharing, cost effectiveness, efficiency, and flexible application interface, respectively.

In addition to identifying aspects of their organization's information sharing, instructions included adding to the existing list applicable strategy not listed. Phase I results were aggregated and new list and questions were formed.

Phase II of the Delphi study consisted of the new aggregated list generated from phase one. In addition to the new aggregated list, the second phase of the Delphi study included open-ended interview questions. Fifteen of respondents returned the second phase of the Delphi interview.

Finally, based on the results from Phases I & II, and the literature, the Delphi study group was able to create rank-ordered list to address the following information sharing and database integration areas:

- Hindrances to information sharing
- Successful information sharing strategies
- Unsuccessful information sharing strategies
- The impacts of secure information sharing, cost effectiveness, efficiency, and flexible application interface

In addition, from this research, a Holistic Model was derived. Chapter V will present a discussion of results, limitations of the research, and future recommendations.

V. Discussion, Limitations and Recommendations

Overview

Based on the IT-21 initiative, the aim of this research was to explore commercial strategies that offer integrated database systems and enterprise information systems alternatives that will identify and provide options for improved operational secure information sharing, cost effectiveness, efficiency, and flexible application interfaces for USMC heterogeneous databases. The design and attributes of established database integration strategies and enterprise information systems that may afford the United States Marine Corps flexibility during a crisis; as well as reliability, security, efficiency, and responsiveness was explored.

Research Questions

With this in mind, is there a commercial solution to meet the Marine Corps information sharing and database integration needs?

In order to answer the research problem, the following questions were investigated:

(1) What are the characteristics of the private sector's incompatible database?

(2) What data sharing strategies were developed in private industries?

(3) What data sharing strategies were successful in private industries?

(4) Under what circumstances was database integration successful in private organizations?

Discussion

The Delphi study began by identifying Impediments to information sharing. Question 1 of Phase I cited a list of possible information sharing impediments, drawn from the literature broken down into two categories: Past Impediments and Present Impediments. The participants were asked to identify items that were applicable to their organization. Table 7 illustrates the results; the Delphi group identified, *Access Control (Data Security),* Integrity of Store Information, and *Business Rules in Corporate Database, Incompatible Databases,* and *Database Interface Language* as the primary impediments to information sharing across their different databases.

Successful information sharing strategies were addressed in Question 2 of Phase I. During phase I, a list citing possible information sharing strategies was provided to the participants and it was broken down into four categories: *Successful, Unsuccessful, Not Used* and *Currently Used*. The participants were asked to identify items that were applicable to their organization. Tables 8 thru 10 illustrate the results.

Successful information sharing strategies are illustrated in Table 8. The Delphi group's identification of successful information sharing strategies include *Enterprise Databases, Enterprise Information Systems*, *Middleware, Collaborative Planning Forecasting and Replenishment, and Internet Databases* as the primary successful strategies to information sharing across their different databases.

Successful and *currently used* information sharing strategies are listed in Table 9. Table 9 illustrates the Delphi group identification of successful and currently used information sharing strategies. The group identified *Enterprise Databases, Enterprise Information Systems*, and *Middleware* as the primary successful and currently used strategies to information sharing across their different databases.

Successful and *not used* information sharing strategies are listed in Table 10. The Delphi group identification of successful and not used information sharing strategies include *Mediation, Federated Databases*, *Exploiting Semantic Constraints, Collaborative Planning Forecasting and Replenishment, Internet Databases, and Persistency* as the primary not used information sharing strategies across their different databases.

Four desired attributes necessary to enhance information sharing are addressed in Question 3. Thus, question 3 cited a list of possible information sharing strategies and four attributes: *Secure Information Sharing, Cost Effectiveness, Efficiency and Flexible Application Interfaces*. Relative to the four attributes, the participants were asked to characterize the strength of their organization's information sharing strategies. Thus, Tables 11 thru 14 illustrates the results.

Secure information sharing and its association with specific information sharing strategies are illustrated in Table 11. The Delphi group identification of the impact of secure information sharing on their organization various successful information sharing strategies. Across their different databases, the group identified E*nterprise Databases, Enterprise Information Systems*, and *Middleware* as the primary information sharing strategies selected based on secure information sharing.

Cost effectiveness and its association to specific information sharing strategies are illustrated in Table 12. The Delphi group identification of the impact of cost effectiveness on the organization various successful information sharing strategies. Across their different databases, the group identified E*nterprise Databases, Enterprise*

Information Systems, and *Internet Databases* as the primary information sharing strategies selected based on cost effectiveness.

Efficiency and its association to specific information sharing strategies are illustrated in Table 13. The Delphi group identified the impact of efficiency on their organizations' various successful information sharing strategies. Across their different databases, the group identified E*nterprise Information Systems, Collaboration Planning Forecasting and Replenishment*, and *Enterprise Databases* as the primary information sharing strategies selected based on efficiency.

Flexible application interface and its association to specific information sharing strategies are illustrated in Table 14. The Delphi group identified the impact of flexible application interfaces on their organizations' various successful information sharing strategies. Across their different databases, the group identified I*nternet Databases,* Enterprise Databases and *Enterprise Information Systems* as primary information sharing strategies selected based on flexible application interface.

Phase I results concludes with analysis of database models currently used in the private sector. Thus, Table 15 illustrates the database models utilized in today's industry. Relational database is the model primarily used.

The next sets of tables focus on the results obtained from phase two of the Delphi study.

Question 1 provided the aggregated list of information sharing impediments obtained from phase one of the study to the participants and asked them to identify and rank their organizations' top three hindrances to information sharing, with one being the biggest hindrance. The Delphi group identified *Data Source across Multiple Systems*, *Business Rules in Corporate Databases*, and *Access Control (Data Security)* as the primary hindrances to information sharing across their different databases.

Question 3 provided the aggregated list of successful information sharing strategies obtained from phase one of the study to the participants, and asked them to identify and rank their organizations' top three successful information sharing strategies, with one being the most successful, Table 17. The Delphi group identified, *Enterprise Databases*, *Middleware*, *Enterprise Information Systems, Data Warehouse, Web Access*, *Internet Delivery*, and *Shared Data Stored* as the primary successful information sharing strategies across their different databases.

Question 5 provided the aggregated list of unsuccessful information sharing strategies obtained from phase one of the study to the participants, and asked them to identify and rank their organizations' top three unsuccessful information sharing strategies, with one being the most unsuccessful, Table 18. Table 18 illustrates the results; the Delphi group identified, *Middleware, Federated Database, Data Warehouse*, and *Exploiting Semantic Constraint* as the primary unsuccessful information sharing strategies across their different databases.

Note that some items were identified as both being successful strategies and unsuccessful strategies.

Question 7 provided the four desired attributes discussed in phase I. The participants were asked to rate each of the attributes: *secure information sharing, cost effectiveness, efficiency,* and *flexible application interface*, from 1 thru 5, with five being the highest. They were instructed to take their organizations' number one information sharing strategy identified in question six. Into consideration and rate each attribute's selection; concerning how it affects the selection. The results are illustrated in Tables 19 thru 22. In addition, after each rating, the participants were asked to explain the rationale behind the rating assigned to each attribute.

Secure information sharing rating relative to each organization's number one sharing strategy is illustrated in Table 19. The Delphi group identified the impact of secure information sharing on their number one successful information sharing strategies across their different databases are *Web Access*, *Data Warehouse, and Enterprise Databases* as primary information sharing strategies selected based on secure information sharing.

Cost effectiveness ratings relative to each organization's number one information sharing strategy are illustrated in Table 20. The Delphi group identified the impact of cost effectiveness on their number one successful information sharing strategies across their different databases are *Enterprise Databases, Enterprise Information Systems, and Data Warehouse, Shared Data Stored, Internet Delivery, and Middleware* as primary information sharing strategies selected based on cost effectiveness.

Efficiency ratings relative to each organization's number one information sharing strategy are illustrated in Table 21. The Delphi group identified the impact of efficiency on their number one successful information sharing strategies across their different databases are *Enterprise Databases, Enterprise Information Systems, Web Access, Data Warehouse, Exploiting Semantic Constraints, and Shared Data Stored* as primary information sharing strategies selected based on efficiency.

Flexible application interface relative to each organization's number one information sharing strategy are illustrated in Table 22. The Delphi group identified the impact of flexible application interfaces on their number one successful information sharing strategies across their different databases, the group identified *Web Access, Enterprise Databases, Data Warehouse, Enterprise Information Systems, Exploiting Semantic Constraints, Internet Delivery and Shared Data Stored as* primary information sharing strategies selected based on flexible application interface.

The above information may prove useful to USMC information system professionals as they strive to devise strategies for better sharing of information across disparate USMC databases. As they do this, they will need to assess the similarities and differences between USMC databases and operational needs versus those of the participants in the Delphi group

Limitations of the research

The research focused on alternatives to generalized information sharing strategies and data integration solutions. The literature review and the Delphi participants' responses addressed the critical concepts that contributed to the derived model. Thus, the limitations of this research were:

(1) Delphi studies are limited by the expertise of the group members.

(2) Difficulty in gaining participation from CIOs; which may contribute to variability in group expertise.

(3) Some members were reluctant to share information that they saw as proprietary.

Recommendation for Future Study

A recommendation for future study includes assessing specific USMC needs to see which strategies identified here might serve USMC purposes. The emphasis could be

(1) *How are the Marine Corps' problems specifically similar to incompatible database problems as it relates to private industries?*

(2) *How well the strategies would identify here work for the Marine Corps database sharing?*

Additionally, research opportunities exist to develop the model to fully satisfy the Marine Corps' collaboration requirements as well as its information sharing and database integration needs.

Conclusions

This study used a Delphi group technique to identify measures identified in the civilian world to enhance information sharing across different databases. The members of the Delphi group were able to identify and rank items that they found to be hindrances to information sharing as well as those items that they found to be successful in enhancing information sharing. They were also able to identify and rank items to help them enhance security, control costs, increase efficiency and make use of flexible application interfaces, all while increasing information sharing. These rank-ordered lists may provide useful information to the USMC as it strives to improve information sharing across its various information systems.

Appendix A - Phase I Data Collection List and Interview Questions

DATA INTEGRATION/INFORMATION SHARING
1st Round

Delphi Interview to Private Sector's Chief Information Officer

1. Which of these data integration and information sharing incompatibilities do you believe characterize past and present impediments your company has encountered?

	Past Impediment	Present Impediment
Access Control (Data Security)		
Business Rules in Corporate Database		
Concurrency Issues		
Client/Server		
Database Interface Language		
Granularity of context		
Incompatible databases		
Incompatible data format		
Integrity of stored information		
Other ___________________		

2. How would you characterize the information sharing strategies your organization has tried?

	Successful	Unsuccessful	Not Used	Currently Used
Exploiting semantic constraints				
Middleware				
Enterprise databases				
Enterprise Information Systems				
Federated databases				
Mediation				
Collaborative Planning Forecasting and Replenishment				
Persistency				
Internet databases				
Other ______________				

3. Of the strategies currently used by your organization, how would you characterize the strength of the information sharing strategies in the indicated four areas?

Information Sharing Strategies	Secure Information Sharing	Cost Effectiveness	Efficiency	Flexible Application Interfaces
Exploiting semantic constraints				
Middleware				
Enterprise databases				
Enterprise Information Systems				
Federated databases				
Mediation				
Collaborative Planning Forecasting and Replenishment				
Persistency				
Internet databases				
Other ____________________				

4. Please select the type(s) of database model(s) utilize in your organization:

___ Flat file

___ Hierarchical

___ Network

___ Relational

___ Object-Relational

___ Object Oriented

5. Title: _______

Name: _______

Organization: ____________________

Appendix B: Phase II Data Collection List and Interview Questions

DATA INTEGRATION/INFORMATION SHARING
Phase II

Delphi Interview to Private Sector's Chief Information Officer

1. In terms of data integration and information sharing, out of all the impediments, which of the following incompatibilities would you classify to be the most hindrance? **Please rank your <u>top three</u> (3) in order of priority** (1 – the most hindrance).

	Top 3
	Ranking of Hindrance
Access Control (Data Security)	
Business Rules in Corporate Database	
Concurrency Issues	
Client/Server	
Database Interface Language	
Granularity of context	
Incompatible databases	
Incompatible data format	
Integrity of stored information	
Reliability of data	
Data Source across multiple systems	
Exploiting semantic constraints	
Incompatible Data Definition Format	
Data Management (Lack of Ownership)	
Not Exploiting Semantic Constraints	

2. What is your reason for selection of the number one information sharing impediment? (Briefly explain.)

3. Which of the following information sharing strategies would you characterize as most successful? **Please rank your top three (3) in order of priority** (1 – the most successful).

Strategic Enablers of information Sharing	Top 3 Ranking of successful strategies
Exploiting semantic constraints	
Middleware	
Enterprise databases	
Enterprise Information Systems	
Federated databases	
Mediation	
Collaborative Planning Forecasting and Replenishment	
Persistency	
Internet databases	
APIs	
Data warehouse	
Internet Delivery	
Objects (O-O)	
Shared Data Store	
Brokered Objects	
Data Exchange	
Web Access	
Quality Validations	
Enterprise Data Management (Owners of Data)	
Other:	

4. What characterize your selection as a success? (Please explain).

5. Which of the following information sharing strategies would you characterize as most unsuccessful? **Please rank your top three (3) in order of priority** (1 – the most unsuccessful).

Constraint of strategic information sharing	**Ranking of unsuccessful strategies**
	Unsuccessful
Exploiting semantic constraints	
Middleware	
Enterprise databases	
Enterprise Information Systems	
Federated databases	
Mediation	
Collaborative Planning Forecasting and Replenishment	
Persistency	
Internet databases	
APIs	
Data warehouse	
Internet Delivery	
Other:	

6. What characterize your selection as unsuccessful? (Please explain).

7. On a scale of 1 – 5, how did each of the following features below impact your selection of your information sharing strategy selection? (1 – No impact).

Features	Ranking Scale 1 thru 5
Secure Information Sharing	

a) How do you characterize "Secure Information Sharing" contributions towards the successful strategy you selected being most successful? (Please explain).

Features	Ranking Scale 1 thru 5
Cost Effectiveness	

b) How do you characterize "Cost Effectiveness" contributions towards the successful strategy you selected being most successful? (Please explain).

Features	Ranking Scale 1 thru 5
Efficiency	

c) How do you characterize "Efficiency" contributions towards the successful strategy you selected being most successful? (Please explain).

Features	Ranking Scale 1 thru 5
Flexible Application Interfaces	

d) How do you characterize “Flexible Application Interface” contributions towards the successful strategy you selected being most successful? (Please explain).

8. Thank you very much for your participation.

Bibliography

Acree Jr., R. G. and Money, W. H. (1995). "Breaking down the stovepipes." *Program Manager* May/Jun95, Vol. 24 Iss. 3, 13, 8p, 1bw

Ambler, S. W. (1998). "Data store system crisis coming soon; Organizations must start to realize that data is only half of the systems integration picture." *Computing Canada.* March 9, 1998.

American Heritage Dictionary of the English Language. Fifth Edition. Copyright 2000 by Houghto Mifflin Company. Published by Houghton Mifflin Company. All rights reserved. http://dictionary.reference.com/search?q=information&r=67

Anderson, R. J. (2001). "Multilevel Security." Security Engineering: A Guide to Building Dependable Distributed Systems. Wiley Computer Publishing. John Wiley & Sons, Inc., New York, NY 2001.

Anonymous. (2003). "Resources: Defining terms." *SDM.* Troy: Sep 2003. Vol. 33, Iss. 9; pg. 57.

Anthes, G. H. (2002). "Bridging data islands." Computerworld. Farmingham: Oct 14, 2002. Vol. 36, Iss. 42; pg. 23, 2 pgs.

Arens et al. (1993). "Retrieving And Integrating Data From Multiple Information Sources." *International Journal on Intelligent and Cooperative Information Systems*, 2(2):127 - - 158, 1993.

Asker, J. R. (2002). "Stovepipe Effect." *Aviation Week & Space Technology.* New York: March 25, 2002. Vol. 156, Iss. 12; pg. 27

Beaty, E. W. (1988). "How to Plan For the Unexpected." *Canadian Datasystems*. Toronto: May 1988. Vol. 20, Iss. 5; pg. 36, 2 pgs.

Bernstein, P. A. (1996). "Middleware: A Model for Distributed System Services." Association for Computing Machinery. Communications of the ACM. New York: Feb 1996. Vol. 39, Iss. 2; pg. 86, 12 pgs.

Boodro, C. J. (2003). "Web Technology pays off for Insurers." National Underwriter: (Property & casualty/Risk & Benefits Management ed.) Erlanger: Aug 25, 2003. Vol. 107, Iss. 34; pg.18.

Brassard, M. (1989). “Affinity Diagram.” The Memory Jogger Plus+ Featuring the Seven Management and Planning Tools. Basic Tools for Process Improvement. Module 4, pp. 1 – 12. Methuen, MA: Goal/QPC.

Brady, F. X. (2003). “USMC Shared Data Environment (SDE) Initiative.” MARADMIN 187/03 R221200Z APR 03. CMC WASHINGTON DC

Brady, F. (2003). "Marine Corps Enterprise Information Technology Services (MCEITS) Engineering Support." *MARADMIN* 568/03 R091700Z DEC 03. CMC WASHINGTON DC

Caeli, W. et al. (1991). Information Security Handbook. London: Macmillan, 1991.

Center for Industrial Technology and Enterprise – 1990. http://www.cite.edu.ph//

Chen, Y. and et al. (1992). "Visualizing Business Rules in Corporate Databases." *Industrial Management and Data Systems.* Vol. 92, Iss. 7, Copyright MCB University Press Limited 1992. Wembley 1992

CJCSI 6212.01B: Interoperability and Supportability of National Security Systems, and Information Technology Systems, Chairman of the Joint Chiefs of Staff, 8 May 2000.

Cleary, M. (1991). “Foundation of the 6555th: The Post War Legacy.” *The 6555th* Missile and Space Launches Through 1970. 45th Space Wing History Office. https://www.patrick.af.mil/heritage/6555th/6555ch1-1htm

Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) *Architecture Framework Version 2.0*, Department of Defense, 18 Dec. 1997.

Codd, E. F. “Mastering the Art of Database Fusion – Part II.” Datamation. Barrington: Nov 1, 1990. Vol. 36, Iss. 21; pg. 81, 2 pgs.

CompInfo “The Computer Information Center.” http://www.compinfo-center.com/tpeis-t.htm

Crook, T. R. et al (2003). "Competitive edge: A strategic management model" Cornell Hotel and Restaurant Administration Quarterly. Ithaca. Jun 2003. Vol. 44, Iss, 3, pg. 44.

Custer, R. L. and et. Al. (1999). “The Modified Delphi Technique – A Rotational Modification.” Journal of Vocational and Technical Education. Digital Library and Archives. Vol. 15, No. 2. http://scholar.lib.vt.edu/ejournals/JVTE/v15n2/custer.html.

Dalkey, N. C. (1972). The Delphi method: an experimental application of group opinion. In N. C. Dalkey, D. L. Rourke, R. Lewis, & D. Snyder (Eds.) Studies in the quality of life. Lexington, MA: Lexington Books.

Defense Information Infrastructure (DII) Common Operating Environment (COE) Integration and Runtime Specification (I&RTS) Version 4.1, Department of Defense, 3 Oct. 2000.

Deogun, J. and Raghavan, V. (1988). "Integration of Information Retrieval and Database Management Systems." *Information Processing & Management*. Oxford: 1988. Vol. 24, Iss. 3; pg. 303, 11 pgs

Department of Defense. (1985). “Trusted Computer System Evaluation Criteria.” DoD 5200.28-STD

Dictionary.com. http://dictionary.reference.com.

DISA “Migrating Stovepipe Systems to Integrated/Interoperable Platforms Using the Technical Reference Model and Object Oriented Operational Architectures.” http://trm.disa.mil/CaseStudy8_NORAD.pdf.

DISA “Migrating Stovepipe Systems to Integrated/Interoperable Platforms Using the *Technical Reference Model and Object Oriented Operational Architectures*.” http://trm.disa.mil/CaseStudy8_NORAD.pdf.

Dori, D. (2003). “Conceptual modeling and system architecting.” *Association for Computing Machinery.* Communications of the ACM. New York: Oct 2003. Vol. 46, Iss. 10, pg. 62.

Electronic Privacy Information Center. http://www.epic.org/ and http://www.privacyrights.org/ar/RFIDposition.htm

Evernden and Evernden. (2003). "Architecting organizational changes." *Information First*: Integrating Knowledge and Information Architecture for Business Advantage. Elsevier Butterworth-Heinemann, Linacre House, Jordan Hill, Oxford OX2 8DP, 200 Wheeler Road, Burlington, MA 01803. pg. 5.

Ibid. pg. 57.

Fabio, C. (2003). "Combination of similarity measures for effective spoken document retrieval." *Journal of Information Science*. Amsterdam: 2003. Vol. 29, Iss. 2; pg. 87.

Finkelstein, R. (1993). "Laying the groundwork." Network World. Framingham: Apr 12, 1993. Vol. 10, Iss. 15; pg. S13, 3 pgs

Flater, D. (2003). "A Logical Model of Conceptual Integrity in Data Integration." *Journal of Research of the National Institute of Standards and Technology*. Vol. 108; No. 5, 395-402, Sep-Oct 2003

Feng, S. C. (2003). "A machine process planning activity model for systems integration." *Journal of Intelligent Manufacturing*. London: Dec 2003. Vol. 14, Iss. 6; pg. 527.

Foster, I. And Grossman, R. (2003). "Data integration in a bandwidth-rich world." *Association for Computing Machinery.* Communications of the ACM. New York: Nov 2003. Vol. 46, Iss. 11; pg. 50.

Gallagher, Jr., P. R. (1991) "Trusted Database Management System Interpretation." NCSC-TG-021. Library No. S235, 625. National Computer Security Center, April 1991.

Groff , J. R. and Weinberg, P. N. (2002). SQL Second Edition: The Complete Reference. McGraw-Hill/Osborne, Berkeley, CA 94710.

Haas, L. M. et al. (2002). "Data Integration through Database Federation." IBM Systems

Journal. Armonk: 2002. Vol. 41, Iss. 4; pg. 578 19 pgs.

Hammer, M. (2001). “Prepare for a Future You Cannot Predict: Institutionalize a Capacity for Change.” The Agenda: What every business must do to dominate the decade. 1st Ed. Crown Business, New York. Oct 2001. Ch. 12, pgs. 243-263.

Hasselbring, W. (2000). “Information System Integration.” Association for Computing Machinery. Communications of the ACM. New York: Jun 2000. Vol. 43, Iss. 6; pg. 32, 7 pgs.

Hay, R. D. (1969). “Use of modified semantic differentials to evaluate formal organization structure.” Academy of Management Journal. (pre-1986); Jun 1969; 12; 000002; ABI/INFORM Research; pg. 247.

Heminger, A.R. and Robertson, S. B. (1998). “Digital Rosetta Stone: A Conceptual Model for Maintaining Long-Term Access to Digital Documents.” IEEE Thirty-First Annual Hawaii International Conference on System Sciences. Vol. 2, p. 158, January 06-09, 1998, Kohala Coast, Hawaii.

Herman, J. (2002). "Making collaborative commerce happen." Business Communication Review. Hinsdale. Oct 2002. Vol. 32. Iss. 10; pg 18, 2 pgs

Her-Kun, C. and Shvan-Ming,Y. (2000). “Optimal binary vote assignment for replicated data. The Journal of Systems and Software. New York: Jul 15, 2000. Vol. 53, Iss. 1, pg 73.

Hovy, E. (2003). “Using an Ontology to Simplify Data Access.” Association for Computing Machinery. Communications of the ACM. New York: Jan 2003. Vol. 46, Iss. 1; pg 47.

Humby, C. (1996). “Digging for information.” Marketing. London: Nov 21, 1996. pg. 41, 2 pgs.

Hunt, J. (1999). “Will your backups still be there when you need them? Fend off data degradation.” Quality. Wheaton: May 1999. Vol. 38, Iss. 6; pg. 52.

Hoffer, J. A. et al. (2002) Modern Database Management. 6th Ed. Prentice Hall. Upper Saddle River. New Jersey, 07458. 2002.

Iansiti, M. (2002). Technology Integration: Making Critical Choices in a Dynamic World.

Iansiti, M. (2003). "Integrating the Right Way, the Wrong Way; Harvard's Marco Iansiti, who has studied the technology strategies of nearly 100 companies, says: Keep your integration expertise in-house." CIO. Framingham: May 15, 2003. Vol. 16, Iss. 15; pg. 1

Joint Technical Architecture Version 4.0, Department of Defense, 2 April 2001.

Jones, C. G. (1975). "A Delphi Evaluation of Agreement between Organizations." In The Delphi Method: Techniques and applications. Edited by Harold A. Linstone and Murray Turoff. Addison-Wesley Publishing Company, Inc. Advanced Book Program. III. B. 4. pgs. 160 – 167.

Kay, R. (2003). Computerworld. Farmingham: Apr 14, 2003. Vol. 37, Iss. 15, pg 44 (1 pgs).

Kerner, J. (2001). "Joint Technical Architecture Impact on Department of Defense Programs." CrossTalk: The Journal of Defense Software Engineering. http://www.stsc.hill.af.mil/crosstalk/2001/10/kerner.html

Kerr, J. (1991). "The IRM Imperative: Strategies for Managing Information Resources." IBM Systems Journal, 1992, 31, 1, New York, 1991. pp 282.

Khan, L et al. (2001). "An Adaptive Probe-based Technique to Optimize Join Queries in Distributed Internet Databases." Journal of Datbase Management. Hershey: Oct-Dec 2001. Vol. 12, Iss. 4; pg. 3, 12 pgs.

Kohler, J. and Schule-Kremer, S. (2002). "The Semantic Metadatabase (SEMEDA): Ontology Based Integration of Federated Molecular Biological Data Sources." IOS Press and Bioinformation Systems. In Silico Biol. 2, 020319 http://www.bioinfo.de/isb/2002/02/0021/>19 March 2002.

Kutler, J. (2003). "Transforming the Board Room Along With the Battlefield." Aviation Week & Space Technology. New York: May 19, 2003. Vol. 158, Iss. 20; pg. 70.

Kohler, J., Lange, M. Philippi, S. Hosfestadt, R. and Schulze-Kremer, S. ( ). "Collaborative Editing of Large Ontologies in SEMEDA." projects.villa-bosch.de/sdbv/people/luca_bernardi/ bioontology/Poster/Koehler.pdf

Lam, K. et al. (2003). "Security Middleware for Enhancing Interoperability of Public Key Infrastructure." Computer & Security. Amsterdam: 2003. Vol. 22, Iss. 6; pg. 535.

Langemo, M. (1988). "An Introduction to Information Resource Management" Records Management Quarterly. Automation-Integration Technologies. ARMA Quarterly. October 1988.

Lin, H., Risch, T. and Katchaounov, T. (2001) "Adaptive Data Mediation over XML Data" To be published in special issue on "Web Information Systems Applications" of Journal of Applied System Studies (JASS), Cambridge International Science Publishing, 2001.

Ling, T. et al. (1996). "Extending classical functional dependencies for physical database design." Information and Software Technology. Guildford. Sep 1996. Vol. 38, Iss. 9, pg. 601, 8 pgs.

Love, R. E. and Collenbore, G. W. (2001). "Integrated Logistics Capability: An imperative for Change." Marine Corps Gazette. June 2001. Vol. 85 Is. 6, p. 16, 4p, 2 diagrams.

Mattison, R. (1996). "Warehousing wherewithal." *CIO*. Framingham: Apr 1, 1996. Vol. 9, Iss. 12, pg. 58, 4 pgs.

McClave et al. (2001). "Methods for Describing Sets of Data." STATISTICS for Business and Economics. 8th Ed. Prentice Hall. Upper Saddle River, NJ 07458, pg 30.

McKnight, W. (2002). "What Is Information Assurance?" CrossTalk, the Journal of Defense Software Engineering. OO-ALC/MASE 6022 Fir Ave, Hill AFB, UT 84056-5820. http://www.stsc.hill.af.mil/crosstalk/2002/07/index.html

Microsoft Developer Network (MSDN) http://msdn.microsoft.com/archive/default.asp?url=/archive/en-us/dnarsqlsg/html/msdn_qryoptim.asp

Myburgh, S. (1998). "Strategic Information Management: Understanding A New Reality." Information Management Journal. Prairie Village. Jan/Feb 2002. Vol 36, 5 pgs

National Computer Security Center. (1991). "Trusted Database Management System Interpretation." NCSC-TG-021. Library No. S235,625

Naval Studies Board. (1997). "Technology for Future Naval Forces: Technologies for Enterprise Processes." Technology for the United States Navy and Marine Corps 2000-2035: Becoming a 21st Century Force. Vol 2: Chap 10:. Naval Studies Board. National Research Council. Commission on Physical Sciences, Mathematics, and Applications. Copyright 1997 by National Academy of Sciences. All rights reserved. http://www.nap.edu/html/tech_21st/tfnf.htm

.
NCSC. (1988). "Glossary of Computer Security Terms." National Computer Security Center. NCSC-TG-004-88. Oct. 1988.

Neuendorf, K. A. (2002). "Contexts." The Content Analysis Guidebook. Sage Publications. International Educational and Professional Publisher. 2455 Teller Road, Thousand Oaks, CA 91320, Ch 9.

Ngwenyama, O. K. and Bryson, N. (1997). "Eliciting and Mapping Qualitative Preferences to Numeric Rankings in Group Decision Making." European Journal of Operational Research 116 (1999) 487 – 497.

Nishiyama, T. (1998). "Using a 'Process Warehouse' Concept: A Practical Method for

Successful Techlogy Transfer." *Second IEEE International Symposium on Object-Oriented Real-Time Distributed Computing*. May 02 - 05, 1999. Saint-Malo, France. Copyright (c) 1998 Institute of Electrical and Electronics Engineers, Inc. All rights reserved.

NIST. (1995). National Information Infrastructure Protection Act of 1996. H.R.3723. The NIST Handbook. NIST Special Publication 800-12. An Introduction to Computer Security. One Hundreds Fourth Congress of the United States of America AT THE SECOND SESSION. 1995.

NMCI. http://www.nmci.navy.mil/Primary_Areas/Policy/Index.htm

Papazoglou, M. and Valder, W. (1989). Relational Database Management: A Systems Programming Approach. Prentice Hall Inc., Englewood Cliffs, NJ 07632.

Park, K. H. and Favrel, J. (1999). "Virtual enterprise – Information system and networking solution." Computers & Industrial Engineering. New York. Oct. 1999. Vol. 37, Iss. 1, 2, pg. 441, 4 pgs.

Paxson, V. and Floyd, S. (1997). "Why we don't know how to simulate the Internet." Winter Simulation Conference Proceedings of the 29th conference on Winter simulation. Sponsors: IEEE-CS: Computer Society. Atlanta, Georgia, United States. Pages: 1037 – 1044. Year of Publication: 1997. ISBN

Plant, T. (1996). "Whose Information is it Anyway?" An Argument for Information Stewardship." Air Force Institute of Technology. Wright-Patterson AFB.

Pravica, D. (1999). "Internet Databases: Next-millennium Flexibility." Computing Canada. Willowdale: Jan 29, 1999. Vol. 25, Iss. 4; pg. 27, 1 pgs.

Rahfaldt, K. (2003). "Access Control: To function as one." SDM. Troy: Sep 2003. Vol. 33, Iss. 9; pg. 50.

Ranade, D. M. (2002). Shared Data Clusters: Scalable, Manageable, and Highly Available Systems (Veritas Series). Wiley Publishing, Inc. Ch 6, ppg 122-125.

Rubin, J. R. and Rubin, I. S. (1995). Qualitative Interviewing: The Art of Hearing Data. Thousand Oaks, CA. SAGE Publications, Inc.

Renner, S. A. et al. (1996). "Data Interoperability: Standardization or Mediation." The MITRE Corporation. Copyright 1996 IEEE. http://www.computer.org/conferences/meta96/renner/data-interop.html

SANS Institute. http://www.sans.org/resources/glossary.php

Scheele, D. S. (1975). "Reality Construction as a Product of Delphi Interaction." The Delphi Method: Techniques and applications. Addison-Wesley Publishing Company, Inc. Advanced Book Program. Reading Massachusetts. 01867. pgs 37-69.

Scheibe, M. et al. (1975). "Experiments in Delphi Methodology." The Delphi Method: Techniques and applications. Addison-Wesley Publishing Company, Inc. Advanced Book Program. Reading Massachusetts. 01867. pgs. 262 – 290.

Sheth, A. and Larson, J. (1990). "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases." ACM Computing, Springer Verlag (Heidelberg, FRG and New York NY, USA)-Verlag Surveys, ; ACM CR 91070567, 1990.

Sidney Jr., H. (2003). "Making the right connections." MSI. Oak Brook: Sep 2003. Vol. 21, Iss. 9; pg. 50

.
Silberschatz et al. (1996). "Data models." ACM Computing Surveys. Baltimore: Mar 1996. Vol. 28, Iss. 1, pg. 105, 4 pgs.

Smith, G. P. (1994). Motivation. In W. Tracey (ed), Human resources management and development handbook (2nd ed).

Sokol, L. (2002). "Creating knowledge from heterogeneous data stovepipes." *Proceedings of the Fifth International Conference of Information Fusion*. FUSION 2002. IEEE Cat No. 02EX5997. (vol 2) 1162-7, 2002 Annapolis, MD, USA, 8-11 July 2002.

Spery, R. (1999). "Geographic Information Spans Borders." *Government Executive.* Washington. May 1999. Vol 31. Iss. 5, pg. 73, 3 pgs.

Stallings, W. (2001). "Business Information." *Business Data Communications.* Fourth Edition. Prentice Hall. Upper-Saddle River, NJ 07458.

Sushil, J. (1996). "Database Security and Privacy" *AMC* Computing Surveys. Association for Computing Machinery. Baltimore: Mar 1996. Vol 26. Iss: 1, pg. 126, 3 pgs

Takashi K. and et al. (2003). "Business process integration as a solution to the implementation of supply chain management systems." *Information & Management.* Amsterdam: Sep 2003. Vol. 40, Iss. 8: pg 769.

Thede, L. Q. "Databases Demystified." Nursing Management. Chicago: Jul. Vol. 22, Iss. 7; pg. 38, 5 pgs

Vesely, E. G. (1990). Strategic Data Management: The Key to Corporate Competitiveness. Yourdon Press Computing Series. Prentice-Hall Inc., Englewood Cliffs, New Jersey 07632

Wikipedia. (2002). "Secure Computing." http://www.fact-index.com/s/se/secure_computing.html

Winstlett, M. (1987). ""Updating Databases with Incomplete Information" STANFORD UNIV CA DEPT OF COMPUTER SCIENCE.

Ye N., and et al. (2001). "A process control approach to cyber attack detection *Association for Computing Machinery*." Communications of the ACM New York Aug 2001

Zachman, J. A. (1999). "A Framework for Information Systems Architecture." IBM

Systems Journal. Vol 38, NOS 2&3, 1999.

Zhou, W. and Zhou, M. (1995). "The Desighn and Application of Sensors Implemented in Java." School of Computing and Mathematics, Deakin Universith, Clayton, VIC 3168, Australia. Blue Cross/BlueShield, Information Systems, 730 Chestnut St., Chattanooga, TN 37402, USA. http://www3.cm.deakin.edu.au/~wanlei/activeai99.pdf

Vita

Gunnery Sergeant Pamela J. Cole from Brooklyn, New York, graduated from Washington Irving High School, New York City in June 1976. She enlisted in the United States Marine Corps in January 1984 as a supply logistic clerk. In 1990, while stationed at Camp Pendleton, California, she earned he under graduate degree in Business Administration from National University.

During Gunnery Sergeant Cole's enlistment, she was stationed at 3rd Supply, Battalion, 3rd Force Service Support Group, Okinawa, Japan. Depot Supply and Women's Recruit Training Battalion, Parris Island, South Carolina; Division Supply, 1st Marine Division, Camp Pendleton, California; Assistant Chief of Staff, G-4, 1st Marine Expeditionary Brigade, Kaneohe, Hawaii; I&I Staff Syracuse, New York; 3rd Material Readiness Battalion, 3rd Force Service Support Group, Okinawa, Japan; Supply School, Marine Corps Service Support Schools, Camp Johnson, North Carolina.

In August 2002, she entered the Graduate Information Resource Management program, School of Engineering and Management, Air Force Institute of Technology at Wright-Patterson AFB, Ohio. Upon Graduation, will be assigned to the Assistant Chief of Staff, G-6, II Marine Expeditionary Force at Camp Lejeune, North Carolina.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
23-03-2004	**Master's Thesis**	Aug 2001 – Mar 2004

4. TITLE AND SUBTITLE

THE UNITED STATES MARINE CORPS
DATA COLLABORATION REQUIREMENTS:
RETRIEVING AND INTEGRATING DATA FROM MULTIPLE DATABASES

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Cole, Pamela J., Gunnery Sergeant, USMC

5d. PROJECT NUMBER
If funded, enter ENR #

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Building 641
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GIR/ENV/04M-04

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
HQMC C4 CP/IA
Attn: Mr. Ray A. Letteer
LFF-1
2 NAVY ANNEX
WASHINGTON DC 20380-1775 703-693-3490/3491 DSN 223 letteerra@hqmc.usmc.mil

N/A)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of this research was to develop a model and suggest a framework to fully satisfy the United States Marine Corps' collaboration requirements; as well as its information-sharing and database integration needs. This research is exploratory; it focuses on only one initiative: IT-21 initiative. IT-21 initiative dictates *The Technology for the United States Navy and Marine Corps, 2000-2035: Becoming a 21st Century Force*. IT-21 initiative states that its information infrastructure will be based largely on commercial systems and services, and the Department of the Navy must ensure that these systems are seamlessly integrated and that the information transported over the infrastructure is protected and secure. . A qualitative method approach was used to develop a model and suggest a framework for information-sharing and database integration. The Delphi Technique was used for data collection, primarily form mid-level to senior information officers, with a focus on Chief Information Officers. Literature review was conducted to gain insight of what was already known from recognized similarities and differences in Strategic Information Management, Information-sharing strategies, and database integration strategies Information-sharing and database integration. It is the hope that the Armed Forces and the Department of Defense overall may benefit from future development of the model.

15. SUBJECT TERMS
Information-sharing, Databases, Strategic Information Management, Data Management, Semantic and Information Retrieval, Secure Information Sharing, Integrity of Stored Information, Integrating Databases, Business Integration, System Integration Strategies, and Information Systems Integration.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
REPORT	ABSTRACT	c. THIS PAGE			Alan R. Heminger, Ph.D (EN)
U	U	U	UU	138	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, X 4097 e-mail: Alan.Heminger@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18