![embarcadero®]

White Paper

Achieving the Goals of the DoD Netcentric Data Strategy Using Embarcadero® All-Access™

By: Ron Lewis, CDO Technologies

April 2010

INTRODUCTION

Several years ago the Department of Defense (DoD) Chief Information Officer (CIO) released a strategy for maintaining "information superiority". The premise was that by changing their fundamental approach to data management, DoD agencies could enhance decision makers' effectiveness and improve our country's overall ability to wage war. The DoD CIO's goals were to essentially take Defense data resources and make them available using a "many-to-many" approach versus a traditional "one-to-one" approach. The new approach would enable commanders to use data in new ways to get a deeper, more robust sense of "situational awareness" in a more expeditious manner. The goals were to make data accessible and available in a netcentric environment by tagging and posting data shared amongst functional communities. Functional communities are also known as "communities of interest".

This paper looks at the specific challenges of meeting these goals leveraging Embarcadero All-Access. All- Access provides a comprehensive, flexible tool set that provides the full range of functionality necessary to achieve the goals set forth in the DoD mandate. This toolset facilitates collaboration between the various members within communities of interest, optimizes the value of data resources throughout the data lifecycle, and allows organizations to meet the goals in a timely and cost effective manner. Embarcadero All-Access flexible licensing allows the government to derive greater value at minimal expense by allowing organizations to share tools across disparate, geographically separate teams, provides access to various versions with ease, and supports all major platforms within one toolset. This paper describes several use case examples for using the tools provided in the by All-Access toolkit: ER/Studio® Enterprise, ER/Studio® Repository, ER/Studio® Portal, DB Optimizer, and Change Manager to solve the challenges specific to the DoD Netcentric Data Strategy.

THE CHALLENGES

This data mandate has been difficult to achieve. The scope of effort has proven itself nearly insurmountable largely due to the sheer magnitude of the amount of data resources, and the hectic pace at which the data environment evolves to meet the demands of today's war fighters. This white paper details how to achieve the goals of the DoD Netcentric Data Strategy by leveraging Embarcadero's All- Access toolset to meet the following four major goals:

- Make Data Visible, Accessible, and Trustable

- Make Data Useable and Understandable

- Enable Data Management

- Facilitate Interoperability

MAKING DATA VISIBLE, ACCESSIBLE, AND TRUSTABLE

There are five basic steps to making data visible, accessible, and trustable: discovering the data resources, creating data catalogs, defining the communities of interest (COI), discovering which

data is tied to which community of interest, and tagging the data with the appropriate COI metadata. *This is the hardest to achieve of the data strategy mandates; the sheer magnitude of the effort makes this impossible without the leveraging the right tools.* Embarcadero's ER/Studio Enterprise and Repository facilitate these needs well. It provides a rich set of model-driven, metadata-rich data management tools for discovering, understanding, documenting, and making data visible, accessible, and trustable.

DISCOVERING THE DATA RESOURCES AND CREATING DATA CATALOGS

Knowing where the structured defense data resources reside is challenging because the data landscape changes so quickly. Database administrators and information assurance personnel often have a good idea where the bulk of the database servers reside, but do not necessarily have up-to-date data maps or schemas for the data residing on these servers. This challenge is simple to solve using ER/Studio. Each database servers can easily be profiled and physical schemas can be derived using ER/Studio. Reverse engineering these physical data structures is the first major step towards establishing the goals.

Once the physical structures have been "discovered" or reverse engineered, the schemas should be published using ER/Studio's facility for generating data catalogs. The data catalogs can be stored into the ER/Studio Repository and merged to create an enterprise master data catalog. The master data catalog can then be carved up into smaller COI data catalogs.

DEFINING COIS AND DISCOVERING WHICH DATA BELONG TO THEM

Defining communities of interest is a multi-step process requiring the use of two tools available in the All-Access toolset: DB Optimizer and ER/Studio. DB Optimizer is used to profile how applications and systems are using the data resources, and ER/Studio is used to associate the usage metrics by recording the derived usage rules using metadata tags.

The first step is defining the applications used by a functional community to manipulate the data. This is a manual effort and requires analysis of which systems are required for members of the functional community to perform their assigned tasks. For example, Human Resources (HR) personnel would use specific systems to facilitate their HR needs. Once the systems have been identified, DB Optimizer would be used to evaluate which systems are relying on which databases, tables, and data elements. Further, the identified systems apply a context around the data be utilized to enable the users to perform specific tasks. Using our Human Resource example, the community of interest would be Human Resources. The HR personnel use specific HR systems. These HR systems rely on specific HR databases and manipulate the HR data is specific ways. Instrumenting a database server using DB Optimizer allows data managers to see which applications (or systems) rely on which specific data elements within a schema and how these elements map back to specific COIs.

TAGGING THE DATA WITH THE APPROPRIATE COI METADATA

Armed with the information from DB Optimizer, the data catalog can be divided into submodels, segregated by COI, and metadata tags can be associated identifying the COI.

DB Optimizer also captures important information as to how the data is being manipulated by these systems. This additional information is critical to define the "consumption rules" the

system uses to put the data in the proper mission context. In order to facilitate the data strategy, consumption rules must be derived and associated with the elements and dependent data objects. ER/Studio allows the creation of custom metadata tags. Consumption rules can be associated to data elements via the custom metadata tags.

Once the tags have been applied the data can be made accessible by publishing the metadata as part of a COI data sub-catalog stored in ER/Studio Repository and granting each COI user access to the Repository.

MAKING METADATA ACCESSIBLE

In order for the investment in metadata management to reap dividends, the metadata must be accessible to appropriate users. ER/Studio Enterprise provides the Portal, a web-based tool that allows authorized users within disparate COIs to browse through the ER/Studio Repository and gives access to metadata resources. It is a powerful web-based metadata discovery tool which allows users to share, publish, and report on the metadata contained with the ER/Studio repository.

MAKING DATA USABLE

Applying the proper context is the key to making data usable. This is a significant challenge due to the current application and data architecture—the context is often articulated as programmatic logic within the application or stored procedures in the database. There is a huge effort to web-enable data by exposing data via web services. This is not a sound approach without first capturing and providing a means of associating consumption rules. A fundamental rule of data management is that raw data can only become information when it is put into a proper and meaningful context. Exposing raw via web services without providing a means of applying the proper business context would be of little value to decision makers.

The netcentric strategy's data approach seeks to migrate away from the traditional one-to-one system interfaces and adopt a "post before processing" approach. In order to accomplish this, data must be tagged with COI, consumption rules, and must be tied to a data glossary which describes each data element in terms of how it is used, contextually, by the COI.

MAKING DATA UNDERSTANDABLE

In order for data to be understandable, it must have a metadata tag describing the data element in its business context. Different terms have different meanings to different communities of interest. Since a primary goal of the Defense data strategy is to make discovered data understandable--it is imperative to associate metadata tags providing a business definition.

ENABLING DATA MANAGEMENT

This white paper has already touched on a few data management requirements but has not discussed the importance of database and metadata change management. In order for data to be trustable, it has to be protected and maintained proactively. This can be a daunting task. Every major database platform provides tools for identifying changes to schema but not necessarily reconciling to enterprise data catalog, or associated metadata tags. A bigger

challenge is identifying and monitoring changes from an enterprise perspective. The DoD data landscape is comprised of a nearly overwhelming number of database servers so effectively orchestrating change management across this landscape is no small task. All-Access provides Change Manager which DBAs can use to perform schema compares and identify deviations from a baseline. Change Manager can also be used to "roll back" and/or report unexpected changes. It is a critical component to both data management and data protection initiatives.

MAKING DATA INTEROPERABLE

Different groups often have different ways of referring to the same type of data. To effectively leverage discovered data to enhance decision making, it is also important to identify synonymous data elements. ER/Studio provides a means for accomplishing this using "aliases". Although the purpose of this capability was to enforce data governance, it is very useful for establishing different data perspectives. ER/Studio's aliases capability can be thought of as a means of recording other means by which data elements have been described. It was included to match up various disparate schemas to a master data catalog. Using the HR system example: separate developers could have built separate data structures to house the same types of data; (e.g., one called social security number SSAN, another Soc_Sec_Num). ER/Studio's aliasing capability lets a data architect build a single structure that replaces both SSAN and Soc_Sec_Num with the master data catalogs standard for this data element. Aliases can be used, though, to establish data elements with similar meanings, and can be used to allow users to discover like data elements within separate COIs. This is critical to meet the vision of sharing data across COIs in a netcentric environment. Aliases, when categorized into an enterprise data thesaurus, can be used to significantly increase the interoperability of "discovered" data resources.

SUMMARY

Achieving the goals of the DoD Net-Centric Data Strategy--making data visible, accessible, and trustable; making data understandable and useable; enabling data management; and facilitating interoperability--is a significant undertaking that cannot be accomplished without leveraging automation. Hereto now, the goals have been largely unattainable, and the DoD has not seen the desired adoption rate of the new data management paradigm—"post before processing" largely due to the lack of availability to the right technology. Embarcadero's All-Access provides the tools necessary to achieve the goals of the DoD netcentric data mandates despite the pace at which the data landscape evolves. Adoption of these tools will help the DoD enhance its wartime decision making effectiveness and ultimately help maintain "Information Superiority".

ABOUT THE AUTHOR

Ron Lewis is an analyst who specializes in application security for CDO Technologies, a systems integrator that delivers technology-based solutions to government agencies and customers in the private sector. He has worked in the government and commercial security arena for more than 15 years identifying and providing guidance for remediating application vulnerabilities. Ron is considered an industry authority, having authored numerous articles on hardening applications and the hacker mindset. He is also actively involved in industry organizations and

efforts such as the Open Web Application Security Project (OWASP) and the Oracle Development Tools User Group (ODTUG).

Embarcadero Technologies, Inc. is a leading provider of award-winning tools for application developers and database professionals so they can design systems right, build them faster and run them better, regardless of their platform or programming language. Ninety of the Fortune 100 and an active community of more than three million users worldwide rely on Embarcadero products to increase productivity, reduce costs, simplify change management and compliance and accelerate innovation. The company's flagship tools include: Embarcadero® Change Manager™, RAD Studio, DBArtisan®, Delphi®, ER/Studio® and Rapid SQL®. Founded in 1993, Embarcadero is headquartered in San Francisco, with offices located around the world. Embarcadero is online at www.embarcadero.com.