Interchange and Interoperability – Modeling Environmental Data with the Common Data Model Framework

Dale D. Miller, Annette C. Janett, Melissa E. Nakanishi, Leo J. Salemann, Timothy W. Miller
Lockheed Martin Simulation, Training and Support
Bellevue, WA
dale.d.miller@lmco.com, ajanett@lads.is.lmco.com, mnakanishi@lads.is.lmco.com, lsalemann@lads.is.lmco.com, tmiller@lads.is.lmco.com

Paul A. Birkel
The MITRE Corporation
McLean, VA
pbirkel@mitre.org

Denise Hovanec, Constance Gray
U.S. Army Engineer Research and Development Center, Topographic Engineering Center
Alexandria, VA
denise.m.hovanec@erdc.usace.army.mil, constance.b.gray@erdc.usace.army.mil

Julio De La Cruz, Todd Kohler
U.S. Army Research and Development Command, Simulation and Training Technology Center
Orlando, FL
julio.delacruz@peostri.army.mil, todd.kohler@peostri.army.mil

ABSTRACT

A logical Environmental Data Model (EDM) specifies the entities, attributes of the entities, and relationships between entities in any environmental domain: terrain, atmosphere, ocean and space. Formal EDMs have been developed for emerging and legacy modeling and simulation systems, data products produced by authoritative data providers, and for systems in the C4ISR domain. The Common Data Model Framework (CDMF) is a collection of tools based on Microsoft Access© that help automate the generation, maintenance and analysis of EDMs. For example, the CDMF automates answering questions like "What percentage of the environmental terrain data required by Objective OneSAF is actually provided by the National Geospatial-Intelligence Agency (NGA) product Urban Vector Map?" Other analyses allow investigation of environmental data interoperability between specific M&S and/or C4ISR systems. The CDMF was developed under government sponsorship and is freely available for both government and commercial use.

Recently, the CDMF has been extended to better support the interchange, interoperability, and use of EDMs in different communities of interest. While originally developed using the Environmental Data Coding Specification (ISO/IEC 18025 Final Committee Draft) as its data dictionary, support for multiple data dictionaries is now provided. Mappings may be defined between equivalent or related concepts in multiple data dictionaries. The mappings between individual concepts may be exact or approximate. The CDMF now supports interchange of EDMs with commercially available data modeling tools through the use of XML Metadata Interchange (XMI).

The CDMF has also been extended to support the forward-engineering of EDMs to physical data models (where the data structures themselves are defined) and realizations using commercial technology. For terrain representation, one physical data model takes the form of an ESRI Geodatabase. ESRI Geodatabase implementations are in current use in a number of Joint and Army systems, all key to the evolving National System for Geospatial Intelligence.

Key words: data modeling, environmental data, terrain data, data dictionaries, modeling and simulation

ABOUT THE AUTHORS

Dr. Dale D. Miller manages the Bellevue, WA office of the Advanced Simulation Center of Lockheed Martin Simulation, Training and Support (LM STS). He was the Principal Investigator for the OneSAF Objective System (OOS) EDM development and is currently leading environmental data modeling activities for the AMSO / TEC Geospatial Data/Information FACT. Dr. Miller received his Ph.D. in Mathematics from the University of Washington in 1976.

Annette C. Janett is a senior software engineer with LM STS Advanced Simulation Center. She previously led in the development of the OpenScene visual system, the STOW Dynamic Terrain and Objects, and Dynamic Virtual

Worlds programs, and STOW SE integration with virtual simulation. She has been the principal developer of the Common Data Model Framework and the EDCS ontology.

Melissa E. Nakanishi is an environmental data modeler with LM STS Advanced Simulation Center. She had lead responsibility for developing and maintaining the OOS EDM for Terrain, and has developed EDMs for several legacy M&S and C4I systems, including Janus, CASTFOREM, BBS, and the Joint Data Model (JDM).

Leo J. Salemann is a senior staff software engineer with LM STS Advanced Simulation Center. He received his Bachelor's of Science in Computer Science & Engineering from the University of Washington in 1993 and has been working for the LM STS Bellevue office ever since.

Timothy W. Miller is a geologist/geomorphologist and GIS applications developer with LM STS Advanced Simulation Center. He has 15 years of experience in all aspects of GIS, application and data development, relational database development and management, and web-based application development. He has been involved in all aspects of simulation Terrain Database generation, including software and data generation development.

Dr. Paul A. Birkel is a senior principal scientist for The MITRE Corporation. He currently provides technical support to the NGA in multiple standards areas including registers, feature data dictionaries and catalogues, portrayal, and modeling and simulation (M&S), as well as support to external standardization activities in the Digital Geographic Information Working Group (DGIWG), International Hydrographic Organization (IHO), and International Organization for Standardization (ISO). He also provides technical support to the Defense Modeling and Simulation Office (DMSO -- SEDRIS-related ISO/IEC standardization activities), Army Modeling and Simulation Office (AMSO -- AGDIMP, GDI FACT), and Army programs (FCS, PD CTIS).

Denise Hovanec is a project engineer at the Engineer Research and Development Center, Topographic Engineering Center (ERDC/TEC). Ms. Hovanec is the data model lead in the Digital Topographic Support System project office.

Ms. Connie Gray is an Associate Technical Director at ERDC/TEC in Alexandria, VA. As Associate Technical Director she is responsible for strategic planning and investment strategy for future system capabilities and program development for 6.1 and 6.2 research. Ms. Gray is the co-chair for the Geospatial Data/Information Focus Area Collaborative Team (GDI FACT).

Julio de la Cruz is the lead for synthetic natural environment and simulation technologies at the Simulation Technology and Training Center, Research Development and Engineering Command (RDECOM-STTC). Mr. DeLaCruz oversees research and science and technology objectives focused in the development of current Database Generation Systems/tools technology to address the defined Army need in Modeling and Simulation.

Todd Kohler is a Principal Investigator for the Research, Development and Engineering Command's Simulation & Training Technology Center (STTC) in Orlando, FL. Mr. Kohler was responsible for the Rapid Construction of Urban Terrain Databases for Training Science and Technology Objective (STO) as part of the Synthetic Natural Environments (SNE) research efforts.

Interchange and Interoperability – Modeling Environmental Data with the Common Data Model Framework

Dale D. Miller, Annette C. Janett, Melissa E. Nakanishi, Leo J. Salemann, Timothy W. Miller
Lockheed Martin Simulation, Training and Support
Bellevue, WA
dale.d.miller@lmco.com, ajanett@lads.is.lmco.com, mnakanishi@lads.is.lmco.com, lsalemann@lads.is.lmco.com, tmiller@lads.is.lmco.com

Paul A. Birkel
The MITRE Corporation
McLean, VA
pbirkel@mitre.org

Denise Hovanec, Constance Gray
U.S. Army Engineer Research and Development Center, Topographic Engineering Center
Alexandria, VA
denise.m.hovanec@erdc.usace.army.mil, constance.b.gray@erdc.usace.army.mil

Julio De La Cruz, Todd Kohler
U.S. Army Research and Development Command, Simulation and Training Technology Center
Orlando, FL
julio.delacruz@peostri.army.mil, todd.kohler@peostri.army.mil

INTRODUCTION

The Environmental Database IPT (EDB IPT) was established in 2000 by the Army Model and Simulation Office (AMSO) at the direction of the Army Model and Simulation Executive Council (AMSEC). Its vision was to develop:

> *An Army capability to generate, maintain, use and re-use realistic Synthetic Natural Environments – cost effectively and on demand – to support a wide variety of warfighter applications.*

Furthermore, the EDB IPT Charter emphasizes the need for terrain and other environmental data for emerging C4I systems, stating:

> *While this effort will result in increased efficiencies for the M&S community, the real purpose behind this Army initiative is **increased capability for our next generation of C4I systems.***

In 2003, leadership of the EDB IPT transitioned from AMSO to the Army Engineering Research and Development (ERDC) Topographic Engineering Center (TEC). Coincidentally with this transition, authoritative data producers (*e.g.,* NGA) have transitioned from a ***product-centric*** to a ***data-centric*** paradigm. Data sets of the future will no longer be produced according to stale specifications, but rather to fresh requirements of the warfighter. In 2004, the EDB IPT was renamed the Geospatial Data/Information (GDI) Focus Area Collaborative Team (FACT).

Environmental Data Modeling Scope of GDI FACT

The GDI FACT identifies and prioritizes critical areas related to interoperability and consistency of environmental (terrain, atmosphere, ocean, and space) data within the Modeling and Simulation (M&S) and Battle Command (BC) communities. The GDI FACT coordinates these priority technology needs with member services, agencies, and offices to resource the development, enhancement and maintenance of capabilities supporting cross-application interoperability of environmental models. These deficient technologies fall in the areas of: data models, data integration, data quality, data sets, data repositories, data distribution, and data updates. The objective of the GDI FACT-sponsored projects is to deliver cost-effective capabilities to support Future Force requirements for seamless environmental representations across the full range of warfighter applications.

In order to effectively resolve these deficiencies, the development, enhancement, and maintenance of processes, procedures, techniques, tools and repositories are required. Resolving these issues will result in significant payoff to the Future Force, establishing interoperable environmental data models for analysis and implementation. Expected benefits include the provision of more realistic training, improved mission planning and mission rehearsal, and enhanced Battlespace Awareness.

Today, the GDI FACT priorities are varied and include: extending and maintaining the Common Data Model Framework (CDMF) to capture and coordinate environmental data requirements across M&S and BC; developing and extending techniques and processes for data quality assurance and quality control; developing methods to define the structure of integrated data set transmittals as Data Set Content and Representation Specification (DSCRS) capable of supporting automated analysis and consumption by applications; enhancing, extending, and maintaining environmental data repositories; developing methods for consuming environmental data in current and developmental M&S and Battle Command applications; and investigating, developing, and promulgating mechanisms for incremental source and integrated data distribution and update.

Environmental Data Models

A logical Environmental Data Model (EDM) specifies the entities, attributes of the entities, and relationships between entities in any environmental domain: terrain, atmosphere, ocean and space. Formal EDMs have been developed for emerging and legacy modeling and simulation systems, data products produced by authoritative data providers, and for systems in the C4ISR domain. (Other categories of data models include the conceptual and physical. A conceptual data model is more abstract than a logical and generally does not fully explicate the features and attributes. A physical data model extends the logical to include definitions of low-level data structures and, for geospatial data, the actual geometry of feature instances.)

As we move toward network centric warfare, embedded training and the Future Force, current requirements for stand-alone M&S and BC systems (*e.g.,* CCTT as a self-contained training simulation) will be replaced with requirements for systems of systems, in which multiple, and potentially disparate, systems must interoperate.[1] Many, if not all of these systems will have representations of the environment. The degree to which these representations align will directly affect the quality of interoperability. Environmental data affects system interoperability and also affects the development of a system in three specific functional areas. First, there are the providers of data, whether observed/measured, interpolated, or generated/modeled. Examples of data providers include NGA[2], AFCCC[3], FNMOC[4], RTV[5] and INE ARP[6]. Second, there are specifiers of functionality, who define the behaviors / capabilities required. Examples include TRADOC[7], NSC[8], schools and consumers. Third, there are developers systems and software who must interact with the data and develop models and behaviors which utilize the environment in a reasonable manner. These include developers of M&S systems, BC systems, even games. Examples of government organizations that develop such systems include CECOM[9], PEO STRI[10], ESC[11] and PMS430[12]. All three of these functional groups need to share environmental data and information about the data.

One of the challenges of the GDI FACT has been to maximize the interoperability of legacy and emerging M&S and BC systems with regard to environmental data. A prerequisite to achieving this goal has been to formally capture the environmental data elements used by these legacy and emerging systems. The approach used for this has been to capture Environmental Data Models (EDM) for these systems. An EDM formally captures the entities (or features), their attributes, allowable values for the attributes, relationships among features, and various metadata. EDMs are discussed in more detail in (P. Birkel, 1999), (A. Janett, 2000), (D. Miller, 2000), (D. Miller, 2001), (C. Stevens, 2001), (A. Janett, 2002), (D. Miller, 2002), (D. Miller, 2003). A sizable inventory of EDMs now exists. These include EDMs for legacy and emerging M&S systems, BC systems and NGA products. These EDMs span the environmental domains of terrain, ocean, atmosphere and space. Figure 1 shows a timeline of the EDMs produced.

The most recent additions to the repository of EDMs include those developed in 2004 for the GDI FACT. These include:

- Geospatial Intelligence Feature Database (GIFD), Spiral 2 and Delivery 3.
- Unified Profile for Mission Specific Data (MSD) Air, Land, Urban, Littoral, Ocean; levels 1-5 – the Military Services' feature requirements for the air, land, urban, littoral, and ocean environments.
- Urban Tactical Planner (UTP) – a geospatial data product produced at the Topographic Engineering Center (TEC).
- Battlespace Terrain Reasoning and Awareness (BTRA) – a suite of tools designed to aid the tactical commander's understanding of the effects of the dynamic Battlespace Environment on personnel and equipment.

[1] For Millennium Challenge 2002, 42 different legacy M&S, C4I and COP systems interoperated on correlated terrain, see (Prager, 2002).
[2] National Geospatial-Intelligence Agency
[3] Air Force Combat Climatology Center
[4] Fleet Numerical Meteorology and Oceanography Center
[5] Rapid Terrain Visualization
[6] Integrated Natural Environment Authoritative Representation Process
[7] US Army Training and Doctrine Command
[8] US Army National Simulation Center
[9] US Army Communications-Electronics Command
[10] US Army Program Executive Office, Simulation, Training, & Instrumentation
[11] US Air Force Electronic Systems Center
[12] US Navy Performance Monitoring, Training and Assessment Program Office

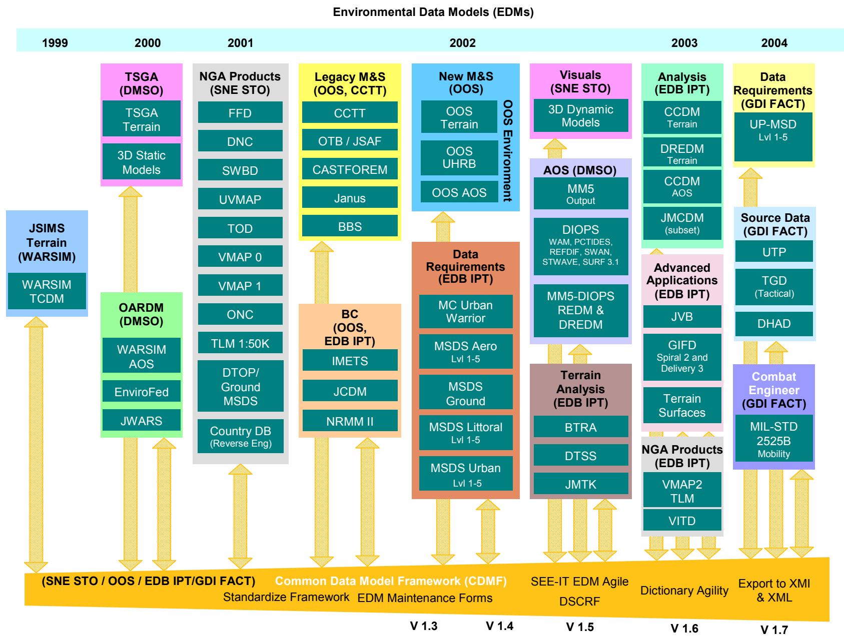

Figure 1. Timeline of EDMs produced

- Terrain Surfaces – this EDM provides a representation of the terrain surface for Grids, Adaptive Grids, TINs, and multi-resolution terrain.

COMMON DATA MODEL FRAMEWORK

Most commercial data modeling software tools (*e.g.,* ERWin®, Rational Rose®) are designed more for the capture of a data model in a graphical notation than for analysis of interoperability among data models. In addition, the prices of such applications tend to be high, putting them in the hands of professional data modelers rather than a larger user community. This section introduces the Common Data Model Framework (CDMF) and demonstrates how it overcomes the limitations of such commercial products while allowing interoperability with them.

The CDMF provides the common framework from which EDMs are initialized, defined, analyzed, reviewed, compared, integrated, maintained and exported. The framework operates from a common database schema and a set of utilities including MS Access® databases, forms, reports, code, and Visual Basic® (VB) applications that allow consistent EDM operations for EDMs compliant with the schema defined for the CDMF.

The CDMF was developed initially for M&S environmental data model (EDM) construction, but in fact could be applied to any system employing a well defined dictionary of terms. It provides for dictionary agility and provides several dictionaries for EDM construction.

The CDMF provides an abstracted paradigm for data modeling. Rather than entering and interacting with features, attributes and relationships exclusively via Entity–Relationship (ER) or UML diagrams from proprietary toolsets, the data are operated on and then stored in an open relational database using data entry forms for interacting with features, attributes, attribute values and relationships. It is this underlying common format using relational database capabilities that permits the rapid development of queries and automated analysis of multiple EDMs. Entity-Relationship and/or UML diagrams remain available via a CDMF export procedure.

The CDMF embodies both a structural specification for EDM development as well as a set of core functionalities as shown in Figure 2 to assist in the construction and analysis of CDMF compliant Environmental Data Models (EDMs).

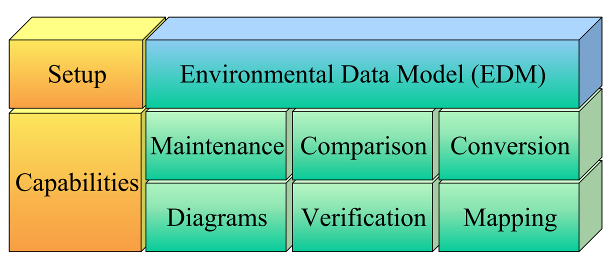

Figure 2. CDMF Capabilities

Key functionalities of the CDMF include:

- The ability to generate an EDM template for new EDMs,
- Dictionary agility allowing an EDM to be created using any one of the available data dictionaries,
- Forms providing for manipulation of features, attributes, attribute domains and feature relationships,
- The ability to generate delta-reports between CDMF compliant EDMs,
- The ability to generate comparisons between CDMF compliant EDMs,
- Standard report-based EDM comparison rubrics,
- The ability to check compliance of a subject EDM to the CDMF,
- The ability to maintain EDM content with configuration management and change histories,
- Automated generation of an EDM representation for import/export of ER diagrams,
- Automated export of an EDM to XMI providing for import into UML toolkits,
- The ability to export an EDM to a physical database representation in XML for ESRI Geodatabases,
- Mapping utilities to aid in mapping environmental content from one EDM to another, and
- Automated conversion capabilities allowing for CDMF compliant upgrades including updates of the underlying database schema or content.

The current capabilities of the CDMF have been implemented in an extensible manner wherein new functional requirements can readily be added to the existing capabilities. The CDMF is freely available for both government and commercial use and may be obtained from ERDC/TEC.

Data Dictionaries

Logical data models are dependent on a well defined set of terms used to describe features and attributes. The CDMF compliant EDMs are also dependent on use of a well defined set of terms – a data dictionary. To ensure consistent use of terms and to aid in creation of a 'well defined' dictionary, the data dictionary is maintained within a separate relational database. Early EDMs were created using the Environmental Data Coding Specification (EDCS) as their data dictionary, and several EDCS versions are available for use. Due to the need for additional terms required in many EDMs, the dictionary tables allow for new terminology to be added. New or modified terms are always flagged as such. In addition to terms found within the data dictionary, each individual EDM may also 'extend' the dictionary with specific terms as required. These extensions are not found in the data dictionary, but remain with the EDM content. Recent additions to the CDMF provide for dictionary agility, meaning that nearly any well defined set of terms may be provided as a dictionary and subsequently used in the construction of an EDM. Currently supported dictionaries include those in Table 1.

Table 1. Data Dictionaries Supported by the CDMF

EDCS 2.7	Based on DIGEST Feature and Attribute Coding Catalog (FACC)
EDCS 2.9	Based on DIGEST FACC, with additions
EDCS 2.9 with EDM extensions	Extended with common additions
EDCS 3.0 (CD) with EDM extensions	Committee Draft of ISO/IEC EDCS with common additions
EDCS 3.0 SEDRIS	SEDRIS release of EDCS 3.0
EDCS 3.1 SEDRIS	SEDRIS release of EDCS 3.1
EDCS FCD with EDM extensions	Final Committee Draft (FCD) of ISO/IEC EDCS with common additions
FACC 2.1	FACC 2.1
FACC 2.1 and JMCDM	FACC 2.1 and where no overlap, JMCDM terms are added
FACC 2.1 with EDCS FCD and EDM extensions	FACC 2.1 and where no overlap, EDCS FCD terms are added and common additions are added
JMCDM	Joint Meteorological Conceptual Data Model
TGD	Theater Geospatial Database

Data dictionaries are expected to include terms to be used for classifications, attributes, attribute enumerants, attribute value metadata, units, unit scale values, and optionally - equivalence classes and classification hierarchy.

The CDMF provides for the linkage of each EDM with its dictionary such that relational database queries can gather and present user friendly labels for reports and operator interactions. Additionally, using the CDMF, existing EDMs that have already been defined may be re-linked to an alternative compatible data dictionary or the EDM content may be converted to a different data dictionary.

CDMF EDM Template

In order to ensure the EDM has the correct database schema, one of the primary CDMF tools is EDM template creation. This tool creates a blank EDM (containing no content) having a relational database schema compliant with the CDMF. The CDMF allows for EDM template creation based on any one of the different data dictionaries as discussed above.

CDMF Compliant EDM Table Schema

In CDMF parlance, a feature has a unique identifier that is defined by a concatenation of three data items:

classification (*i.e.,* feature designation), feature type (*e.g.,* point, line, area, grid), and a variant code. The rationale for this is that the attribution schema for a given classification may depend upon the feature type or other feature variation. For example, a building as a point feature may have attributes of length, width and orientation angle, whereas a building as an area feature may have none of these because they are implicit in the geometry. An attribute is defined only in the context of the feature to which it applies – so an attribute is the following quadruple of information: classification, feature type, feature variant, and attribute.

CDMF EDM Maintenance

The CDMF provides tools for EDM Maintenance. Specifically, this is a set of forms and associated Visual Basic for Applications code allowing for operations on an existing EDM or newly created EDMs. These operations include:

- Add, modify or remove features,
- Add, modify, remove or replicate attributes,
- Add, modify or remove attribute allowed values and identify attribute default value,
- Adjust feature and attribute source requirements,
- Add, review or remove feature relationships,and
- Add, review or remove qualified attributes.

The CDMF maintenance provides for formalized revision management if desired and logs all changes made to features and attributes along with user identifier and change rationale.

Another maintenance form allows the operator to add a relationship specification to the EDM. The operator selects two features from list boxes and the desired relationship verb phrase and cardinality values for both the forward and reverse directions of the relationship.

EDM Comparison and Analysis

The CDMF provides for analysis of two or more EDMs and generates reports and statistics to evaluate the results of either a union analysis, an intersection analysis, or a comparison with respect to a target EDM.

Comparison and intersection algorithms may (upon operator direction) make use of data dictionary ontology and equivalence mappings (Miller, 2003) to perform "fuzzy" comparisons or intersections. A variety of specific reports are available upon completion of automated comparison and analysis. In addition to the reports, the operator may choose to create a new EDM representing the intersection or union of several EDMs.

Based on the information obtained from the comparison and analysis capabilities of the CDMF, graphical results such as those shown in Figure 3 may be created to provide insight on areas in which source data requirements may or may not be met or in areas in which system interoperability may be problematic. For instance, a comparison was performed based on the OneSAF Objective System Terrain (OOS-T) EDM and several source products that might be used to populate the data requirements. Another comparison was performed between OneSAF Objective System Ultra High Resolution Buildings (OOS-UHRB) and available source products. The resulting graphs (scaled from 0 to 100% compliance) from the CDMF automated comparison quickly demonstrate that while the required features do appear in many of the source products for OOS-T, there is less compliance for required attributes, and still less compliance for the requirements of OOS-UHRB.

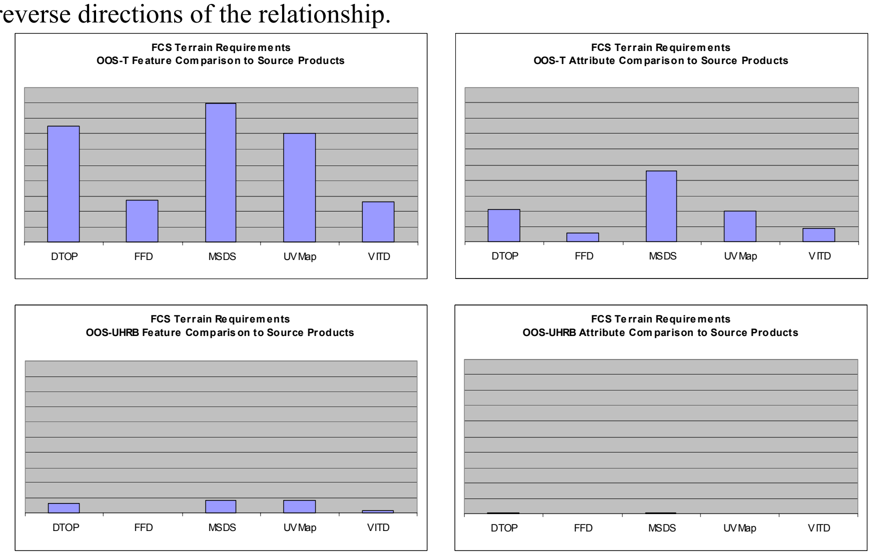

Figure 3. Results of Automated Analyses Comparing the Content Requirements of OOS against Data Sources

Reports on the Union of EDMs
All features and attributes from all of the indicated EDMs are gathered into a union. A set of reports is available to quickly identify lineage of the content:

- Features and Attributes with EDM Lineage,
- Features with EDM Lineage, and
- Features and Attributes with EDM.

Reports on EDM Intersections
A group of reports describes the potential intersecting features and attributes between the selected EDMs. When performing the intersection analysis, the operator may allow the system to use ontology and equivalence classes to attempt to find 'close' content, or the operator may decide that exact matches are required. Intersection reports include:

- Features Intersection,
- Features and Attributes Intersection,
- Features Union and Intersection Summary,
- Intersecting Features Query Use,
- Non-Intersecting Features Query Use,
- Intersecting Features,
- Non-Intersecting Features,
- Intersecting Attributes,
- Non-Intersecting Attributes, and
- Intersecting Features, Attributes, and Values.

Reports on EDM Comparison
The comparison reports all content based on one subject EDM that was identified by the operator as the primary EDM. These reports include

- Features and Attributes Comparison Summary,
- Features Comparison,
- Features and Attributes, and
- Features Comparison with Details.

EDM Verification Tools

The CDMF provides for extensible EDM verification capabilities. An operator-driven verification utility allows for verification of EDM content by finding content that breaks certain constraint rules. For instance, each feature must have at least one attribute, and an attribute coded as NULL is an acceptable attribute. When EDM content is found to break this rule, the operator has an option to add a NULL attribute. Early EDMs included significant content that was manually inserted and subject to error. All EDMs are based on an open relational database, and the underlying tables may be manually manipulated, which can result in errors or inconsistencies. Many of these errors are found and corrected automatically. The EDM verification conditions for which the CDMF searches and corrects include:

- Referential integrity,
- No attributes for a feature,
- No features for an attribute,
- Non-conforming attribute identifier,
- Non-conforming feature identifier,
- Duplicate attribute identifiers,
- Illegal length of feature or attribute identifier,
- No coverage specification defined (optional),
- No coverage identified for feature (optional),
- Missing attribute specification content,
- Missing qualified attribute specification,
- Missing allowable values,
- Undefined allowable attribute value,
- Undefined attribute,
- Undefined classification,
- Duplicate attribute addition code,
- Duplicate classification addition code, and
- Duplicate enumerant addition code.

Other Modeling Tools

IDEF1X
The CDMF provides for the automated generation of a database template of any EDM that can be reverse-engineered to produce ER diagrams in tools such as Microsoft Visio® or ERWin®. The ER diagram tools may also be used as a basis for a new EDM given a semi-automated process to import content from the exported ER diagram content. An ER diagram using the (Integrated DEFinition eXtended) IDEF1X notation can be generated.

UML
The CDMF provides for the export of any EDM to XMI 1.0 which may then be worked with in UML toolkits that support an XMI import capability. These tools include Rational Rose® and Poseidon UML®. The utility is provided as a CDMF tool written in VBA. The operator is presented with a simple GUI in order to select the source EDM, perform the XMI translation and output the results. We have found that UML is most suited to camel case[13] labels for classes and attributes, and, in the implementation of this capability, data dictionaries were enhanced with additional labels in a camel case form.

The underlying mechanism is based on the following components:

1. A XMI template definition in which all of the tags and sections of the XMI content are identi-

[13] The first letter of an identifier is lowercase or uppercase and the first letter of each subsequent concatenated word is capitalized, *e.g.,* "numberOfDays", "IsValid".

fied along with keywords indicating content that is to be replaced with EDM-derived values.

2. A set of SQL queries that may replace specific keywords (indicating EDM-derived values) within the XMI template with EDM or data dictionary content and place the output in a temporary table.
3. A temporary output table in which results are retained and later sorted and output to a physical XMI file.

A sample UML layout from Rational Rose is shown in Figure 4 below.

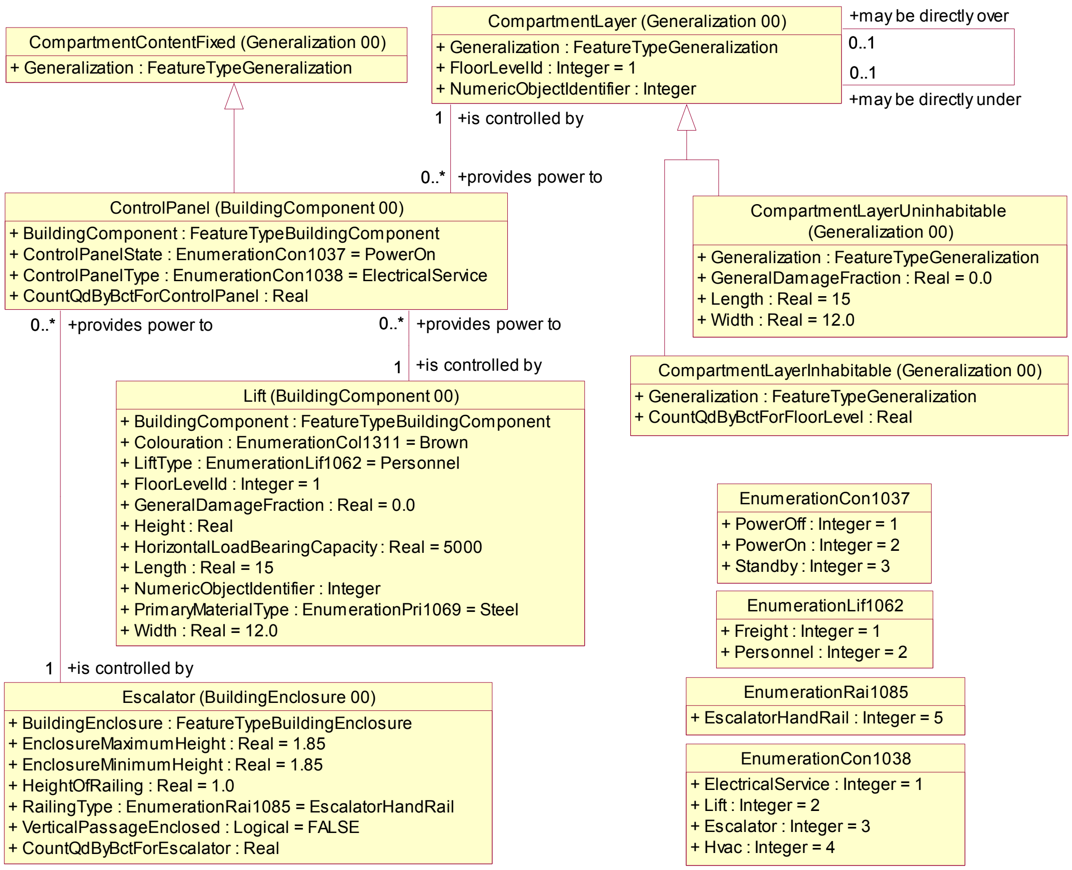

Figure 4. UML of "Control Panel" in the UHRB EDM

EDM Conversion and Mappings

The CDMF toolkit allows an operator to convert an EDM structure to a new CDMF compliant schema and to re-link EDM content from one common dictionary to another, providing there are mappings for the dictionary terms.

CDMF support for mapping of EDMs continues to be an area of on-going development. Capabilities developed early in the CDMF implementation include the automated conversion of any given classification to any other classification, and automated conversion of any given attribute to any other attribute. This allows for conversion of an EDM where there are one-to-one mappings for the terms. One-to-many and many-to-one mappings are a larger issue. To facilitate enhanced EDM mappings, the data dictionary provided with the CDMF contains several tables in which mapping entries may be maintained.

Future development is expected to be performed in which the mapping content may be used to facilitate comparison of EDMs that use different data dictionaries; today one can only compare EDMs that utilize the same data dictionary.

Fuzzy Comparisons for Semantic Equivalence

One of the operations supported by the CDMF is analysis of multiple EDMs including creation of union EDMs and intersection EDMs. Given a set of systems under consideration for interoperation, the Data Requirements EDM (DREDM) is formed as the union of the systems' respective EDMs. It represents the set of environmental data elements for which at least one system has declared a requirement. Similarly, a Reference EDM (REDM) is formed as the intersection of the subject EDMs, representing the set of environmental data elements that are common to all systems. Conceptually, one would like to tune the EDMs of the subject systems to *maximize* the content of the REDM.

Focusing on the terrain domain, a number of EDMs were selected for this REDM/DREDM analysis. These EDMs included OneSAF (OOS) Terrain (emerging high-resolution M&S), JSIMS Terrain Common Data Model (TCDM), JointSAF / OneSAF Testbed Baseline (OTB) (legacy M&S), Joint Data Model (JDM) (BC representative), Vector Product Interim Terrain Data (VITD) (NGA product for terrain analysis), Digital Topographic Data (DTOP) (terrain analysis), and various Mission Specific Data (MSD) sets representing Army requirements for high resolution geospatial data (Aeronautical MSD Level 5, Littoral MSD Level 5. Urban MSD Level 5, and Ground MSD Level 5).

A substantial number of environmental data elements are contained in these EDMs. The DREDM (union EDM) contains nearly 2,000 features and 30,000 attributes.

Before describing the intersection EDM, or REDM, we remind the reader that until recently, the CDMF was only capable of performing syntactic comparisons. For example, one EDM may represent a lighthouse as a BUILDING classification with a BUILDING FUNCTION attribute of LIGHTHOUSE. Another EDM may represent the lighthouse using a LIGHTHOUSE classification. While they are semantically equivalent, they are syntactically different, which the CDMF was unable to resolve.

That said, we expected the intersection EDM to be small but non-empty. Certainly all of these EDMs represent roads, rivers and other common features. We were surprised when the intersection EDM had zero features and zero attributes. So why did not even the ROAD classification appear? The JDM represents a road with the TERRAIN TRANSPORTATION ROUTE classification, and an attribute TERRAIN ROUTE TYPE to distinguish roads, railways, super highways and trails. Because the JDM does not utilize the EDCS as its native data dictionary, and because we tried to preserve the "look and feel" of the JDM by defining new "EDCS–like" concepts (which the other EDMs had no hope of sharing), we hypothesized that the JDM was an outlier and was skewing the results. This was partly true, but we found there was still not much commonality. As seen in Figure 5, zero features belong to nine EDMs, one feature belongs to eight EDMs, four belong to seven, and twelve belong to six. In fact 692 features belong to only one EDM. The commonality of the source EDMs was far less than we had hoped.

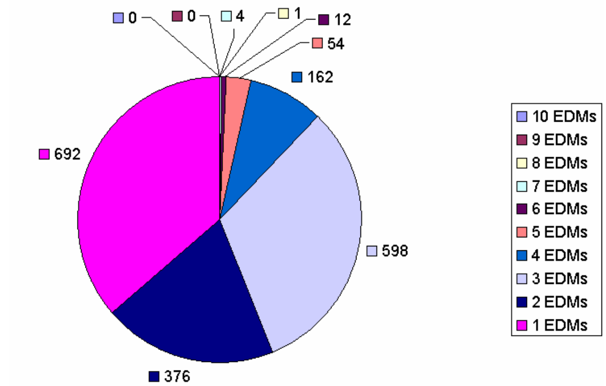

Figure 5. Features in Intersecting EDMs

The fact that this REDM turned out to be empty is also due to the generality or specificity of the concepts. Two EDMs might represent similar concepts, but one is more generic and the other more specific, *e.g.,* BARRIER vs. CROSS COUNTRY BARRIER vs. ENGINEER TRENCH.

A methodology was needed for performing "fuzzy" comparisons that better capture the semantic rather than syntactic similarities. This leads us to the two next subsections, Equivalence Classes and Ontology.

Equivalence Classes

We have provided additional contextual content to the EDCS in the area of classification to attribute enumerant value equivalences. This content captures equivalences including the often-used example of the classification LIGHTHOUSE being equivalent to the classification BUILDING having the attribute BUILDING FUNCTION value of LIGHTHOUSE. We have currently identified more than 650 equivalences.

Some potential "equivalents" may not provide an exact match. For example, an AQUEDUCT classification may be equated to any feature having the attribute TRANSPORTATION USE with the value AQUEDUCT. We cannot be certain that two concepts match completely. For this reason, we've included a Confidence Level field that is used to indicate the confidence of the match. The CDMF has been extended to allow comparisons based on equivalence classes.

An Ontology for the EDCS

Ontological relationships among EDCS classifications were developed to support fuzzy comparisons. Based initially on the EDCS classification definitions, we have derived a set of hierarchical relationships between classifications and, where indicated, the component relationships as well. Working from an initial ontology automatically derived from the hyperlinks in the definition of one EDCS concept to another concept, we have refined it with the goal of identifying common concepts hierarchically. At times, this called for adding classifications to help separate terms that were previously grouped.

The full ontology defined for the EDCS is very large and is impossible to show fully in this paper. It consists of over 1500 fully specified hierarchical relationships and 61 component relationships. From this, we can then derive over 6500 inherited relationships. Many EDMs capture data as per the legacy system or data product specification, and this includes variations of data resolution and different syntaxes. Using the EDCS ontology, we have built automated queries to capture these different syntactic content and resolutions across EDMs.

The CDMF has been extended to perform "fuzzy" comparisons of EDMs based on Equivalence Classes and this Ontology. Using it to build a REDM for the ten EDMs selected above, significantly more commonality is found.

A quantitative analysis using this new methodology is shown in Figure 6. Now, 28 features are common to all EDMs, 37 are common to nine, and only 66 are unique to a single EDM.

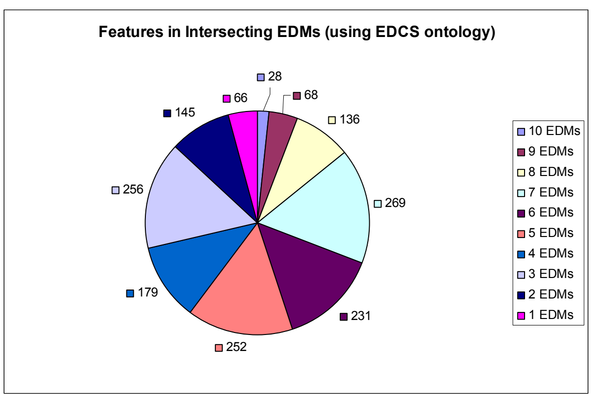

Figure 6. Features in Intersecting EDMs (using EDCS Ontology)

Dobey and Eirich (Dobey, 2004) argue that better automated syntactic comparisons of EDMs would be facilitated by using higher forms of normalization in the EDMs themselves. We agree but also point out that there is no canonical normalization of a logical data model, and no degree of normalization will render syntactic and semantic comparisons equivalent.

FORWARD ENGINEERING: LOGICAL TO PHYSICAL DATA MODELS

Logical versus Physical

Logical data models describe the relationships and attributes of an object or entity, but do not describe implementation parameters such as data storage. Physical data models describe the implementation of logical data models and are concerned with issues like data storage and access, indexes, and optimization. Logical data models are often normalized, while physical data models may not be normalized to optimize storage or access resources. Physical data models are typically tied to software systems and the data formats they use, such as a relational database or Geographic Information System (GIS).

An EDM is an example of a logical data model. It describes entities, attributes and relationships, but has no definitions for spatial data storage or coordinate systems. A lake areal feature in an EDM can include attributes for depth, temperature, and water quality, but would not have any description of vertices, lines, or polygons.

ESRI Geodatabases and XML Schema Specification

The ESRI Geodatabase

The ESRI Geodatabase is the main data repository for the ESRI ArcGIS system. Geodatabases are being used in all levels of government, from cities, counties, states, to the federal government. ArcGIS and Geodatabases are central to Army Battle Command System[14], the Commercial Joint Mapping Toolkit (C/JMTK), and software systems used by the National Geospatial-Intelligence Agency (NGA).

An ESRI Geodatabase is an example of a physical data model. While the EDM can be thought of as a recipe for what attributes a lake can have, the Geodatabase can represent a set of instantiated lake features, each with its own vector of attribute values, as well as a specification for the geometry of each lake polygon.

The ESRI Geodatabase provides a generic object-oriented vector data model where individual entities are modeled as objects, and objects with spatial information are called features. Objects and features are then collected into object classes (*i.e.,* tables) or feature classes. Feature classes can then be organized into feature datasets or be standalone like object classes. Feature classes within a feature dataset share a common spatial coordinate system and can participate in topological and network relationships. Objects and features

[14] *E.g.,* DTSS, BTRA

can participate in behaviors and relationships. The Geodatabase also supports metadata and data validation through attribute domains and subtypes. Attribute domains can be created for range data, coded values or enumerations, and text strings.

ESRI Geodatabase Designer

Geodatabase Designer® is an ArcCatalog extension that can be used to import and export Geodatabase schema in an XML representation. The extension was created by the ESRI Application Prototype Laboratory and is available at the ESRI ArcScripts website (http://arcscripts.esri.com/). The XML representation supports the core components of the Geodatabase including Object Classes, Feature Classes, Feature Datasets, Coded Values/Enumeration, Range, and String Domains, Subtypes, Topology, Geometric Networks, Relationships, and Metadata.

EDM to XML Converter

Our EDM to XML Converter, a component of the CDMF, can be used to read an EDM and create an XML representation of it. Two schemas are supported:

- XMI 1.0, which can be imported into a UML tool such as Rational Rose or Poseidon UML, and
- ESRI/Geodatabase Designer, which can be imported into an ESRI Geodatabase.

A set of queries and tables are used to read various EDM structures and format the appropriate XML/XMI nodes. The use of the converter is straightforward, allowing the user to identify the basic inputs and outputs of the process.

For the Geodatabase Designer schema, more user input is required. Recall that an EDM is a logical data model, while the Geodatabase Designer XML schema is a physical model. Whenever a logical model is converted to a physical model, a second set of inputs is required to supply the spatial component (as well as other inputs such as specification of the data types and data structures).

In the case of the EDM to XML Converter, the physical component is referred to as a spatial reference. (ESRI uses the same terminology to refer to a geospatial data set's coordinate system (projection), its origin and precision.) The EDM to XML Converter provides a drop-down list of most coordinate systems recognized by ArcGIS, as well as support for wildcard string searches.

Population with Instance Data

Once an EDM has been forward-engineered into a Geodatabase, the Geodatabase will have a set of empty feature classes into which instance data (geometry and attribution) can be loaded. ArcGIS includes several tools and techniques for loading instance data into Geodatabases. These include the ESRI Object Loader, an ArcMap edit session, and ArcObjects for creating custom data loading software.

Automated Data Set Conversion

The ESRI object loaders operate on a per-Feature Class basis. They are able to perform simple mappings from source fields to destination fields, but do not perform data-type conversions.

For a more automated approach, LM STS has developed EDM mapping and data set conversion tools. Used together, a data set based on one data model (e.g. UVMAP) can be re-mapped and loaded into a Geodatabase based on another (e.g. OOS-Terrain). The DataSet converter systematically iterates over all feature classes in the source data and populates the appropriate feature classes in the destination Geodatabase, without user intervention.

In addition to mapping a logical EDM to a Geodatabase, a semi-automated approach to the reverse mapping was also developed.

DATA MODEL–AGILE APPLICATIONS

Confirmation of our hypothesis that CDMF–based data models are more amenable to analysis than data models produced by more traditional applications depends, in part, on the proliferation of application software that accesses EDMs and performs analysis without innate programmatic knowledge of them. To date, a number of such applications have been developed. The WARSIM terrain compiler is fully data-driven in that it reads an EDM and a SEDRIS transmittal, producing a WARSIM run-time terrain database. Enhancements to the EDM (and associated SEDRIS transmittal) are transparent to the compiler software. The OneSAF Objective System is developing a similar data-driven approach (Stevens, 2001).

LM STS has developed a number of EDM–agile applications. For example, a data model mapping tool has been developed that creates, in a semi-automated fashion, mappings from one data model to another (even when using different data dictionaries). Based upon the XML output of this tool, a data model conversion tool is able to map instance data from one EDM schema to another.

EDM–Compliant Repositories

Most recently, the GDI FACT has developed a web-based geospatial repository infrastructure that is data model agile. The agility of the system stems from two core areas: (1) the generic physical data base structure of the geodatabase; and (2) data model-driven, plug-in architecture of the web-based access software.

Repository Import

The GDI FACT has developed a SEDRIS Transmittal Format (STF) import capability into a repository geodatabase. The import process is data-driven from an EDM and is structured in a wizard interface that operates within the ArcGIS software system. One application imports a STF into an intermediate geodatabase. A second application builds the repository geodatabase based from the contents of an EDM. A third application wizard converts the intermediate geodatabase into the repository geodatabase. All three applications are accessible through the ArcGIS ArcCatalog application.

Repository Web-Based Access

The web-based access has been implemented through the Apache/IIS web server, Apache Tomcat served engine[15], and SOAP[16] technologies. The web server and served engine provided a web-based communications protocol for the client and the server applications. The client is a JSP, Java Script, or HTML web page that communicates through simple URL base parameters. The served engine on the server accepts the request and triggers the export process that accesses the geodatabase and extracts the requested data as shown in Figure 7.

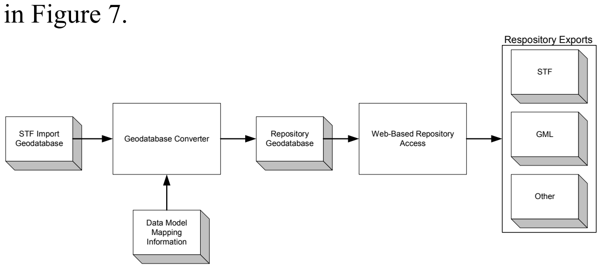

Figure 7. Repository data flow

FUTURE WORK

The Army Future Force requires an integrated capability to support embedded training and mission planning /rehearsal within 72 hours after deployment notification. Resolution of critical issues relating to interoperability and consistency of environmental (terrain, atmosphere, ocean, and space) data within and between the M&S and BC domains is necessary to satisfy the Future Force requirement. The objective of the GDI FACT is to develop, enhance and maintain capabilities supporting cross-application interoperability of environmental models to include: data models, data integration, data quality, data sets, data repositories, data distribution, and data update. The goal is to deliver cost-effective capabilities to support Future Force requirements for seamless environmental representations across the full range of warfighter applications. To effectively resolve these issues, the development, enhancement and maintenance of processes, procedures, techniques, tools and repositories are required. Resolving these issues will result in significant payoff to the Future Force, including the provision of more realistic training, improved mission planning and rehearsal, and enhanced Battlefield Awareness by providing interoperable environmental models.

The GDI FACT will support the following FY04 activities through AMSO's AMIP/SIMTECH program:

- Develop, update, and maintain a data dictionary, CDMF, and EDM components in support of the Future Force and evolving commercial infrastructure standards,
- Develop procedures for QA/QC of source and environmental data sets and enhance automated QA/QC capabilities, and
- Enhance, extend, and maintain environmental data repositories.

ACKNOWLEDGMENTS

The authors thank AMSO, the GDI FACT, the ERDC TEC and RDECOM STTC for support of this work.

REFERENCES

S. Prager, D. Miller and D. Gafford, Terrain Interoperability in Large Federations: Correlation and Consistency in MC02, I/ITSEC, Orlando, December, 2002.

P. Birkel, D. Miller, T. Miller, J. Rourke, E. Root, The Terrain Common Data Model (TCDM): a Joint Effort of JSIMS, STOW, and JWARS, Proc. 8th Computer Generated Forces & Behavioral Representation, Paper No. 8TH-CGF-004, May 1999.

A. Janett, D. Miller, J. Nordstrom, P. Birkel, Extending the JSIMS Terrain Common Data Model to Support Virtual Simulations, Simulation Interoperability Workshop, Paper No. 00S-SIW-112, March 2000.

D. Miller, A. Janett, M. Kruck, R. Schaffer, P. Birkel, B. Gajkowski, P. Woodard, Environmental Data

[15] Java based software system that provides a component-based, platform-independent method for building Web-based applications.
[16] Simple Object Access Protocol

Modeling for Simulation System Requirements Specification, I/ITSEC, Orlando, December, 2000.

D. Miller, A. Janett, L. Salemann, S. Farsai, P. Birkel, E. Miller, An Environmental Data Model for 3D Models, Simulation Interoperability Workshop, Paper No. 01S-SIW-071, March 2001.

C. Stevens, D. Miller, S. Prager, C. Campbell, J. Watkins, E. Root, J. Oneal, Data Modeling Implications for a Data Driven Environmental Software Sub-system, Simulation Interoperability Workshop, Paper No. 01F-SIW-024, September 2001.

A. Janett, D. Miller, S. Prager, Capturing Environmental Data Requirements – The Common Data Model Framework, Simulation Interoperability Workshop, Paper No. 02F-SIW-087, September, 2002.

D. Miller, A. Janett, M. Nakanishi, R. Schaffer, D. Wilbert, An Environmental Data Model for the OneSAF Objective System, Simulation Interoperability Workshop, Paper No. 02F-SIW-082, September, 2002.

D. Miller, A. Janett, M. Nakanishi, Environmnetal Data Modeling: REDM, DREDM, Ontology and Metrics, Simulation Interoperability Workshop, Paper No. 03S-SIW-132, March, 2003.

V. Dobey and P. Eirich, Environmental Data Models: Necessary but Not Sufficient for Interoperability, Simulation Interoperability Workshop, Paper No. 04S-SIW-114, March, 2004.