Data Management Challenges and Development for Military Information Systems

Marion G. Ceruti, *Senior Member*, *IEEE*

Abstract—This paper explores challenges facing information system professionals in the management of data and knowledge in the Department of Defense (DOD), particularly in the information systems utilized to support Command, Control, Communications, Computers, and Intelligence (C^4I). These information systems include operational tactical systems, decision-support systems, modeling and simulation systems, and nontactical business systems, all of which affect the design, operation, interoperation, and application of C^4I systems. Specific topics include issues in integration and interoperability, joint standards, data access, data aggregation, information system component reuse, and legacy systems. Broad technological trends, as well as the use of specific developing technologies are discussed in light of how they may enable the DOD to meet the present and future information-management challenges.

Index Terms—Command and control, data access, data aggregation, data and knowledge management, data mining, integration and interoperability, military information system, network-centric warfare, software reuse, standards.

1 Introduction

Some of the most significant challenges today in the US Department of Defense (DOD) are in the design, integration, upgrade, and maintenance of information systems, particularly for Command, Control, Communications, Computers, and Intelligence (C^4I) [11]. As the tactical emphasis in the DOD shifts from platform-centric toward network-centric warfare, the issues in information systems integration, interoperability of new and legacy systems, data mining, aggregation, standardization, and reuse become more important. Similarly, the need for efficient, cost effective, technical solutions becomes more urgent.

DOD laboratories and agencies are aware of the problems in information systems and are engaged in programs at various levels. For example, at the joint level, the Defense Information Systems Agency (DISA) promotes engineering practices aimed at sharing data among the services [34]. At the service-specific level, the Navy has designated the Space and Naval Warfare Systems Center, San Diego, (SSC SD) as the Navy's lead laboratory in command, control, and ocean surveillance. SSC SD has focused for many years on a wide variety of research, development, test, and evaluation programs in command and control. (See, for example, [24] and [41].) This paper describes major information-management issues in the DOD using Naval and Joint systems and programs as examples. It describes technology the DOD is developing and using to address information system challenges.

This paper is organized as follows: Section 2 describes standards and reuse of joint information systems. Section 3 addresses issues in integration and interoperability. Section 4 describes problems and guidelines associated with data access. Section 5 discusses network-centric warfare. Section 6 describes challenges in data aggregation. Section 7 addresses new technologies and research in military information systems. Section 8 suggests directions for new research and DOD systems. The paper concludes with Section 9.

• *The author is with the Space and Naval Warfare Systems Center, San Diego, Code D4121, 53560 Hull Street, San Diego, CA 92152-5001. E-mail: ceruti@spawar.navy.mil.*

Manuscript received 10 July 2000; revised 2 Jan. 2001; accepted 25 July 2001. For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number 112420.

2 Joint Information System Standards and Reuse

2.1 Standards

The autonomy of the component systems in a federation and the cooperation between them are conflicting goals. A balance between these goals is necessary [43]. This is also the case in the area of data standardization. For example, a major challenge in data standardization is to make standards general enough to apply to all services, and still meet the requirements of the individual specific services. This tradeoff is readily apparent with data- and metadata-element naming conventions. Metadata standards for the Navy's tactical systems are described by the Naval Warfare Tactical Database (NWTDB) data-element naming conventions. The Army also has its data-element naming conventions, which are not necessarily the same as those of the Navy. The Air Force still has other standards. Each service had evolved standards according to its specific needs without regard to joint considerations, until the reductions in the defense budget forced the services to reexamine their policies and practices.

Because the services have come to realize that they no longer can afford to continue the duplication of effort, they have become very interested in joint data standards [34] and data metrics [33] to promote and measure progress toward interoperability, especially in the area of C^4I. Each service has efforts in progress to achieve this goal. For example, the NWTDB is not actually a database itself, but rather a management process and framework

for database standardization. NWTDB provides a forum for Navy and Marine Corps database and data administrators to register their data-element formats with the goal of evolving all systems that use these data toward the DOD approved standard data-element formats [45]. The Army, with its C^4I technical architecture, has made similar efforts toward joint C^4I information standards [3].

One pitfall of standardization is to standardize on a particular vendor's commercial products, rather than to rely on open, consensus-based standards. Vendor-specific "standardization" is practiced in an effort to bring all aspects of software into a common environment, regardless of how much this impedes future integration.

Open standards promote reuse and interoperability across a wider variety of platforms and systems. Examples of open standards are the American National Standards Institute (ANSI) Structured Query Language (SQL) [9], and the International Standards Organization (ISO) Remote Data Access (RDA) [25], [26], [43]. The Object Management Group's (OMG) Common Object-Request Broker Architecture, (CORBA) [1], [8], [32] is a de facto industry standard.

The opposite problem of selecting standards that are not sufficiently universal is to select standards that are not detailed enough to address all of the functions and capabilities that the users of software applications demand. This is particularly true when attempting to conform to international standards that have long lead times for addressing the changes that come with the advancement of new technology. One example of this problem is standard ANSI SQL, the original version of which did not address the recursion or object-oriented design that many database administrators want to use in their applications. By the time SQL:1999 (formerly called "SQL3") was introduced, numerous database applications already were using SQL with object-oriented and recursive features. (See, for example, [16] and [31].)

Finding a balance between standards that are too general and those that are too specific is an open research issue. Middleware and flexible architectures can begin to span this gap. New technology can be applied to the standardization process. For example, expert systems can be used to standardize data elements [28].

Standards are emerging for the exchange of knowledge with the goal of making knowledge bases, ontologies, and knowledge base development tools more interoperable. Examples of standards in this category include the Knowledge Query Manipulation Language (KQML) [19] and the Open Knowledge Base Connectivity OKBC [12], [13], [19]. The Knowledge-Interchange Format (KIF) [12], [17], [23] also has been used for knowledge base integration in DARPA's High Performance Knowledge Base program. (See, for example, [6].) A major challenge with standardizing knowledge bases and ontologies is to define standards that do not restrict their inherent expressiveness that also will be sufficiently bounded and well-defined to yield useful integration results.

2.2 Defense Information Infrastructure (DII) Common Operating Environment (COE)

Technology has achieved more uniformity and interoperability with network protocol than it has with databases and knowledge bases because of the many ways in which databases can differ and also because of the multiple knowledge representation methods that can be used to construct knowledge bases. The growing size of these information resources also does not contribute to the tractability of the problem.

Aware of this disparity in the case of data, DISA has instituted a shared-data environment (SHADE) initiative [34]. SHADE is the data access, management, and administration part of the DII COE. SHADE is an extension of the DII COE that specifically addresses improved portability and data interoperability. Focused on the client-server portion of the DII, the SHADE architecture addresses how key technical components and services can be engineered to maximize the sharing of data that reside on COE-compliant servers.

The Defense Advanced Research Projects Agency (DARPA) Information Exploitation Office (IXO) is addressing the disparity that has developed in the case of knowledge reuse. The development of methods to integrate heretofore stand-alone ontologies and knowledge bases is an active research area [6], [13], [19], [38], [40].

A significant effort in the DOD that has a potential for substantial payoff is to combine the products of DARPA projects with DISA's DII COE so that the most advanced information technology can be incorporated into joint systems where it can be reused. This is addressed through the DISA/DARPA Joint Program Office's Leading-Edge Services.

2.3 Information-System Component Reuse

Software and information reuse is important to the joint forces because of the implications for the common operational picture, and also for cost savings through the reduction in duplication of effort. Standardized data segments provide some degree of interoperability [39]. Standards are the key to reuse and can affect the efficiency of reuse [9]. Similarly, reuse can affect the efficiency of software development. This relationship between standards, reuse, and software development is recognized in the joint standards for database and software reuse. Thus, a question arises: What is the best way for the Navy to design and implement information systems that the Army or the Air Force also can reuse efficiently?

Reuse has two aspects: the design of new systems for future reuse and the reuse of legacy software components in existing systems. Many legacy systems were not designed for reuse, which makes reuse much more challenging, and in many cases, not cost effective. (See, for example, [9].) An active topic in defense software engineering is to determine when reuse makes economic sense and to retrofit legacy information systems and software to make them more reusable.

Candidate approaches to the reuse problem have been considered in the migration of legacy database systems and applications. For example, with object technology, not only schema integration and transformation are possible, but also object abstraction and the encapsulation of entire systems. (See, for example, [8], [32], [43].) A major benefit of the object-oriented paradigm with regard to software reuse is the production and development of reusable components

for the assembly of new systems [15]. The encapsulation approach is one of several aspects of object-oriented technology that enables a legacy system to be reused in an object-oriented environment.

Software reuse is the key to the maturation of a technology. If a technology is not used widely, it could be because industry has not found a systematic way to reuse the underlying software or information that is based on that technology. For example, widespread access to Web-browser technology was impossible 25 years ago because no one knew how to implement Web standards and the reuse they enabled. Now, sufficient expertise and commercial products are available to reuse Web-based software. (See, for example, [27].) Intelligent agents are not yet common place because sufficient reuse of knowledge-based systems has not yet been achieved.

3 Integration and Interoperability

Closely related to joint data standards and reuse are data and knowledge integration and interoperability, which become possible and more efficient through the use of these standards. Standards are necessary, but insufficient [37] for interoperability. More interfaces need to be developed between COE databases and the legacy systems with which they share data [34]. Appropriate and consistent application of standards, along with a resolution of heterogeneity at all levels are prerequisites for seamless data and software integration in an information system. Interoperability is an active area of research among many DOD components, DOD contractors, and industry in general. (See, for example, [18], [37], [39], and Section 7.2 of this paper.)

Metadata integration is a key step toward interoperability for DOD. (See, for example, [7], [18], [34], [36], [39].) Data mediation is also a valid approach to the interoperability problems caused by differences in data definition [36] and representation. A mediator acts as the second tier in a three-tiered architecture that limits the number of translators necessary to convert data from one format to another [8], [15].

Databases can differ in many ways. The categories of heterogeneity are summarized below [7]:

- Platform heterogeneity—e.g., differences in database management system (DBMS) versions and vendors,
- Data-model heterogeneity—e.g., different data models, query languages, integrity constraints, and schemata,
- Semantic heterogeneity—e.g., conflicts in metadata specifications, relation and attribute names, levels of precision, levels of abstraction, units of measure, and data inconsistencies.

These differences can result in conflicts that pose obstacles to interoperability. The single most challenging obstacle in DOD data management is the seamless integration of legacy information systems at all levels of heterogeneity. This obstacle is significant because integrating multiple legacy databases or developing interoperable information exchanges between legacy databases is slow, difficult, and expensive [18].

As indicated in the previous section, more progress has been made in the integration of software other than in databases. What passes for good systems integration results in interoperability only at relatively coarse levels of granularity. The DOD has hardware platform interconnectivity, and sometimes software platform interconnectivity. It also has message communications, Web sites, user interfaces with a common look and feel, standard support applications and application-program interfaces (APIs), "canned" queries, and client platforms that can access a variety of databases. To the casual observer, these DOD systems look integrated (particularly during demonstrations). Missing from this picture is integration at the finer levels of granularity, including the ontological, semantic, and data levels [6].

It has been suggested, and with considerable merit, that systems integration may be achieved best through levels of interoperabilty [7], [42]. In this approach, heterogeneity is resolved in stages that are identified clearly. Different user groups will require different levels of integration. For example, even if database integration is not needed to support users in the C^4I community, the users of modeling and simulation systems may require a finer level of integrated granularity. These interoperability challenges apply to tactical and nontactical systems alike.

3.1 Integration in Tactical Systems

Within the charter of DISA is the oversight of the development of joint C^4I systems such as the Global Command and Control System (GCCS). GCCS, the overarching architecture for new information technologies, will allow commanders to share data in multiple formats not only among senior commanders, but also among lower-level military commanders. The GCCS worldwide network is intended to coordinate operations not only between the joint forces, but also with allies. Many of the subsystems that compose GCCS were developed prior to the establishment of the DII COE and are not programmatically connected to DII COE per se. However, current GCCS software is built to comply with DII COE standards. Achievement of various levels of interoperability between GCCS components from all of the services represents a major systems integration effort that uses these standards [14].

The NWTDB is the Navy's tactical standard and focal point not only for DII COE compliance and coordination, but also for propagating proposed changes to the DII COE standards and documentation to reflect the development and insertion of new technology. Under the NWTDB umbrella are standard databases that are updated periodically for tactical users and applications. The process is flexible enough so that additional C^4I databases can comply with the SHADE and NWTDB specifications, and users can obtain them on a regular basis.

The Joint Maritime Command Information System (JMCIS) has become the maritime configuration of the GCCS, or "GCCS-M," serving primarily the Navy, Marine Corps, and Coast Guard. NWTDB and the GCCS-M federated database (FDB) together provide a good case study in legacy-systems integration because the NWTDB has good metadata documentation and the GCCS-M FDB includes

some data sets maintained using the NWTDB process. (See, for example, [7], [11], [22], [29], [30].)

Like many other integrated systems, GCCS-M is derived from a collection of legacy C^4I systems, the principal components of which have evolved into the afloat and ashore variant forms of GCCS-M [8], [39]. In this context, a variant is a version of a system composed of parts and modules assembled to support the specific applications of a major division of the system's users. For example, the Operations Support System (OSS) became the ashore variant, supporting ashore command centers, and Navy Tactical Command System Afloat (NTCS-A) became the afloat variant, which was installed on ships. (See, for example, [5], [8], [9].)

Prior to GCCS-M, OSS, for instance, was composed of a collection of legacy component systems supported by databases that were not fully integrated at all levels, particularly at the data and semantic levels. This was also the case for the other variants. Thus, the data-integration task for GCCS-M is one of fully integrating components databases into the GCCS-M FDB when these components themselves consist of data sets that never have been fully integrated. This constitutes a significant challenge, particularly due to the growth of GCCS-M and the added complexity each new component brings to the task.

Another problem in DOD information systems is a lack of complete documentation of the metadata. For example, some C^4I systems have entity-relationship diagrams for their databases, but no system-wide integrated data dictionary, or vice versa. By contrast, NWTDB has a well-developed and well-supported data-element dictionary that can serve as a model and starting point for the creation of metadata specifications for these systems. The expanded use of the NWTDB and SHADE data-element naming conventions in C^4I systems is one way to approach better database interoperability. A more long-range, cohesive view of the GCCS-M FDB can be accomplished with this approach. (For example, see [7].)

The NWTDB standards and data-integration tools and techniques, such as the Data Analysis Reconciliation Tool (DART) [22] and some of the more advanced Computer Aided Software Engineering (CASE) tools have the potential to improve data interoperability in C^4I systems. The database systems that support C^4I software may not be sufficiently integrated to provide the best support to the applications and to avoid duplication of effort. This uncertainty exists because, until recently, tractable methods and tools to accomplish a systematic database-integration analysis were not implemented. DART provides a "microscope" for a database analyst to examine conflicts internal to a single data set as well as conflicts between different databases.

Currently, available data-integration tools can assist an analyst in the identification of integration problems in databases. These tools also can be used to group similar data elements together to facilitate conflict resolution. (See, for example, [7].) Tools also support more advanced data-integration methods, including methods that use artificial intelligence (AI). (See, for example, [6], [28] and Section 7.2.) The intent is to streamline a systematic study of potential areas of conflict that could result in the identification of integration opportunities and eventually, to automate the conflict-resolution process. Alternately, integration engineers can alert users about the conflicts that cannot be resolved, a situation that can occur with multiple accesses to independent information sources available on a network.

Managers and developers need to become fully aware of the capabilities, strengths, and limitations of data-integration tools and techniques. Both SHADE and NWTDB can provide better support to deal with database integration challenges in C^4I systems. Previously, these problems were ignored partly because DOD has so many database and metadata inconsistencies that correcting all of them was impossible. (For other reasons, see Section 6.) A better approach to database integration for C^4I needs to be considered at the program-management level.

3.2 Combining Tactical and Nontactical Systems

Historically, the tactical (i.e., artillery officer and fighter pilot) and the nontactical (i.e., contracts specialist and shipyard manager) defense communities have developed as separate specialty areas or groups of areas, requiring different training and experience. Therefore, it is not surprising that the computer systems supporting these different communities have evolved separately. A significant trend over the past 15 to 20 years is the establishment of connections between the systems that serve these communities. This trend has accelerated recently because specialists in each of these communities have come to realize that they not only require information from the other community to increase their job efficiency, but also because the technology has become more available to implement communication and integration of legacy systems.

Nontactical systems can affect tactical systems and vice versa. For example, tactical specialists who plan schedules need to consider the availability of scarce resources (e.g., weapons, personnel, platforms, fuel, etc.), the status of which is tracked in nontactical systems for acquisition and maintenance. In determining the operational employment schedules for Navy ships, planners must consider the shipyard maintenance schedules, which can affect when the ship is available for tactical deployment and when it is in dry dock or undergoing an overhaul. The ultimate result of efficient shipyard scheduling (e.g., a nontactical concern) is an increase in fleet readiness (e.g., a tactical concern). Similarly, the procurement and availability of advanced communications equipment may affect the planning of an Army exercise.

These examples underscore the need for an interface between tactical information systems and their nontactical counterparts. As a result of this trend, the clear boundary that previously existed between tactical and nontactical systems and the databases that support them has become diffuse and ill-defined. Tactical and nontactical systems should be combined [24] to provide a comprehensive information resource for strategic and tactical purposes. Conceptually, this recommendation is easy to understand; technically, it is no more challenging than the integration of other legacy systems. Whereas the main obstacles to

implementation are economic, political, and resistant to change, this is also a data-access issue.

4 Data Access

One of the most serious and long-standing problems, not only in DOD, but also in industry, is that no one has devised a robust, comprehensive solution for end-to-end data-access requirements. The problem is that database engineers want to provide efficient access to users who need to view and use data objects and also to deny access to unauthorized individuals [21]. Solutions that favor more universal access tend to be deficient in the area of database security, whereas solutions that implement some of the more trusted and approved security controls can be cumbersome and, in some cases, inefficient even for authorized users. Thus, problems exist on both "ends" of this data access issue due to conflicting goals. (These conflicts are similar to those pertaining to the autonomy of component systems in a federation and the cooperation between them, as described in Section 2.)

Data and metadata availability, speed of access, format compatibility, reliability of information, and data-translation tractability are among the data-access challenges to the DOD. The use of technology can create new challenges, even as it solves other problems. Specific technological advancements can act as catalysts for policy change. For example, the growth of networks has prompted changes in policy concerning the way the DOD deals with security. To meet the challenge, data access in legacy systems has been the subject of numerous investigations. (See, for example, [34] and [35].)

Middleware is a key component in solving the data-access challenge. The World Wide Web (WWW) has emerged as a form of data-access middleware in some applications because of its efficiency and generality, owing to a common data-transfer protocol. (See, for example, [6], [20], [27].) It also has become a tool for software reuse [27]. Databases, knowledge bases, data-mining tools, images, textual documents, standards, and variety of software tools are all accessible on the WWW.

Other technical developments that facilitate data access are CORBA, the Open Group's Distributed Computing Environment (DCE) [1], [2] and ISO's RDA [25], [26]. These standards and specifications are used not only in interactive applications, but in compiled applications as well. For example, CORBA-compliant object-oriented middleware tools are available commercially.

Technological advancements notwithstanding, data access challenges remain. For example, the following questions need to be addressed.

- How does a data repository manager communicate with the users-at-large to provide them enough of the information they need to decide which data sources to use?
- What is the best way to integrate the features of CORBA, DCE, and the WWW to provide applications versatile and robust data access?
- How does data mining contribute to more accurate and comprehensive modeling and simulation?

To improve information access in the DOD, data managers should support the following actions:

- Construct reusable data and knowledge bases rapidly,
- inform potential users about data availability,
- ensure reasonable authenticity and assurance of information described in metadata repositories,
- provide information sources via the WWW,
- promote affordable and capable commercial-off-the-shelf (COTS) middleware usage, and
- comply with relevant metadata standards and assure access to authorized browsers or data-mining tools.

5 Network-Centric Warfare

Closely related to data access is the topic of information networks, the use of which has revolutionized warfare concepts and applications. Information is at the heart of the C^4I, surveillance, and reconnaissance vision [36]. Information superiority depends to a large extent on the warrior's ability to access the right data at the right time, regardless of the location of the data source, and with enough bandwidth to transfer a sufficient aggregate of information faster than one's adversaries. Information warfare has emerged as a significant area of interest due to the development and widespread usage of computer-communications networks. As a result of this trend, platform-centric warfare has become obsolete in the information age. The importance of battlespace-information superiority is forcing a change in paradigm in the DOD. The new emphasis on network-centric warfare shifts the burden from a mainly hardware-centric military to a military that is both hardware and software centric. For this reason, the growth of network-centric warfare in general and database-centric C^4I in particular continues to pose new challenges in data management.

When computing equipment was introduced into the military, it was installed in shore-based, centralized headquarters and support facilities. In the case of tactical systems, this equipment was tested at engineering facilities and installed in command centers. At that time, widespread use of computers in the DOD was impossible due to their large size, limited processing power, limited (if any) networking, and primitive software capabilities.

By contrast, today's warfighter wants light, powerful, and portable computers that can be worn on the belt or transported in the pocket of a backpack. Today's fighting forces need a computing network that includes the information producers who can maintain contact with the warfighter and contribute to the common operational picture. Specifically, warfighters want to obtain the latest specific, dynamic, operational data on current and emerging events in their area of interest. They also can contribute updates to the common operational picture to assist personnel in other units.

Therefore, networks have been installed on numerous ships, Naval shore facilities, and on Army and Air Force bases. Local Area Networks (LANs) rely on dependable computers (servers), routers, repeaters and bridges, as well

as a host of internet software protocol layers. The hardware technology is much more mature than the software that it supports. (See, for example, Section 7.4.) As a result, software engineers in general and database engineers in particular are challenged to find innovative ways to meet the growing demands of an increasingly sophisticated and better-equipped fighting force. Network-centric warfare is one of the military's implementations of collaborative computing.

6 Data Aggregation

As discussed above, conflicting sets of requirements, such as the trend toward wider information access and the need to control access for security reasons, can pose unique challenges. The conflicting goals and requirements of data sharing and database security have produced a situation that has yielded few, if any, practical solutions that can address both areas comprehensively and efficiently.

This particularly difficult problem in information management concerns the aggregation of data and its implications for database security, not only in the DOD, but also in many other areas. The problem is illustrated with the following example: Suppose A and B are data sets that are classified separately at classification level X. Also, suppose that a data engineer wants to integrate data sets A and B into the same database. How does one classify the aggregate of A and B? Are these data still classified at level X, or is the aggregate now classified at level X+Y, where Y is an integer greater than or equal to 1?

A Subject-Matter Expert (SME) must assist the data engineer in assessing the integrated database to avoid a compromise of information that could result if the aggregate were underclassified at level X when it should be classified at a higher level. If the SME recommends classification of the integrated database at level X+Y, users who need the data but are not cleared at level X+Y are denied access to the aggregate due to what could be inappropriate and unnecessary overclassification.

This problem is a specific example of the data access issues discussed in Section 4. Moreover, the aggregation problem is not limited to database engineers designing and implementing systems for DOD use. Computer users in many domains, such as the medical and manufacturing domains, also can access multiple databases on networks to download information to their local sites. The aggregates of data thus formed could constitute sensitive information, for example, about patient medical conditions or proprietary information on a company's manufacturing processes.

The DOD would have fewer stand-alone information systems (called "stovepipes") if DOD policy provided a comprehensive, systematic, and tractable method to handle the aggregation of data components, including data elements, multimedia data objects, databases, knowledge bases, and models. The DOD has no such policy to determine when an aggregate of information warrants protection at a higher level of security.

Consequently, for example, when a database integration effort is mounted, a security risk arises that many DOD personnel think is unacceptable. This concern is reasonable in light of the difficulties with aggregated classification.

Although DOD has some of the technical elements of a solution in place (e.g., user views, read/write permissions), nothing is available that draws it all together. The technology to address this issue is in a preliminary, experimental form that is not widely available in commercial products. Initially, knowledge-based systems that fall into the category of AI could assist an analyst in making more rapid determinations about which aggregates produce potential security risks and which ones do not. Currently, the DOD does not have a systematic, automated way to determine this, but it could have such a system using AI in the future. If a security expert can determine the classification of an aggregate of data, this process can be captured in a knowledge base, automated, and applied across many systems. The following are challenges to automation using expert-systems technology.

- Deciding what to automate (policy), what can be automated (technology), and how to automate it (engineering) and, then,
- assessing the deficiencies in the particular method of automation in question (testing).

7 New Technologies and Research in Military Information Systems

One of the main challenges in the implementation of new technology is the speed with which new developments are made operational. Even with rapid prototyping, technology changes so quickly that a system may be obsolete (or may be based on obsolete technology) before it is completed. The following questions persist as challenges in implementing new technology.

- What is the best way to insert new products and systems into the operational environment before the technology becomes obsolete?
- What is the best point at which to upgrade to newer technology considering the capabilities-cost trade off?

If an upgrade is implemented too soon, the cost per item will be unaffordable across the many platforms that need the upgrade. If an upgrade is delayed to the point where the technology becomes considerably less costly, operational systems will become obsolete. Rapid prototyping has provided some answers to this dilemma, (see, for example [9]). But these technical issues have political and economic ramifications. Here again, specific technological advancements can act as catalysts for policy changes and vice versa.

7.1 Rapid Development and Reuse of Knowledge Bases

Expert systems and knowledge-based information representations will play an increasingly important role in C^4I. Because both databases and knowledge bases are used to represent the relevant parts of an application domain and to allow convenient access to stored information, the interactions between database and knowledge bases will continue to be of interest to DOD information-system engineers. Engineering with AI is more experimental and complex than the software engineering that has resulted

in mature, commercial products. This is because AI languages and information representations are more expressive than traditional software languages and information representations. These trends suggest implementation questions, such as:

- What are the obstacles to AI technology insertion into operational C^4I systems?
- How can engineers make AI software and knowledge bases more reusable and more efficient to build?
- What are the main challenges to providing a complete set of standards for the development and interoperation of knowledge bases in support of the warfighter?
- What are the best metrics to use to evaluate AI software and knowledge bases, and their utility in DOD information systems?

7.2 Semantic Integration

As the number and size of information sources grow, global semantic integration and semantic interoperability are becoming a growing concern and a major challenge to DOD information managers [40]. Advances in the field of AI have paved the way for more comprehensive information-systems integration, particularly at the semantic level [6], [19], [38], [40], some of which engineers are applying to DOD information systems. For example, Fowler et al. have summarized several difficult problems associated with the Environmental Data Exchange Network (EDEN), in which the DOD has collaborated [19]. These problems, many of which relate to ontology mapping and development, include the following [19]:

- Different contexts affect the manner in which users want to query and display information.
- Ontologies used for semantic mapping must be abstracted adequately from available resources to allow new information sources to map to the same ontology. Exact mapping will not be achieved very often because this abstraction is very difficult to perform (owing to the incomplete and incorrect nature of some necessary information resources.)
- An efficient method is needed to convert concepts, attributes, and values taken from one ontology and transferred to another. (This relates, in part to various levels of granularity and expressiveness in ontologies.)
- Ontologies are incomplete for various reasons. They can contain incomplete, uncertain, or evolving concepts. At present, the available resources and technology are insufficient to ensure that these ontologies will be complete by the time they are needed for semantic integration. A semantic integration can be only as good as the ontologies on which it is based.
- Data cleansing is not performed adequately prior to introducing information from various sources into multidatabase systems. This adds to the difficulty posed by imprecise and incorrect data that may have been abstracted incorrectly, and mapped and correlated in an approximate manner using ontologies that do not fit the information sources precisely.
- Different versions of the same data source that previously were available only on stand-alone systems now are available across networks, thus confusing users when they access what could be inconsistent information.
- Pedigree of information is becoming increasingly important to users to enable them to evaluate the significance and reliability of their results, particularly when these results are generated from multiple information sources that required multiple steps in semantic integration and/or data fusion. InfoSleuth, an agent-based system for integrating heterogeneous, distributed information sources, uses common ontologies to provide these details to users upon demand, among other functions [19].
- A challenge for the DOD information-system engineers is to find a method to apply multidatabase integration approaches like InfoSleuth [19], OASIS [38], and that of Smith and Obrst [40] to operational tactical and business systems in a cost-effective, comprehensive, and easily applied manner without degrading user access during daily operations.

7.3 Data Mining

Like AI, data mining is emerging as a potentially significant technology for C^4I systems. Data mining will be an integral part of the information system deployed on future Naval surface vessels. Data mining can provide a link to previously undiscovered trends in legacy databases [4], [44]. These trends can be used to shape future war games, models, simulations, exercises, and battle plans to provide commanders with a historical perspective and to improve the design of future sensor or information systems. Used as inputs into the databases and knowledge bases of future decision support systems, the products of data mining constitute another step towards information superiority for the military. (See, for example, [43] and [44].)

Like software reuse, data mining has two aspects, performing data mining on legacy databases and designing future databases to support data mining. Data-mining challenges include:

- How can the DOD systems perform pattern recognition on sketchy and sparse data, which also includes detecting rare events using available data?
- What are the best data-mining algorithms to use for the identification of missing data and for the integration and correlation of this missing information with observable data?
- What is the best way to combine the results of various data-mining algorithms with data-fusion products?
- Data mining is, by definition, not a planned use of legacy databases. However, this does not preclude design considerations of future systems from facilitating the application of data mining in a way that legacy systems did not. Thus, a question remains, what is the best way to design databases and data warehouses to support future data mining and fusion?

7.4 Real-Time Systems and Computational Speed

Military information systems, particularly command and control systems rely increasingly on object-oriented software in client-server environments. In an effort to provide timely information to the warfighter, as faster hardware becomes more economical, these systems are upgraded with better platforms to avoid costly delays and maintenance problems, which are particularly critical on Naval vessels. In spite of the advances made in hardware speeds over the last two decades, software architectures have not fully utilized the new hardware capabilities that are now available [10]. As many layers of middleware are present in client-server environments, some software executes too inefficiently even with better hardware. Moreover, as data mining becomes more widespread in the DOD, the increased size of databases and data warehouses will require more enhanced computational capabilities to achieve tractability for some queries and algorithms. Bringing these systems into the realm of real-time execution is a major challenge.

Complementary processing (CP) is a scheduling algorithm that is implemented on top of operating systems to increase the speed of software execution [10]. Designed to eliminate interrupts, stacks, heaps, and the inefficiency they introduce, CP schedules tasks concurrently on uniprocessor systems with the result of much more rapid execution times. However, CP is a new technology that needs further investigation. The following challenges remain in introducing CP into military information systems.

- Where is CP best utilized in command and control systems?
- What are the main obstacles to introducing CP into DOD information systems?
- What modifications to existing systems are necessary to implement CP for various applications?
- What is the impact of utilizing CP on nodes in a network?
- What are the effects of using CP with a relational database management system or with an object-relational database management system?
- What data-mining methods can CP enable in the future that cannot be done now?

8 Research Directions and Systems

DARPA is an important source of DOD research systems that are focused on future capabilities, as opposed to the operational information systems that the military uses today. Research programs aimed at advancing the state of the art in information systems include the Information Assurance and Survivability program and the Dynamic Database program. (See, for example, [46].) Each program had research systems associated with it.

8.1 Information Assurance and Survivability

The DARPA Information Assurance and Survivability program was motivated by a desire on the part of the warfighters to be able to trust the data they use for decision making in critical warfighting functions, such as mission planning, status of forces evaluation, precision engagement, and logistics. The warfighter must access much of this information over networks in an environment that faces threats such as network intrusions, malicious code, and insider attacks. Thus, the issues of multilevel security, protective mechanisms, adaptive survivable architectures, intrusion detection, intrusion assessment, and cryptology came into sharper focus here. Although the focus of this program was mainly from a DOD standpoint, the risk is at the national level and could affect many more information systems throughout several levels of government and in industry.

The program's six objectives included cyber command and control, intelligence strategic intrusion assessment, autonomic information assurance, dynamic coalitions, intrusion-tolerant systems and fault tolerant networks, and information assurance science and engineering tools. The approach to scientific experimentation included five types of experiments: field experiments, red (hostile) team lab exercises, laboratory experiments, interdisciplinary white boarding, and component specific testing.

The program had three new research directions, each of which is discussed below. First, the cyber sensor grid is intended to monitor the attack space to detect intrusions. Technologies that support this new direction include Bayesian techniques, neural networks, statistical analysis, graphical analysis, hidden Markov model detection, and signature-based detection.

Second, malicious-code mitigation addresses the problems that give rise to and allow the use of malicious code, such as mobile code, insider attack, vulnerable architectures, the use of mobile code, the inability to detect malicious code, and the lack of a useful policy and its enforcement. The task of combating malicious code is complicated by the military's increase in reliance on commercial off-the-shelf products, an increase in connectivity, and an increase in use of and reliance on systems. The strategy for this research direction was to detect and expunge malicious code "on the fly," to develop new architectural concepts and to address the policy gap.

Third, reliable mobile agents are programs that can migrate from machine to machine under their own control. Code mobility is advantageous because it provides better functionality and presents survivability opportunities. For example, reliable mobile agents can install new functionality on remote machines and provide remote processing of data for better efficiency because they promote the exchange of smaller programs and smaller data sets, thus reducing network congestion. Since reliable mobile agents can be replicated at several remote sites simultaneously, they avoid the problem of a single point of failure and, thus, provide better survivability.

More ideas for future research in this area are needed in denying denial of service, self-healing systems, proof-carrying code, track back, dynamic defense, and metrics and science-based designs.

8.2 Dynamic Database

The goal of the DARPA Dynamic Database program was to convert large quantities of sensor data efficiently into useful information for tactical commanders. In their quest for early threat identification, commanders today must contend with

data from a large number of partially overlapping sensors; hundreds of reports, thousands of images per minute, and a very high false alarm rate. Commanders would rather have timely situation knowledge, comprehensive coverage of more than a thousand targets per Km^2, accurate target locations with small circular error probabilities, and a low-burden geo-referenced database. Other capabilities that commanders want that they do not have today are multi-sensor analysis integrated across platforms in the battle-space and a method to avoid missed opportunities that result from too much data that have not been exploited, while still maintaining a low false-alarm rate.

The solution includes, but is not limited to, a common geo-registered database tied to wide-area terrain data, fusion across sensors using model-based evidence accumulation, and the ability to track targets and features at the object level. The research database architecture overcomes the limitations of data ownership and enhances data sharing and interoperability, thus enabling sensor data access and technology growth. This architecture takes advantage of object-oriented databases and the open-systems approach to leverage commercial products where they are useful while exploring military needs that exceed most commercial interests.

Specifically, the architecture includes a sensor data store with sensor history data in scalable space-time indexed databases, a user interface, applications for all-source track fusion, model-based classifiers, change detectors, etc., and a situation history that can handle high input-output rates of multimedia data. The dynamic data services include spatial temporal indexing, rapid search and retrieval, space-time queries, probabilistic queries, databases mediation, and query and data distribution.

This architecture enables two capabilities:

- Registration of all sensors to a common targeting grid and
- normalcy models that provide wide-area change detection.

The ability to combine multisensor data is necessary to define the state of normalcy and for comprehensive spatial and temporal pattern analysis. Techniques for change detection that apply to a wide variety of intelligence areas are under investigation in this program.

Dynamic database also featured data-driven collection management, the goal of which is to close the loop between data-collection management and situation estimation. For example, the coupling between collection management and situation estimation is a manual process that limits the use of uncertainty information in task valuation. In the future, automated processes will perform task valuation based on uncertainty estimates of the current situation.

The Dynamic database program has demonstrated the following accomplishments: Change detection over a wide area, all-source track and identity fusion, and dynamic data services, such as a common targeting grid and registered sensor histories over space and time.

More ideas for future research in this area are needed to increase track continuity more reliably, to reduce errors in positioning and false alarm rates, and to provide better multisource fusion engines that can feed data stores and tie each data set to the common grid automatically.

9 Conclusion

This paper describes major data-management issues and challenges in the DOD, including standards, software reuse, database integration, interoperability, data access, network-centric warfare, and data aggregation. New technology as well as new combinations of existing technologies can provide possible solutions. This paper also discusses some of the new technologies and research directions considered most important to information systems in the DOD, such as networks, object-oriented design, artificial intelligence, data mining, information assurance, and dynamic databases for sensor fusion that future military systems will use. However, these technologies will need to gain political acceptance and financial support to be successful.

Acknowledgments

The author thanks the Defense Advanced Research Projects Agency for funding for this work. This paper is the work of a US Government employee produced in the capacity of official duty and no copyright subsists therein.

References

[1] T.J. Brando, *Comparing DCE and CORBA.* MP 95B0000093, Bedford, Mass.: The Mitre Corp., 1995.

[2] D. Brown, *DCE Open Client and Open Server Support of the Distributed Computing Environment.* Emeryville, CA: Sybase Corp. Technical Paper Series, 1995.

[3] J.D. Burke, "Achieving Joint Interoperability Through Information Standards," *Proc. Dept. of Defense Database Colloquium '95,* pp. 37-52, Aug. 1995.

[4] P.L. Carbone, "Data Mining: Knowledge Discovery in Data Bases," *Handbook of Data Management,* pp. 611-624, Boca Raton, Fla.: CRC Press, LLC, 1998.

[5] M.G. Ceruti, "Development Options for the Joint Maritime Command Information System (JMCIS) Specialized Data Servers," *Proc. Dept. of Defense Database Colloquium '96,* pp. 217-227, Aug. 1996.

[6] M.G. Ceruti, "Application of Knowledge-Base Technology for Problem Solving in Information-Systems Integration," *Proc. Dept. of Defense Database Colloquium '97,* pp. 215-234, Sep. 1997.

[7] M.G. Ceruti and M.N. Kamel, "Semantic Heterogeneity in Database and Data Dictionary Integration for Command and Control Systems," *Proc. Dept. of Defense Database Colloquium '94,* pp. 65-89, Aug. 1994.

[8] M.G. Ceruti, M.N. Kamel, and B.M. Thuraisingham, "Object-Oriented Technology for Integrating Distributed Heterogeneous Database Systems," *Proc. Dept. of Defense Database Colloquium '95,* pp. 79-98, Aug. 1995.

[9] M.G. Ceruti, S.D. Rotter, K. Timmerman, and J. Ross, "Operations Support System (OSS) Integrated Database (IDB) Design and Development: Software Reuse Lessons Learned," *Proc. Armed Forces Comm. and Electronics Assoc. Database Colloquium '92,* Aug. 1992.

[10] M.G. Ceruti, R.C. Trout, and T. Lee, "Complementary Processing and Its Impact on Software Performance," *Proc. Fourth Int'l IEEE Workshop Object-Oriented Real-Time Dependable Systems, WORDS '99,* pp. 110-116, Jan. 1999.

[11] M.G. Ceruti, "Challenges in Data Management for the United States Department of Defense (DOD) Command, Control Communications Computers and Intelligence (C^4I) Systems," *Proc. 22nd Ann. Int'l Computer Software and Applications Conf., IEEE COMPSAC '98,* pp. 622-629, Aug. 1998.

[12] V.K. Chaudhri, A. Farquhar, R. Fikes, P.D. Karp, and J.P. Rice, *Open Knowledge Base Connectivity 2.0.* Knowledge Systems Laboratory Technical Report No. KSL-98-06, http://ksl-web.stanford.edu/KSL_Abstracts/KSL-98-06.html, 1998.

[13] V. Chaudhri, A. Farquhar, R. Fikes, P.D. Karp, and J.P. Rice, "OKBC: A Programmatic Foundation for Knowledge Base Interoperability," *Proc. Am. Assoc. Artificial Intelligence Conf., AAAI '98,* July 1998.

[14] P. Cooper, "GCCS Gives Allies Unique Data Sharing Powers," *Defense News,* Oct. 1996.

[15] R.T. Due, "Object Technology Essentials," *Handbook of Data Management,* CRC Press, LLC, Boca Raton, chap. 14, pp. 187-203, 1998.

[16] A. Eisenberg and J. Melton, "SQL:1999, Formerly Known as SQL3," *SIGMMOD Record,* vol. 28, no. 1, pp. 131-138, 1999.

[17] R. Fikes and A. Farquhar, *Large-Scale Repositories of Highly Expressive Reusable Knowledge.* Stanford Univ. Knowledge Systems Laboratory, Report No. KSL 97-02, Mar. 1997.

[18] R. Foss and R. Naleen, "An Environment for Metadata Engineering," *Proc. Dept. of Defense Database Colloquium '97,* pp. 75-91, Sept. 1997.

[19] J. Fowler, M. Nodine, B. Perry, and B. Bargmeyer, "Agent-Based Semantic Interoperability in InfoSleuth," *SIGMMOD Record,* vol. 28, no. 1, pp. 60-67, 1999.

[20] M. Frank, "Database and the Internet," *DBMS Magazine,* vol. 8, no. 13, pp. 44-64, 1995.

[21] J.N. Froscher, "Security Information Through a Replicated Architecture," *Handbook of Data Management,* chap. 13, Boca Raton, Fla.: CRC Press, LLC, pp. 173-183, 1998.

[22] R. Gressang, G. Michaels, E. Harris, J. Mathwick, and J. Lu, "Database Integration Using NWTDB Procedures," *Proc. Dept. of Defense Database Colloquium '95,* pp. 99-105, Aug. 1995.

[23] T.R. Gruber, *A Translation Approach to Portable Ontology Specification.* Technical Report No. KSL 92-71, also in *Knowledge Acquisition,* vol. 5, no. 2, pp. 199-220, 1993.

[24] W.J. Holland, Jr., "The Race Goes to the Swiftest in Commercial, Military Frays," *Signal,* vol. 52, no. 7, pp. 68-71, Mar. 1998.

[25] *Information Technology—Open Systems Interconnection—Remote Database Access Part 1: Generic Model, Service and Protocol.* Int'l Organization for Standardization and Int'l Electrotechnical Commission (ISO/IEC), (ISO/IEC 9579-1, 1993(E)), Geneva, Switzerland: ISO/IEC Copyright Office, 1993.

[26] *Information Technology—Open Systems Interconnection—Remote Database Access Part 2: SQL Specialization.* Int'l Organization for Standardization and Int'l Electrotechnical Commission (ISO/IEC), (ISO/IEC 9579-2, 1993(E)), Geneva, Switzerland: ISO/IEC Copyright Office, 1993.

[27] R. Lawson, "Developing Web-Based Internet Applications with Reusable Business Components," *Proc. Dept. of Defense Database Colloquium '96,* pp. 503-520, Aug. 1996.

[28] J. Little, "Using Expert Systems to Standardize Data Elements," *Proc. Dept. of Defense Database Colloquium '95,* pp. 205-217, Aug. 1995.

[29] J.E. Mathwick, "Database Integration, Practical Lessons Learned," *Proc. Department of Defense Database Colloquium '97,* pp. 31-38, Sept. 1997.

[30] J. McKee, "JMCIS: 'The Big Picture,'" *Surface Warfare,* vol. 20, no. 4, pp. 10-11, 1995.

[31] J. Melton, "ANSI SQL3 Update," *Database Programming and Design,* vol. 8, no. 11, pp. 61-63, 1995.

[32] *The Common Object Request Broker: Architectureand Specification.* Object Management Group (OMG) and X/Open Company Limited, Revision 1.1, OMG Document Number 91.12.1 Revision 1.1, 1992.

[33] P. Piper, "CIM/EI Data Metrics," *Proc. Dept. of Defense Database Colloquium '95,* pp. 351-357, Aug. 1995.

[34] P. Piper, "Defense Information Infrastructure (DII) Shared Data Environment (SHADE)," *Proc. Dept. of Defense Database Colloquium '96,* pp. 407-418, Aug. 1996.

[35] D.C. Reilly and J.A. Flemming, "Lessons Learned in Legacy Data Access," *Proc. Dept. of Defense Database Colloquium '95,* pp. 623-629, Aug. 1995.

[36] S.A. Renner, "Organization, Process and Technical Foundation for Data Interoperabilty," *Proc. Dept. of Defense Database Colloquium '97,* pp. 39-48, Sept. 1997.

[37] S.A. Renner, A.S. Rosenthal, and J.G. Scarano, "Data Interoperability Between C^3I Systems," *Proc. Dept. of Defense Database Colloquium '95,* pp. 107-116, Aug. 1995.

[38] M. Roantree, J. Murphy, and W. Hasselbring, "The OASIS Multidatabase Prototype," *SIGMOD Record,* vol. 28, no. 1, pp. 97-103, 1999.

[39] A. Rosenthal and M.G. Ceruti, "Toward Data Administration in the Large," *Proc. Dept. of Defense Database Colloquium '96,* pp. 37-51, Aug. 1996.

[40] K. Smith and L. Obrst, "Unpacking the Semantics of Source and Usage to Perform Semantic Reconciliation in Large-Scale Information Systems," *SIGMMOD Record,* vol. 28, no. 1, pp. 26-31, 1999.

[41] *SSC San Diego '97 Brief.* SPAWAR Systems Center, San Diego, corporate publication, SSC SD TD 2949, June 1997.

[42] B. Thompson and A.K. Anderson, "Achieving 'Jointness' Through Improved Levels of Information System Interoperabilty (LISI)," *Proc. Dept. of Defense Database Colloquium '97,* pp. 49-61, Sept. 1997.

[43] B.M. Thuraisingham, *Data Management Systems: Evolution and Interoperation.* Boca Raton, Fla.: CRC Press LLC, 1997.

[44] B.M. Thuraisingham, *Data Mining: Technology, Techniques, Tools, and Trends.* Boca Raton, Fla.: CRC Press LLC, 1999.

[45] J. Wineland, "The Naval Warfare Tactical Data Base (NWTDB) Data Standardization Process," *Proc. Dept. of Defense Database Colloquium,* pp. 3-8, Aug. 1994.

[46] http://www.darpa.mil/DARPATech2000/presentation.html, 2002.

Marion G. Ceruti received the BS degree from the State University of New York at Stonybrook in 1973 and the PhD degree from the University of California at Los Angeles in 1979. For 21 years, she has been employed as a scientist at the Space and Naval Warfare Systems Center, San Diego, CA,. Her professional interests include information-systems research and analysis, knowledge management, artificial intelligence, data mining, and real-time systems. Dr. Ceruti is the author of more than 70 journal articles, conference proceedings, monographs, and book chapters on various topics in science and engineering. She received four Navy publication awards for her work concerning object-oriented technology and joint command and control database systems. Dr. Ceruti is a senior member of the IEEE, and a member of the IEEE Computer Society, the IEEE Systems, Man and Cybernetics Society, the Armed Forces Communications and Electronics Association, and the New York Academy of Sciences.