

SIGNAL

AFCEA'S INTERNATIONAL JOURNAL • JANUARY 2023 • \$5.00

THE
CYBER
EDGE

EXPANDED



CYBER'S HUMAN ELEMENT



PERFECT FIT FORM FACTOR

The Intelligence Advanced Research Project Activity's SMART ePANTS program seeks the perfect form factor for smart clothing to suit intelligence community needs. [PAGE 27](#)

THE CYBER EDGE

8 NSA Texas Is Open for Cyber Business

The National Security Agency is growing its U.S. cybersecurity workforce, particularly in the Lone Star State. *By Kimberly Underwood*

11 U.S. Cyber Command Means to Magnify Cyber Intelligence

The command pushes to create a cyber intelligence center. *By George I. Seffers*

14 Zero Trust Is About Priorities and Thinking the Unthinkable

Zero-trust architecture is both elusive and inevitable. *By Diego Laje*

17 Achieving Cyber Skills Through Gaming

CyberStart America lets students try the cybersecurity field.

20 Microelectronics Offer Case Study for Securing Defense-Critical Supply Chains

Securing this critical industry presents significant challenges. *By Howard D. Grimes, Gabriela F. Ciocarlie, Robert J. Butler and Wayne E. Austad*

METaverse TRAINING

24 Training in the Metaverse Never Looked So Realistic

Enhanced technology makes it easier for trainees to make war in a digital world.

SMART FABRICS

27 SMART ePANTS May Be the Perfect Fit

IARPA is developing comfortable clothing integrated with smartphone-like technologies. *By Nuray Taylor*

INSIDER THREAT

29 CISA Expanding Insider Threat Toolbox

New resources to be available in the coming weeks.

SPONSORED CONTENT

SPECIAL INTEREST EDITORIAL

32 Unlocking the Power of Unstructured Data

Unlocking the Power of Unstructured Data

BY PATIENCE WAIT

Governments, industries and individuals alike all report feeling overwhelmed by the rising tide of data.

How big is the challenge? One research firm estimated that the volume of data/information “created, captured, copied and consumed worldwide” would reach 97 zettabytes in 2022 (one zettabyte is the equivalent of 1 billion terabytes)—and the figure is projected to almost double, to 180 zettabytes, by 2025. Compounding the problem, most of that growth will be of unstructured data, so-called because it doesn’t adhere to structured data models.

“The growth in the volume of information is increasing at a rate that it overwhelms the ability to digest, categorize and make it actionable,” said Jeff Casale, CEO of MarkLogic. “And the largest explosion has been in unstructured data, which is growing at an exponential rate and doesn’t want to adhere to structured models.”

Breaking the structured data model is one problem. Casale says the bigger problem is that unstructured data shifts the user paradigm. “You need to be agile and flexible enough to gain insights from all the data without having to go through categorizing it first,” he said.

In some fields, the usefulness of the data is in its timeliness. For instance, much of the value of satellite images is based on their capture of what is happening in a particular location at a particular time. U.S. military operations might be tracking the movements of equipment and personnel, or the intelligence community may be following the movements of a hostile leader. Putting these images in front of human eyes can enable officers to know what is happening in near-real time—but think of the man-hours lost to examining images that do not have useful information.

In other areas, the data is useful in its current form, but limited in its applicability. If a bolt on a piece of equipment fails, it is not difficult with today’s inventory tracking systems to find and inspect the bolts on the same piece of equipment elsewhere. But what if that bolt is used as a standard part in many types of equipment? Can the bolt manufacturer locate all of those, and send the appropriate notice to all those users?

The key: Searching data while gathering it

Casale suggested there are searchable database systems that address the problem of ingesting such massive amounts of unstructured data.

“The issue is categorizing. If you’re looking for a bunch of photos, you’re not going to go into the filing cabinet holding taxes,” he said. “It needs to be categorized to be actionable, [and] that’s where metadata comes in.”

He pointed out that one of the most prolific forms of unstructured data is the PDF. Since a PDF uses a file format that provides an electronic image of text or text and graphics, “historically, that’s dumb data. Other than classify the heading, you can’t do much to search for it,” Casale said.

To address that limitation, modern database systems give organizations the flexibility to conduct instant searches by using NoSQL, which means “not only SQL.” That is, it sets up databases that are not based on the tabular relations (rows and columns) used in relational databases, and can provide high throughput, fault-tolerant and scalable data storage and retrieval.

“NoSQL is a powerful capability. You can categorize these millions of PDFs, and extract from them the specific mission needs, the text, images, etc., that you can categorize and group so you can search,” Casale explained. “With

this process of being able to actively pull data [as it streams in], you can search at the same time.”

The Department of Defense and intelligence community (IC) requirements for these kinds of capabilities introduce other, unique constraints. For example, the flow of data may be intermittent or may have limited communications capabilities, given that many mission locations are globally located. That also translates into how to keep a mission going when the data that they rely on isn’t coming in consistently. The needs are dynamic in the field and in a global scenario.

Use case: Intelligence agency improves manpower management for mission readiness

One big advantage of using structured and unstructured data together is the ability to craft new kinds of searches, recombining data over and over to find new insights from differentiated data.

For example, a U.S. intelligence agency tasked with providing military intelligence to warfighters, policymakers and force planners in the Defense Department and IC did not have a true, cohesive view of the contributions, expertise, work products and activities of the workforce. As policies and missions changed, such as a strategic pivot from the Middle East to the Asia Pacific regions, current evaluations, metrics and set methods to assign staff and teams became much less relevant.

The situation was putting the agency’s mission at risk, through staffing delays and inappropriately staffed projects. Misallocating workers—wrong person in the wrong assignment—could lead to unhappy employees and high turnover, creating higher costs to find replacements. And sticking with the traditional methods of updating internal resume databases and skills lists couldn’t move fast enough in the rapidly evolving environment.

To address these quickly developing shortcomings, the agency sought to enable a unified view of workforce talent across multiple fragmented



data sources and data silos, whether internally or externally located. Agency leadership wanted managers to more quickly and accurately find the right personnel to support any given mission's requirements, based on being able to link to individuals' actual knowledge and experience with ongoing and new project priorities and competency requirements.

Leadership also saw this kind of search as a way to provide "look-ahead" capabilities, identifying developing skill and position gaps while giving employees opportunities to find new positions that match their skillsets and provide greater job satisfaction.

The agency selected MarkLogic's Data Platform to meet its needs. It offered integration of internal data—resumes, annual reviews, self-appraisals, training history and other HR information—with travel history and unstructured information such as work products created on particular topics. The IT staff developed a harmonized data model to communicate between different formats and used the platform's built-in query capabilities to support better search, discovery and analysis across the entire enterprise.

Once in place, the agency had advanced workforce search, discovery, semantic and geospatial analysis capabilities. Filling mission-critical positions accelerated, with better matching of resources to assignments. It also cut costs by maximizing the use of existing human assets and improving project outcomes.

Looking ahead, the agency plans to make use of the searchable database to improve workforce strategic planning,

based on mission needs, organizational budgets, billets and policy directives.

Digital "twinning:" Another pioneering concept for databases

Organizations are looking to use digital systems to mirror physical realities. These "real-world" digital representations, sometimes called "digital twins," transform how manufacturers create products, enabling them to produce goods better and cheaper than before.

Currently, most manufacturers are burdened by complex systems with interdependent components, accompanied by demand for shorter development cycles. These combine to create the need for more efficient, less error-prone development approaches—something that digital twins can address.

Having a digital twin gives engineering teams a better understanding of the products they build and can improve quality and development time. The challenge is that traditional IT is not designed to model reality—it doesn't support digital twins.

This leads to model-based systems engineering (MBSE), a methodology where a digital model depicts and relates the components and architecture of a system and the requirements and restrictions on that system to perform a function or for it to be safe.

In a real-world manufacturing scenario, as when a manufacturer builds an aircraft with more than six million parts, it means that teams can finally understand how a single small change in a requirement ripples through all the

parts, processes, manufacturing and delivery of the entire aircraft—saving time and money while making sure the plane is safe and airworthy.

The key to making MBSE work is managed data on a unified platform. Over the years, all the information—from parts specifications to performance measurements—may have been gathered into many different databases, siloed from each other. Moving to a multimodel, semantic database means the manufacturer can create an ontology—a set of relationships and objects—that draws on all the available data to create the digital twin.

Just as important, the engineers can track the digital provenance of each bit of data—which database(s) it came from, changes made to it, where it is used—providing visibility and trust in the data quality, and trust in the digital twin's representation of real-world conditions and results.

Conclusion

Rather than drowning in an accelerating flood of data, trends in data creation, management and use are creating the tools to surf that flood and pull out valuable information to make decisions in real time.

MarkLogic's Casale said these tools are still maturing, moving to interact with artificial intelligence and machine learning capabilities to enable even faster trend detection and decision-making. The incorporation of more visual object-like front ends will allow a greater number of users to capitalize on their new-found dynamic data capabilities.

The data remains the same. What is evolving is our ability to harness it, to use it in ever more varied ways to generate new insights and earn the benefits of flexibility and agility.



For more information about unlocking the potential of unstructured information:
<https://www.marklogic.com/national-security>