Making Modeling & Simulation Reuse Attractive

Dr. Gary Allen
DoD M&SCO
Alexandria, VA
gary.allen@osd.mil

Mr. John Daly, Mr. Michael Heaphy
DoD M&SCO
Alexandria, VA
john.daly.ctr@osd.mil, michael.heaphy.ctr@osd.mil

Dr. Shep Barge, Mr. Bob Halayko
OUSD(P&R)
Alexandria, VA
walter.barge@osd.mil, robert.halayko.ctr@osd.mil

Mr. Kevin Gupton
ARL:UT
Austin, TX
kgupton@arlut.utexas.edu

ABSTRACT

Modeling & Simulation (M&S) is the keystone of effective conduct of Live-Virtual-Constructive events across the spectrum of M&S community activities (training, test & evaluation, experimentation, acquisition, etc.). The widespread and efficient application of M&S technology can become more effective and cost-efficient by leveraging past M&S investments. As noted in the *Live-Virtual-Constructive Common Capabilities: Asset Reuse Mechanisms Implementation Plan* (Riggs 2010), one way of accomplishing this is through discovery and reuse of extant M&S concepts of operations (CONOPs), data, scenarios, and lessons-learned. With available data elements numbering in the tens of thousands, however, simple search techniques do not provide the technical sophistication necessary to find the right data, if it exists at all. There is a need to provide the incentive and reward necessary to outweigh the difficulty of potential reuse. To reverse this situation, the Department of Defense (DoD) has to make better use of structured search techniques that are currently being employed by organizations such as Google and Facebook. These techniques can extract 'context' from the search data or can apply graphic search techniques to develop purpose driven 'relationships,' and may be the difference between making reuse a viable option or perpetuating redundant development and inefficient use of data.

This paper explores the re-use proposition in the DoD modeling and simulation community. Discovery metadata attributes, data quality indices, and search technologies for M&S data are explored in representative use cases (including a business-case for the archiving, attribution, and discovery of priority M&S data). It also explores the applicability of existing standards and specifications like Amazon's OpenSearch, as well as the potential for leveraging current DoD efforts (e.g. Intelligence Community and Department of Defense Content Discovery and Retrieval Integrated Project Team) that support the potential business-case.

ABOUT THE AUTHORS

Dr. Gary Allen is the Associate Director for Data at the Office of the Undersecretary of Defense (USD) for Acquisition, Technology, and Logistics (OUSD(AT&L)) Modeling and Simulation Coordination Office (M&SCO) and a Project Manager for the Live, Virtual, Constructive Architecture Roadmap Implementation (LVCAR-I) project at the Program Executive Office for Simulation, Training, and Instrumentation (PEO-STRI). He has over 30 years of experience in various aspects of modeling and simulation, including the Training Simulation Center for I Corps at Ft. Lewis Washington; Director, Simulation Training Branch at the US Army Intelligence Center and School, Ft. Huachuca, AZ; and Project Director, Tactical Simulation (TACSIM) Intelligence Simulation and the design group that initiated the Aggregate Level Simulation Protocol (ALSP). Dr. Allen spent 29 years on active duty in the US Army finishing his career as Commander of the International Technology Center at the US Embassy, Germany. His DoD Civilian Career began in 2002 as the US Army Liaison Officer to the German Military Research and Development Agency in Koblenz, Germany, where he served until 2009. He holds a Master of Science from the School of Engineering, University of Colorado; a doctorate in Instructional Technology from the University of Kansas; and is a graduate of the US Army War College.

John J. Daly is an engineer with Booz Allen Hamilton in Washington, DC, and works with various DoD clients in the areas of: Modeling and Simulation (M&S) technology and policy, military concept of operations (CONOPS); Command and Control, Information Technology, and computing infrastructure technology. He currently is the program manager for support to the USD AT&L M&SCO. He has also worked at the Naval Research Laboratory in operational CONOPS; advanced technology implementation; C2 research, development, and implementation; C2/M&S interoperability; and C2 embedded training technologies. Mr. Daly previously served on the staff of the Director, Defense Information Systems Agency (DISA), as well as in the DISA Modeling and Simulation Directorate on C4I and simulation interoperability.

Michael Heaphy is a systems engineer with Booz Allen Hamilton and works with various DoD clients in the areas of: Modeling and Simulation (M&S) technology and policy, systems analysis and systems engineering, military training and the rapid fielding of technology. He currently is an M&S data analyst for the M&SCO. He served as a Navy Submarine Warfare Officer and nuclear engineer, and continues to support strategic and theater-level exercises, training and readiness in the Navy Reserve. He holds a Bachelor of Science in Systems Engineering from the US Naval Academy, and a Master of Engineering Management from Old Dominion University.

Kevin Gupton is an Engineering Scientist in the Modeling and Simulation Information Management Group at Applied Research Laboratories at the University of Texas (ARL:UT). He has over 10 years of experience in enterprise system engineering, data modeling, and knowledge management. Mr. Gupton has developed net-centric applications and data services using C4I and M&S common data standards and ontologies. He holds a Bachelor of Science in Mathematics and a Master of Science in Computer Science from Texas A&M University.

Dr. Walter 'Shep' Barge is director of the Joint Assessment and Enabling Capability (JAEC) within the Office of the Under Secretary of Defense (Personnel and Readiness)(OUSD(P&R)). He has more than 20 years of experience as an analyst, academic, and team leader applying operations research methods and models to inform strategic decisions that affect national security. He has interests in many areas of operations research, policy analysis, and decision making strategy.

Mr. Bob Halayko is currently an Operations Research Analyst in the Joint Assessment and Enabling Capability (JAEC) within the Office of the Under Secretary of Defense (Personnel and Readiness) (OUSD(P&R)). Prior to working at JAEC he spent many years with the US Army Corps of Engineers and the Joint Staff performing analytic studies, applying operations research methods, and building models. He has also represented his organizations' interests on both the DoD M&S Integrated Process Team and its predecessor, the DoD M&S Working Group.

Making Modeling & Simulation Reuse Attractive

Dr. Gary Allen

DoD M&SCO

Alexandria, VA

gary.allen@osd.mil

Mr. John Daly, Mr. Michael Heaphy

DoD M&SCO

Alexandria, VA

john.daly.ctr@osd.mil, michael.heaphy.ctr@osd.mil

Dr. Walter Barge, Mr. Bob Halayko

OUSD(P&R)

Alexandria, VA

walter.barge@osd.mil, robert.halayko.ctr@osd.mil

Mr. Kevin Gupton

ARL:UT

Austin, TX

kgupton@arlut.utexas.edu

REUSE: IF ONCE WAS GOOD, MORE MUST BE BETTER

In the findings of a report on Live-Virtual-Constructive asset reuse mechanisms produced by the Johns Hopkins University/Applied Physics Laboratory, the authors noted that:

> "The reuse of software, data, and other assets in Department of Defense (DoD) Modeling & Simulation (M&S) development is neither as frequent nor as effective as it could be, and as a consequence, the potential benefits of reuse to the DoD enterprise are not being fully realized." (Riggs 2010)

The report went on to explain that successful reuse has two components: technical and social; therefore the purpose of this paper is to address both. The content of the paper looks at various technical aspects that are intrinsic to increasing the probability of reuse. At the conclusion, the reader should have a raised awareness of reuse benefits and barriers, technical capabilities that increase the probability of reuse, and some potential solutions for consideration. Though the focus of this paper is on the training community, the points made are applicable across DoD. Further, the circulation of this paper provides a venue to advertise reuse and its potential for DoD M&S.

While the advantages of reuse may seem obvious to some, there is a need to establish what is meant by reuse in the context of modeling and simulation, and the case for supporting reuse as a business model. As used in this paper, reuse means:

> "the process whereby an organization defines a set of systematic operating procedures to specify, produce, classify, retrieve, and adapt software artifacts for the purpose of using them in its development activities." (Frakes 1996)

This definition is particularly useful because it recognizes a key aspect, "retrieval," which is critical to the success and benefit of reuse.

REUSE PROPOSITION

Cost avoidance and return-on-investment (ROI) are of perennial interest to industry, and DoD is no different in that regard. Modeling and simulation are touted by the Services as essential capabilities that can provide savings through a reduction in operational costs. ROI was the theme for the Fall 2012 issue of the *M&S Journal*, published by the DoD M&S Coordination Office (M&SCO), in which the articles identified the potential for significant cost savings and increased return-on-investment arising from the use of modeling and simulation. Though the authors in the *M&S Journal* articles also noted the difficulties in making accurate estimates of savings, there remains ample evidence that reuse can be a significant source of savings if implemented properly.

The benefits of an effectively-implemented, practical reuse program include:

- Reduced developer time and risk by reusing others' concepts, algorithms, and experience
- Adoption of a common component structure that enables modularity, interoperability, and component reuse
- Use of accredited components and more reliable code
- Cost savings by avoiding new code development

A 1993 report by the General Accounting Office (since renamed the Government Accountability Office) estimated that DoD spent $23 billion annually on all software. The amount DoD spends specifically on M&S is estimated by industry experts to be billions of dollars a year, as well. The precise number is difficult to determine because spending on M&S is usually a smaller element of a larger program, and may be identified as software, processes, or other tools in budgets. The Institute for Defense Analyses (IDA) placed a figure of $10 billion on identifiable FY05 M&S projects and activities (Henninger 2005), but this figure did not necessarily represent the total M&S cost. As DoD faces declining funding for modeling and simulation, it is M&S capabilities that are often targeted as a means to lessen budgetary impacts. If the Department spends less on these resources, reuse of existing models and simulations is one way to do more with less.

Reuse is a familiar concept in computer-based activities. In the early 1950s, the new subroutine construct was, among other things, to "reuse" code. That era was of course pre-web (i.e., little need for interoperability and discoverability), and duplicate code stole valuable space from applications operating in relatively small addressable core memories. Today, essentially unlimited random access memory and virtual memory management techniques make code economy a moot issue. Yet reuse remains just as important today. Reuse is endemic to programming; Microsoft's Dynamic Link Libraries, Linux's kernels, C libraries, and Smalltalk's development environment, among others, are all examples of code reused at various levels and transparency.

The reuse addressed in this paper is reuse of the software and data specific to DoD modeling and simulation activities. As DoD became increasingly dependent on computers, not only for modeling and simulation purposes but also for information technology (IT) in general, senior DoD officials questioned why seemingly duplicate software for analysis, command and control, and even personnel management systems had been developed. When standalone monolithic computer mainframes gave way to ubiquitous Intel and AMD microprocessor-driven computers, transfer or adaptation of applications from one environment to another became an even larger issue. The presumption was that functionality should be introduced to applications by merging or integrating other code. Using existing software to develop new applications should save time, money, and resources; in practice, however, DoD's experience implementing software reuse has been problematic.

DoD modeling and simulation assets are in many ways unique. In practice, the following circumstances have on occasion stood in the way of implementing a Department-wide reuse process:

- Adapting software is not "plug-n-play." Not only may different coding languages and inadequate documentation limit reuse, but reuse of existing resources is unlikely unless the effort to adapt code is clearly better and more cost effective than doing it in-house.
- Lack of common formal descriptions or languages to describe software unambiguously frustrates discoverability. It may be difficult to determine whether the software is fit-for-purpose, or to find evidence that the code was verified, validated, and accredited.
- The original developmental environments may be incompatible, even aside from security issues. While there may be justification (because DoD allows different application architectures), the consequence is duplicative software.
- True interoperability is a multi-level criterion. In distributed, late-binding environments, syntactic is easier than semantic interoperability. Designing compliant software is not easy.
- Whereas technical problems are frequently mastered over time, the same is not true of legal issues. Acquisition regulations, intellectual property rights, or technical data issues can frustrate reuse attempts, and are slower to adapt to technology changes.
- Coding in reusability at inception is costly, and after the fact, even more so; standards may be helpful, but the organization or project manager must believe that something useful will endure by making his or her software more widely usable.

DoD has previously sought to address challenges associated with reuse of M&S in other forums. The Intelligence

Community (IC) and DoD Content Discovery and Retrieval (CDR) Integrated Project Team was formed in 2010 with the purpose to "develop and publish a set of architecturally driven standards and specifications to enable content discovery and retrieval from all IC and DoD data collections." (ODNI 2013) To this end, the CDR Integrated Project Team developed a Reference Architecture (RA) to provide conceptual and contextual guidance to architects, engineers, and developers tasked to implement CDR specifications. The Office of the Director of National Intelligence (ODNI) describes the RA as follows:

> "The CDR Reference Architecture (RA) is the keystone artifact for the overall set of guidance artifacts. The primary content of the CDR RA is the definition of an extensible set of capabilities and components that are realized via service specifications. It also describes various architecture interaction patterns and the specific technical framework to enable robust, scalable, and repeatable implementations across varying business and mission boundaries." (ODNI 2013)

Furthermore, two primary capabilities, content discovery and content retrieval, are defined in the CDR RA. This effort is applicable to the problem set described in this paper, but additional efforts will be necessary as technology advances, and as we learn to implement capabilities that will provide for ease of reuse.

Whatever the architecture used, to increase reuse success DoD must promulgate a reuse system that provides potential "reusers" with enough information and a compelling value proposition. It must enable the reuser to make an informed decision on whether the artifact satisfies their needs, and equally important, how difficult it might be to assimilate. This would best be served by having an intelligent query capability that exceeds wiki- or Google-levels of interaction and a descriptive information set that defines the value of each asset to the user.

DISCOVERY: THE MOTHER OF REUSE

The foundation of a useful intelligent query capability is built on metadata, data quality, and structured search techniques. Even the seemingly miraculous Google search has limitations. Those limitations revolve around how the data you are looking for is described (metadata), how the search is being conducted (e.g., query terms, cross referenced data, and hierarchy of sources), and the various aspects of data asset quality (e.g., whether the source is complete, pertinent, its format, and its availability).

Metadata

Metadata is a description of data, or more commonly, "data about data." DoD Directive 8320.02, concerning "Data Sharing in a Net-Centric Department of Defense," defines metadata as "information describing the characteristics of data; data or information about data; or descriptive information about an entity's data, data activities, systems, and holdings." There are conventionally three types of metadata, each of which will be described below: descriptive metadata (applicable to discovery and identification), structural metadata (identifying data format and structure), and semantic metadata (providing meaning, context, and understandability).

As a preliminary matter, though, it is important to clarify the definition of "data." For the purposes of discussing metadata, "data" is any collection of facts or statements which may have been obtained through measurement, declaration, observation, perception, or other means. Data is typically established for some purpose, though the data content may be truthful, false, fictional, or notional. Data may be machine-computable or human-cognizable. Additionally, "[a] data asset also includes a service that may be provided to access data from an application." (DoD Directive 8320.02)

Descriptive Metadata
Regardless of the contents of the data, data producers and consumers have a perpetual need to describe it, identify it, and archive it. Descriptive metadata concerns the identification of the contents of one or more data assets. Descriptive metadata may identify a data asset with a title and unique identifiers or locators for managing the data in an archive, and for retrieving the data at a later time. Descriptive metadata may also identify other archival aspects of data, including: the creator, author, publisher, manager, or release authority; the date of creation, publication, copyright, or (security) declassification; and the physical location of the data. While not a complete list, the characteristics listed above facilitate the management and archiving of data assets.

A special type or application of descriptive metadata is "discovery metadata." Discovery metadata is used to enable "data assets to be found using enterprise search techniques." (DoD Directive 8320.02) Virtually any descriptive metadata may be used to enable search and discovery of data assets through search techniques. However, the characteristics listed previously merely identify a data asset, but generally do not describe its contents. For this, descriptive metadata (and thus discovery metadata as well) may also include summary information about the data asset. This summary information provides a distillation of the data asset to a set of facts that is more easily searched than searching the data asset itself. Summary information may include representative keywords from or about the data asset, categories or topics for the specialized domain to which the data asset applies, "coverage" information like geospatial extents or temporal extents, or specific entity types ("mechanized infantry brigades") or named individuals ("CEO of Amazon.com"). Each of these additional characteristics may be used to describe a data asset for easier search and discovery.

Structural and Semantic Metadata
The two other types of metadata are structural and semantic metadata. Structural metadata is defined as the "information provided about a data asset that describes the internal structure or representation of a data asset" (DoD Directive 8320.02), and includes XML schemas, database table column structures, and formal grammars. Semantic metadata is defined as the "information about a data asset that describes or identifies characteristics about the asset that convey meaning or context." (DoD Directive 8320.02) The term "semantic metadata" is typically associated with data dictionaries, taxonomies, and ontologies, but is meant to ultimately align to authoritative bases of meaning and common understanding like military doctrine (e.g., JP 5-0, FM 3-0, MIL STD 2525C) and systems architecture agreements established in DoD Architecture Framework (DoDAF) views (e.g., AV-2, all OVs, DIV-1, and DIV-2). Structural and semantic metadata are essential to making data *readable* and *understandable*. But their impact on *discoverability* is just as great: a data asset's discovery metadata has structure and semantics too.

The Dublin Core metadata terms are an open, standard set of metadata "tags" established by international digital library communities as a basis for making resources discoverable. Dublin Core takes the form of a controlled vocabulary (semantics) and a standard syntax through the Resource Description Framework (RDF) (structure). Dublin Core arguably is the most widely used and most influential metadata standard today; it is the basis for other metadata efforts, including the DoD Discovery Metadata Specification (DDMS) and the M&S Community of Interest Discovery Metadata Specification (MSC-DMS). DDMS and MSC-DMS go far beyond Dublin Core to incorporate DoD and M&S-specific requirements, such as security markings, distribution controls, and many other elements, to drive better metadata interoperability and precision within DoD.

Standard formats and semantics enable discovery metadata to be searched more systematically and with greater precision than basic free-text searches or searching actual data, since data can be stored in a variety of inconsistent formats across an enterprise. Enterprise discovery metadata may require key attributes (who, what, where, how, and why) that standardize data attribution across the enterprise, speed searches, and promote understanding in a common, familiar way.

Once data assets are "tagged" with discovery metadata, including additional coverage details that adhere to common structure and semantic specifications, the metadata can then be shared to make the data visible to an enterprise and subsequently reused. Data consumers need to find data assets hosted by other organizations and services, and the data producers must publish their data assets in a way that makes them visible to prospective users. This is the essence of the DoD Net-Centric Data Strategy and resultant policies: to enable users to easily find and understand data asset uses, and then to allow them to incorporate those assets into their own data and software environments for use. If attributed sufficiently with proper structural metadata, this can often be accomplished by software tools.

Structured Search Techniques

Many different user interfaces are available on the internet. Whether shopping, looking for a movie schedule, or ordering a pizza, the user is working through some sort of interface that provides the ability to enter requests, receive information, and perform some sort of action (buy, sell, download, etc.). There are a myriad of ways in which these interactions are presented to the user; one appreciates well-designed websites, but poorly conceived ones can be

aggravating. The best interfaces are as intuitive as possible (no user manual required) and their output is useful. To that end, multiple design patterns have emerged for making search of remote repositories possible (see Table 1).

Table 1: Structured Search Techniques

Structured Search Techniques			
Search Type	**Advantages**	**Disadvantages**	**Example Technologies**
Simple: A data repository search service that data consumers access directly (either through web services or a search application).	Repositories can enable search of metadata and actual data assets, and may provide specialized search interfaces for the specific data they store.	Each repository must be visited and searched separately. Repository services must scale to support the frequency of searches from users.	• OpenSearch • ATOM Search API
Brokered: A centralized "search broker" takes search parameters from users and "brokers" the query to repository search services. The broker then aggregates responses into a single reply for users.	Repositories can enable search of metadata and actual data assets. Multiple data sites can be searched from single broker, capable of using a variety of repository search services.	Repository services must scale to respond consistently to brokered search requests with low latency.	• OpenSearch • ATOM Search API
Publish and Subscribe: Repositories or search brokers enable users to subscribe to data assets (a.k.a. "saved query"), and be notified of data assets as they become available.	Users are automatically notified of new search results without having to search manually.	Requires users to know what they are searching for; often established in conjunction with other search patterns.	• SMTP • AMQP • ATOM Search API • RSS
Metadata Providers and Catalogs: One or more centralized catalogs aggregate discovery metadata from repositories, creating a cache that can be searched by users.	Users receive rapid search results from a centralized search service with a common interface. Repositories are not burdened with search requests. This search type can be brokered.	Search functionality may be reduced based on what all repositories can support.	• OAI-PMH • AMQP
Semantic Web and Graph-based: A complex search based on terms *and relationships among terms* that enables more precise, intelligent searches across multiple repositories.	Greater querying functionality than simple search. These searches may use synonym sets, taxonomies, and ontologies to enable varying levels of machine reasoning capabilities.	Connecting existing repositories to semantic or graph-based technologies can be challenging.	• OWL • SPARQL • Tinkerpop Gremlin • Lucene/Solr

Unstructured Search

Structured search techniques, while powerful, have an obvious Achilles' heel they require properly designed and completed metadata for each data asset in the search pool. Modern data analysis tools have powerful inference engines capable of searching through data without structure. These kinds of tools have led to two generalizations. First, unstructured data is sufficient if the right data analytics tool is used to "pull" the metadata attributes out of an information pool of unstructured data. Second, with these tools, additional discovery metadata may be an unnecessary expense or superfluous effort to expose a data asset to discovery.

The problem with these premises is that these tools are designed for unstructured, free-text, hyperlinked data assets such as documents, presentations, and other complex information sources. Search engines such as Google make

heavy use of natural language processing (NLP) and machine learning to "make sense" computationally of tomes of textual content. In the end, the data made searchable is not computable; it is only readable by humans, and not ingestible by systems including M&S systems. These techniques break down when analyzing structured datasets because of the lack of inherent context in the data and an expectation of higher quality search results. For example, it is not enough to find data sets relating to a particular country or military unit; rather, analysts seek data sets produced by specific trusted sources within a specified timeframe and for a particular purpose. This latter goal is much harder, if not impossible.

Data Quality Indices

Data quality is "the correctness, timeliness, accuracy, completeness, relevance, and accessibility that make data appropriate for use." (DoD M&S Glossary 2011). Developing indices of data quality attributes is one method of documenting quality characteristics or attributes useful in determining the reuse value of an asset. Such indices would enable users and owners of catalogs and repositories to conduct better assessments of the reuse value of legacy assets. These quality indices would provide information for structured search that is not documented in structural and semantic metadata, but is a type of descriptive metadata. The indices would include information in traditionally captured metadata and extend beyond; it would add information such as data currency, relevancy, accuracy, and pedigree. Once an asset is found (with discovery metadata), and it is determined that the data structure and meaning meet requirements (with structural and semantic metadata), this additional quality metadata can help establish the contextual relevance and fit-for-purpose of an asset.

Adding data quality attributes to metadata to enable structured search, though, is an additional expense that is most efficiently performed when the data is produced or shortly thereafter, before it is stored and exposed for structured search. Ideally, either incentivized by policy or an appreciation for the entire user base, data producers would create the necessary metadata as a routine part of their data production process. Even if all data producers instituted this practice immediately, though, a large amount of insufficiently attributed data already exists in repositories, databases, LANs, and file cabinets. This data would need to be attributed with the approved metadata schema in order to be exposed to structured search. Such an effort is daunting, particularly in light of the expense of attributing that data vice the potential economic benefit realized from that effort. This begs a "business case" for the project of attributing older repositories of data that lack the metadata schema desired to optimize an enterprise structured search.

A business case for developing new metadata for older data repositories necessarily includes a quality evaluation of the data assets. One must address the question, "Are these data assets potentially reusable, and if so, in what order of precedence?" Similarly, a potential reuser must determine the reusability of the data assets available. Developing simple data quality indices and implementing them would aid in both evaluations. A sample index of some potential data quality attributes is provided in table 2.

Table 2: Sample Data Quality Index

Sample Data Quality Index		
Data Quality Attribute	**Description**	**Significance**
Currency	Date of creation, last use, or last validation	Some data loses value over time
Subject/Topic	Description of subject (what is the data about?)	Some data is important just because of what it describes or is related to
Pedigree	Information regarding whether data is authoritative for purpose/ mission at hand (e.g., name of data source and/or user)	Some data is important because of who/what created it, or used it
Relevancy	Information regarding relevancy of data (e.g., is the software/organization that used this data still in existence and/or relevant to the problem-at-hand?)	Some data may fit structurally, but because of the context of the data creation or the purpose for which it was generated, it may not be fit-for-purpose
Accuracy	VV&A documentation, records of data usage/results	Is the data from a source that is known for accuracy
Conceptual Value	Description of the asset's key concepts, contributions, processes	While the data itself may not be literally relevant, it may have key conceptual value as a reference

When making the decision to attribute data assets with a metadata schema to support our enterprise efforts, a key activity must be triage; DoD must maximize its effort to achieve the highest return on investment. Data quality indices applied in an evaluation process to create "classes" (first, second, third, etc.) will assist in creating a viable enterprise pool of data assets prepared for potential reuse.

REUSE ROAD TO SUCCESS

What can DoD do to increase reuse success? One solution could be to establish a comprehensive reuse approach that addresses:

- the financial and cultural challenges to reuse,
- the technical challenges, and
- the benefit to the user, possibly in functional areas.

This potential approach could use structured search capabilities, with metadata schemas that incorporate quality indices optimized for each functional area, as subsets of an overall meta-schema. This would address the problem that the value proposition for reuse is not necessarily the same between users in functional domains. For instance, for users of distributed simulation-based training, where training scenarios have wide visibility and reuse is well-established, the value of reuse is well-known; however, for users of physics-based simulations, access to a variety of scenarios may not be as valuable.

There are currently programs that are addressing facets of this reuse problem for specific functional domains, and DoD should provide the guidance, planning, and oversight to steer these programs to an efficient solution. For example, the M&SCO-funded Rapid Data Generation (RDG) project has an objective of building an environment that enables discovery, retrieval, and reuse of data assets from across DoD communities. To this end, RDG is developing the Common Data Production Environment (CDPE) software solution. The project is initially focused on high-value data reuse candidates of the training community, with initial capability releases providing order of battle (OOB) and environmental representation data. The CDPE solution includes data, as well as data services, repository and cataloging with structured search capabilities. The data services include data enhancement and transformation tools to help prepare data for reuse. In a broader context, the CDPE design is extensible to other domains; thus, it supports the RDG vision by providing data and data services enabling reuse across the spectrum of DoD communities enabled by modeling and simulation.

In addition to technical solutions like RDG's CDPE, the Department could adopt policy to encourage implementation and use of a "functionally-tuned" reuse system. Furthermore, DoD might promulgate a charter that promotes development of reusable components, maintains a reuse catalog of discoverable metadata and actual software artifacts, adopts structural search technologies, and establishes and funds an organizational structure that will encourage compliance and maintain currency. A key component to this plan would be incentivizing project managers to build reusable components. It is not enough to direct developers to reuse and build in reuse without also recognizing that investment is required to change the current culture. Rather than a single organization managing the process and resources, a framework along functional lines (similar to net-centric enterprise communities of interest) would probably be most advantageous to ensure that functional domain values are included in the reuse equation.

While the above solution outlines steps to encourage reuse in new data products, there is a wealth of legacy data assets and data asset attributions as described above. In addition to the recommended triage evaluation by data quality indices to focus limited resources on those assets that have the highest potential for reuse and cost savings, DoD could establish a "preferred data" catalog of discoverable, functionally relevant, and accessible candidates of data assets. This catalog could be tested initially within a single community or functional domain, such as scenario reuse for distributed training exercises, as a proof of concept in a limited investment. In this way, a realistic, evaluation of that archive can be based on the level of reuse when users are targeted by quality reuse candidates.

Further effort will be required to identify the unique set of quality indices for other domains. Each community or functional domain will have a unique index of key quality attributes based on needs and requirements according to its functional purpose and the software requirements of their models and simulations. For example, for distributed training exercises, key quality attributes may include OOB, force-level modeled, environment (terrain or location), architecture and run-time infrastructure (RTI), and the source of the data. Functional domains and communities could develop indices that support the efficiency in their business process, and target "fresh" data in these archives for attribution and discovery that has the most value to potential reuse users of that functional domain or community.

CONCLUSION

DoD must continue to emphasize reuse of M&S software, data, and other assets to reduce costs and generate return-on-investment gains. Present practical challenges for DoD M&S, including inadequate documentation, lack of common formal standards, and technical interoperability issues, must be addressed for reuse of these resources to be maximized across functional domains and communities. Efforts such as the CDR Integrated Project Team and RDG's CDPE are presently among DoD's major initiatives to promote reuse, but additional thrusts in the areas of discovery metadata, structured search techniques, and data quality indices are necessary. In this way, the Department could engender a "functionally tuned" reuse system to leverage all existing M&S assets, capabilities, and resources for the next generation of defense initiatives.

REFERENCES

DoD Directive 8320.02 (2007), Data Sharing in a Net-Centric Department of Defense.

DoD Discovery Metadata Specification (2013). Retrieved May 19, 2013, from http://metadata.ces.mil/dse/irs/DDMS/.

DoD M&S Glossary (2011). Retrieved May 22, 2013, from http://www.msco.mil/MSGlossary.html.

DoD Net-Centric Data Strategy (2003). Retrieved May 19, 2013, from http://dodcio.defense.gov/Portals/0/documents/Net-Centric-Data-Strategy-2003-05-092.pdf.

Frakes, W., & Terry, C. (1996). Software Reuse: Metrics and Models. *ACM Computing Surveys, Vol 28, 2*, 416-435.

Henninger, A., et al. (2008). Live Virtual Constructive Architecture Roadmap (LVCAR) Final Report, *Institute for Defense Analyses*, *M&SCO Project No. 06OC-TR-001,* 55.

M&S Community of Interest Discovery Metadata Specification (2012). Retrieved May 19, 2013, from http://www.msco.mil/resource_discovery.html.

Office of the Director of National Intelligence (2013). *IC CIO Enterprise Integration & Architecture*. Retrieved May 19, 2013, from http://www.dni.gov/index.php/about/organization/chief-information-officer/cdr-reference-architecture.

Petty, M., et al. (2011). A Reuse Lexicon: Terms, Units, and Modes in M&S Asset Reuse, *M&S Journal Fall 2012*, 20-27.

Riggs, W., Coolahan, J., Morse, K., & Gustavson, P. (2010). Live-Virtual-Constructive Asset Reuse Mechanisms – Final Report, *Johns Hopkins University/Applied Physics Laboratory, NASD-R-2010-023*, 1-55.

Government Accountability Office (1993). *Software Reuse: Major Issues Need to Be Resolved Before Benefits Can Be Achieved* GAO/IMTEC-93-16.

Wilkes, M., Wheeler, D., & Gill, S. (1951). *The preparation of programs for an electronic digital computer, with special reference to the EDSAC and the use of a library of subroutines*. X + 170 S. m. 2 Abb.