

Rapid Data Generation: A Flexible Data Discovery and Access Architecture

Kevin T. Gupton, Bruce Carlton, Roy Scudder
Applied Research Laboratories
The University of Texas at Austin
Austin, Texas

kgupton@arlut.utexas.edu, bcarlton@arlut.utexas.edu, roy.scrudder@arlut.utexas.edu

Dr. Rob Cox
PEO Simulation, Training, and Instrumentation
Orlando, Florida
robert.m.cox14.civ@mail.mil

Ralph O'Connell
Joint Staff J6
Suffolk, Virginia
ralph.m.oconnell.civ@mail.mil

The Defense Office of Prepublication and Security Review has cleared this document for public release (Distribution A) (Case No. 14-S-2334).

ABSTRACT

Obtaining and preparing the right data for M&S-based activities is a huge consumer of resources, regardless of the activity supported by M&S (training, testing, etc.). The Rapid Data Generation (RDG) project, sponsored by the US Department of Defense (DoD) M&S Coordination Office, on behalf of the Under Secretary of Defense for Acquisition, Technology and Logistics, was therefore established to improve the visibility and accessibility of data, as well as to reduce the time and effort necessary to integrate the necessary data for an M&S event.

This paper presents the RDG Common Data Production Environment (CDPE) system architecture. This architecture defines a service-oriented design that specifies how data provider and data consumer systems integrate to enable net-centric discovery, assessment, and retrieval of M&S-relevant data. The architecture has been implemented in “order of battle” data capabilities in addition to the “environmental representation” data capabilities. These two capability releases focus on the sharing of military force structure datasets, such as orders of battle, scenarios, and entity-type enumeration data, as well as geospatial imagery, elevation, feature, and weather effects datasets and 3D models. The CDPE system architecture design makes use of DoD enterprise standards with industry best practices and design patterns to achieve a solution that is agnostic to the types of data exchanged. Through the use of reference architectures with implementation-independent and -specific designs, the design is resilient and adaptive to evolving technologies. The architecture also incorporates design alternatives that mitigate the variety in data producer and consumer system architectures. As a result, the architecture can be applied by others to develop capabilities for data discovery and sharing across diverse, loosely connected communities. The CDPE system architecture enables the rapid use and improved reuse of the data necessary for simulation-enabled training and mission readiness exercises for multiple tiers of training, all while incorporating and enabling data sharing with peer communities.

ABOUT THE AUTHORS

Kevin T. Gupton is an Engineering Scientist in the Modeling and Simulation Information Management Group at the Applied Research Laboratories, The University of Texas at Austin (ARL:UT). He has over 14 years’ experience in enterprise system engineering, data modeling, and knowledge management. Mr. Gupton has developed net-centric applications and data services utilizing 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.

Bruce Carlton is an Engineering Scientist for the M&S Information Management Group at the Applied Research Laboratories, The University of Texas at Austin (ARL:UT). Mr. Carlton served as a US Army officer for over 16 years primarily as a systems automation officer (FA 53). He has over 30 years’ experience with military command and control systems and models and simulations, concentrating the last 15 years in information management for M&S. Mr. Carlton holds a Bachelor of Science degree in Aerospace Science from Central Washington University.

Dr. Rob Cox is currently an Assistant Program Manager (APM) for Rapid Data Generation. Prior to becoming an APM he was a member of the US Army PEO STRI Engineering Directorate supporting the SE Core program in PM ConSim as the Production Lead Engineer. Before that, he supported PM Instrumentation Target and Threats Simulators (ITTS) and was responsible for the engineering design and testing of full motion and tracking simulators. Before PM ITTS, he was a member of the Future Combat System (FCS) Lead System Integrator (LSI) Training IPT. His primary responsibility was the development of the common environmental representation for FCS. He led the SAIC Synthetic Natural Environment (SNE) Research and Development team developing core SNE technologies to include the US Army OneSAF Environmental Runtime Component (ERC) and the SEDRIS project. Prior to joining SAIC, Dr. Cox was a member of the USAF, where he held Program Manager Positions at the Defense Threat Reduction Agency (DTRA), the National Defense University (NDU), and with Air Force Weather (AFWA). Dr. Cox has an earned Doctorate from Texas A&M University.

Ralph O'Connell is a senior systems engineer with the Joint Staff J6 Deputy Director for Command, Control, Communications, Computers, and Cyber Integration (C5I). He has a BSEE from Virginia Tech, a Masters in Systems Analysis from the Naval Postgraduate School, and is certified DAWIA Level III Systems Planning, Research, Development and Engineering - Systems Engineering. Mr. O'Connell has over 30 years' experience supporting various acquisition programs. He is currently working to advance reuse of DoD Enterprise data and data services in support of the M&S Steering Committee through the fielding of the Rapid Data Generation Common Data Production Environment.

Roy Scudder is the Program Manager for the M&S Information Management Group at the Applied Research Laboratories, The University of Texas at Austin (ARL:UT). Mr. Scudder served as the Associate Director for Data at the DoD M&S Coordination Office from 2005-2009. He has over 30 years' experience in information systems analysis and development, concentrating the last 20 years in information management for M&S. Mr. Scudder's professional experiences are in the areas of data management and data engineering with a specialization in metadata. Mr. Scudder holds a Bachelor of Science degree in Applied Mathematics from the University of Tennessee.

Rapid Data Generation: A Flexible Data Discovery and Access Architecture

Kevin T. Gupton, Bruce Carlton, Roy Scrudder
Applied Research Laboratories
The University of Texas at Austin
Austin, Texas

kgupton@arlut.utexas.edu, bcarlton@arlut.utexas.edu, roy.scrudder@arlut.utexas.edu

Dr. Rob Cox
PEO Simulation, Training, and Instrumentation
Orlando, Florida
robert.m.cox14.civ@mail.mil

Ralph O'Connell
Joint Staff J6
Suffolk, Virginia
ralph.m.oconnell.civ@mail.mil

The Defense Office of Prepublication and Security Review has cleared this document for public release (Distribution A) (Case No. 14-S-2334).

INTRODUCTION

Hundreds of modeling and simulation (M&S) supported military events occur each year, ranging from individual and team, through Joint, and multinational events. These events support a wide range of military communities from Training (the most prevalent) to experiments, analysis, testing, and acquisition. While objectives and the level of detail vary across the spectrum of activities, the assimilation and preparation of data is increasingly time and labor intensive. Warfighters require on demand high-resolution high fidelity synthetic training environments. RDG will assist in alleviating the amount of time and labor which goes into preparing for a simulation event. As the time and labor costs go down the number of events can increase thus meeting more objectives over time than prior to RDG development.

Simulation planning and preparation includes the consolidation of data for describing the start conditions of the operational environment. Building from event objectives and mission tasks for a specific audience, simulation databases require detailed descriptions of friendly, hostile, and neutral organizations, equipment, and personnel. Supply and munition holdings for each organization or entity must be accurately configured. Operational plans and orders are defined for the specific scenario, forming "master scenario event lists" (MSELs) that support the event objectives. The geospatial environment includes terrain, elevation, geographic and manmade features, and weather features and conditions. There must be a consistent environmental representation across the audience presentation, virtual 3D renderings of the synthetic environment, and calculations of visibility, line of sight, and adjudicating physical interactions between entities within the environment. That environment is further complicated with communication network, utility networks, road networks, social networks, and most recently the inclusion of cyber threats and offenses.

Integrated data needed for simulation-based events often originates from more than one authoritative data source and multiple data production processes. The right data is not always known, easily attained, or readily usable. In the best cases, exercise planners and facilitators must search multiple websites to find, retrieve, and manually process data for a scenario. The net result is a sub-optimal M&S data production value-chain that demands significant time and manual resources that are inflexible to dynamic changes in the training scenario.

The DoD M&S Steering Committee has sponsored the Rapid Data Generation (RDG) initiative to incrementally develop a service oriented architecture that enables data reuse from M&S correlated data repositories, automated access to authoritative source data, and implementation of data enhancement services to rapidly create M&S data sets. The RDG Common Data Production Environment (CDPE) is emerging as a virtual one-stop-shop for distributed data resources and M&S scenario data production services that are designed to mitigate the time and man-hour costs of preparing data for simulations.

This paper describes the current and evolving RDG CDPE service oriented architecture and technologies in terms of operational, reference, and solution architectures. Other programs and projects may also benefit from the RDG systems engineering methods, design patterns, standards, and technology choices of the CDPE. Once the CDPE architecture is described, we highlight how data consumers interface with data providers through the CDPE and how 3rd-party tool and service developers may integrate in CDPE through Web application programming interfaces (APIs). In closing, we describe future goals for broader automation in data preparation and reuse.

WHAT IS RDG?

The Rapid Data Generation (RDG) project emerged from the premise that data reuse and data process optimization are the keys to reducing time and cost for not only training, but across testing, evaluation, experimentation, etc.. By making data more readily available, elevating existing tools and services for broader “enterprise” adoption, supporting new tools and services to fill gaps, and promoting data and service standards, RDG exists to address the data preparation needs of military simulation activities.

Motivation behind the RDG Effort

Warfighters greatly benefit from modelling and simulations (M&S) for acquisition, testing, experimentation, analysis, planning, and training. To realistically represent today’s dynamic operational environment requires new and agile processes and data sources.

Multiple scenario data generation systems exist, however, it is difficult to find, obtain, and reuse data to support M&S events. If you can find current data, it is difficult to obtain the data. If you can obtain the data, it is difficult to understand and transform the data for your use.

RDG enables the discovery, assessment, retrieval, and transformation of data to support events across the DoD M&S Enterprise. RDG significantly reduces the time and cost to generate and reuse M&S scenario-specific datasets.

The RDG vision is to implement a CDPE as a virtual environment of distributed capabilities. The objective is to field service-oriented architecture (SOA)-based web services to enable data and data services sharing across multiple organizations that produce and consume integrated, fused, and correlated data. The RDG approach requires the tagging of data in these systems with discovery metadata and the creation of integrated search capabilities that allow users to discover, assess, and obtain reusable data and services.

RDG does not envision replacing any existing data generation systems, but enabling the effective integration and interoperability of these systems. The result will be an integrated environment of data production capabilities.

What is “M&S-Ready” Data?

The initial focus of the RDG CDPE is to enable the sharing and reuse of “M&S-ready” data. M&S-ready data is defined as data that has been processed or prepared for use within a model or simulation application. The source data from a typical data producer may have utility across multiple mission areas. However, an “M&S data integrator” integrates, correlates, and translates data from multiple data producers to provide data for use within a specific model, simulation application, or federation.

Describing each type of data involved in an M&S event and the correlation points between data elements are beyond the scope of this paper but the initial focus of our efforts has been on data that is common across the communities with the most potential for reuse across the M&S Enterprise. RDG uses the catch-all term “order of battle (OOB) data” to mean the detailed force structure of friendly, hostile, and neutral organizations, equipment, platforms, sensors, weapon systems, munitions, supplies, and personnel; characteristics and performance parameters; operational environment, tasks, orders, training objectives, and scenario events. Environmental representation (ENV) data covers the physical environment of geographic and manmade features, terrain, sea, air, and space; imagery, elevation, and maps; static and moving visualization models; and the effects of weather upon the entities that are interacting within the environment.

M&S-ready data are produced from many authoritative and operational data sources. Some M&S-ready data are termed “actual” or “geo-specific” if a real operational environment is reflected (e.g., precise locations of buildings; ground-truth force structures of friendly or hostile forces). Other data are termed “typical” or “geo-typical” to denote that the data was supplemented or otherwise modified to reflect realistic—but not actual—operational environments for purposes such as training (e.g., typical opposing force structures (OPFOR), typical village, or terrain layout). Users of M&S systems must understand the differences in data sources and the purposes for which data are generated. To ensure proper use and reuse of data, data providers must mark their datasets to communicate the purpose and scope (aka applicable mission) of the data they produce so that data consumers can discern whether another organization’s data are suitable for their M&S applications.

BUSINESS AND ENGINEERING APPROACH

What is Our Approach, and Why is it Novel

The CDPE is a confederation of data providers, service providers, and data consumers. Systems are connected using technical standards, industry convention, and a limited common infrastructure; thus providing the user with a one-stop shop for data access. While many business plan alternatives and engineering approaches were evaluated, RDG project leadership adhered to the following guiding tenets.

Build upon existing infrastructure, but improve data management practices.

Trainers et al. rely on established data production centers today, utilizing processes that are not agile enough to support emergent training requirements of a dynamic threat environment. Data management practices (creation, curation, distribution, and utilization) within the training enterprise can be improved, yet the DoD budgets and patience cannot afford a “big bang” replacement of existing infrastructure. Change must be deployed incrementally, affecting existing systems, services, and business processes—data management must be institutionalized.

Make existing data visible and accessible.

Exercise planners, data consumers, and even data producers/integrators need the ability to discover and obtain available data and M&S-ready datasets. Discovery of existing data, notification/announcement of when new data is available, and awareness of when published datasets change are cues—or triggers—to data preparation processes. Through metadata, consumers must be able to assess the applicability of data to their needs. Finally, immediate retrieval must be possible—or delivery for datasets too large for network transfer—all pending security and authorization. Prior solutions that rely on helpdesks, emails, and phone calls for discovery or retrieval act to redirect manual effort, rather than eliminate it.

Make data understandable and usable.

Making data usable in their exercise planning systems, simulation, and other training-support systems involves interoperability of formats as well as the processes that generate data—data provider publish datasets for a particular purpose or context. As mentioned before, metadata helps exercise planners to assess the applicability of data to their needs, but datasets are rarely “ready to use” as-is without some customization. Users require tools and services for validating data, converting between formats, combining and aligning datasets, and otherwise enhancing, thinning, thickening, or supplementing datasets. Failing to address the usability of data to the customers’ data processes would stop short of fulfilling their data needs.

Prioritize solutions that reduce manual effort

Process automation is a priority guiding principle for future architectures. As data products and sources grow, so will the mundane, time-consuming tasks that users must perform to use the data—unless a system’s architecture attends to the automation of processes. To reduce the demand for touch labor, web services and web APIs as well as semantically unambiguous data formats require as much attention as web portals and user interfaces. Many tasks can be automated or semi-automated—data validation, search, subscription, metadata assessment, retrieval, data fusion, translation—leaving trainers and exercises planners to focus more on their primary responsibilities. Many phone calls and emails from data consumers to data or service providers (i.e., time-consuming manual effort) are avoidable through automation. To achieve automation, the services, data, and metadata must be computable.

Be responsive, agile, and accept compromise

Reducing time and effort for event data preparation requires business process optimization. New capabilities for data discovery and retrieval as well as data enhancement services are technically achievable, but the community-wide data management challenges are complex, nuanced, and distributed. The specific needs of individual user groups—exercise planners, scenario builders, terrain database managers, et al.—must be addressed specifically and not entirely as a generic enterprise data problem. Thus, as an understanding of the training, testing, planning, and other communities improves, the systems engineering process for a project like RDG must be responsive to those needs.

Current and Near-term Implementation Status

In November 2013, RDG fielded an initial operating capability that provides discovery and retrieval of unclassified order of battle data from two sources—the Joint Staff J7 Joint Training Data Services (JTDS) and the U.S. Army Training and Doctrine Command (TRADOC) Intelligence Support Activity (TRISA) Threat Force Structure data. While the data from these sources was initially created for training applications, proof of principle demonstrations have shown that this data is of value to other communities employing M&S, such as operational planning. By December 2014, RDG will be fielding a second capability release: classified order of battle data from JTDS data will be discoverable and accessible; the Training Brain Operations Center's (TBOC) holdings of Training Support Packages will be available for reuse; and discovery of environmental (geospatial) data from Army and Navy sources. Discovery of weather effects data is scheduled for early 2015.

The current RDG capability is accessible through a web browser at <https://cdpe.jten.mil/> and web APIs. Additional interfaces will be provided that support the use of RDG capabilities through the Ozone Widget Framework.

Challenges – What Problems or Setbacks had to be overcome by Our Approach

Data Provider issues

Coordinating with each data provider presented the greatest challenges because of the variety in differences with each. Each provider prioritized data reuse and cross-community data sharing differently. Each provider had a different technical maturity—ranging from the Joint Training Data Services (JTDS) with a structured database system with distributed user access, to TRISA Threats who managed data in analyst-readable spreadsheets and PDF documents.

Fielding RDG CDPE capabilities is based on external dependencies of integrating data providers—something that seemingly many programs defer for the programmatic challenges they present. Each provider is managed as a subproject with an integration concept/vision, schedule, requirements, design concerns, and success criteria. The overarching CDPE project management and systems engineering incorporates the many integration points to generalize the solutions as appropriate (e.g., no point-to-point solutions) and enables continuous delivery of capabilities as each data provider integration is completed.

Business issues

From its inception, the RDG program has been a cross-organizational, cross-community integration effort. By advocating for users in multiple communities, establishing a business plan and sustainment “home” for RDG has been an ongoing challenge. Many of these business challenges are being resolved or have been resolved, but the concern for sustainability has been a major influence on the RDG Architecture.

Recognizing that every standard published and every system built must be maintained, the RDG CDPE systems architecture prioritized a small operational and maintenance footprint. RDG relies almost entirely on open industry or DoD standards. The CDPE “core” services that coordinate data providers and data consumers was designed as a small installation with minimal administrative overhead. By focusing on a common, open architecture, many of the RDG goals are achievable through coordinating new and existing programs to align to standards and incorporate requirements of users in additional M&S user communities (see Standards & Technologies below).

THE ARCHITECTURE

In this section, we present the RDG CDPE systems architecture design. The architecture design is expressed from multiple viewpoints, requiring multiple diagrams to define the layers and aspects of the CDPE. The architecture

design attempts to adhere to DoD Architecture Framework (DoDAF) v2 by way of the Unified Modeling Language (UML) with the Unified Profile for DoDAF and MoDAF (UPDM). In spite of that, the RDG architects have taken effort to make the architecture useful to most interested stakeholders and to reflect the design decisions made to address the RDG user requirements and design constraints.

Overview (Parts)

The CDPE systems architecture is designed in layers, starting with expression of the “Operational” perspective. The Operational View is found in the RDG Concept of Operations [APM RDG, 2013], RDG System Requirements Specifications [APM RDG, 2013], and supporting annexes—Annex A: Order of Battle (OOB) Data Requirements Supplement [APM RDG, 2014] and Annex B: Environmental Representation (ENV) Data Requirements Supplement [APM RDG, 2014].

To satisfy the operational needs, a Reference Architecture and Solution Architecture were designed. The CDPE Reference Architecture defines the abstract elements (e.g., Data Providers, Data Consumers), design patterns, vocabulary, and technical standards that guide the integration of provider and consumer systems.

The CDPE Solution Architecture is the result of applying the Reference Architecture to satisfy the operational requirements. Specifically, the solution architecture shows how the CDPE organizations connect and interoperate. Subsequent sections explain the web API definitions and information exchange schemas that provider and consumers must implement to integrate in the CDPE.

Operational View

The data providers and data consumers integrated in the CDPE thus far support the following operational scenarios.

Operational Scenario #1

Simulation database managers must obtain detailed military force structure data and geospatial/environmental data for initializing a synthetic operational environment in simulation systems. The force structure (“OOB”) data includes hierarchies of military units, platforms, personnel, weapons sensors, and supplies. The environmental (“ENV”) data includes elevation, geospatial features, imagery, maps, models, and weather effects data for presenting geospatial intelligence to exercise participants and for simulation systems’ calculations about interactions in the physical environment. Both OOB and ENV data must be in “runtime formats” compatible with the systems to be used in the simulation-based exercises. The sources of these OOB datasets are JTDS OBS (US military forces) and TRISA Threats (typical opposing forces). The sources of ENV runtime datasets are the Army Synthetic Environment Core (SE Core), NAVAIR Portable Source Initiative (NPSI), Army Geospatial Center Joint Experimentation, Test, and Simulation Lab (AGC Jets Lab), and Air Force Environmental Data Cube Support System (EDCSS).

Operational Scenario #2

US Army units preparing for training must prepare Training Support Packages (TSPs) [TRADOC, 2012]. TSPs identify the training audience units, training objectives, training tasks, operational environment, training venue, opposing forces, “road to war”, PMESII-PT (Political, military, economic, social, infrastructure, information, physical environment, and time) factors, Operations Orders (OPORDs), and much more. The Training Brain Repository (TBR) is the system used to draft and publish TSPs; once TSPs are published, other exercise planners may utilize past TSPs as a basis for planning future training events. The operational environment elements defined in a TSP are the basis for OOB data and ENV data required for initializing simulations for training (see “Operational Scenario #1”). In addition to US Army training, the old TSPs and the TBR services show value for test and evaluation (T&E) event design, US Joint training exercise planning, and US Army experimentation and analysis activities. Figure 1 below depicts a baseline of operational resource flows (i.e., information “need lines”) of “order of battle data” between data providers and data consumers.

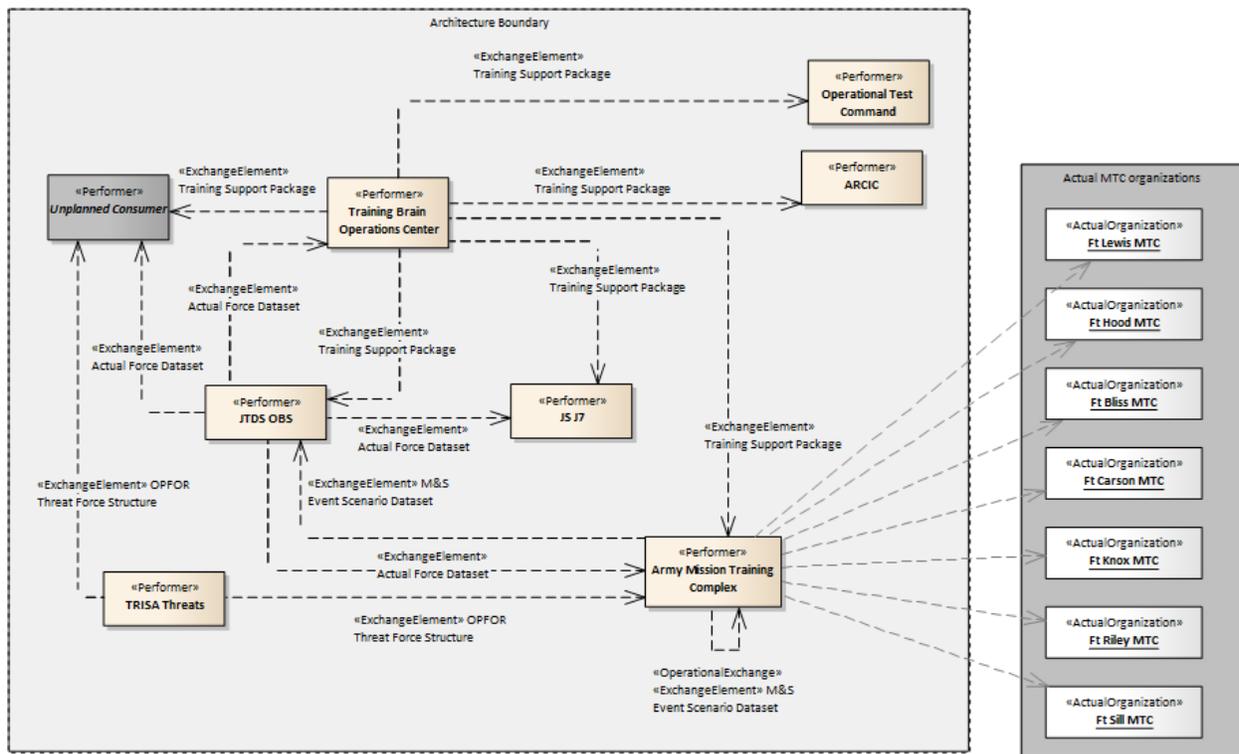


Figure 1 Operational information flows among OOB providers and consumers (OV-2)

Operational Scenario #3

Environmental data production centers such as SE Core, NPSI, and AGC JETS Lab prepare datasets for use in operational and simulation applications. ENV data production centers refine and correlate data from NGA, commercial source, and various other sources to generate products composed of correlated imagery, geospatial features, maps, elevation, and model data. While these production centers primarily generate runtime data products to support system initialization (see “Operational Scenario #2”), the production centers also share with each other. The production centers require continual awareness of and access to each other’s data holdings so correlated datasets may be reused and repurposed across services and M&S communities. In this operational scenario, data products are too large to support over-the-network retrieval so physical delivery of datasets is a necessity. Figure 2 below depicts a baseline of operational resource flows (i.e., information “need lines”) of “environmental representation data” between data providers and data consumers.

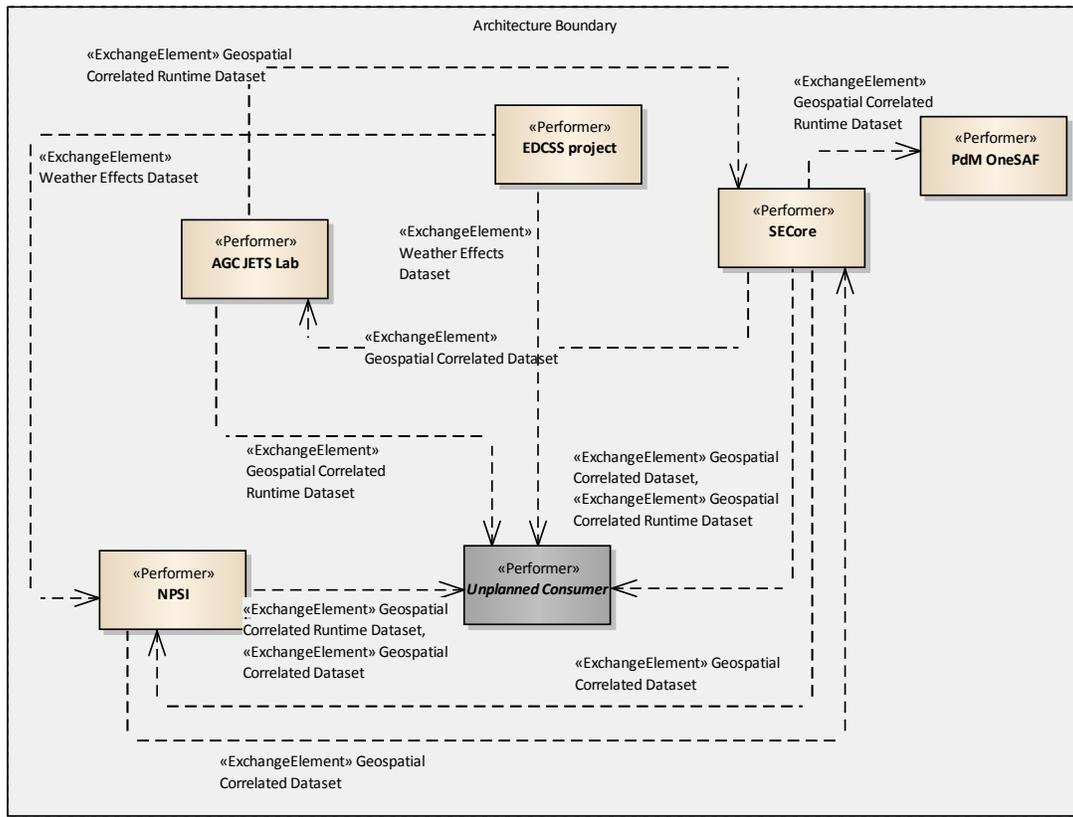


Figure 2 Operational information flows among ENV providers and consumers (OV-2)

Operational Scenario #4

Joint Staff J7 Joint Coalition Warfighting (JCW) is evolving their infrastructure to align with the Joint Information Environment (JIE) [DoD CIO, 2013] and is collaborating with US armed services to establish a Joint Training Enterprise Architecture (JTEA). Pursuant to those objectives, JCW is evolving system user interfaces—including those for exercise planners, simulation administrators, and trainers—to the “Ozone Widget Framework” (OWF). The activities described in other operational scenarios, the functionalities of JTDS OBS, TBR, and other services supporting joint training must additionally support web APIs and OWF-based widgets to provide a unified, common user interface (UI) for training support activities.

Reference Architecture

The CDPE Reference Architecture generalizes the operational requirements to define an abstract architecture that guides the integration of each data provider and data consumer in the CDPE. The RDG SyRS establishes five principal capabilities: Resource (data and services) Discovery, Data Retrieval, Data Enhancement, User Collaboration, and Access Management.

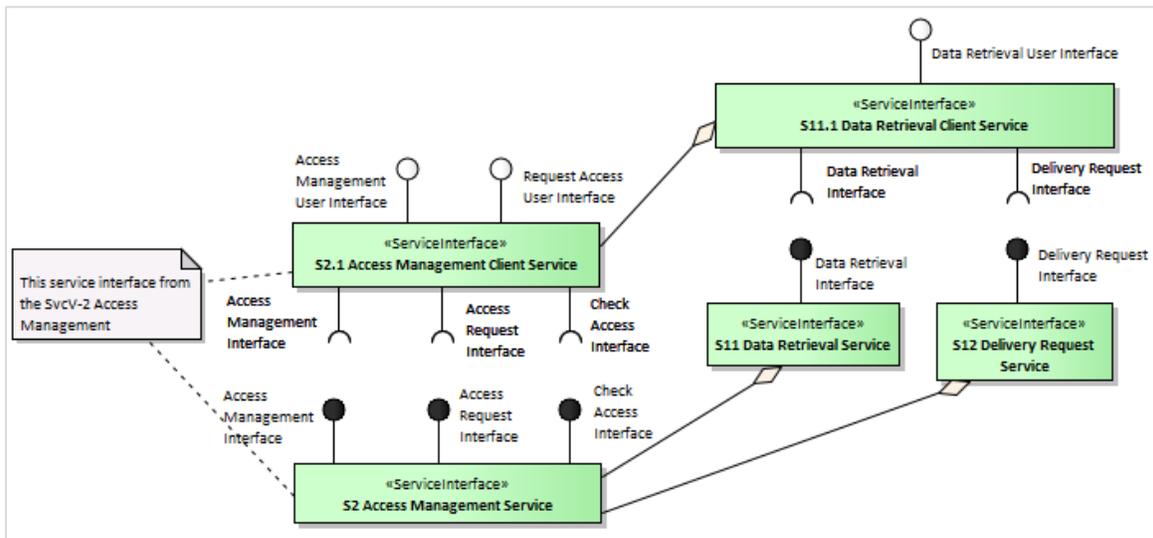


Figure 3 Data Retrieval services defined in the CDPE reference architecture

The RDG SyRS defines the services needed to enable the capabilities. Resource Discovery capabilities involve metadata management, metadata provider, search, query management, and publish/subscribe services. Data Retrieval capabilities include data retrieval and delivery request services, and so on. Access Management capabilities are essential to the other four capabilities as security and authorization control are needed for discovery, retrieval, data enhancement, and collaboration. Figure 3 depicts the Data Retrieval and related Access Management services.

The CDPE Reference Architecture is greatly dependent on the IC/DoD Content Discovery and Retrieval (CDR) Reference Architecture [IC/DoD CDR IPT, 2011]. The CDR Integrated Product Team (CDR IPT) established the CDR specifications to fulfill many of the same technical needs as RDG: to standardize the interaction patterns, service definitions, and interface protocols for the basic activities of search, brokered search, retrieval, and delivery of information resources.

Once the operational requirements and service requirements were considered in the CDPE SyRS, design alternatives were evaluated and a unifying systems architecture pattern was defined for data providers and consumers to connect in the CDPE. The following systems view shows the primary information flows for data discovery and retrieval.

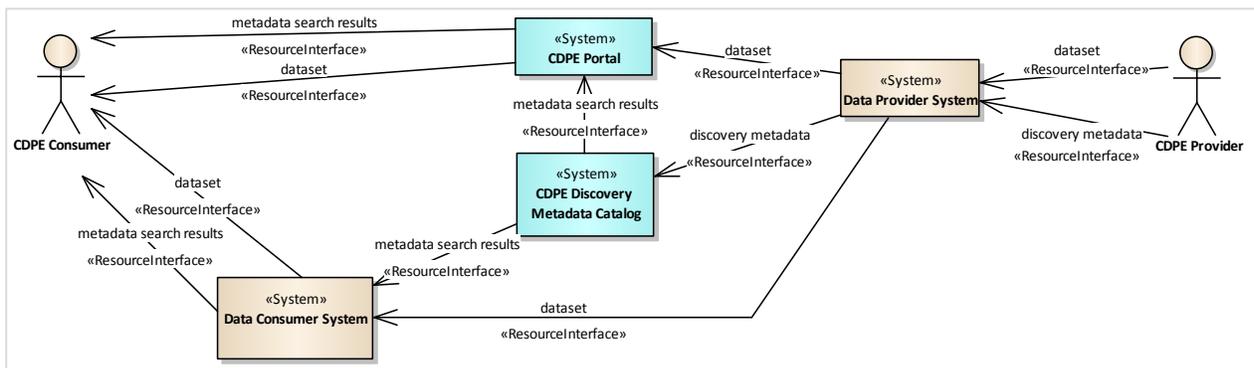


Figure 4 CDPE SV-1, data discovery and retrieval

Figure 4 depicts the pattern followed for integrating OOB data providers in the initial CDPE deployment in Fall 2013. The information flows may be read as follows:

- CDPE Providers users publish data and respective metadata in their external Data Provider Systems.

- Discovery metadata is *automatically* harvested from the Data Provider System by the CDPE Discovery Metadata Catalog; this harvesting occurs on a scheduled frequency. Metadata is processed/indexed to support search based on new, changed, or removed datasets at the data provider sites.
- CDPE Consumer users may then discover and retrieve the providers' datasets by searching the metadata at the CDPE Discovery Metadata Catalog. The consumers may interface in two ways: CDPE Portal or Web APIs.
- Option 1: CDPE Portal. The CDPE Portal is accessed with a web browser that enables a data consumer to search for datasets, assess the available metadata, and directly retrieve the datasets.
- Option 2: Web APIs. Instead of using the separate CDPE Portal website, data consumers may use their own data preparation or simulation systems to search and retrieve data using the CDPE web APIs, resulting in a more seamless integration of providers and consumers.

Standards & Technologies

The architecture presented so far is largely implementation and technology independent—that is, the interaction patterns are not restricted to a particular protocol and data format. This separation of high-level design from the vendor tools and protocol/data standards enables CDPE to change technical approach or support multiple technical approaches (e.g. REST and SOAP) without invalidating the entire design.

The primary standards used in the CDPE architecture are as follows:

- For metadata harvesting, the Open Archives Initiative Protocol for Metadata Harvest (OAI-PMH) is a widely supported standard from the international digital libraries community [OAI Technical Committee, 2002].
- For search and retrieval, the IC/DoD CDR REST specifications for search and retrieval are used [IC/DoD CDR IPT, 2012] [IC/DoD CDR IPT, 2012]. The CDR specifications are based on OpenSearch, basic HTTP-based data retrieval, and other open industry standards and practices.
- For exchange of discovery metadata—or “metacards”—the DoD Discovery Metadata Specification (DDMS) is used with extensions to support specialized OOB and ENV data search and assessment requirements.
- The OOB metadata extensions to DDMS utilized Universal Core (UCORE), the Global Force Management Information Exchange Data Model (GFMIEDM), and additional elements found in the Multilateral Interoperability Programme (MIP) Information Model (MIM). Future revisions will phase out UCORE and evaluate applicability of NIEM and MilOps.
- The ENV metadata extensions to DDMS are based on the National System for Geospatial Intelligence (NSG) Metadata Foundation (NMF).
- Viewing and exchanging geospatial data necessitates use of several Open Geospatial Consortium (OGC) standards including Catalogue Services for the Web (CSW), Web Map Service (WMS), Keyhole Markup Language (KML), and the Defence Geospatial Information Working Group (DGIWG) WMS Profile.
- For dataset formats, the CDPE *does not* restrict or specify what data formats should be provided by data provider systems. The CDPE architecture does, however, require that all data formats are registered in the DoD Data Services Environment (DSE; formerly the DoD Metadata Registry, MDR).

Solution Architecture

Several Data Provider Systems have been integrated into the CDPE based on the CDPE Reference Architecture while commitments are in place for more data providers to join in 2015 and 2016. Two CDPEs are deployed today—one on the US DoD unclassified “NIPRNet” (CDPE-U) and the other on the US DoD secret SIPRNet (CDPE-S). On CDPE-U, both OOB and ENV data are discoverable. On CDPE-S, only OOB data are discoverable, but ENV data providers will be added in 2015. Cross-domain data sharing (i.e. from NIPRNet to SIPRNet or vice-versa) is out of scope for RDG and currently not supported.

The CDPE web APIs and CDPE Portal are common interfaces for both OOB and ENV data, but data providers tend to be categorized as OOB-related or ENV-related providers. Thus, in the next two sections, the solution architectures for OOB data and ENV data are presented. Readers should note that the CDPE Portal, web APIs, and CDPE Discovery Metadata Catalog are the same for both OOB and ENV data.

Order of Battle (OOB) Data Discovery

The Joint Training Data Services (JTDS) Order of Battle Service (OBS) system and the Training Brain Repository (TBR) have been integrated into the CDPE based on the CDPE Reference Architecture. TBR's training support packages (TSPs), JTDS OBS correlated force structure data, and TRISA Threats typical OPFOR data are available in this solution, satisfying Operational Scenario #1 and Operational Scenario #2. Figure 5 below depicts the components and connections for discovery and retrieval of OOB data described here.

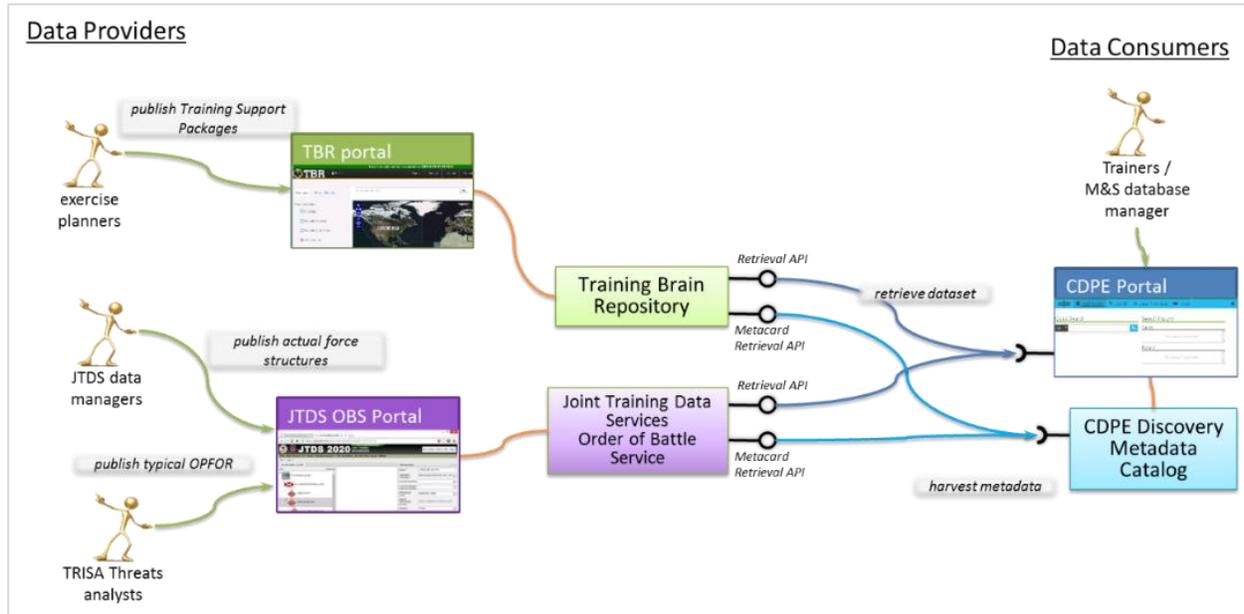


Figure 5 CDPE solution architecture for OOB data

Prior to CDPE integration, the TRISA Threats OPFOR datasets were managed offline in spreadsheet and document form, but now TRISA analysts manage OPFOR force structures as a distinct authoritative database within the JTDS OBS system. The OPFOR datasets are now available in XML form and accessible through the CDPE.

Environmental Representation (ENV) Data Discovery

The second major release of CDPE builds upon the OOB data solution by incorporating support for ENV data from SE Core, NPSI, AGC JETS Lab (Common Map Background, CMB), and EDCSS. The datatypes include full correlated data collections, imagery, features, maps, elevation, models, and simulation/system-ready runtime data products.

In order to address ENV data requirements—such as map based search and display of results, feature-type search, and mission-impacts of weather on platforms, weapons, and sensors—two new components have been added. The CDPE Metadata Geoserver is a component where metacards containing geospatial information (such as geospatial coverage, named locations and features, thumbnails, and overlays) are rendered as map layers and used for spatial searches.

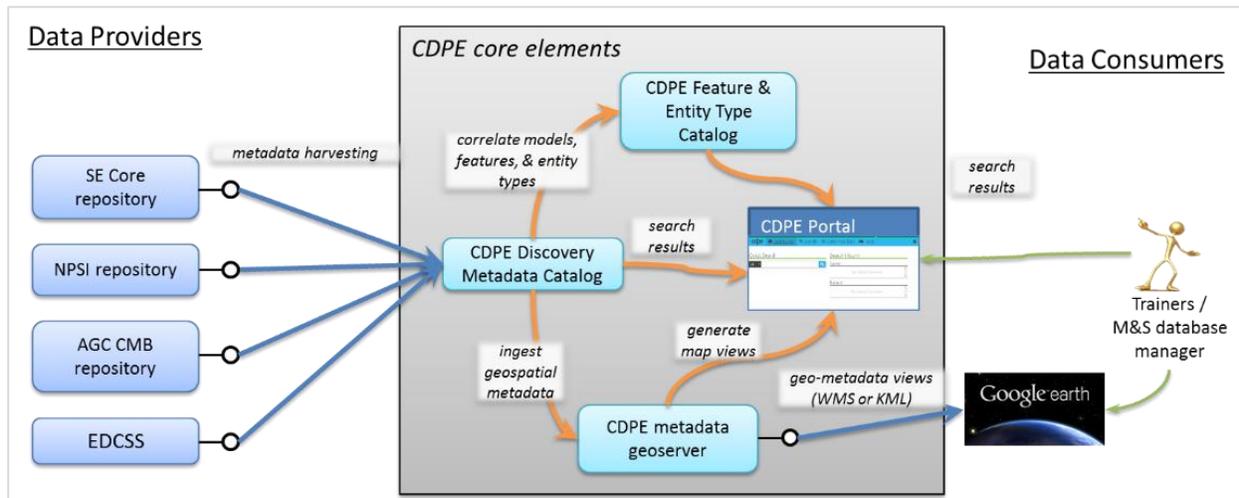


Figure 6 CDPE solution architecture for ENV data

The CDPE Feature & Entity Type Catalog manages the controlled vocabularies of geospatial feature type codes and platform, weapon, and sensor entity type compositions. In recognition that many coding schemes exist for feature and entity types, the CDPE architecture does not mandate any code list. The CDPE Feature & Entity Type Catalog also manages mappings between code lists, enabling future functionalities for semantic search and conjoining datasets that are based on different coding schemes. Figure 6 depicts how Operational Scenario #3 is supported for end-users of ENV data as well as the mutual data sharing among the data producers (i.e., SE Core, NPSI, and AGC are providers and consumers in the CDPE).

HOW PROVIDERS JOIN THE CDPE

Integrating a Data Provider System into the CDPE is relatively simple, though some data management changes may be expected. The CDPE architecture requires that data providers serve metadata in one of the formats established by the CDPE architecture. Data Provider Systems must support automated harvesting, either through OAI-PMH web services or through a simple publication of metacards on a website per the OGC CSW standard. Either direct data retrieval must be supported (per the CDR RESTful Search specification) or data providers must make clear how requests for physical delivery of datasets may be submitted. The data provider systems must conform to the user authentication and authorization schemes established by DoD. Of course, the architecture is adaptable and compromises are made for industry data sources or established data sources such as NGA, MIDB, or GFM DI services.

The CDPE Reference Architecture includes guidance for data providers to offer services for searching data or metadata directly from the data provider systems (making federated search possible). The reference architecture also specifies how data providers may optionally serve data and metadata through standards-based map layers viewable in Google Earth or other OGC-compliant tools; providers may serve summary information (e.g., thumbnails, summaries in KML, or overviews as 3D PDF files); providers may also include hyperlinks with metacards that reference related data products, source data products, additional metadata products, or tools and services for operating on the data (for example, TBR has chosen to include links to their website within metacards in addition to links for direct retrieval).

HOW CONSUMERS USE THE CDPE

As already stated in the Solution Architecture, trainers, exercise planners, and database managers may use the CDPE Portal through a web browser for basic search and retrieval access to the CDPE data providers. Additionally, developers of data preparation and simulation systems may integrate with the open web APIs available in CDPE. APIs for search, retrieval, map layer viewing, and more are detailed in the CDPE SAD [APM Rapid Data Generation, 2014] and can be implemented by data consumer systems to seamlessly integrate access to CDPE data providers.

PLANS FOR THE FUTURE

As previously discussed, RDG has already established an initial operating capability for unclassified force structure data, and will expand this capability to classified force structure data, training support packages, and initial environmental data by early next year. However, those capabilities are far from the planned full capability. In upcoming years, environmental data discovery and access capabilities will be expanded to include data from all the Services, and discussions are ongoing to expand this capability to coalition partners. The expansion to coalition partners will also likely include force structure data, as well as more complete scenario descriptions. Additional types of data are planned for upcoming years, including data that enhances the usability of force structure data such as equipment characteristics and performance and weapons effects. Planning discussion for potential RDG integration with providers for additional types of data are ongoing, to include out-year project force structure, static and moving 2- and 3-dimensional models, logistics, and readiness data. All of this data is critical to successful employment of M&S across communities, and highly reusable.

However, expansion of data providers is just one direction in which RDG will evolve. Just as important are capabilities in terms of tools and services that enable the automation and coordination of data preparation capabilities. Once the M&S-relevant data are visible and accessible through the RDG CDPE, improvement of data generation services and processes will soon follow. Development activities are underway within the RDG program, as well as at other US Joint Training, Army Training, and research laboratory programs, to create data “enhancement” tools and services that will enable and eventually automate much of the desktop work done today by data managers. Those tools and services will be coordinated through business process services managed by the trainers and exercise planners. Not only will data reuse potential improve, but also the value chain for data production can be optimized. As with data sharing, open standards and tools are available from industry to orchestrate business processes while establishing pluggable architecture for the many existing and emerging community tools to work together better.

RECOMMENDATIONS

In this paper, we have described an established and rapidly evolving capability that will greatly reduce timelines and level of effort for producing data for M&S-supported events across multiple communities. The potential community of providers can directly engage in multiple ways. First, those who would like to participate in driving the future requirements and priorities for RDG capabilities are encouraged to contact the authors. Secondly, those who currently expend the effort to prepare and integrate data products for M&S use, and believe their data has reuse potential beyond their initially targeted users, are also encouraged to contact the authors about becoming an RDG data source.

But just as importantly, this paper has described a highly reusable and scalable architecture for solving the more general problem of increasing data visibility and data access, and orchestrating data generation processes. That architecture, while developed for enterprise-wide potential employment, can be easily employed within individual communities, nations, commands, and Services. Through use of the open industry standards and simple design patterns—all of which are supported by multiple open source projects—the automation of data management of other communities, coalitions, organizations, and domains can also improve through data visibility, accessibility, and reuse.

REFERENCES

- APM Rapid Data Generation. (2014, January 24). Common Data Production Environment (CDPE) Systems Architecture Design (SAD) v1.1. Non-public domain reference.
- APM RDG. (2013, April 29). Rapid Data Generation (RDG) Concept of Operations (CONOPS) v1.0. Non-public domain reference.
- APM RDG. (2013, April 29). Rapid Data Generation (RDG) Systems Requirement Specification (SyRS) v1.1. Non-public domain reference.
- APM RDG. (2014, February 3). Rapid Data Generation (RDG) Systems Requirement Specification (SyRS) v0.9 Annex B CDPE Environmental Data Services Requirements. Non-public domain reference.

- APM RDG. (2014, April). Rapid Data Generation (RDG) Systems Requirement Specification (SyRS) v0.9 Annex B CDPE Order of Battle Data Services Requirements. Non-public domain reference.
- DoD CIO. (2013, September 18). Department of Defense Strategy for Implementing the Joint Information Environment (JIE). Retrieved from DoD Chief Information Officer website:
http://dodcio.defense.gov/Portals/0/Documents/JIE/2013-09-13_DoD_Strategy_for_Implementing_JIE_%28NDAA_931%29_Final_Document.pdf
- IC/DoD CDR IPT. (2011, February 25). IC/DoD Content Discovery and Retrieval Reference Architecture v1.1 25 February 2011. Retrieved May 15, 2014, from Office of the Director of National Intelligence:
<http://www.dni.gov/index.php/about/organization/chief-information-officer/cdr-reference-architecture>
- IC/DoD CDR IPT. (2012, October 3). IC CIO Enterprise Integration & Architecture, CDR: Retrieve. Retrieved May 15, 2014, from Office of the Director of National Intelligence:
<http://www.dni.gov/index.php/about/organization/chief-information-officer/cdr-reference-architecture>
- IC/DoD CDR IPT. (2012, October 3). IC CIO Enterprise Integration & Architecture, CDR: Search. Retrieved May 15, 2014, from Office of the Director of National Intelligence:
<http://www.dni.gov/index.php/about/organization/chief-information-officer/cdr-search>
- OAI Technical Committee. (2002, June 14). The Open Archives Initiative Protocol for Metadata Harvesting, v2.0 of 2002-06-14. (C. Lagoze, H. Van de Sompel, M. Nelson, & S. Warner, Eds.) Retrieved May 5, 2014, from Open Archives Initiative: <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- TRADOC. (2012, February 24). TRADOC Pamphlet 350-70-1 Training Development in Support of the Operational Domain. Retrieved from United States Army Training and Doctrine Command (TRADOC):
<http://www.tradoc.army.mil/tpubs/pams/tp350-70-1.pdf>