

# OMEGA DATA ENVIRONMENT

## Drowning in Data, Starving for Knowledge

*A White Paper on Securing Meaningful Access to the Information Stored in Our Vast Data Warehouses*

---

Keith Coble

Global Director of Sales and Marketing, Telemetry and Data Systems

Wyle Laboratories, Inc.



**Telemetry and Data Systems**

## **ABSTRACT**

The quantity Test and Evaluation (T&E) data has grown in step with the increase in computing power and digital storage. T&E data management and exploitation technologies have not kept pace with this exponential growth. New approaches to the challenges posed by this data explosion must provide for continued growth while offering seamless integration with the existing body of work. Object Oriented Data Management (OODM) provides the framework to handle the continued rapid growth in computer speed and the amount of data gathered and legacy integration. The OMEGA Data Environment (ODE) is one of the first commercially available examples of this emerging class of OODM applications.

## INTRODUCTION

Moore's Law states that the rate of technological development in the semiconductor industry doubles every 18 months. Technology roadmaps predict that Moore's Law will continue for at least another 10 years, offering another 100-fold improvement in computer speed as shown in figure 1. A similar progression has held for hard-disk storage—in fact, the rate of progression in disk storage over the past 10 years has actually been faster than for semiconductors.

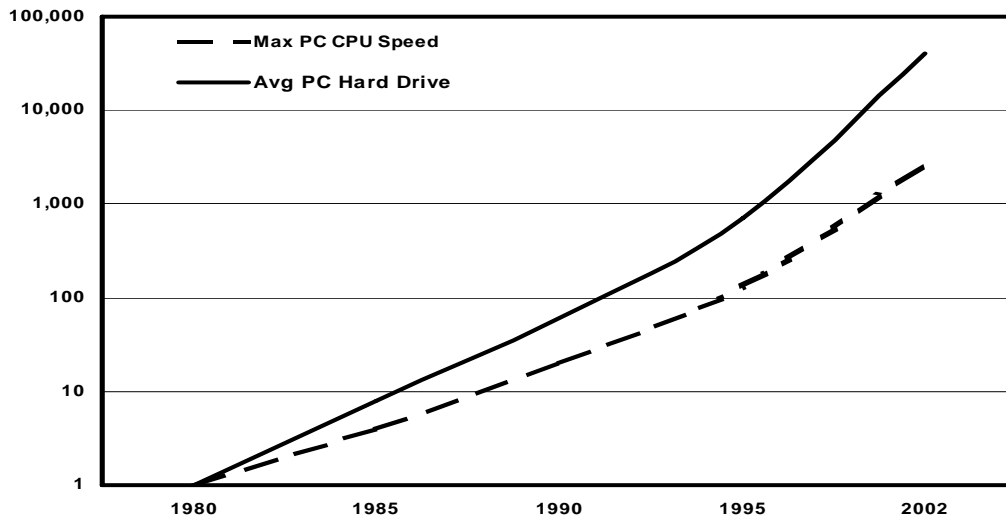


Figure 1

Some of the laws governing data include:

- **Parkinson's Law**—data will expand to fill the space available for storage.
- **Wood's Law**—if it is possible to measure, image, or capture a data point, it will be done.
- **Johnson's Law**—the time it takes to locate information is the inverse of how urgently it is required.

Executives, managers, and researchers are witnessing the convergence of these laws at an accelerating pace. Technology has delivered to our doorstep gigahertz notebook computers, petabytes of mass storage, gigabytes of memory on a stick, and high-resolution displays less than half-an-inch thick—super computer power at appliance prices.

High performance computing has become pervasive, and with the advent of the Internet, the expectation of instantaneous and accurate access to selected information has grown accordingly. In parallel, the expectation of virtually limitless storage, either locally or over networks, has become the bedrock of our information infrastructure.

# OMEGA Data Environment

## Implications for Test and Evaluation

This growth in computing power and deep storage has transferred directly to T&E. In the twenty years since the introduction of the PC, clock rates and the size of the average PC hard drive have grown by six orders of magnitude. In that same period, the amount of data gathered per hour of flight test has grown by seven orders of magnitude. The growing body of evidence suggests that this order of magnitude gap between the commercial sector and T&E will be sustained and will perhaps widen over the next twenty years.

The key challenges for the T&E community associated with absorbing this explosive growth in information volume fall into three categories:

- **Storage and Infrastructure**—Maintaining physical storage and related infrastructure
- **Data Management**—Managing, formatting, and processing data
- **Data Exploitation**—Locating specific data elements within the enterprise data superset

The world produces between 1 and 2 exabytes of unique information per year, which is roughly 250 megabytes for every man, woman, and child on earth. An exabyte is a billion gigabytes, or 10<sup>18</sup> bytes.<sup>1</sup>

Analysts forecast that disk shipments as measured by storage density will continue to grow by a compound annual growth rate of over 60% in the foreseeable future. In 2002, the global demand for storage for just enterprise software applications was over 80 petabytes. Other markets including personal computing, digital appliances, engineering, and scientific research increase this figure by almost an order of magnitude. The result is that over 36 billion gigabytes of new digital information, which will require storage, will be created over the next two years.

The total market for T&E-related storage infrastructure accounts for a fraction of a percent of the global storage infrastructure market. The expectation that unique T&E-related R&D in this sector will deliver a better return on investment than the commercial sector is no longer valid. The technical requirements of T&E are not significantly different from the commercial sector with regards to bandwidth, data integrity, and density. Significant differences arise only regarding in situ environmental test conditions and are handily met by specialized packaging.

For this reason, physical infrastructure is the simplest area for the T&E community to address. In responding to the demands of their enterprise customers, the computer and datacom industry have developed a robust pallet of technologies that the T&E community can leverage. For example, from 1996 to 2001, available network bandwidth in the U.S. grew by roughly two orders of magnitude.<sup>2</sup>

Adoption of commercial-off-the-shelf (COTS) technologies enables the entire community to ride the wave of innovation created by billions of dollars of commercial research and development.

---

<sup>1</sup> [www.sims.berkeley.edu/research/projects/how-much-info](http://www.sims.berkeley.edu/research/projects/how-much-info); senior researchers Peter Lyman and Hal R. Varian

<sup>2</sup> NRC-CRC Information and Telecommunication Technology Briefs; <http://cisti-isist.nrc-cnrc.gc.ca/ittb/2000-01e.html>

## **Data Management**

Traditionally the T&E community has responded to information technology (IT) management challenges with domain-specific T&E solutions, which include:

- **Standards (IRIG)**—T&E unique standards ranging from file formats to hardware specifications.
- **Nonrecurring Engineering (NRE)**—One-time, task, or program-specific solutions to solve a point problem.
- **Synthetic Enterprises (SE)**—NRE that transitions from a point solution to an internally funded entity in order to recoup sunk cost by marketing a point solution to other T&E users.

One needs to look no further than his or her own organization to find unique file formats, software applications, and data management solutions. Interestingly, the degree to which an organization attempts to market its unique solution to other organizations via a synthetic enterprise is directly proportional to the amount of money invested in the solution. Doing so results in organizational heat loss as personnel focus on recovering costs at the expense of their core T&E competency.

The result is a landscape of independent point solutions that long outlive their useful lives. Contemporary budgets and technology demand a re-engineering of this approach. They demand the application of IT solutions from general industry to solve IT problems. This preserves capital and provides more effective use of scarce and expensive T&E resources.

As enterprise-level computing continues to obey Moore's Law, unique point solutions become less compelling. Moreover, as budgets become constrained, focus on non-core activities becomes virtually impossible. Specifically, the problems facing T&E today require pervasive IT solutions from industry, not unique one-of-a kind domain specific T&E solutions.

## **Data Exploitation**

While numerous commercial vendors are addressing the challenges of data storage, infrastructure, and management, the challenge of data exploitation is just beginning to appear on the technology horizon. Data exploitation in the context of binary data sets, or binary large objects (BLOBs), has three components:

- **Finding the Data**—Locating one or more specific pieces of information out of hundreds, thousands, or even millions of binary data sets across distributed computing environments.
- **Presenting the Data**—Reconstituting the data with the appropriate and properly versioned processing and visualization engines and their associated support files and information.
- **Sharing the Results Through Collaboration**—Creating resultant data products, retaining their associations with the spawning data set, and sharing the resultant output.

### **Finding the Data**

Industries facing this task include not only T&E but also genomics, proteomics, and imagery. Specifically, how does one find, review, and share the golden needle in so many haystacks?

Once again, commercial technology leads the way. Commercial internet search engines enable users to casually explore terabytes of globally distributed text documents located in heterogeneous computing environments.

The resultant subset enables users to drill quickly down to locate information of value. Returns can be culled by iterative fine-tuning of queries. Unfortunately, BLOBs do not lend themselves to the type of indexing and easy retrieval that commercial search engines employ. And because of the unique nature of BLOBs (size, format, symbology), it is unreasonable to expect that commercial vendors will build BLOB-specific indexing and search technologies.

The solution to the BLOB search challenge lies in another family of emerging technologies centered on XML. XML metadata can provide a framework enabling BLOB consumers to build programmatic bridges between raw data and text representations so that commercial search technologies can be used for search and retrieval functions.

### **Presenting the Data**

After the data of interest is found, it must be reconstituted via processing and visualization engines. Significantly, these engines often require applications, variables, and associated files to be present in order to control data presentation, data definitions, and parameter processing. Specific subroutines, libraries, and text loaders may also be required. Hence the second challenge of data exploitation is the holistic association and delivery of all relevant digital elements associated with the core binary data set.

Object Oriented Programming (OOP) offers a suitable solution model. Encapsulation of related applications, files, and variables within a common framework or object model provides comprehensive integration and access to the data elements. Abstraction of the raw data and support elements utilizing this object model yields an elegant set of methods and properties. These methods and properties then provide standardized programmatic interfaces to the raw data, processing engines, and support elements based on their high-level class definitions irrespective of the underlying structures.

### **Sharing the Data**

The third step, which follows data presentation, is to create data products in the form of data slices, reports, and analysis. Often resultant data products are separated from the original source data, negatively impacting the extended enterprise's ability to leverage the efforts of all users in the extraction, collaboration, and exploitation of data. Anecdotes have users writing essentially the same report multiple times against a single set of data or repeating test points because of the difficulty in finding historical data of interest.

Most reports are created with COTS tools including Microsoft Excel, Word, and MATLAB. Inclusion of these documents in the object model enables users to view, review, and collaborate on the efforts of other extended enterprise participants.

## Object Oriented Data Management (OODM)

OODM is a new application class that combines text-based indexing and search capabilities with standardized interfaces. OODM provides a solution set that enables standard search technology to find and return data objects, coupled with an underlying object model that allows legacy and future data formats and processing engines to be accessed via standardized interfaces.

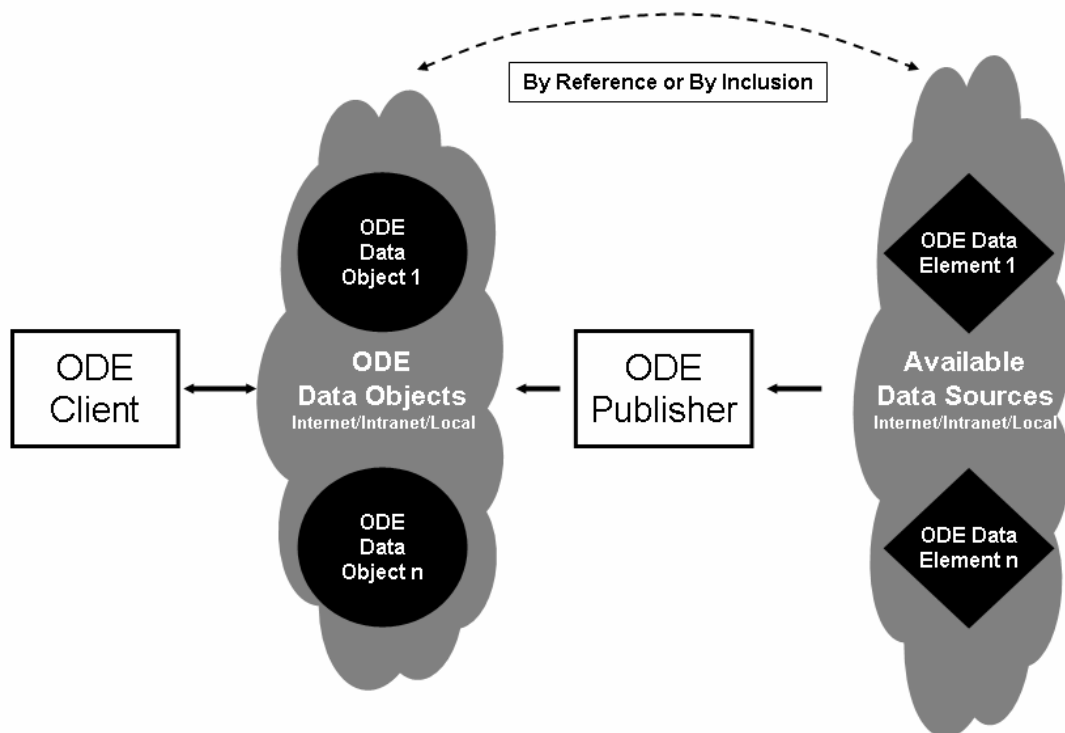
At the core of OODM is the data object. Data objects are the collection of necessary and sufficient data and support elements required for any given raw data set to be self-describing and self-instantiating. Data objects are accessed via published methods and properties allowing standardized programmatic access to the underlying data, processing engines, and resultant output elements.

Because data objects and services can be instantiated on either the client or server, they lend themselves to enterprise-wide distribution using web services technology.

## OMEGA Data Environment (ODE)

ODE is one of the first applications designed from the ground up to be OODM compliant. It is designed to provide a contemporary IT solution to the challenges presented by Moore's Law in the T&E community. By leveraging proven architecture and technologies from the commercial sector, it offers a robust, scalable solution to the problems associated with managing and exploiting the exponentially growing volume of T&E data. ODE adopts contemporary storage and infrastructure technologies to store, manage and exploit data.

ODE is comprised of three entities: ODE Publisher, ODE Data Objects, and ODE Client as shown in Figure 2:



**Figure 2**

## **ODE Publisher**

The ODE Publisher allows users to aggregate and interrelate the data elements used to generate ODE data objects. As a server side application, it uses the ODE Data Object Model (ODE DOM) to provide ODE data object assembly, formatting, versioning, and publishing services.

## **ODE Data Objects**

ODE Data Objects (ODOs) are composed of three components:

- XML Metadata
- Standardized Interfaces
- Data Elements

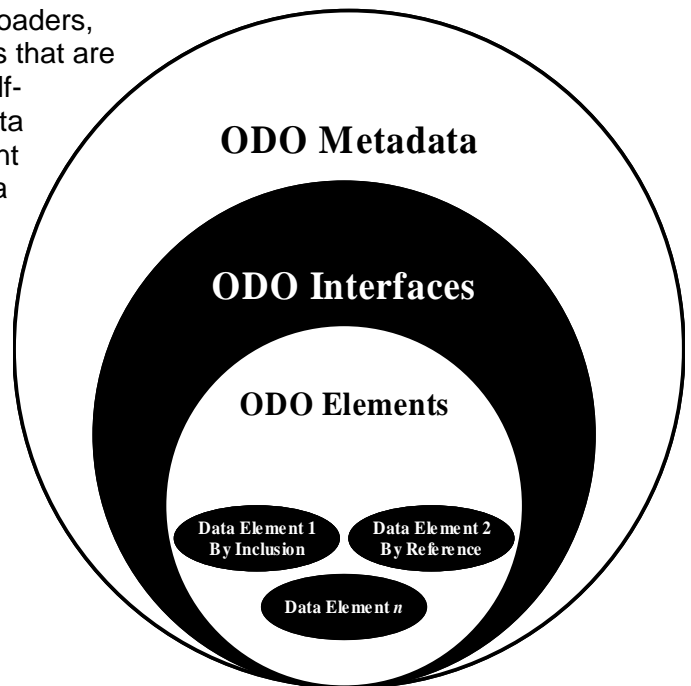
ODE Data Object Metadata is an XML wrapper that describes in detail the ODO elements, interfaces, and sub-element specifics for the elements included and referenced by a given data object. ODO metadata provides a searchable index of data elements and sub elements within ODOs.

Based on an Object Oriented Programming class model, ODO interfaces are the programmatic framework exposed by the ODE Data Object model for access to, and interaction with, the data elements contained within the data object. The interfaces consist of a series of methods and properties that control I/O and the manipulation of the encapsulated elements. (Figure 3)

ODE data object elements are the specific files, loaders, executables, documents, tools, and raw data sets that are required for an ODO to be self-describing and self-instantiating. These elements are attached to data objects by inclusion or by reference. If an element is attached by inclusion, it is held local to the data object and has a static link to current object instantiation. If an element is attached by reference, a logical path is held internal to the ODO as a pointer to the element with a dynamic link. Both methods have implications for object size and integrity.

ODO elements are divided into four groups:

- Source Elements
- Processing Elements
- Output Elements
- Variables



**Figure 3**

## *Drowning in Data, Starving for Knowledge*

Source elements include raw data and baseline reference data sets for processing elements. Source elements can be local or remote. Processing elements include processing, visualization, output, and formatting engines. Processing elements can be local or remote.

Output elements are the products created by users as a consequence of interaction with source elements via processing elements that are deemed worthy of retention. Output elements can be local or remote. They typically consist of the documents and files accessed by post-processing applications such as Microsoft Word, Excel, Acrobat, Probe, or MATLAB.

Variables are the data used to vary the settings in processing elements as well as properties of ODOs. Variables are held locally and are exposed as part of the ODO DOM via the metadata XML schema.

ODE data objects are dynamic. Specifically, the ODE Publisher creates the first instance of a data object. The object is subsequently accessed by one or more users via browsers or other user interfaces utilizing one or more of the included processing elements. During interaction, users may create additional output elements, source elements, or revised variables.

Examples of new output elements might be revised reports and templates relating to the source elements. New source elements could be slices of the source data in the time or frequency domain, data versus data, reprocessed data, or even decimated data.

### **OMEGA Viewer**

The ODE Client is a web service rendered on the user's desktop. The client enables users to peruse all available ODE data objects within their network horizon. ODOs are accessed via named paths or are located using queries and COTS search engine technology.

Queries are run against the ODO metadata contained within the ODO XML wrapper that is indexed by the search engine. A list of matching ODOs is returned to the user by parsing selected ODO XML tags through Extensible Stylesheet Language Transformations (XSLT). See figure 4 on the following page.

Using the ODE Client, ODOs of interest are located and called. Another page is rendered with a more exhaustive visualization of the ODO XML metadata representing the elements contained within the ODO. ODO elements are presented in a well-structured, logical fashion for ease of navigation and instantiation.

Elements that are called from the viewer can be run on the client, on the server, or both depending on the element type. For example, if a user wants to replay a raw data recording of a telemetry downlink archive, the raw data source element, which may be hundreds of gigabytes in size, would remain on the server and a replay processing element (not unlike Adobe Acrobat Reader) would be called from the ODO and instantiated using processing-on-demand (POD) technology. This replay element would stream raw data from the server and render data visualization locally on the client.

To create an Excel export file of selected parameters over a specified time slice, a user would use the local instance of the replay processing element to export the data to a local Excel file. This local Excel file would be reviewed using a local copy of Excel, and the user would have the option to add that file to the ODO by inclusion or by reference. The resultant file is then available for review by others who might access the same ODO in the future. Collaboration in this most basic form is sometimes referred to as an adhococracy.

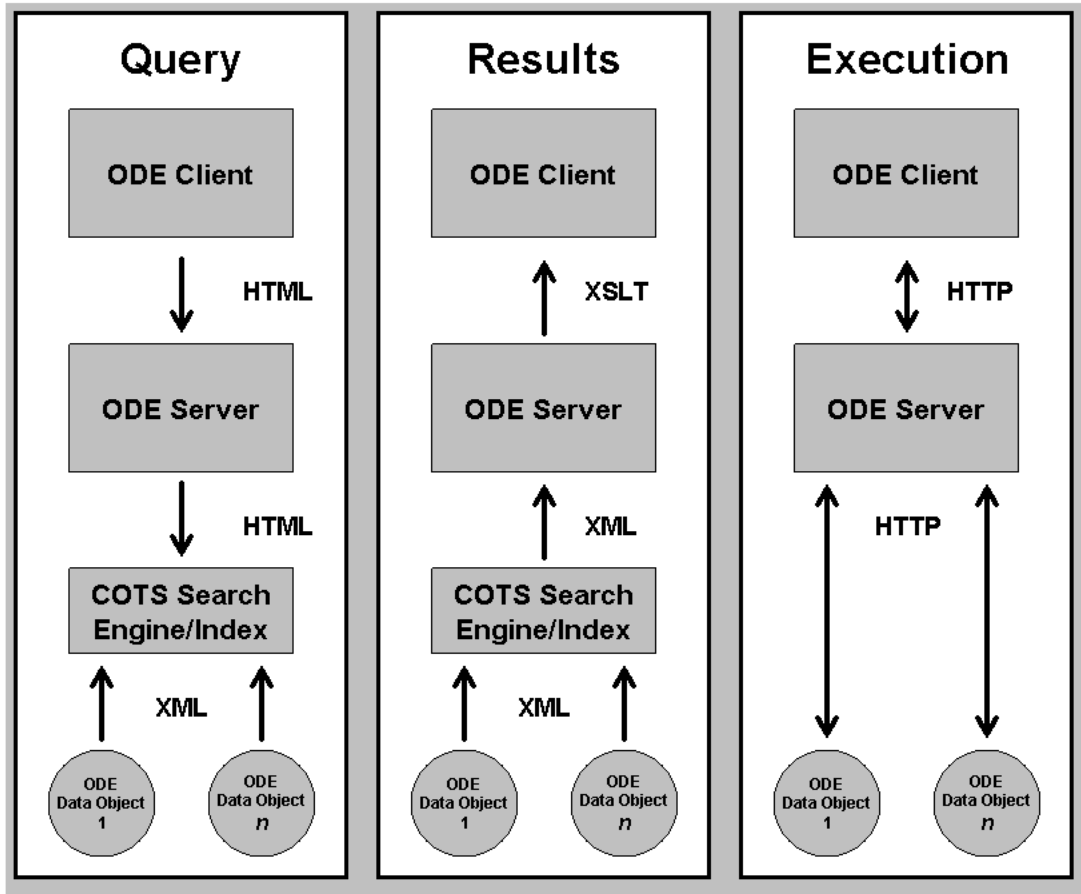


Figure 4

## Collateral Benefits

### Legacy Integration

ODE provides standardized open interfaces to the underlying data elements. These data elements can include legacy and future file formats, applications, and output templates. ODE mitigates the challenges associated with legacy integration by providing common programmatic interfaces for all underlying data elements regardless of the data's heritage. This abstraction of underlying data elements specific to a common object model with open interfaces allows users to continue to leverage their existing investment while building scalable systems for future growth.

### Historical Perspective

As users interact with ODE data objects, they have the opportunity to add information to those objects. The information resulting from these interactions is captured in the object metadata, which is indexed and made available to other users. This organic growth of data objects through user interaction provides a powerful vehicle for enterprise-wide collaborative knowledge growth. The result is a heretofore unrealized ability to leverage the collective knowledge of the enterprise enabling users to learn from history rather than unknowingly and unintentionally repeating it.

### **Time to Results**

As the pace of commercial product development continues to accelerate, T&E customers will demand results in less time and for less money. ODE offers a solution to this problem by offering a standardized framework for information management and exploitation using (and reusing) COTS tools. This allows testers to focus on performing quality testing instead of building one-of-a-kind tools and technologies. Most importantly, ODE can deliver an order of magnitude increase in productivity when searching and managing BLOB data.

### **Centers of Excellence**

ODE is collaborative at its core. Iterative interaction with the enterprise knowledgebase provides for the emergence of centers of excellence within the enterprise. Enterprise dynamics are Darwinian in nature. As each group of users interacts with the enterprise knowledgebase and its underlying data elements, groups will emerge as de facto centers of excellence. Some will excel at building high performance, real-time rendering engines; others as posttest, output tool creators. The actual function is immaterial. What is impactful is that ODE provides a platform for the enterprise marketplace to select the best in class through data element use and reuse.

### **Information Pull versus Data Push**

The information universe is changing from push to pull. A few years ago, banks sent monthly hard copy statements to customers regarding their bank accounts — push. Today's user logs on 24/7 via the Internet to see the real-time status of their accounts — pull. T&E customers expect the same. They expect access to their data, processing engines, and output formats from any machine, 24/7. They also expect to see meaningful interaction with the information by other team members. ODE provides the mechanism for this to happen.

## **CONCLUSION**

The tremendous increase in data acquisition bandwidth, storage depth, and processing capability has the potential to produce increased fidelity in test results. It also presents a rapidly mounting challenge in how data is stored, distributed, managed, and exploited. The OMEGA Data Environment is one of the first of the class of emerging OODM applications available for commercial consumption. By utilizing XML metadata to expose native data elements, it provides a rich framework for programmatic search and interaction with complex data. It also provides a solution set for the incorporation of legacy data while extending a flexible coupling to future data types. Because ODE can encompass all data types, from raw data to end results, collaboration is enhanced and meaningful results are generated in significantly less time, with less effort, and for significantly less money.

*Drowning in Data, Starving for Knowledge*

For More Information

Wyle Laboratories, Inc.  
Telemetry and Data Systems  
44417 Pecan Court  
California, MD 20619

301-737-1555  
[www.wyletds.com](http://www.wyletds.com)