Establishing Sharing for Geospatial Environment Data

Mark Faulk
Cornerstone Software Solutions
Oviedo, Florida
mfaulk@cssflorida.net

Robert Cox
U.S. Army PEO STRI
Orlando, Florida
Robert.M.Cox14.civ@mail.mil

Bill Reese
U.S. Army PEO STRI
Orlando, Florida
William.C.Reese16.civ@mail.mil

ABSTRACT

Live, virtual, constructive, and gaming (LVC-G) integrated training environments bring challenges to data providers with increased target formats for geospatial environment data and visual models. New sharing points along the processing pipeline support consumer expectations of both managed correlation for interoperability and fair fight and optimized runtime content. Sharing points now include cleaned source, intensified source, confederate differentiated source, and runtime formats providing benefit of increased interoperability and fair fight through managed, well-defined levels of correlation. While datasets available from a single provider have increased, so has the number of providers bringing differing standards and conventions. With this increased sharing comes the complexity of managing the number and variety of datasets and providing efficient search and retrieval by data consumers. Oftentimes consumers ask for data from one sharing point without fully realizing the intended use for that share, resulting in poor reuse performance and consumer frustration. Maximum reuse requires incorporating externally developed, value-added data submitted with a variety of formats, data models, dictionaries, fidelity, and specialization levels. Provider reuse policies must balance between accepting un-validated data, risking contaminating their repository and full data validation which may be as costly as using raw source data. Effective data sharing across this vast set of available data possesses potential for improved approaches to managing the acquisition of geospatial environment data for the M&S community. Multiple initiatives have been established or proposed to address the standardization of metadata, exchange protocols, and data product formats toward improved interoperability both between sharing sites and with consumers. This paper describes how some of those efforts are converging to support improved human and machine discovery and selection and interoperability between providers. We describe real world experiences solving these problems from the perspective of a large data provider and propose future direction for effective data sharing.

ABOUT THE AUTHORS

Mark Faulk is a Systems and Software Engineer at Cornerstone Software Solutions, Inc in Orlando, Fla. He is currently the Lead Systems Engineer for the Army Synthetic Environment (SE) Core Common Virtual Environment Management (CVEM) program. Mr. Faulk received his bachelor's degree in computer science from the University of Central Florida. He has over 20 years of experience with simulation and training systems including the Close Combat Tactical Trainer (CCTT), Virtual Combat Convoy Trainer (VCCT), UK Combined Arms Tactical Trainer (CATT), Common Training Instrumentation Architecture (CTIA), and Engagement Skills Trainer (EST) 2000.

Dr Robert 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. He was responsible for development of the common environmental representation for the Future Combat System (FCS) Lead System Integrator (LSI) Training IPT. 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. Dr. Cox was also a USAF Program Manager 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.

William Reese is currently the Chief Engineer for the Synthetic Environment Core (SE Core) program, as a member of the US Army PEO STRI Engineering Directorate in support PM ConSim. Mr. Reese has his Bachelors of Science in Electrical Engineering from Virginia Tech, and a Masters in Computer Engineering from the University of Central Florida. During his 30 years within the modelling and simulation community, he was the lead engineer for several Army programs including the CCTT, M1 Tank Driver Trainer, Aviation Combined Arms Tactical Trainer (AVCATT), and the Conduct of Fire Trainer (COFT).

Establishing Sharing for Geospatial Environment Data

Mark Faulk
Cornerstone Software Solutions
Oviedo, Florida
mfaulk@cssflorida.net

Robert Cox
U.S. Army PEO STRI
Orlando, Florida
Robert.M.Cox14.civ@mail.mil

Bill Reese
U.S. Army PEO STRI
Orlando, Florida
William.C.Reese16.civ@mail.mil

PROBLEM STATEMENT

The Department of Defense has developed and utilized training devices for decades. These training devices have steadily grown in complexity and sophistication. Current training devices accurately replicate all aspects of vehicles, aircrafts, and weapon systems including cockpit layouts, physical feedback systems, realistic enemy forces, and immersive 3-D visual environments. During this evolution, training programs were typically designed as stand-alone systems. As systems later became networked to allow several Warfighters to train and interact, the idea of interoperable systems became a realistic goal.

However, since most systems were designed without interoperability in mind, the DoD is faced with several challenges before true interoperability can be achieved: differences in network communication protocols, damage effects, semi-automated forces, time management, and environment, particularly 3-D visual environments. Correlation between the 3-D visual environments (simulated terrain) poses the most obvious problem, as the trainee notices visual anomalies immediately. Differences in simulated terrain between systems cause significant training distractors and degraders such as in roads missing or misaligned, buildings existing in one system but not another, mobility mismatch, and elevation differences (Z axis) causing vehicles to float or burrow underground (see Figure 1) and significant targeting errors. Combined, these miscorrelations make systems non-interoperable for training purposes.

Figure 1. Miscorrelated elevation artifact

A Move Towards Common Terrain

Historically, simulated terrain developed by each program was not only miscorrelated, it was downright expensive. A small 100 km x 100 km database cost $1M - $3M to produce. Exacerbating the problem, the same geographical location would be built by multiple systems, within multiple agencies, by multiple DoD services. For example, the National Training Center (NTC), Iraq, and Afghanistan have been built dozens of times, for millions of dollars, and most do not correlate with another. This represents a great expense without the benefit of interoperability between systems. Even when programs had good intentions to interoperate with one another, the lack of coordination between programs, agencies, and services made it impossible to align schedules and budgets. As a result, few databases correlate.

As the Army began to focus on interoperability, it established common network protocols (DIS and HLA), common semi-automated forces (WARSIM and OneSAF), and common infrastructure (Live-Virtual-Constructive Integrating Architecture (LVC-IA)). These were necessary, but not sufficient. Correlated terrain is key to interoperability. To that end, the DoD established initiatives within each service to standardize terrain production processes. For example, the Army established the Synthetic Environment Core (SE Core) and the Navy established the NAVAIR Portable Source Initiative (NPSI) to accomplish those goals. This paper uses SE Core as the exemplar data producer.

Customer Growth Compounds the Problem

Typically, a majority of the effort producing databases is spent collecting raw source data, comparing that data against requirements, cleaning and attributing that data to meet the requirements, and placing that final data set into configuration management storage, named the Master Database (MDB) within SE Core. Once the MDB is populated with a geographical location (e.g., Afghanistan) that data is used to produce the dozens of correlated runtime formats needed for each customer program. For example, meeting the needs of the thirteen (13) LVC-IA core programs requires generation of over forty (40) correlated runtime outputs. Table 1 illustrates the number of program-specific runtime format-content products produced by SE Core. Orange cells indicate which products are required for an LVC-IA Integrated training Environment (ITE) terrain database delivery.

Table 1. SE Core Format Combinations

Format	AVCATT	CCTT	LVC-IA										LCT	NCM3	OneSAF Progr	Game For Traini	CDT	CFFT	JCV	NPSI	FSXXI	AME
			AARS	FIRSIM	HITS	VBS2	OneSAF	VARSIM	JDLM	JCATS	Muse3											
											VRSG	NVIG										
CADRG MIL	CS■			CS■	CS■			CS■	CS■					CS■								CS■
CADRG PFPS	CS■													CS■								CS■
CIB MIL	AV■							WS■						AV■								AV■
CIB PFPS	AV■													AV■								
CTDB (update on hold)																						
DAF										JC■												
DAT			ITE■																			
DRLMS	AV■													AV■								
DTED	AV■							ITE■					LCT■	AV■					AV■		AV■	AV■
DTED/DMED (Source)	ITE□													ITE□						CS■		
Environment Manager	AV■	CT■												AV■	CT■							
EPM CM2	AV■	CT■												AV■								
EPX-50 Terrain	AV■	CT■												AV■	CT■							
EPX-5000 Terrain													LCT■									
F1B				ITE■																		
File GeoDatabase																				CS■		
GeoTIFF Elevation		CT■													CT■					CS■		
GeoTIFF Imagery		CT■			ITE■										CT■					CS■		
GRID ASCII (Update on hold)																						
JPEG			ITE■																			
Metadesic (MDS) Terrain											ITE■											
MMRS													LCT■									
NVIG Terrain												OS■										
OpenFlight Static Models																				CS■	CS■	CS■
OpenFlight CM2	AV■	CT■												CS■				CS■		CS■	CS■	CS■
OpenFlight Terrain	AV□	CT□						WS□						AV□	CT■		CT■	CT■		CT■	AV■	AV■
OTF Terrain (Fielding)	AV■	CT■											LCT■	AV■	CT■		CT■	CT■				
OTF Terrain (Trunk)		CT■					OS■	WS■					LCT■		CT■							
P3D CM2		CT■																				
P3D Static Model		CT■																				
pbo Terrain		CT■				OS■									CT■	CT■						
Personal GeoDatabase	AV□	CT□											AV□	AV□	CT□							
PDF Paper Map	CS■													CS■				CS■		CS■		CS■
Project File		CT■						WS■							CT■							
PVD		CT■													CT■							
S2 Focus					ITEΔ																	
Shape w/FAC	AV□																					
Shape w/FACC		CT■	WS■					WS■							CT■				CS■	CS■		CS■
Text file (BMC)	AV■													AV■								
TIREM	AV■	CT■												AV■	CT■							
Tune Town	AV■	CT■												AV■	CT■							
Word Doc (DBDD)	AV■	CT■											LCT■	AV■	CT■							

LEGEND			
AV - AVCATT Specific content	ITE - Integrated Training Environment	OS - OneSAF Specific content	□ - Deliver to CM only
CT - CCTT Specific content	JC - JCATS Specific content	WS - WARSIM Specific content	Δ - Future format
CS - Common Source	LCT - Longbow Crew Training	■ - Deliver to Confederates and CM	"Orange" shaded color is for ITE format

"Runtime" refers to the final terrain product, tailored for a specific training system (similar to text tailored for Microsoft Word). Generating products for each runtime format requires specialized software, so as the customer base grows, so do the number of unique formats required. In addition to format differences, customer training systems bring unique constraints such as performance sensitivity to the number of polygons, graphics state changes, models, and levels of detail, which require further tailoring of content for each product delivery. Supporting this variety of differing outputs adds complexity to the architecture, as well as time and cost to produce the correlated formats and also complicates data sharing

The Need for Sharing the Data

As SE Core matured its processes and successfully began delivering correlated terrain products to customers, other groups within the DoD started asking for the cleaned and attributed data within the MDB. Leadership realized they could avoid the significant cost to collect and clean the data themselves. As SE Core began distributing the terrain data, the program recognized the enormous benefit to the DoD and tax payers to share the processed data. Soon, the number of clean data requests outnumbered the final runtime product requests.

At that point, SE Core began a systems engineering effort to understand the requirements and processes needed to effectively and efficiently share data products. Upon analysis, it became evident that sharing common data to a disparate set of customers is not a trivial task. The remainder of this paper addresses those challenges and solutions using SE Core as the exemplar.

A COMMON END-TO-END PROCESS

Providing consistent, correlated products across programs as varied as LVC-IA requires a reliably repeatable production process. This paper does not attempt to provide a comprehensive look into that process or the pipeline details. It does provide an introduction as a basis to define and discuss the resultant shared products and how they relate in content and provide reuse opportunity. At the basic level, the production process consists of:

- Define the requirements
- Acquire and ingest available source data
- Process the source to common standards
- Generate runtime specific output
- Perform product testing, acceptance, and delivery

Performing these basic steps becomes somewhat more complex when providing correlated products across many target runtimes.

Repeatable and Architected for Change

Figure 2 shows the top level SE Core production architecture (Moore, Nguyen, Bowman 2014). This production architecture must repeatably generate high quality, correlated products while allowing for the inevitable and seemingly constant changes including:

- New consumer programs requiring new runtime formats and expecting correlation with existing programs and existing products
- Changing and concurrent runtime versions to include complete swap-out of runtime technology by existing consumers
- Commercial and Government tools and formats changing between versions and reaching end of life
- Changing and expanding requirements for database extents and content fidelity

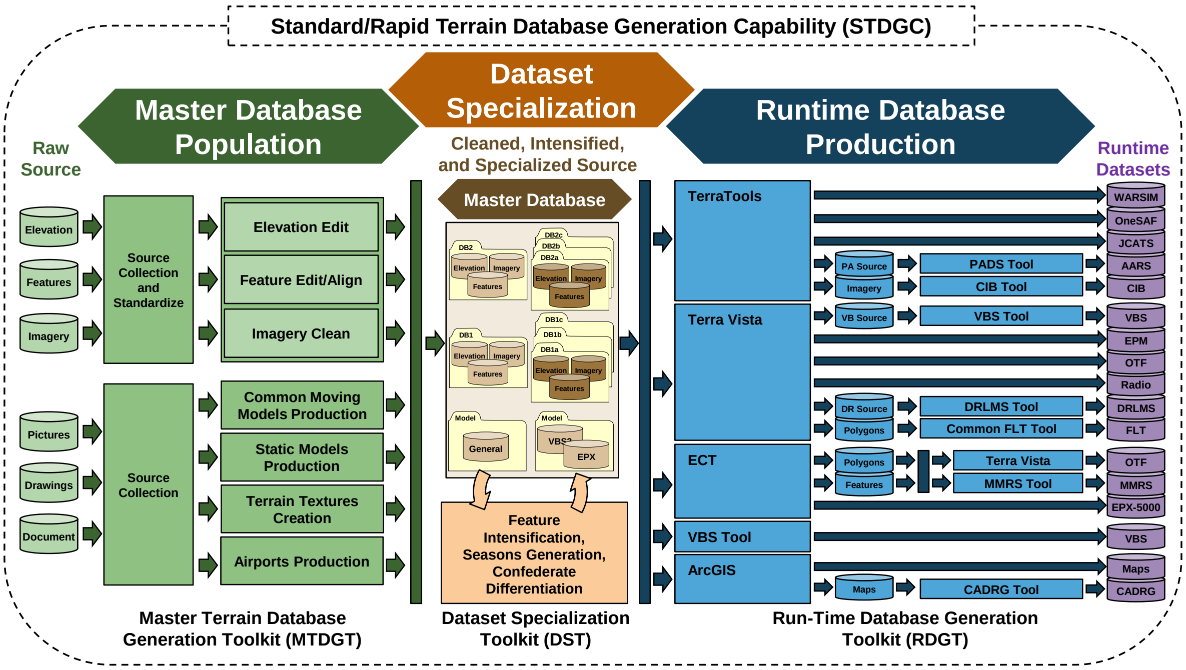

Figure 2. SE Core Production Architecture

This adaptability as a key architectural goal suggests additional architectural and design goals as follows:

- Maximize industry standards and de-facto standards for all major data exchanges. Most often this results in standard file formats, as more tools are intended to operate on files. The Esri Spatial Database Engine (SDE) database is a heavily used exception to file rule, but is an industry de-facto standard data storage container.
- Align component boundaries to typical COTS functional boundaries. This, in conjunction with standard formats, greatly simplifies the incorporation and chaining of COTS software components both by the provider and by reusing organizations.
- Minimize custom software to where truly needed for performance or required functionality. This also encourages standard output products for sharing.
- Support multiple database generation toolsets. There is little chance that a single tool will be the best solution for producing all runtime formats. As stated earlier, some final formats can only be generated using a specific custom tool. Using standard file formats further eases the ability to incorporate multiple externally provided toolsets with no custom wrapper or adaptor software.

This provides a flexible architecture into which we can easily incorporate new tools and technologies to support ever-changing customer needs and also promotes producing shareable data products adhering to industry and consumer tool expectations.

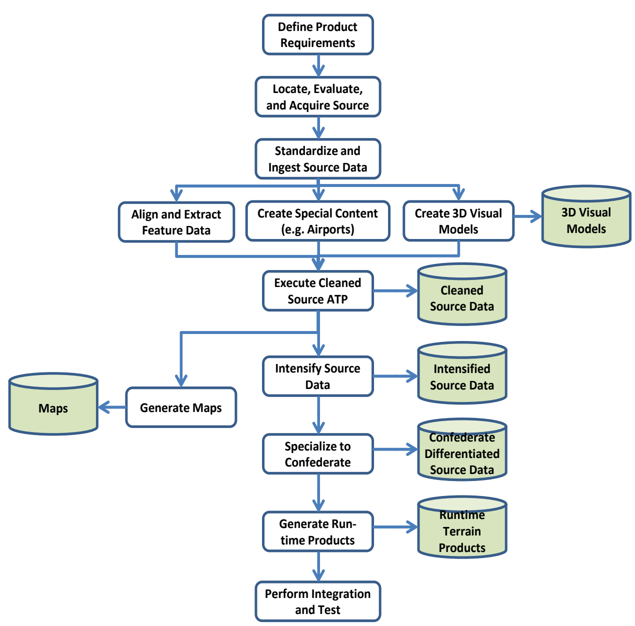

Figure 3. SE Core Production Process

DATASET SHARING AND SHARING POINTS

Opportunity for data sharing exists throughout the terrain development pipeline. Figure 3 shows the top level SE Core terrain development process flow. The

data stores shown in green indicate where the shareable datasets are produced.
The sharable datasets are:

- Source Data at the cleaned, intensified, and confederate differentiated processing share points. These three source sharing points will be described later in this paper. Each source share includes:
  - Terrain elevation as raster data
  - Terrain imagery and synthetic imagery
  - Feature data in vector format
- Static and moving 3D visual models
- Simulation-specific runtime formats
- Common sharing transmittal in OpenFlight
- Maps

Advantages of Sharing

Sharing of data has obvious savings benefits through shared cost of acquiring and processing data and through shortened schedules. Sharing from any sharing point saves the recipient the effort for the down-stream terrain development processing steps. Table 2 lists the estimated cost savings from reusing data at the different sharing points. While actual results will vary based on a number of factors, this provides a notional comparison.

Table 2. Estimated Savings from Sharing Points

Reuse Sharing Point	Typical Savings %
Cleaned Source	50%
Intensified Source	60%
Confederate Specialized Source	80%
Common Sharing Transmittal	80% - 100%

The benefit of increased interoperability and fair fight provided by a managed, well-defined level of correlation between resulting datasets is often more valuable than the cost savings. Accepting contributions from external developers and re-sharing that data, provides further benefit through value-added improvements, similar to the open source software model. When successful, these provide significant cost avoidance, reduced schedules across programs, and increased dataset fidelity beyond what a single consumer could afford.
Additional reuse benefits, as listed below, may be realized through the reuse of processes and tools.

- Rapid stand up of new source processing and terrain database generation
- Minimizes rework through pre-defined product quality standards and checks
- Eliminates tool market survey and selection
- Eliminates trial and error configuration of the tool chain
- Increased consistency across products from different terrain developers

Data Sharing Points

To understand why multiple sharing points are necessary, we first define our concept of Managed Correlation. Managed Correlation is the method/process/concept of identifying and quantifying the measure/degree of correlation between alternate run-time representations of common geospatial data, and allowing the customer(s) to determine if these differences will impact the intended use and select the most appropriate dataset. Managed Correlation acknowledges that there are tradeoffs necessary to maximize fidelity for each interoperating confederate while maintaining an acceptable level of correlation. Databases may have disparate fidelity, but still sufficiently correlate to support training. The alternative to managed correlation is to reduce to the lowest common fidelity which essentially brings all runtime content down to the least capable interoperating system. Figure 4 shows the four main sharing points along the terrain development process, which are described below. The three source data sharing points in particular can cause confusion and so are defined in more detail below.

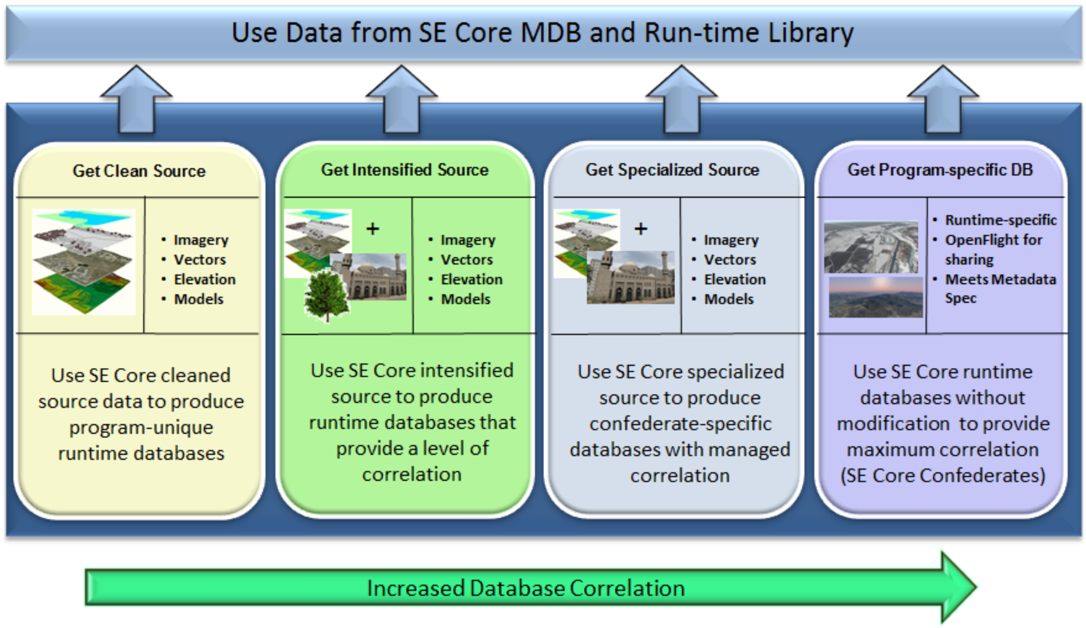

Figure 4. SE Core Sharing Points

Cleaned Source is the first processed sharable dataset. The initial source data is acquired from various sources, evaluated and prioritized for processing, and transformed to meet common specifications for format, data model, attribution, and dimensional units. Imagery is compiled, prioritized, georeferenced, orthorectified, and organized into an image server. Elevation is aligned and formatted. Vector feature data is aligned to or extracted from the imagery. Features are extracted from the imagery per the SE Core Scale Content Specification (SCS) which defines the features, attributes, and extraction criteria for each Area of Interest (AOI) scale. SE Core uses variable fidelity scales for cleaned source population. This ensures that the training Areas of Interest receive the most significant amount of detail. Currently four scales of fidelity have been defined with feature content established for each scale. The current scales include 1:250K (aligned with JOG-A maps), 1:50K (aligned with TLM 1:50K maps), 1:25K and 1:12.5K scales. Additionally, the SCS is implemented as a Scale Specification Interactive Reference (SSIR) for easy, two click access to all scale specification information. The feature extraction process is costly so effective reuse across providers generates significant savings. Detailed airport feature and model data is added based on published data. The cleaned source data is both a sharable dataset and the foundation for building the remaining source shares and runtime terrains. As such, a formal acceptance test procedure (ATP) is executed with full quality assurance and government participation.

The intensified source data builds on the clean source and is enhanced with additional features providing improved feature fidelity. This includes expanding aggregate features into individual features, increasing fidelity of existing features, generation of features beyond what is extracted per the collection standards, and building model assignment. Aggregate feature expansion includes scattering of individual tree point feature placement over a forest areal feature and individual geo-typical buildings within a built-up region areal feature. An example of increased fidelity of existing features is the addition of road lanes. Generated features include urban clutter and traffic controls and signage. Best fit building models are matched and assigned (by name since the final runtime models may vary). This intensification provides greater geo-typical representation of the database area without the added cost of collecting detailed geospecific features, particularly in lower AoIs. The intensified source is the data share with the highest fidelity of features. Use of intensified source will provide a medium level of correlation for the same database across different confederates.

Confederate differentiated source has been further processed for a specific target system or group of systems, is tailored to address those specific performance and storage constraints, and may have format translations applied. Sharing from this dataset provides near total feature content correlation across systems but is still in source data formats allowing for use by differing database generation system toolsets to generate differing runtime formats.

The final processed datasets are the runtime specific and Common Sharing Transmittal. Runtime specific formats provide near or total correlation but are of use only to exact duplicates or very similar systems. Common sharing transmittal provides a common, (OpenFlight for SE Core) format based on the selected runtime. This provides a shared output with geometry near to a corresponding runtime but in a more widely usable format. The common sharing transmittal closely matches the Aviation Combined Arms Tactical Trainer (AVCATT) runtime.

As indicated in Figure 4, correlation is managed along these datasets. Correlation decreases moving left in the drawing while fidelity and richness of the database increase. Maximum correlation comes from reuse of the actual runtime, as this provides polygon for polygon correlation of terrain skin and features. However, the customer must be able to execute with this same format and must be willing to live with the database built to the original target. Table 3 defines the levels of correlation provided by the different sharing points. A notional example of sharing across SE Core confederates is show in Figure 5. This shows only the reuse from planned confederates where SE Core produces the runtime products. Additional sharing takes place by request from all available sharing points by external runtime developers.

Table 3. Levels of Correlation

Dataset	Level of Correlation	Advantages
Cleaned Source	Same feature classes and aggregate features	No simulation geotypical content added. Real-world only data.
Intensified Source	Inclusive correlation (includes any feature found in any generated runtime)	Most feature-rich content. Content likely a superset of any generated runtime.
Confederate Specialized Source	feature correlation. Generated terrain geometry may vary based on toolset and any pre-processing.	Provides good correlation with support for differing runtimes.
Common Runtime	Near polygon for polygon terrain skin and atomic feature match.	OpenFlight available, built to match confederate runtime.
Confederate Runtime	Maximum geometry and feature correlation	Near or perfect match. Limited opportunity to reuse the exact runtime across systems

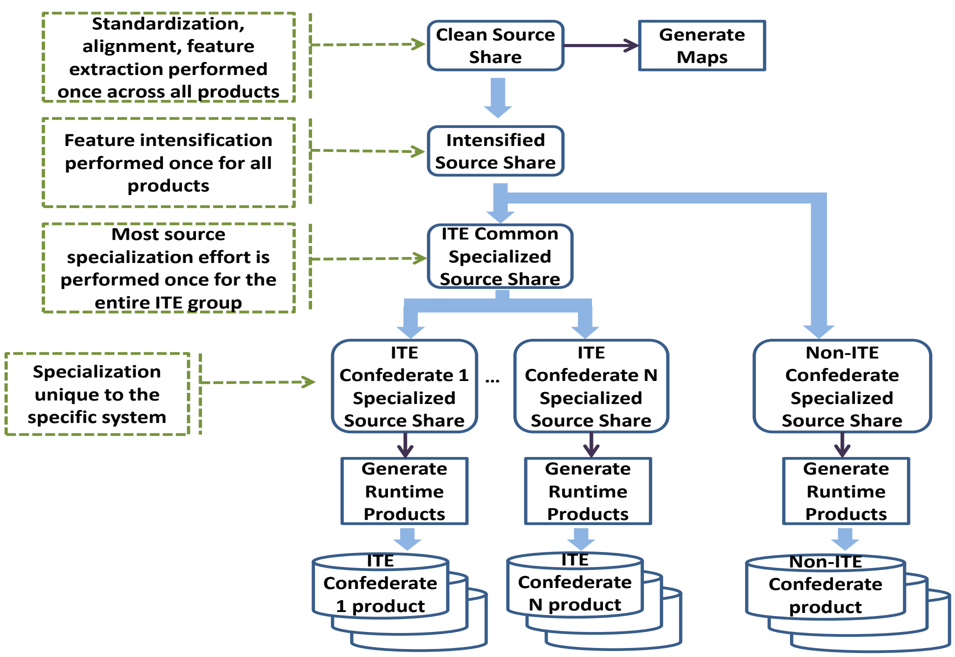

Figure 5. Notional Data Reuse Across ITE

Return on Investment

Database growth in size, complexity, and number of target runtimes all drive up the total costs of database production. This can obscure the fact that cost to confederate per area is significantly reduced through sharing. I.e. the cost increased but the size of the gaming area increased far more. Figure 6 illustrates the growth in database area size has far exceeded the cost growth for SE Core generated products, in large part through sharing.

It should be noted that savings from data sharing requires some overlap in geographic need within a reasonable time period. If the customer required a unique area or the provider's available data has become stale, there is limited or no cost sharing and the savings is limited to production setup cost and schedule savings.

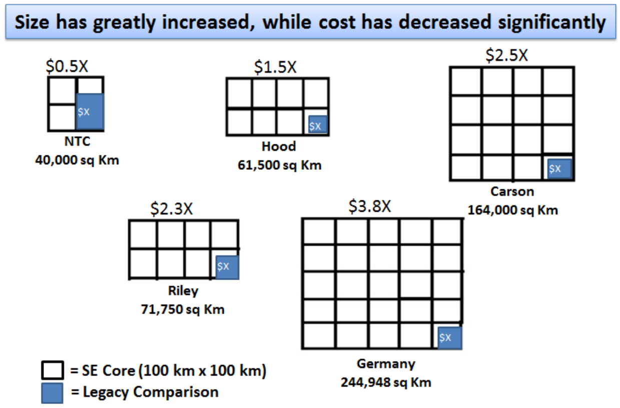

Figure 6. Cost per area decrease

HOW TO SHARE

Data must be shared using consistent formats, conventions, and quality standards. A reuse consumer must have confidence in the data and must be able to consistently process the data toward their final products. Commercial industry accepted standards, de-facto standards, and conventions are preferred as they allow for easy use of COTS tools for processing. We discuss the following categories:

1. Data Formats
2. Data Model and Dictionary
3. Dataset Organization and Granularity
4. Exchange Mechanisms
5. Metadata

Data format standards and conventions include file formats, data organization, and use of non-specific data fields. Many file formats exist and are commonly used in industry. We recommend limiting to a small number of these formats. Open formats and formats with no-cost APIs are preferred. Table 4 lists the recommended file formats for sharing data.

Table 4. Recommended Sharing File Formats

Data	Recommended Sharing Formats	Additional Specifications
Elevation	GeoTIFF, DTED, Geopackage (future)	
Imagery	GeoTIFF, Geopackage (future)	
Vector Features	Shapefile, ESRI File Geodatabase, Geopackage (future)	EDCS FACC
Visual Models	FilmBox (.fbx) COLLADA OpenFlight	SE Core Static Model Spec SE Core Gen 1 CM2 Spec SE Core Gen 2 CM2 Spec
Common Runtime	OpenFlight	Unique usage of comment
Maps	CADRG, GeoPDF	MIL-STD-2402 (Symbology)

Feature data models, dictionaries, and usage conventions can vary widely across publicly available source, making it difficult to reuse without significant special handling. Even then, necessary guessing at the usage and enumeration intent causes inconsistencies. SE Core has standardized on an internal feature data model based on the SEDRIS Environmental Data Coding Specification (EDCS) and closely aligned in structure to the OneSAF EDM and uses the EDCS dictionary. When sharing feature source data, we translate to the Feature and Attribute Coding Catalogue (FACC). The EDCS dictionary had been tailored to well suit the needs of the simulation community; however, FACC is more widely supported by COTS toolsets. As the National System for Geospatial Intelligence (NSG) Feature Data Dictionary (NFDD) becomes more common for COTS toolsets, migration to NFDD is likely.

Shared datasets may be grouped to facilitate ease of acquisition, based on the anticipated needs of the consumers. This can range from individual features to all data for an entire region. Currently, SE Core consumers typically look for a specific region based around a well-known training area or geo-political area, such as Ft. Bragg or Afghanistan. Combining all data for such regions results in an extremely large dataset for retrieval. Serving individual features or arbitrary areas requires automated mechanisms for filtering and metadata generation to be cost effective. Figure 7 illustrates the sharing folder structure recommended by SE Core (Garcia 2014). Each folder is a separately retrievable shared dataset. Metadata is generated at each folder level to facilitate consumer selection.

Additionally, static and moving visual models and textures are posted in this files and folders structure for reuse independent of a specific source or runtime dataset.

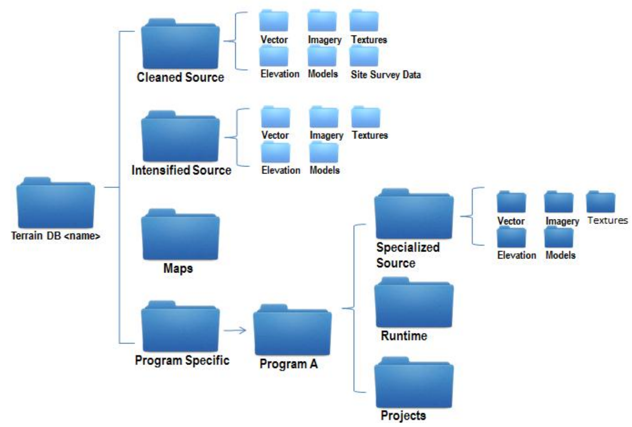

Figure 7. Shared Dataset Folder Organization

Exchange Mechanisms

Multiple exchange mechanisms are recommended due to differing consumer capabilities and dataset size. Historically, most SE Core data requests have been made through requests processed by hand by the program team. To be efficient with the high growth of requests, the data must be available on-line, discoverable, and downloadable. Electronic file download provides simple access when network bandwidth is sufficient. Organizing as bundled (Zipped) files at each of the lowest folder levels shown in Figure 7 is recommended with a mechanism to download an entire branch of the tree for user convenience. Support for online request of physical media is necessary to support large datasets and consumers with insufficient network capabilities. There is also an increasing consumer expectation for world-wide coverage catalogue and web streaming or download access for arbitrary geographic areas. Commercial and open source server solutions exist and increasingly support standards, such as the Open Geospatial Consortium (OSG) web standards for data exchange. At a minimum, providing metadata to external data brokers can provide consumers effective search and filtering capabilities.
Information Assurance regulations present significant challenges to setting up an online presence for data discovery and retrieval. Differing interpretations of the Information Assurance (IA) regulations hinders reuse of working environments. The Rapid Data Generation (RDG) program effort to provide a common data discovery portal, discussed later in this paper, will provide some relief, at least from the consumer side.

Geospatial metadata definitions are provided by multiple organizations. This paper does not describe the details of the metadata but rather the necessity and effect on sharing. SE Core is actively working with the RDG program to define a minimum acceptable metadata set by use case analysis and coordination across organizations, as shown in Table 5 to identify a balance of metadata production effort with discovery capability. RDG is further investigating automated metadata generation from source to reduce generation effort.

Table 5. RDG Metadata Analysis Data Sources

Geospatial Stakeholder	Discovery Metadata Document
MSC DMS (M&S COI Discovery Metadata Specification)	MSC-DMS-v1_4.pdf
SOFPREP (COCOM Operation Force Planning, Reh., Env. Prep)	SOFPREP_MIS_Metadata_SOF.xls
GWG (Geospatial Intelligence Working Group)	NFM_v1_final.doc
SE Core (Army Synthetic Environment Core)	2011_10_07_SE Core Metadata Analysis-0.91 Report.doc
NPSI (NAVAIR Portable Source Initiative)	NPSI_RSDM_Schema_V2.4.XX.2.docs
EDCSS (Environmental Data Cube Support)	coreCatalog.xsd
AFCD (Air Force Simulation DB Facility Common Data Set)	NPSI_RSDM_Schema_V2.4.XX.2.docs

Producer Certifications

To share effectively, the consumer must have confidence in the quality and consistency of the shared data. Third party certifications of process and product quality provide additional assurance. The Army Geospatial Center (AGC) Co-Producer Certification Program (Collins, Huiett 2014) is defining a program to assess select U.S. Army geographic data producing organizations to become certified data producers of Geospatial Foundation data.

SE Core, is participating in this pilot program which will evaluate both the products and the processes. This provides assurance to potential consumers and is expected to increase sharing across Army programs.

Cross-Organizational Sharing

Cross-organizational sharing of this data is absolutely important in these times of restricted resources. Currently, the DoD has no method to discover and reuse high-value, labor-intensive, and time-consuming data produced for Modeling and Simulation (M&S) applications, even though the majority of this data is government owned and is authorized for DoD-wide use. Consequently, each of the DoD Services has established its own data production or repository capabilities. The Service's data production centers are eager and willing to share their data holdings, but lack the technology and infrastructure to do so. Therefore, the Modeling and Simulation Coordination Office (MSCO) chartered the RDG program to enable M&S users in their sharing of their data.

The goal of the RDG project is to minimize the time and cost required to produce (clean, refine, thicken, integrate, and format) data for M&S applications. The approach is to enable the discovery and reuse of data that has already been produced for M&S applications and to facilitate the integration of source data through web services. The objective of RDG is to incrementally build a service-oriented DoD enterprise Common Data Production Environment (CDPE) that will enable the rapid discovery, retrieval, and reuse of data and services across the spectrum of Communities enabled by M&S (Acquisition, Analysis, Experimentation, Intelligence, Planning, Test and Evaluation, and Training). Details of the cross-organization participants are provided in the I/ITSEC hosted Geospatial Environment Database Standards special event.

The RDG CDPE will provide five major capabilities (Scrudder et. al. 2013):

1. Data Access Management – Provide controls to ensure authorized access to and retrieval of the data provider's metadata and datasets.
2. Data Discovery – Provide capabilities to search for and discover data products across multiple data repositories.
3. Data Retrieval – Provide access to download datasets or data products from the data provider repository or, delivered them on physical media.
4. Data Enhancement Services – Provide access to services and tools that enhance or add value to a dataset.
5. User Collaboration – Allow data providers and data consumers to interact, communicate, and collaborate in user-defined, domain-specific, on-line forums.

Details of the cross-organization participants, including the current and planned integration of data providers with the RDG over the next 12-18 months, are provided in the I/ITSEC hosted Geospatial Environment Database Standards special event.

Incorporating Externally Developed Data

The sharing described thus far has been outward facing. That is, a consumer comes to the data provider requesting data products. Additional savings and increased data quality can be achieved through external contributions of both new and value added datasets. External programs can contribute datasets to the data provider for sharing across the community. This is a limited, controlled version of crowd sourcing. Figure 8 shows an example of how value-added content can increase source data quality. The initial dataset was built by the data provider to the combined confederate requirements. The external developer added the new high fidelity areas within the overall geographic area. The contribution may be a new geographic region, new format, or a previously shared dataset with value added content or processing. Figure 9 describes how external contributions are accepted at the four main sharing points.

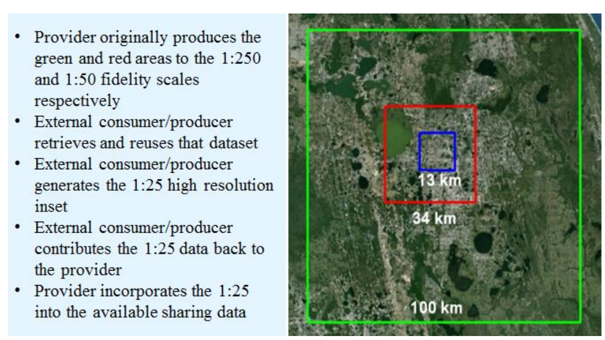

Figure 8. External Value-Added Fidelity

The level of reuse opportunity increases from runtime program specific runtime to cleaned source. The provider reuse policy must maximize this reuse within affordable bounds. Un-validated data cannot just be incorporated and mixed into provider datasets with specific, verified quality standards or the provider risks contaminating that repository. Full data validation is expensive so processing and validating contributed source donations with unknown standards may be more costly than using raw source data. At a minimum, the data can be reposed for potential reuse as-is, with metadata clearly defining the heritage. At best, data from a well-known external provider can be fully incorporated.

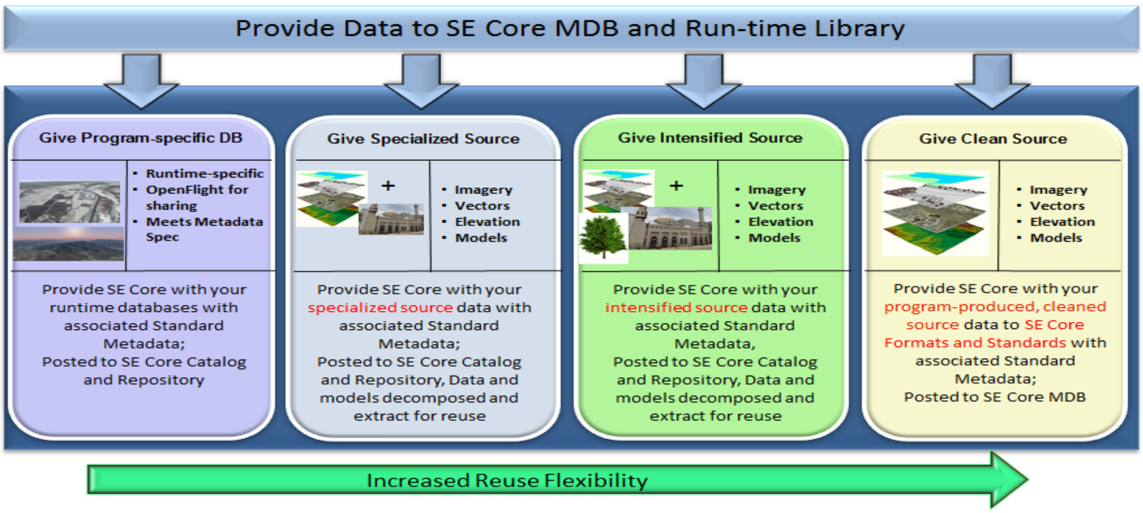

Figure 9. New and Value-added Contributions

Typically, runtime, confederate differentiated source, and intensified source with be reposed as-is with potentially some model extraction. Cleaned source will be reposed for reuse by external consumers and as another potential for the provider's future internal needs.

Other sharing opportunities

Beyond data, shareable items include production processes, detailed work instructions, standards (both identified industry and program-defined), tools and tool configurations, each with its own benefits. As example, SE Core has documented processes listed in Table 6. Sharing of these processes can save significant time and risk when standing up a terrain generation capability. The cost of tools is a major consideration when making this reuse choice. Major providers have strict quality requirements and can spread the tool cost across many customer products. A single system external program may not need or afford the higher end toolset or process standards.

Table 6. Sharable Processes

Processes	Work Instructions
Production Process (Overview)	Source data collection, standardization, and ATP
Product Requirements Analysis	Source cleaning processing and ATPs for elevation, imagery, and vector feature data
Product Integration and Test	Airfield creation and ATP
	Confederate runtime database
Specifications	Common sharing transmittal
GIS Source Data Environmental Data Model	Map development
Scale Content Specification	
Static and moving 3D Visual model specifications	

CONCLUSION

Growth of database size and complexity and an increase in the need to interoperate between training systems is continuing. Sharing has shown to provide cost and schedule savings, particularly when terrain development schedules are coordinated across programs. Maintaining an architecture emphasizing industry standards for data exchanges promotes production of truly shareable products. Sharing of mature process and tools can significantly reduce stand-up costs and maturation timeline for new producers and consumers needing to modify or enhance shared data.

Multiple data processing sharing points are required to satisfy differing correlation and training needs for interoperating systems. Managed correlation across the shared datasets maximizes the product fidelity and quality to each interoperating program while maintaining training essential correlation.

Data may be shared at differing levels of aggregation from individual features to training area datasets. Online discovery and retrieval is becoming the expected norm. There are technologies available and maturing to help but obstacles such as IA, sufficient, affordable, and standardized metadata, and support for compatible web protocols and services still present a challenge. Programs such as the RDG are working to address these challenges.

Cross-organization sharing utilizing commercial discovery and exchange standards is the next leap in sharing efficiency across the Army. Certifications such as the AGC co-producer program will help identify valid, reliable sharing sources.

REFERENCES

Collins, Timothy & Huiett, Chris (2014), Army Geospatial Center Data Co-Producer Certification Pilot Program.

Garcia, Juan (2014). SE Core AME Sharing Dataset Organization Briefing, Program Executive Office Simulation, Training and Instrumentation (PEO STRI) 12350 Research Parkway, Orlando, Florida 32826-3276.

Moore, Ronald, Nguyen, Phuong, & Bowman, Lewis (2014). SE Core Standard/Rapid Database Generation Capability (STDGC) Re-architecture. *Proceedings of the IMAGE Conference*, The IMAGE Society, pp. na.2014.

Scrudder, Roy, et. al. (2014). Rapid Data Generation (RDG) Common Data Production Environment (CDPE) System Requirements Specification (SyRS), Program Executive Office Simulation, Training and Instrumentation (PEO STRI) 12350 Research Parkway, Orlando, Florida 32826-3276.