NPS-CM-03-006

ACQUISITION RESEARCH SPONSORED REPORT SERIES

Auto-Redact Toolset for

Department of Defense Contracts

30 September 2003

by

Ron B. Tudor, J.D., Lecturer

ACQUISITION RESEARCH
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

The Acquisition Chair, Graduate School of Business & Public Policy, Naval Postgraduate School supported the funding of the research presented herein. Reproduction of all or part of this report is authorized.

To request Defense Acquisition Research or to become a research sponsor, please contact:

NPS Acquisition Research Program
Attn: Jim Greene, RADM, USN, (Ret)
Acquisition Chair
Graduate School of Business and Public Policy
Naval Postgraduate School
555 Dyer Road, Room 332
Monterey, CA 93943-5103

Tel: (831) 656-2092

Fax: (831) 656-7699

e-mail: jbgreene@nps.edu

www.nps.navy.mil/gsbpp/acqn

ACQUISITION RESEARCH
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Abstract

This paper addresses the Auto-Redact initiative associated with the compilation of electronic copies of awarded Government contracts. The advancement of electronic systems allows for unlimited data storage capability; it also allows for the quick and easy access to all the stored data, and can make that data immediately available to the public. However, data stored by the Government is subject to statutory guidelines. Chief among these is the Freedom of Information Act (FOIA). By creating these databases, the Government has created records that are subject to release to the public under the Electronic Freedom of Information Act (EFOIA). In doing so, the Government must take care to safeguard information that may not be otherwise releasable. Under FOIA, if an Agency decides to not release information that it has within its databases, it must submit that decision to not release information to an Initial Denial Authority.

With the depth and breadth of electronic databases or data warehouses, an ability is needed to automatically identify and classify data so that it can be automatically redacted (Auto-Redact) and not be released under FOIA. The solution for protecting critical operational data while making all other data available to the public is to create an architecture for recognizing the data within the various documents used in the contracting process. To do so the data must be characterized as to its nature, whether it is operational (requiring protection from release), or otherwise protected from release under a FOIA exemption or another statute, and then the data must be homogenized so that it is readable, or capable of being protected, across any document or data warehousing system. Doing this with data also converts the data into a form that allows the data to be manipulated and used for various official purposes.

The proposed solution within this paper is a non-traditional approach to data characterization and handling. The resources to establish the architecture are relatively minor and can be accomplished in a relatively short time.

About the Author

Ron Tudor has been on faculty teaching government contracting, contract law and fiscal law at the Naval Postgraduate School in Monterey, California in the Graduate School of Business and Public Policy since December, 2000. He has also been the Principal Investigator for researching advanced procurement methods for the Federal Government; and for advanced database management, data recognition, and characterization. Ron started his career with the Government in 1984 when he entered active duty with the U.S. Army at Fort Bliss, Texas in the Office of the Staff Judge Advocate. While there he was the contracts and administrative law attorney. In 1995, he transferred to Germany where he was the contracts attorney for the southern half of Germany for the United States Army Contracting Command, Europe, which serviced the European Command, the United States Army, Europe, and V Corps. In 2000, he became the contracts attorney for the Italian theatre for the U.S. Army and the Southern European Task Force. Ron is currently a Lieutenant Colonel in the U.S. Army Reserves in the Judge Advocate General's Corps and is serving as the Deputy Staff Judge Advocate at the 5035th Garrison Support Unit, Fort Bliss, Texas.

Table of Contents

NPS-CM-03-006

ACQUISITION RESEARCH SPONSORED REPORT SERIES

Auto-Redact Toolset for

Department of Defense Contracts

30 September 2003

by

Ron B. Tudor, J.D., Lecturer

I. Introduction

This paper addresses the Auto-Redact initiative associated with the compilation of electronic copies of awarded Government contracts. The advancement of electronic systems allows for unlimited data storage capability. It also allows for the quick and easy access to all the stored data, and can make that data immediately available to the public. However, data stored by the Government is subject to statutory guidelines. Chief among these is the Freedom of Information Act (FOIA). By creating these databases, the Government has created records that are subject to release to the public under the Electronic Freedom of Information Act (EFOIA). In doing so, the Government must take care to safeguard information that may not be otherwise releasable. Under FOIA, if an Agency decides to not release information that it has within its databases, it must submit that decision to not release information to an Initial Denial Authority.

With the depth and breadth of electronic databases or data warehouses, an ability is needed to automatically identify and classify data so that it can be automatically redacted (Auto-Redact) and not be released under FOIA. The solution for protecting critical operational data while making all other data available to the public is to create an architecture for recognizing the data within the various documents used in the contracting process. To do so the data must be characterized as to its nature, whether it is operational (requiring protection from release), or otherwise protected from release under a FOIA exemption or another statute, and then the data must be homogenized so that it is readable, or capable of being protected, across any document or data warehousing system. Doing this with data also converts the data into a form that allows the data to be manipulated and used for various official purposes.

Although the Auto-Redact project is a subpart of the Navy – Air Force Interchange (NAFI), the concepts of database management apply to any other system used to maintain data. The NAFI is an attempt to load all data from all publicly awarded contracts into an electronic database so that there is complete visibility to everyone.

The Small Business Administration is pushing this project so that small businesses will have greater opportunities. The intent is to increase visibility of projects so that small businesses can obtain a greater percentage of Government business. However, some issues are created by this public release of contract information.

Current data technology does not support the concept of a pure "auto-redaction" solution. However, software is evolving away from monolithic, hard-to-maintain masses of code toward smaller components that communicate with each other to complete particular tasks. This migration potentially provides the solution for "auto-redaction". Flexibility among various data sources can provide an effective use of these types of software components, and also provide for wrapping of applications that by themselves will not support this concept. Substantially, this need for flexibility and componentization is driving the increasing adoption of object-oriented technologies that can support software applications and objects written in any language on any platform. These applications and objects are bound only by the common Data Access Language (DAL) of the underlying software infrastructure. In addition, the trend in software is to "hotlink" documents or systems together thereby exponentially expanding the available data in any given "system". Setting a standard for data access across all data systems is the solution. Although it is not yet available, the solution is very close and is exactly what is called for in the E-Government Act of 2002 (H.R. 2458).

With the vast differences among "data systems", software applications, and software integrations, a single database, or even a data ware-house, cannot hope to encompass all the potential data available through the interconnected systems of the future. All these current, and future, components must work together in a "Network-Aware" environment. In a network-aware solution that brings multiple systems together there are more components than in the standard three-tier system model: network definitions, database definitions, data source definitions, data type definitions, rules, transactions, data sets, interfaces and user interfaces. Providing a standard method for defining these components, and a standard way of describing their interaction will

ensure both short and long term flexibility. It will also structure all data and all systems to comply with the mandates of the EFOIA and other data-mining requirements.

II. The Freedom of Information Act

Regardless of the complexities associated with data handling, the EFOIA mandates that all data within electronic systems be released, or denied under one of its exceptions. It is important to understand FOIA, its history, and its legislative history to understand the EFOIA.   FOIA was passed in 1966 by Congress (Congress revised the FOIA in 1974, 1976 and 1986 before it enacted the electronic amendments in 1996).   It created a philosophy of full disclosure of information that was enforceable by the courts. ((But see U.S. Dep't of Justice v. Reporters Comm., 489 U.S. 749, 774-75 (1989) (holding that the "central purpose" of the FOIA is to disclose only those records that directly shed light on the operations of government.)) The Act applied to all "records" held by Federal Agencies, and required that they be made available to the public and placed the burden of justifying nondisclosure on the Government. (See 5 U.S.C 552(a)(4)(B)(b) (1994). See also National Labor Relations Bd. v. Robbins Tire & Rubber Co., 437 U.S. 214, 234-236 (1977); Environmental Protection Agency v. Mink, 410 U.S. 73, 79, 87-88 (1973)).

The FOIA recognized that citizens in a democracy need access to information within Government records so that the citizenry can make informed decisions. (H.R. Rep. No. 89-1497, pt. 1 (1966), states, "A democratic society requires an informed, intelligent electorate, and the intelligence of the electorate varies as the quantity and quality of its information varies... [The FOIA] provides the necessary machinery to assure the availability of Government information necessary to an informed electorate.")

The FOIA prevents politicians and Government employees from being the decider of what information the public is given access to. Congress also recognized that there were rightful reasons to keep some information secret. As such, Congress created nine exemptions, under which Federal agencies could refuse to disclose information. (See 5 U.S.C. 552(b)(1)-(9)(1994). Briefly stated, the FOIA does not apply to matters that fall under the categories of (1) classified information and national

security, (2) internal agency personnel information, (3) information exempted by other Congressional statutes, (4) trade secrets and other confidential business information, (5) agency memoranda, (6) disclosures that invade personal privacy, (7) law enforcement investigation records, (8) reports from regulated financial institutions, and (9) geological and geophysical information.)

The four amendments to FOIA (in 1974, 1976, 1986, and 1996) also deserve a general discussion because the amendments showed Congress' intent for the FOIA to represent a broad policy of full disclosure. Congress amended the FOIA in 1974 with the intention of strengthening the statute because there was a general reluctance by agencies to comply with the law's policy of full disclosure. Federal agencies had been interpreting the exemptions broadly to justify withholding documents, and officials often used various ploys to discourage use of the FOIA, including high fees for copying documents, long delays, and claims that they could not find the documents requested.

The 1974 amendments required agencies to respond to information requests within ten days or face a lawsuit, (See 5 U.S.C. 552(a)(6)(A)(i) (1994)) and directed each agency to issue FOIA fee regulations for the recovery of only the direct costs of search and duplication. (See H.R. Rep. No. 93-1380, at 7 (1974).) A key revision authorized federal judges to conduct an in camera review of classified information in order to confirm that the requested materials actually fell within the guidelines of Exemption 1, the national security exemption. (See 5 U.S.C. 552(b)(1)(B) (1994)).

In response to a 1973 Supreme Court decision, Congress revised Exemption 1. See Environmental Protection Agency v. Mink, 410 U.S. 73 (1973). In deciding Mink, the Supreme Court interpreted Exemption 1 broadly and held that classified documents were exempt from judicial review. Congress overrode the Mink decision because legislators believed the Court's ruling conflicted with the general philosophy of full disclosure. (See H.R. Rep. No. 93-1380, at 11-12 (1974); S. Rep. No. 93-1200, at 12 (1974)).

Congress amended the FOIA for the second time in 1976 because legislators wanted to clarify Exemption 3 (in response to Administrator, FAA v. Robertson, 422 U.S. 255 (1975), which held that the FAA had wide discretion to withhold Government records). This exemption provided that the FOIA did not apply to information clearly exempted by other laws previously passed by Congress. These revisions to Exemption 3 created guidelines that strictly limit the discretion of an agency's executive to withhold information from the public. (H.R. Rep. No. 94-880, at 23 (1976)). This change is worthy to note because, by expressly limiting agency discretion for withholding, the amendment reflected a congressional FOIA policy that favored disclosure. (H.R. Rep. No. 94-880, at 23 (1976))

In 1986, Congress revised the FOIA for the third time when legislators amended the Act by passing the Freedom of Information Reform Act of 1986. (See 5 U.S.C. 552(b)(7) (1994)). The amendment provided broader exemption protection for law enforcement information and added new exclusions for law enforcement records under Exemption 7. A larger impact from this amendment was the change to the fee structure. Under the new fee guidelines, the Government could only recover a portion of the true cost of responding to the FOIA request. (See Long v. Internal Revenue Service, 596 F.2d 362, 366-67 (1979) holding that the expenses of editing computerized records cannot justify an agency's decision to refuse to segregate disclosable materials subject to the FOIA).

The three amendments in 1974, 1976 and 1986 clearly show that Congress intended to open up Government files to the public, and that the exemptions were to be strictly construed. Administrative secrecy is not tolerated, and the interests of the public in accessing Government information are paramount. The EFOIA amendments in 1996 continue this broad policy, but also apply it to electronic records, something that the original FOIA in 1966 could not have contemplated.

III. The Electronic Freedom of Information Act Amendments

There has been an explosion of computing power since the 1970s. The concept of the "mainframe" has given way to powerful desktop computers connected by the Internet. In addition, data systems have grown beyond what anyone could have conceived in 1966 when FOIA was created.

In the 1970s, some FOIA requests were denied for information stored in electronic formats. (See Dismukes v. Department of the Interior, 603 F.Supp. 760 (D.C. 1984); SDC Dev. Corp. v. Mathews, 542 F.2d 1116 (9th Cir. 1976)). The requests were denied as not qualifying for disclosure under the Act. The 1996 amendments established that the rules for public access under FOIA apply equally to electronic records and paper records, and a search request for electronic records using software is to be treated the same as a paper search. (See Electronic Freedom of Information Act Amendments of 1996, Pub. L. 104-231, 110 Stat. 3048, 3049, 4 (1996), amending 552(a)(2)). The new law stated that a "record" that is subject to the FOIA comprises information maintained by an agency in any format, including an electronic format. (See Electronic Freedom of Information Act Amendments of 1996, Pub. L. 104-231, 110 Stat. 3048, 3049 3(2) (1996), amending 552(f)). Under the EFOIA, agencies must make reasonable efforts (1) to provide a record "in any form or format requested by the person if the record is readily reproducible by the agency in that form or format," (See Electronic Freedom of Information Act Amendments of 1996, Pub. L. 104-231, 110 Stat. 3048, 3050, 5(B) (1996), amending 552(a)(3)), and (2) to maintain records "in forms or formats that are reproducible" so that requests for the information can be honored. (id.) THE LAW ALSO MANDATED THAT WHEN AGENCY OFFICIALS REDACT PARTS OF AN ELECTRONIC RECORD BECAUSE THE INFORMATION IS DETERMINED TO FALL WITHIN ONE OF THE NINE EXEMPTIONS, THEY MUST NOTE THE LOCATION AND THE EXTENT OF ANY DELETIONS MADE ON THE ELECTRONIC

RECORD. See Electronic Freedom of Information Act Amendments of 1996, Pub. L. 104-231, 110 Stat. 3048, 3053, 9 (1996), amending 552(b).

With the 1996 amendments, the EFOIA specifically applied FOIA to electronic data systems. This change was significant in that it recognized the evolution of technology and the means by which data was stored using automated systems. The section states that the FOIA is amended as follows:

"(f) For purposes of this section, the term --

(1) "agency" as defined in section 551(1) of this title includes any executive department, military department, Government corporation, Government controlled corporation, or other establishment in the executive branch of the Government (including the Executive Office of the President), or any independent regulatory agency; and

(2) "record" and any other term used in this section in reference to information includes any information that would be an agency record subject to the requirements of this section when maintained by an agency in any format, INCLUDING AN ELECTRONIC FORMAT. (Emphasis added.)"

Initially, the courts took the position that electronic data storage systems did not fall under FOIA. The seminal case in this area is SDC Development Corporation v. Mathews, 542 F.2d 1116 (1976). This case involved an electronic data system established by the National Library of Medicine. The Agency established the database to aid research and charged fees for access to it. SDC Development Corporation submitted a FOIA request to obtain the database in its electronic form. It was clear that SDC wanted to use the database for commercial purposes and the FOIA fees were far less than the use access fees the National Library was charging. The ninth circuit sided with the National Library and ruled that the electronic database was not a record under FOIA, and particularly noted that FOIA did not define what a "record" was. The court also recognized that SDC was attempting a commercial use of a government database

and it should pay the charges that all other users were paying. Congress submitted the EFOIA act amendments largely in response to SDC.

There have been a number of court decisions since enactment of EFOIA in 1996. Of these, the most recent FOIA case is R&W Flammann GmbH, v. United States, 2003 U.S. App. LEXIS 16171, August 7, 2003. This case concerned the release of pricing under a FOIA request by a competitor company on a new solicitation. Flamman was the incumbent contractor and the District Court ruled that a contracting officer erred when he released pricing information on Flammann's contract base year and option years after the Government decided that it did not want to exercise the option years. The contract was resolicited and a competitor filed a FOIA request to obtain all the pricing information on the previous contract. Flamman initially filed an Agency protest, which was denied, but then proceeded to the District Court and surprisingly obtained an injunction against award of the new contract. The Circuit Court reviewed the case and reversed the District Court. It ruled that the release of pricing information under the FOIA request by the competitor was in accordance with FOIA and the Trade Secrets Act. (The Trade Secrets Act, a criminal statute, bars government officials from disclosing or making known to any extent not authorized by law numerous categories of information, including confidential and trade secret information. 18 U.S.C.S. § 1905.) The Court specifically noted "that when a sealed bid was available to the public, it entered the public domain and was therefore not confidential under Exemption 4 of FOIA." Under Flamman, it is clear that pricing information on awarded contracts is releasable.   This should cause an additional concern to arise among activities that maintain electronic data bases of price information such as the DoD EMALL (The EMALL is operated by the Defense Logistics Agency (DLA) and provides the ability to order goods and simple services through an electronic system.) Although it has not been challenged yet, the pricing structures of competitors that the EMALL protects may be releasable under EFOIA.

Since EFOIA was enacted, other than Flamman, four cases have addressed it in ways that relate to the auto-redaction capability. The courts have ruled that electronic records are subject to the EFOIA, O'Kane v. U.S. Customs Service, 169 F.3d 1308, that internet addresses themselves are not "records", Essential Information v. U.S. Information Agency, 134 F.3d 1165, that agencies have the authority to mandate submission of data and documents in electronic form, United Transportation Union v. Surface Transportation Board, 132 F.3d 71, and that agencies are required to comply with EFOIA provisions, Public Citizen v. Raines, Civ. no. 96-1194 (NHJ) (DDC Nov. 27, 1996). It is this last area that potentially causes the greatest difficulty for the Navy - Air Force contract data system. With the increase in network usage and data systems, EFOIA will require ever increasing access for people that want information.

In this last post-EFOIA suit, filed in the U.S. District Court for the District of Columbia, Public Citizen, a public interest group, filed suit against seven federal agencies asking the Court to order the agencies to comply with Section 11 of EFOIA. This section directs federal agencies to make reference materials and guides available to the public on the Internet. The goal was to enable FOIA users to discover exactly which agencies possessed records that users were seeking and to understand how to request the desired records from these agencies' systems. The section specifies three distinct categories of reference information that can make FOIA access easier and faster: (1) A FOIA handbook that explains how to obtain information from an agency, (2) an index of all major information systems maintained by an agency, and (3) a description of any major record-locator systems maintained by an agency. The first of these categories, the handbook, is clear and self-explanatory. The second requirement, the index, is a listing of the various types of information held within information systems. In other words, this is a content listing. This requirement applies directly to NAFI or any other type of system used to store government records. The content of any system must be indexed so that users may clearly identify what information is stored in the systems. The third requirement mandates a description of the various locator systems.

This section makes it absolutely clear that systems must be visibly available to searchers, and can best be thought of as an organizational listing of systems.

While these requirements might be viewed as overly exposing government information systems, they can actually assist agencies in terms of savings. FOIA requests are often filed across several organizations or agencies. These duplicate requests consume time and money, neither of which agencies have in abundance. Having content-based and organizational-based FOIA search systems can actually save time and money for the agency. Regardless of whether there are savings for agencies under EFOIA, the openness required by the statute makes operational security critically important.

IV. SOFTWARE APPROACH

Under the NAFI, all contracts and data within it are releasable through FOIA. This potentially creates a major security problem. DoD's mission is world-wide and dependent on contractor support. If all DoD contracts are loaded into an electronic database, those contracts can be data-mined for operational information. It is entirely possible for an enemy to chart purchases by various major, unrelated subordinate commands and connect that information. For example, an out-of-the-ordinary purchase of plywood by DLA (plywood is the construction material of choice for contingency operations) combined with increased buying by a major installation or unit is a good indicator that activity is bound for a contingency operation. Applying world events as a third data element provides a clear picture of where the unit is going. Knowing the operational mission of the unit provides the overlay for the scope of operations; thereby, putting our soldiers, sailors, airmen, and marines at risk.

The NAFI, and any potential database of contracting information across DoD, will contain an extremely large number of contracting actions. Further, if Indefinite Delivery / Indefinite Quantity contract task orders or delivery orders are considered, the scope of the database is so large that no single activity can hope to redact all elements of data that should be protected. Auto-redaction provides the solution for preventing this disclosure of information. By identifying what blocks, or types, of information should not be released, an automated system can redact that information across all known documents. However, the primary methods for inserting copies of contracts into an electronic database are methods that use files such as Adobe PDFs (or other specific file formats such as .doc, .tiff, .jpg, .txt, etc.), or transmit documents directly from the Standardized Procurement System (SPS) or other electronic contract writing systems.

However, current data recognition technology does not have the capability to interact with level I data. (Level I data is a flat file with data that is not interactive. For example, the information on a credit card statement identifies how much money was

spent, where it was spent, and what day it was spent on. Level II data potentially provides additional data but it still cannot be electronically accessed or manipulated. Level III data provides the "bar code" for every data element and makes each element system identifiable and suitable for manipulation. To conduct auto-redaction functionality across all the potential data elements, the data must be Level III. If this format is applied to an Adobe PDF (which is widely used in DoD), the PDF becomes readable across all its potential data elements. Each data field becomes recognizable to the auto-redaction process and is protected from release.

A. THE CHALLENGE OF ACHIEVING THE GENERAL SOLUTION

We are more likely to achieve a general approach to the problem by defining the general solution, and identifying restrictions to be applied either because of specific attributes of contracting or to manage overall solution cost, than by defining a contract-specific solution, and then trying to see how to extend that specific solution to the general case. The challenge, of course, is to do a good job up front of defining this more general problem space, and the attendant solution.

Fortunately, the challenge of meeting the general solution has many attributes shared by the specific solution. A key challenge is that contracts exist in a wide variety of systems, in a wide variety of formats, with a wide variety of methods of access. Ideally, a single, coherent framework for accessing contract information would be available. This fundamental problem of disparate systems with different types of data access and different formats of equivalent data is one that is shared across all documents and data in all systems within any large organization, and is certainly true of data outside of contracts within the DoD and Federal Government. It is also fortunate that addressing this disparity has been identified as a key initiative for the Federal Government: the E-Government Act of Dec. 2002 (H.R. 2458) and the DoD Transformation Guidance Planning Act of 2003 state clearly a demand for a near paramount focus on bringing coherence to the data managed by systems and applications throughout the government, military and intelligence community.

There are two traditional approaches to bringing together data from multiple systems: integration and "Data Warehousing". The former involves creating intermediate systems that either accept or grab data out of one system, and push it into another. Integration solutions are effective for homogenizing data within different systems, but do not solve the redaction problem. Data Warehousing solutions collect subsets of data and compile them into a summary database from which reporting can be accomplished. Although the basic approach of collecting data from multiple systems and presenting it is certainly applicable, specific data warehousing applications are not. Still, the basic techniques required to address the general solution of EFOIA requests, etc., will be closer to a data warehouse solution than an integration solution. However, no Commercial Off-The-Shelf (COTS) products seem to provide a complete, or near complete solution to the problem.

B. SOLUTION CHARACTERISTICS

Given the absence of a COTS product, the next logical step is defining the characteristics that must be present in a unifying environment. Surprisingly, though, the challenge of finding a unifying set of characteristics that can act as the basis for the solution does not appear as daunting as it might be. By looking at the commonality among a representative group of electronic systems, we can readily piece together the fundamental concepts and mechanisms that necessarily underpin a solution that provides both extensibility and coherency. Fortunately, technology and standards have advanced substantially in the last few years, and may very well be at a point now where a solution can be cost-effectively developed based on these concepts and mechanisms, a solution that can simultaneously aggregate data from multiple, disparate systems, as well as provide a platform that provides easy, coherent, and consistent access to all of their associated information.

1. Separation of Data and Presentation

a. Fundamental to providing a comprehensive, coherent solution is a rethinking of the concept of "data". Data is more than just the information collected by a particular

application: it is also the glue that brings different applications together in a coherent fashion. However, data by itself is not useful: it must be presented in a way that promotes problem resolution. Moreover, when one is looking at data from different systems or different applications, particularly when it will be put to yet another use, then a particularly important principle is the separation of data and presentation (or form). This separation is a central principle for achieving extensibility. Unfortunately, many of the most common applications available on the market today treat content as a single object (e.g., .doc, .tiff, .jpg, .txt, etc.). A far better approach for achieving this architecture is to separate the data comprising the body of the document and the instructions that define how the data is represented, and to create some other method of bringing these together. While this concept is touched upon in specifications such as XML, there appears to be no application available on the market today that treats data and presentation as two separate and equally viable components of providing information to users. Even applications like ERP and CRM that have data, and present that data in multiple ways, have the presentation programmed and tied explicitly to the types of data on which it can function. Again, this usable, or Level III, data exists within very few systems. Mostly, if it exists at all, it is a result of the second tier, or logic (software) process of systems. It is rarely a function of the database itself.

b. Fundamentally, any software designed to address the critical issue of auto-redacting DoD or Federal contracts, and ultimately any other document covered by the EFOIA, must meet three key requirements:

1) Cost Savings – It must decrease the cost of meeting EFOIA requests,

2) Access Limitations – It must ensure that all releasable, and nothing but the releasable information, is made to available to the public under the EFOIA, and finally,

3) Flexibility – It must provide for an environment that can be changed to ensure that both requirements (1) and (2) continue to be met, even as the definition of what constitutes (2) may change due to new or revised legal opinions or Congressional legislation.

Meeting these three requirements is the most critical characteristic of any solution. However, the contract data (and other electronic information collected by the government) has great potential value that goes well beyond meeting EFOIA requirements. For instance, by properly indexing, aggregating, and presenting contract data in a controllable manner, government agencies can streamline and improve their functioning in other ways: (a) by providing the means to meet other types of information requests, e.g., Congress; (b) by enhancing government procurement by collecting information on previous, similar engagements; and (c) by enabling more effective coordination between vendor and agency, etc.

The reality, though, is that EFOIA does not apply simply to contracts, but rather applies to all government documents, and the same extensions described above that make sense for contracts also make sense for any other kind of document, whether it be patent applications, or EFOIA requests themselves. The ideal is an approach than not only meets these requests for contracts, but can also be extended to form a general approach for managing government documents for purposes of EFOIA, information requests, and optimization.

Providing for these additional uses imposes additional requirements on the ideal solution:

4) Ubiquitous access – it must be accessible from anywhere, ideally over the Web,

5) Search – it must permit identification of contracts or collections of contracts meeting specific criteria, including key word searches,

6) Multiple access specifications – it must provide for flexible specification of control requirements and redacting rules to support information control for purposes other than the EFOIA,

7) Single point of entry – it must be able to aggregate contract information so that searches and queries can be made from the multitude of servers and solutions currently containing documents,

8) Presentation alternatives – it must provide the ability to customize presentation formats so that the documents presented provide information in a way that is consistent with alternate usages,

9) Electronic and Paper media – it must support aggregation of both electronic and paper documents, and

10) Security – it must authenticate each user to prevent unauthorized access.

Although these appear to be a number of additional requirements, numbers (4), (5), (7) and (9) are implied requirements for the overall system, given the case law and legislation, (6) is a requirement for meeting (3), and (10) is an implicit requirement for meeting (2). In effect, the only "extra" capability needed to support multi-use contract information access is (8) extending the presentation capabilities of the system. Given the full value that multiple uses could represent, requiring this extended presentation capability is the only logical choice. Thus, items (1) through (10) are the high-level system requirements. A more detailed look at these requirements is presented in the following paragraphs:

C. DETAILED REQUIREMENTS OF THE GENERALIZED SOLUTION

That the system meet requirement (1) is fundamental: the cost of manually redacting information is the primary driver for moving to an automated system. There are a number of expense drivers that must be considered when looking at the cost of the system:

1) Cost Management

a. Cost of adding, changing, or removing contracts - maintaining a single, duplicate contract database is not a viable solution. Rather, contracts must be maintainable within their native systems, and then be automatically aggregated into a central index for analysis and collection;

b. Cost of administration - it is particularly important that adding new users, and granting the limited access specified by the EFOIA, be performed without human intervention. In addition, users with less limited access must be easily set up;

c. Support for electronic generation of data, in addition to hard copy - providing data electronically is often a more effective alternative, particularly when large volumes of data are required; and

d. Ease of transforming paper documents into electronic documents - since redacting documents in paper format cannot be automated, a cost effective way of putting paper documents into a suitable electronic format is a long-term requirement.

Providing limited, secure access is at the very foundation of software capability. The specific needs are as follows:

2) Access Limitations [also (6) and (10) above]

a. Contract and contract elements access limitations

Minimally, the system must be able to identify that certain elements of the contract, e.g. pricing and trade secret information, cannot be viewed by anyone except authorized personnel. However, in the context of multiple uses, control of contract elements has to be more flexible: certain elements have to be visible to some users, and not others. Moreover, the degree of access may depend on the contract itself. For example, a government contractor should get full access to the contracts on which he is a principal, but only the EFOIA-level of access to contracts in which he has no role.

b. Access Limitations Definition

Requirements (1), (2), and (3) collectively imply that access limitations be defined administratively (by an administrator), not programmatically (by a change to the program or software), and that changes to access can be made without changes to the underlying data. Moreover, to aid in minimizing the cost of this administration, a set of user classifications and access specifications per those classifications is required, along with a way of specifying the relationship between the organization and role to which a user belongs, and the organizations and roles that are participating in a contract.

c. Authentication

Once the system provides differential access to differential users, users must be authenticated to determine their level and breadth of access. Users who do not have any access defined for them will automatically be limited to the access provided under the EFOIA.

d. Access Rules

Simply limiting access to contracts or elements of contracts or both is insufficient to provide all the redaction necessary. The actual content of certain elements can drive redaction decisions. A method for specifying the characteristics of elements of contracts, and limiting access to both the elements and the contracts based on those characteristics, is required.

Providing a cost effective solution which meets the flexible redacting rules imposes a requirement for certain capabilities on the system.

3) Flexibility

There has been a substantial amount of standardization imposed on contracting and contract documentation over the last few years. Despite that, redacting standards are always subject to change by the Courts and Congress, and by decisions made by government agencies. This potential change requires flexibility in a number of ways:

a. Multiple Contract Styles

Two of the major award formats are Standard Form (SF) 33 and SF 1442. The existence of these two formats, and others, implies that [1] the system must be able to support multiple contract formats, [2] given requirements (1) and (2) above these formats must be managed administratively, not programmatically, and [3] the flexibility to support other types of contracts administratively is also required.

b. Interchangeability of Contract Styles

Not only must multiple contract styles be supported, but behaviors associated with contracts must behave consistently across styles. Thus, common content in a SF 33 or a 1442 must be consistently identified, even when the underlying form uses inconsistent terminology. Moreover, field level searches must approach common, inconsistently named content as though fields were defined consistently.

c. Changing Standards

This demand for interchangeability must be considered within the reality of changing standards. Thus, flexibility implies two requirements: (1) the ability to specify contract form styles and (2) the ability to define relationships between differently-named common elements in different contract styles administratively, not programmatically. Moreover, the ongoing evolution of standards can deliver contract management advantages by providing for the standardization of certain types of attachments. The ability to create specific forms to represent these types of attachments will further enhance the value of the system as a multiple-use contract information management platform.

Ideally, information provided under the EFOIA and for other uses would be easily accessible:

4) Ubiquitous access

Given the penetration and ease of use of browser-based capabilities over the Internet, providing data through these means is the only logical alternative. Such access should provide for on-screen presentation, printing, and downloading of accessed data.

Given modern standards for information access an effective search capability is required, both because of the mandate of EFOIA, established by Congress and the Courts, but also for the other uses to which the data could be applied. Note: this requirement has strong implications given requirement (7).

5) Search

a. Keyword Search

Providing capabilities to find out information about contracts based on keywords is mandatory.

b. Element Search

For many other uses the ability to find information based on specific information within the contract is also required, e.g. being able to find all contracts for buying torpedoes or any other commodity. This requirement differs from the requirement in (a) above, in that a user might be interested in torpedo contracts, but would not be interested in sonar detection systems for torpedoes that one might find using a torpedo keyword search.

c. Attachments and Search

Basic contract information must be searchable, but attachments associated with the contract must also be searchable.

Currently, any request that covers more than one organization within a single agency is difficult to fulfill. From the EFOIA viewpoint, this difficulty is problematic, but it is even more so for other uses of contract information. For instance, if one were to try to analyze all purchases of a particular component by the DoD, one might have to search hundreds or thousands of servers. Thus, a capability to aggregate data on diverse systems is required:

7) Single Point of entry

a. Cross-Server Aggregation

Data stored on different servers must be accessible through a single point of entry. Given the requirement for search capabilities (5), this requirement cannot be met by simply allowing the user to navigate through a series of links between these systems: an indexing scheme that brings contracts and contract elements that meet search criteria together for analysis is required.

b. Transformation

Unfortunately, these disparate systems also have differential element definitions, and different names for the same element. For instance, "Lockheed" might be "Contractor" 1091 on one system, and "Vendor" LH202 on another system. Thus, any search across systems requires special processing, and special capabilities within the system. For purposes of keeping costs under control, the system must support this requirement administratively, not programmatically.

c. Common representation

Government documents are in a variety of systems in a variety of formats: contracts are stored in databases, file systems, document management systems, etc. Thus, this system must be able to aggregate data from multiple, disparate types of systems.

Different uses implies different types of presentation, e.g. a single contract being provided under the EFOIA might take the form of a formal contract document, but all of the contracts for torpedoes would be more usefully presented as a tabular listing suitable for import into Microsoft Excel:

8) Presentation alternatives

a. Document Presentation

For certain collections of information, e.g. a single contract, a document representation is required.

b. Listing or Table Presentation

For other collections of information, e.g. all torpedo purchases over the last year, a tabular presentation is more appropriate.

c. Automatic Indexing

All areas containing information must be indexed and all redacted information areas must likewise be identified.

Although a lot of the information required to meet the FOIA is electronic, contracts still exist in paper form. Rather than hand redacting these items, a better approach is creating an electronic representation of this data:

9) Electronic and Paper media

A capability for scanning paper documents and then interpreting their content and form in order to put them into an electronic contract structure is a requirement for the system.

Network-Awareness

A centralized (centralized from the perspective of the user) "console" must bring together data from systems from multiple places on the network, and do so seamlessly. Today, network capabilities are sufficiently robust that the technical challenges of network-awareness appear to be behind us: the Internet is everywhere. Now, the challenge is to structure data (at least Level III) so that all systems, including legacy systems, can interface with all other systems without costly third party integration systems.

Global Uniqueness

This example also illustrates another critical characteristic of the solution: it must provide a structure for uniquely identifying on a global scale the actual data and documents. We cannot have global access to contract information if contract numbers from one system supercede unrelated contracts in another system. The solution is a standardized format or structure for all data.

Open Standards

In the last few years, interoperability has been accelerating at an unprecedented rate. Much of this acceleration can be attributed to one of the unexpected developments of the past decade: the unprecedented spread of TCP/IP, XML, MP3, and other nonproprietary networking and data communications standards known collectively as the Internet. Despite the dotcom meltdown, open data communications standards continue to evolve and take hold, creating the foundation for some of the most exciting new applications of over-the-horizon computing. Taking advantage of this rapid outgrowth in capabilities is the only logical direction.

Inherent Security

External threats to systems security, coupled with growing terrorist threat concerns, must figure prominently in this solution. The security requirements must have the following characteristics:

1) It must ensure that users who access the system are properly identified, and then accorded the actual privileges to which they are entitled;

2) It has to provide for security over the actions that the user can perform. Not all users have equal authority, but programming a unique set of capabilities for each type of user is expensive in both the short and long term.

3) It must limit access to components of documents and data (the EFOIA allows everyone to see most of the documents that the government collects, but it is well-established that access to a document does not imply access to the whole document);

4) It must restrict access to the particular documents and data that the user is entitled to see (EFOIA recognizes that some information has so great a security or privacy implication that access to the document or transaction as a whole must be limited);

5) It must make secure any communications that occur over public or semi-private networks to ensure that information that is properly accessed by users is only accessed by those users (encryption techniques are fundamental to this process); and

6) It must also support "rules-based" redaction, analysis of the actual data that thus requires redaction.

Flexible, "Task-Oriented" User Interface

Although historically the standard, hard-coded "one-size-fits-none" user interfaces cannot achieve the results required to support the rapidly changing systems and legislative environment in which this solution must function, the ability for users, in addition to administrators, to customize and personalize their experience, not only on their office computer, but ultimately from their home computer, cell phone, or PDA should be an inherent characteristic of this solution. Although current technologies do not fully support these capabilities, much of the systems-related development that is occurring is directed at just this issue. Thus, the solution should support seamlessly taking advantage of these new capabilities as they are released. However, in the short-term, office-computer-based flexibility should be delivered.

V. Summary

The analysis above has established the requirements and capabilities of a new data architecture. Rather than establishing a new "system" or software package, it presents the concept for a new way of looking at and manipulating data. By changing the architecture or format of data, rather than changing the software or logic approaches of "systems", all such "systems" can be incorporated into a single network-aware environment wherein all legacy systems, data warehouses, and internet applications can read all data sources. This is not merely a recommendation for a standardized format for data, it is a description of an entire new architecture for recognizing, reading, storing, and manipulating data across diverse systems. While this type of architecture does not currently exist in a COTS application, the effort associated with establishing such an environment is relatively low. It is not an issue of building something new; rather, it is a changing of a viewpoint on how data is recognized, characterized, homogenized, and used.

VI. Recommendations for further research:

a. Determine capabilities of the market place to meet the requirements as described above.

b. Determine willingness of software companies to develop this proposal into a viable architecture at no cost to DoD under a Cooperative Research and Development Agreement under the authority of the U.S. Federal Technology Transfer Act of 1986 (Public Law 99-502, 20 October 1986, As Amended.)

c. Develop a pilot demonstration project.

FY 2004 Sponsored Acquisition Research Products

Sponsored Report Series

NPS-CM-04-001   Update of the Navy Contract Writing
December 2003

NPS-CM-04-002   Marine Corps Contingency Contracting MCI
December 2003

FY 2003 Sponsored Acquisition Research Products

Sponsored Report Series

NPS-AM-03-003   Centralized Control of Defense Acquisition Programs:
A Comparative Review of the Framework from 1987 – 2003
September 2003

NPS-AM-03-004   Reduction of Total Ownership Cost
September 2003

NPS-CM-03-006   Auto-Redact Toolset for Department of Defense Contracts
September 2003

Working Paper Series

NPS-CM-03-002   Transformation in DOD Contract Closeout
June 2003

Acquisition Case Series

NPS-CM-03-005   Contract Closeout (A)
September 2003

Other Sponsored Research

NPS-CM-03-001   Transformation in DOD Contract Closeout
MBA Professional Report
June 2003

Copies of the Acquisition Sponsored Research Reports may be printed from our website www.nps.navy.mil/gsbpp/acqn/publications