Data Sharing: The Standard Specification is Just the Start

Robert F. Richbourg, George E. Lukes

Institute for Defense Analyses

Alexandria, Virginia 22311

rrichbou@ida.org, glukes@ida.org

ABSTRACT

Data sharing across multiple lines of effort is an often-cited component of reducing costs, improving efficiency, supporting interoperability, and providing other potential benefits. However, achieving a state where data can be readily shared is far from trivial and, as a first step, requires standards to be universally accepted among the data users. To achieve real success, many other steps must follow.

The Multinational Geospatial Co-production Program (MGCP) is an international cooperative effort where 32 nations together are coordinating the production and sharing of digital geospatial data that will eventually provide high-resolution vector data at a scale equivalent of 1:50,000 or 1:100,000 for much of the world's landmass. The MGCP is a successful data sharing program that continues to provide benefits for all member nations. As an example, much of the 1:50,000 data that was used in Afghanistan was produced by 7 different MGCP nations. All of the multi-purpose Atlas data the United States used to provide humanitarian relief in Haiti following the 2010 earthquake was produced from MGCP data. While there are many other success stories, the enduring value of the MGCP extends beyond its ability to provide timely, accurate geospatial data. The MGCP is a role model exemplifying the potential benefits of standards that are fully supported throughout the enterprise.

This paper describes key components of the MGCP effort, starting with the MGCP standard development processes and the importance of the supporting technologies that the MGCP has put in place to complete the standards. These include mechanisms for standards evolution, adjudication, compliance assessment, and enforcement. After developing these elements, the paper describes how they could be extended to provide similar benefits to other problem areas and thus form a domain-independent model for successful data sharing.

ABOUT THE AUTHOR

Robert F. Richbourg is a member of the Research Staff in the Joint Advanced Warfighting Division (JAWD) at the Institute for Defense Analyses. He is a retired Army officer who earned his Ph.D. in computer science in 1987. In his last active duty assignment, he was an Academy Professor and Director of the Artificial Intelligence Center at the United States Military Academy, West Point. He has been working in the area of simulation environments for the last twenty years under sponsorship of DARPA, DMSO, M&SCO, JFCOM, STRICOM, and the NGA.

George E. Lukes is a member of the Research Staff at the Institute for Defense Analyses. From 1994 to 2000, he served as a Program Manager at the Defense Advanced Research Projects Agency where his responsibilities included the Synthetic Environment Program for the Synthetic Theater of War. Previously, he led research and development efforts at the U.S. Army Topographic Engineering Center in advanced distributed simulation that included generation of terrain databases for SIMNET, ODIN, 73 Easting, Warbreaker, STOW-Europe and Bosnia.

Data Sharing: The Standard Specification is Just the Start

Robert F. Richbourg, George E. Lukes

Institute for Defense Analyses

Alexandria, Virginia 22311

rrichbou@ida.org, glukes@ida.org

INTRODUCTION AND BACKGROUND

"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay!" (Sir Arthur Conan Doyle, 1892.)

Despite its age, the wisdom of Sherlock Holmes still applies quite well today. Modern computer systems can be remarkable in their capabilities, but they must be fueled with high-quality data to reach their fullest potential. In a real sense, providing systems with suitable data represents a necessary evil that must be overcome before realizing any benefits from analysis (or other uses) of the data. However, overcoming this necessary evil by primary acquisition of appropriate data is usually expensive, regardless of the metric one might apply (production time, skilled personnel labor cost, direct cost, and so on). As a result, the motivation to share, reuse, and repurpose existing data has been great for an extended period and will likely remain so well into the future. A persistent barrier to effective data sharing is that the costs paid by both the data producer and consumer must be appropriate for the benefits each can accrue. Thus, achieving a state where data sharing is more routine than exceptional poses its own unique set of challenges and has proven to be an elusive goal.

There are multiple application domains where both the value and difficulties of sharing existing data are important issues. This paper uses the specific domain of sharing geospatial data to both frame the problems and point to potential solutions. The US Government has a long-standing interest in sharing geospatial data for modelling and simulation. There are several examples, including one that dates back to the middle 1980's. The US Air Force's "Project 2851" was an early effort, intended to permit sharing of visual databases among flight simulators. Other similar data-sharing efforts followed and built on the lessons from Project 2851. One of the most prominent was the Synthetic Environment Data Representation and Interchange Specification (SEDRIS) which began in 1994 (see http://www.sedris.org/ab_2trpl.htm) and embraced several objectives including "to provide a standard interchange mechanism to distribute environmental data and to promote database reuse among heterogeneous systems" (see http://www.sedris.org/ab_1trpl.htm). A basic premise of the SEDRIS effort was that sharing environmental data across simulation platforms would not only avoid the repeated government cost of building multiple databases of the same location, but was a prerequisite to achieving meaningful interoperability between different types of simulators. While several of the key SEDRIS technologies are still in meaningful use today, it would be difficult to claim that using the SEDRIS technologies to share simulation-ready environmental data is much more than an infrequently-occurring achievement. Sharing such data is rare, despite long-standing government desire, multiple well-funded programs, and the dedicated efforts of highly skilled and knowledgeable researchers. Clearly, this can be a difficult problem.

To some degree, the difficulties are magnified because of the level at which early data-sharing efforts had been focused. The environmental run-time data that is used in simulations results from a time-consuming, labor intensive, and complex chain of transformation. Generally, once collected, original source data, such as satellite imagery or digital elevation data, must be subjected to intensification, coordinate transformation, integration, correlation, and format translation. Each step is typically subjected to quality control evaluation and modification as well. As an example, Figure 1 provides an overview of the complex process used to create the visual form of the environmental data used in SOFPREP simulations (Miller and Black, 2013). Sharing data that has been subjected to this type of transformation process is desirable so that the entire cost of producing the data can be amortized across multiple uses of the data and to offer the best potential support for improved interoperability. However, each step in the process can also be directly and fundamentally influenced by optimizations appropriate for the intended use of the final run-time data. The type of simulation that will be supported and the end-use simulation system itself impose constraints throughout the process. As a simple example, an environmental database that supports air

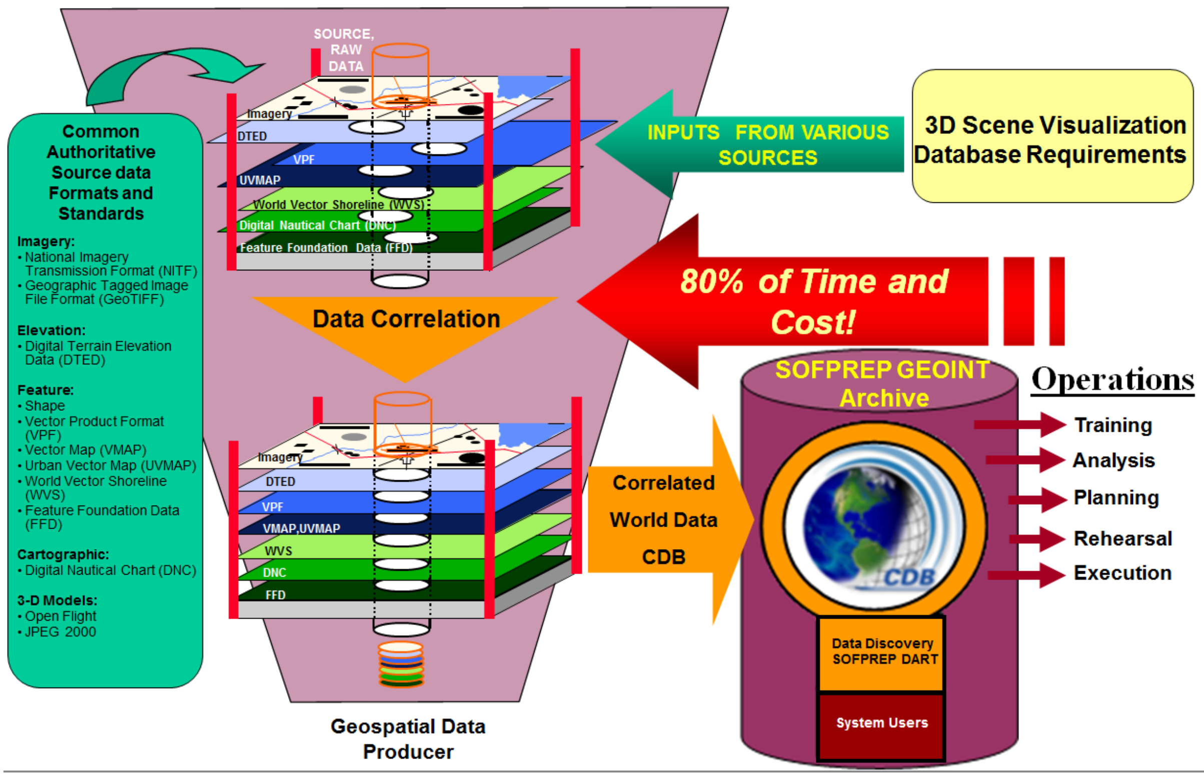

Figure 1. General Process for SOFPREP Visual Database Production (Miller and Black, 2013)

operations will typically cover a large physical extent and be produced at relatively low levels of detail appropriate for fast-moving aircraft operating at altitude. Conversely, an environmental database intended to support training or operations of dismounted infantry would typically cover a much smaller physical extent, but at a greatly increased level of detail. Clearly, directly sharing these two environmental run-time databases is very unlikely, even though they may have been created from similar transformation processes applied to equivalent source data. In either case, significant processing would likely be necessary to remove the influences and optimizations derived from the original intended use (e.g., large extent and low-resolution) and include those necessary to support the new context (e.g., small extent and high resolution). This somewhat extreme example is not intended to imply that no data sharing could be achieved in this case. Some data, perhaps the shoreline of a large body of water as an example, might be appropriately and directly used in either context. However, attempting to take an environmental database constructed for one context of use and use it directly in a new context adds even more difficulty to an already complex problem.

Targeting Data Sharing At A Different Level

As described above, several data sharing attempts that focused on the fully-transformed and improved run-time form of environmental data have met with limited success at best. However, sharing and reuse of environmental data still presents a lucrative goal. Follow-on efforts changed the level at which data was targeted for sharing, partly to overcome the problems that derived from the need to tightly incorporate context of use into the environmental database construction processes and partly to make greater use of existing capabilities. The U.S. Government-sponsored Portable Source Initiative (PSI) (Nichols, 2004) focused on sharing data much closer to the source level and on using de facto, industry-standard formats (such as *OpenFlight* – to obtain and understand the specification, see http://www.presagis.com/products_services/standards/openflight/more/openflight_specifications/) to serve as the interchange mechanism. Approaches depending on the existing capabilities of systems to ingest industry standard data require applying much of the data improvement and transformation to the original source data while deferring the production of the actual run-time data as late as possible in the data creation process. This change in strategy led to greater realization of the potential benefits. In fact, PSI proponents cite a specific reuse example where an original database, created at an expense of $700,000 in labor costs, was ported to a new system while incurring a labor cost of only $20,000 (Nichols, 2004, p. 3).

This approach has been sustained. The Portable Source Initiative is now known as the NAVAIR Portable Source Initiative (NPSI) and has been included in the U.S. Navy Aviation Simulation Master Plan (NASMP). The approach has also been further extended. Currently, the U.S. Department of Defense is sponsoring the Rapid Data Generation (RDG) project "to reduce the time required to discover, integrate, and correlate M&S data and promote data sharing through common metadata and services" (Scrudder, 2011). The entire RDG program is ambitious and intended to extend capabilities for sharing data across several functional domains. One of the targeted domains is geospatial data where the same basic approach used in the NPSI initiative has been adopted. That is, geospatial data that has been stored in one of several "de facto" industry standards will be targeted for discovery using the RDG services.

This is an important concept because the format for the data to be shared is well-defined and broadly accessible with commonly-used tools. Refer back to Figure 1 where many of these formats are listed in the inset box on the left side of the figure. Almost all geospatial data producing facilities rely on these standard formats to ingest the original data from which the run-time environmental data will be constructed. The formats that have been selected for the RDG environmental data are indicated in Table 1 below.

Table1. RDG Selected Data Interchange Formats

Type of Data	Format	Reference
Terrain surface elevation model ("raster" of point elevation values)	Digital Terrain Elevation Data (DTED)	MIL-PRF-89020B, Performance Specification: Digital Terrain Elevation Data (23 May 2000)
Terrain surface elevation model ("raster" of point elevation values)	GeoTIFF	http://trac.osgeo.org/geotiff/
Point, line, and area geometry (vector) feature data	ESRI Shapefile	http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
Imagery and texture data	GeoTIFF, RGB or JPEG 2000	As above, and http://www.jpeg.org/jpeg2000/
Static and dynamic models	Presagis OpenFlight	http://www.presagis.com/products_services/standards/openflight/more/openflight_specifications/

It is useful to note two characteristics about these interchange formats. First, both the NPSI and the RDG programs offer an existence proof that interchange of data much closer to the source format (and thus further from the run-time database) is a viable strategy. Second, not all of the interchange formats are fully-formalized standards having the stature of an ISO standard. Instead, some are "de facto" standards that are being widely used by many systems and organizations. The main point here is that having a large community of acceptance can be far more important than having an internationally vetted and agreed standard. Reliance on these de facto standards allows data-sharing to proceed without any changes to the normal work flow for either the data producer or the data consumer.

Benefits And Costs

As noted in the first paragraph of this paper, effective data-sharing implies that the costs paid by both the data producer and the data consumer must be appropriate for the benefits they can receive from participating in data-sharing efforts. Great strides towards achieving this state are made when data sharing occurs within the framework of an established work flow. However, other factors can influence the cost drivers as well. Perhaps chief among these is the suitability of available data from the data consumer point of view. That is, suppose that an environmental data producer participates in a data sharing program and absorbs the cost of making their data available to other programs. This producer must expect to benefit by acquiring suitable data from other data producers who have also decided to participate in the data-sharing effort. If, in the end, a program participant discovers an unfavorable quality mismatch (measured by geometric rigor, topology, content, and so on) between contributed data and available data, the cost – benefit ratio becomes less favorable.

The small number of participants in the RDG environmental data sharing effort know each other well. Each knows *a priori* what to expect about any data, given some knowledge of the data producer. Thus, the potential suitability of any shared data can be fairly accurately estimated prior to acquisition and the cost-benefit ratio is consistent. Data sharing programs that include larger numbers of participants may not share this same luxury.

Indications: What Can Be Learned From These Experiences?

There are several key observations that derive from recent experience in attempts to share environmental data within our simulation enterprise.

- Sharing environmental data that has been through the entire process of transformation for use in simulation systems is an attractive goal so that all of the cost of producing that data can be amortized across multiple uses
- Sharing fully-transformed data may imply the need to work with non-standard data formats and data that has been optimized for use in a specific context
- Sharing data from earlier in the transformation process can be useful and allow transfer of many of the improvements applied to create the correlated, simulation-ready data
- Using de facto standards is a viable alternative – neither program-specific data-sharing standards nor widely-vetted international standards are absolute requirements
- Using de facto standards for data transfer does not impose additional costs on the data producer to transform the data to a specific transfer format, improving their return on investment
- Using de facto standards allows data consumers to accept data without changes to the normal work flow, improving their return on investment
- Data quality and content can fundamentally influence return on investment and these data characteristics must be managed to assure a favorable return

A SUCCESSFUL INTERNATIONAL MODEL

The Multinational Geospatial Co-production Program (MGCP) is an international consortium of 32 nations who have jointly agreed to produce and share digital geospatial data. An individual MGCP dataset may originate from any one of a large number of different data producers; each of the 32 nations can employ an arbitrary number of commercial firms under nationally-governed data production contracts. Some nations also have dedicated, government-owned production facilities. Thus the MGCP is sufficiently large so that a potential participant cannot pre-determine a cost – benefit measure based solely on familiarity with a specific nation member. Accordingly, the MGCP has emplaced a tightly crafted set of standard specifications, processes, and capabilities to allow each participating nation to accurately judge their return on investment from participating in the program.

Figure 2. The MGCP

Overview of the MGCP Data Production Process

There is no required workflow that member nations must respect to produce acceptable MGCP data. Participating nations are free to use systems of their choice, as long as the end product conforms to program requirements and is able to pass through the established quality assurance and acceptance testing processes. These are fundamental to the program; the MGCP focuses on ensuring that only high-quality data is included in their inventory and this goal influences every step of the data's lifecycle.

The geographic coordinate system divides the globe using latitude meridians and longitude parallels, as exemplified in the MGCP logo in Figure 2. The MGCP relies on this division of the globe to designate spatial regions for data production. Member nations produce data in one degree cells (i.e., one degree of latitude by one degree of longitude). Once produced, each cell is "contributed" to a large store of one-degree data cells maintained in the MGCP's International Geospatial Warehouse (IGW). Producing nations are given credits for each cell they contribute to the IGW and may spend their credits to "withdraw" many more cells covering other regions and produced by other MGCP partners. Thus, each MGCP participant can benefit by acquiring comparatively large amounts of new data without incurring any new production cost (exclusive of nation-specific needs to intensify or otherwise improve the MGCP base data) and this provides the key motivation for member nations to produce and contribute new data to the IGW. It should also underscore the MGCP need to ensure the IGW data is uniformly high-quality data. If it could occur that a nation contributed high-quality data, but could only withdraw data of lower quality, the motivation for continuing participation in the program would likely be forfeit. To prevent this possibility, the MGCP relies on standards, assessment tools, and adjudication and enforcement processes that collectively assure each participating MGCP nation can have a favorable and pre-determined return on their investment.

THE MGCP STANDARDS: THE ENABLERS OF SUCCESS

The MGCP uses a variety of standards. In some cases, their standards have been developed by augmenting existing de facto industry or multi-national standards. In other cases they have fully developed original standards within the MGCP organization to meet their specific needs. Universal acceptance across the MGCP community is the key and common characteristic among these different standards. Each standard is also supported by an MGCP document, maintained by one or more participating nations (the "custodian" of the document) but available to each nation for both comment and use. Each standard is also a living document as the MGCP includes an active change management process enabled by semi-annual meetings of member-nation committees to vote on potential changes and revise the standards accordingly. The set of standards documents is collectively the MGCP Technical Reference Documentation (TRD).

The MGCP Feature and Attribute Catalogue

The Feature and Attribute Catalogue is a key component of the MGCP TRD. While it defines the features, attributes, and attribute values that can be found in MGCP data, it is also more than a data dictionary. The feature catalogue provides a comprehensive listing of the geometric form (e.g., point, line, or polygon) and set of attributes that are associated with each feature as well as any constraints or enumeration of values that can be assigned to the attributes given the specific feature and geometry. The catalogue is derived from the DGIWG Feature Data Dictionary (DFDD), a standard developed and maintained by the Digital Geographic Information Working Group (DGIWG). The DGIWG is a multi-national body that has long provided dictionaries of environmental concepts (such as the widely-used Feature and Attribute Coding Catalogue (FACC)). All participating MGCP nations use or plan to use at least some part of the DFDD within their existing work flow, so the MGCP decision to use a subset of the DFDD is a very natural choice. France is the custodian for the MGCP Feature and Attribute Catalogue.

The MGCP Extraction Guide

The MGCP participants all produce digital data by "extracting" the feature and attribute information from recent satellite imagery, printed maps, or other source materials. The purpose of the MGCP Extraction Guide is to provide a consistent approach to feature extraction. It includes guidance such as "all roads with length greater than 300 meters, those necessary to complete a road network, or isolated roads that serve as landmarks" must be extracted from the source material and included in the digital data. It also lists some logical, relationship-derived constraints for values assigned to attributes. As an example, it specifies that no road feature can have a value assigned to its width attribute that is greater than the value assigned to the width attribute of a bridge feature that carries the road. Since extraction guides like this one are used by all data producers, the constraints it imposes are very familiar. Program wide use of a single guide assures a commonality and consistency of the data that originates from diverse production facilities. The United Kingdom is the custodian for the MGCP Extraction Guide.

The MGCP Semantic Information Model

The purpose of the Semantic Information Model is to provide a conceptual view of essential relationships between the different feature types in the MGCP Feature and Attribute Catalogue. It is a somewhat unique document for the MGCP in that it attempts to describe in general terms the implicitly understood relationships between features that exist in the view of most data producers. As an example, the documents states "A Helipad may be within a Land Aerodrome or a Heliport, but it may also exist on its own." This document also explicitly describes some specific relationship requirements between features: "An Apron must always be joined to a Runway, Stopway or Taxiway." Both of these kinds of specifications are important to the quality of MGCP data. The former example identifies the expected-case relationships requirement for Helipad features. Exceptions to the expected case may exist and are allowed, but they typically require verification. The latter example provides an explicit rule that must always be followed in MGCP data. Sweden is the custodian for the MGCP Sematic Information Model document.

The MGCP Metadata Specification

The MGCP Metadata Specification defines the feature, cell and sub-region metadata required by the MGCP. It also defines the conformance rules to be satisfied by those metadata. The requirements are derived from those listed in the international standard, ISO 19115. The requirements expressed in this document are intended to allow discovery of data according to a uniform set of data characteristics such as dates for the source information, extents of the data coverage, the producer of the data, and so on. The United States is the custodian for the MGCP Metadata Specification.

Other Documents

The TRD also includes several other documents. These provide descriptions of standard MGCP processes, such as how and when data stored in the IGW must be updated, procedures for matching type and geometries of features and values of attributes at the boundaries of different cells, and the procedures for reviewing data before it can be accepted into the IGW. Other documents provide specific rules such as how data will be packaged when it is submitted to the IGW and the allowed form of data encoded in ESRI Shapefiles (the format of MGCP data files). These documents complement each other so that all of the program processes and requirements are documented. As a set, these standards specifications enable success for the MGCP in that they provide a mechanism to establish accurate expectations for the data that can be shared through participation in the program. However, these standards specifications do not stand alone in the MGCP.

THE MGCP ASSESSMENT TOOL: THE GUARDIAN OF SUCCESS

The MGCP TRD is a valuable resource, in and of itself. Similar comprehensive sets of specifications and descriptive documents for geospatial data are simply not available and the MGCP TRD serves as a one-of-a-kind model for other data-sharing programs to study. The TRD is useful in that it supports expectation management for MGCP participating nations. However, expectations are not always matched by achievements, sometimes by design and sometimes as an unintentional consequence. The MGCP includes a capability that can help ensure that expectations are met.

The United States contributes much to the MGCP program. It is not only a document custodian but has also assumed responsibility for operation of the IGW as well. Further, the United States provides the Geospatial Analysis Integrity Tool (GAIT), an analytical software utility that can consume MGCP data sets and produce reports (and other information) detailing the degree of the data's compliance with the expectations for that data as specified by the TRD.

The National Geospatial-Intelligence Agency (NGA) represents the United States in the MGCP organization. The NGA also sponsors continued development of the GAIT utility so that it remains in step with revisions to the TRD specification. GAIT now implements the current and several previous versions of the TRD feature catalogue and the inter- and intra-feature requirements that have been extracted from TRD documents. The common practice within the MGCP is that explicit requirements expressed in the TRD, such as the example above describing the required relationship between Aerodrome and Runway features, will always be implemented in GAIT so that the software will flag violations of those requirements. As exemplified above, the TRD also includes description of expected-case constructions that are not hard and fast requirements. GAIT also implements tests to identify any exceptions to the expected-case descriptions. This provides an attention-focusing mechanism for data reviewers who can verify that the data is an accurate reflection of reality, despite the conflicts with expected norms.

Just as the MGCP includes a nation designated as the custodian for each of its standards documents, the NGA is responsible for managing modifications to the GAIT software. Just as each member nation can suggest changes to the standards documents using the change management processes, any member nation can suggest changes to GAIT using a GAIT change management process. Just as any nation can be represented on the MGCP committees and groups that actively recommend actions for each of the standards documents, any member nation can participate in the MGCP GAIT Working Group to discuss potential changes and vote to make recommendations concerning potential modifications stemming from the GAIT change management process. This group also has the key responsibility of overseeing or conducting assessments and experiments that can result in software realizations of ill-specified requirements. As an example, the MGCP TRD prohibits geometric constructions referred as "kinks" in line features. The GAIT Working Group assesses parametric descriptions of these geometric constructions (i.e., a kink occurs when the angle formed by consecutive line segments is less than 10 degrees) and recommends representation in GAIT. In a real sense, GAIT is managed as though it were simply one more of the MGCP standards specifications rather than a software embodiment and approximation of those specifications.

The NGA sponsors continued development of GAIT and has unlimited distribution rights for the software; they can distribute it freely as deemed appropriate. NGA installs the software on their corporate workstations, provides it to NGA contractors who produce their data, furnishes it to MGCP member nations, and allows MGCP member nations to further distribute the software as desired. This management policy of free and wide distribution results in the creation of a common measure and perception for data that meets and does not meet the desired quality threshold. It establishes an "open-book test" where surprises for data producers during the acceptance process are minimized since all parties have access to the common metric and measure of success. Rather than being reserved as a private capability to be used only during data acceptance, GAIT is distributed as widely as possible to help ensure data is accepted the first time presented. This approach lowers cost, increases quality, and improves return on investment for all parties.

THE MGCP DATA ACCEPTANCE PROCESS: THE GUARANTORS OF SUCCESS

As described above, the MGCP has defined clear specifications of standards that enable the production of high-quality data. The program also includes an automated assessment capability (GAIT) that can characterize the degree to which data complies with those standards. A third component of the program is a clearly-defined process for moving the data from the workbench into the warehouse of accepted data approved for sharing.

Every time that a member nation completes a cell of data and uploads that cell to the IGW, the cell will be characterized as "in review" until the cell has been reviewed and approved by representatives of a second MGCP member nation. The MGCP considers *quality control* of each cell to be the responsibility of the producing nation, and *quality assurance* to be the responsibility of the reviewing nation. It is incumbent upon the producing nation to comply with the standards and local findings during GAIT assessment to the greatest extent possible. It is the duty of the reviewing nation to attest that the producing nation has achieved an acceptable level of compliance. The procedures that support these requirements define three levels of quality assurance testing for newly produced data.

When a producing nation is submitting their first cell of MGCP data, the reviewing nation is required to perform a "Quality Assurance Benchmark" evaluation. This is a very detailed evaluation of the data that includes manual comparison of the digital product to ortho-rectified imagery as well as review of the GAIT tests and other manual evaluations of the features and attributes in the data. The metadata describing the cell is manually reviewed as well. The extensive test of the first cell produced by any member nation is intended to provide assistance to that nation during their initial involvement with the MGCP. There are many MGCP standards, processes, and procedures that require compliance and the benchmark process is designed to help familiarize new MGCP participants.

After a producing nation has successfully submitted the first cell of data to the IGW, most subsequent submissions will be evaluated using the procedures defined as the "Quality Assurance Standard" test. These tests are primarily automated although there is a manual review and comparison of reports generated by the automated systems. Every fifth cell of data submitted by a producing nation is required to pass the "Quality Assurance Extended" set of tests. These include all the automated tests and reviews included in a standard test as well as further manual evaluation of the data using a Geographic Information System (GIS). Evaluations using the GIS are performed to complete a checklist of required inspections. Both the standard and extended tests include examination of the cell metadata as well.

The data acceptance process also provides for adjudication of characteristics that remain in the data even after they have been identified as potential problems by either the automated assessment or by the reviewing nation. Such characteristics may well remain in the data when the producer feels the data is ready for submission to the IGW. Recall that GAIT has been designed to focus reviewer attention on atypical constructions within a dataset. Atypical does not mean impossible, so a producer is able to enter comments on the GAIT-generated reports to explain any entries that remain in the report after completion of all quality control actions. These "exceptions as noted" must be validated by the reviewing nation. Typically, these exceptions to the expected case are resolved by discussions between the producing and reviewing nations. In the rare circumstances when the two national representatives cannot reach agreement, other nations are asked to intervene and reach a group consensus to resolve the issue. The possible outcomes of the adjudication process are: 1) the producing nation must repair the fault; 2) the reviewing nation must accept the data as a reasonable characterization of the physical environment; or 3) the data will simply not be accepted into the IGW.

THE MGCP: A MODEL TO BE EMULATED

The MGCP is a successful data sharing program that continues to provide benefits for all member nations. As an example, much of the 1:50,000 scale feature data that was used in Afghanistan was produced by 7 different MGCP nations. All of the multi-purpose Atlas data the United States used to provide humanitarian relief in Haiti following the 2010 earthquake was produced from MGCP data (Dellagnello, 2012). The success of the MGCP is largely due to the participating nations' knowledge that they will achieve a high return on their investments in the program; contributing one cell of data provides access to many more. This potential return is founded on and validated by the existence and availability of uniformly high-quality data in the IGW. The high-quality data is assured by standards, assessment capabilities, and procedures that enforce the assessment of standards. This three-axis approach to ensuring a successful data-sharing effort is a model that could be extended to other programs.

The historical examples of attempts to share geospatial data within the simulation community have relied on standards, but not on a rigorous, three-axis model of standard specification, compliance assessment, and standards enforcement that govern operations throughout the program. Also, the standards that have been used have primarily been designed to ensure that data could be transferred between programs. This is a very different use of standards than exemplified by the MGCP where the goal is to ensure availability and transfer of ***high-quality*** data by employing standards reinforced with assessment and enforcement capabilities. It also seems fair to note that while the historical examples have attained various levels of success, none have come close to the levels of wide-spread success established by the MGCP. Is use of the MGCP approach a viable alternative within the simulation community?

The RDG program is the best current effort underway to share geospatial data within the defense simulation community. Today, RDG is chartered to provide a discovery and acquisition capability for several types of data, including independently-produced geospatial data. As noted previously, participants in the RDG geospatial data-sharing effort rely on personal knowledge or collegial relationships to understand the value they can expect when acquiring data originally produced by another program participant. The geospatial-data producers have agreed to make their data visible to other program participants and the data can be shared, as-is. Given the current scope of the data-sharing effort, it could be that nothing else is required to achieve success. However, it is difficult to believe that familiarity and altruism would suffice as enablers of success for larger-scale, department-wide data-sharing programs. Table 2 (below) compares several characteristics of the RDG and MGCP programs for geospatial data sharing.

As described in Table 2, the geospatial data providers in the RDG program and the MGCP have some commonality in that both are sharing vector feature data. In fact, much of the vector feature data used by the RGD geospatial data producers could have originally been produced by a MGCP participant. An MGCP-like program to jointly produce and share raster elevation data is currently emerging in the international community (the TanDEM-X High Resolution Elevation Data Exchange Program) and already includes many of the nations represented in the MGCP. (The MGCP is unlikely to incorporate either imagery or visual databases.) These two efforts also use the same format for vector feature data transfer. Beyond these two characteristics, the programs have few exact matches.

Table 2. RDG and MGCP Key Characteristics for Geospatial Data Sharing

Characteristic	RDG	MGCP
Data to be shared	Vector feature, raster elevation, imagery, visual databases	Vector feature data
Transfer format	Shapefiles, DTED, OpenFlight, GeoTiff	Shapefiles
Schema	Specific schema for each producer	Program-wide: all producers use the MGCP TRD specification
Quality Requirements	Specific to each producer	Program-wide: all use the MGCP TRD specifications and the GAIT assessment tool
Motivation	Voluntary participation in the program, data posting, and unmanaged data acquisition	Voluntary participation in the program, credit – debit system for data posting and withdrawal
Meetings	Periodic, informal	Regularly scheduled formal meetings
Review Process	Specific to each producer	All follow the same process which may require producer rework for acceptance

The two programs do have near matches. The differences described in Table 2 are primarily focused on level of applicability, whether a characteristic is shared by all program participants or the characteristic is non-uniform so that it exists, but could be described differently for each data producer. The MGCP sets program-wide specifications that all participating nations must follow. The RDG geospatial data producers have specifications for schema, quality requirements, and so on, but these are local to each producer. However, a global specification for an RDG geospatial feature "schema" could readily be constructed by forming the intersection of those used locally by each current producer. The same holds for a statement of quality requirement; each RDG geospatial data producer has quality control requirements and these could be analyzed to form a common set of requirements. Also, some of the RDG data producers already routinely use GAIT. In general, one could form the quality-related specifications by crafting a statement of the common requirements from each program. This starting point could then be improved by blending in the RDG-focused specializations of the existing feature data specifications used by the MGCP. While this approach would initially be limited to feature data, it would serve to start the process. In fact, it seems that it should be easier to enact the same quality-control mechanisms across the RDG geospatial data producers than it was to start the MGCP process. After all, the RDG geospatial data producers are small in number and are all within the U.S. Department of Defense. Compare this to the much larger number of participants in the MGCP, consisting of sovereign nations with independent financing.

CONCLUSION

Expanding the geospatial data sharing program to encompass larger numbers of participating agencies within the Department of Defense simulation community would offer several benefits. Cost avoidance could be achieved by reusing improved and correlated source data rather than completely rebuilding databases of the same location when required for use in new contexts of application. And, as noted originally by the SEDRIS program, the potential for valid system interoperability could be improved. Such expansion will be hard to achieve without introducing a more formalized program than currently exists with the RDG effort (which is, after all, chartered to focus on discovery and acquisition, not production of a particular quality of sharable data). The MGCP offers an existence proof that using standards specifications buttressed by compliance-assessment tools and enforcement policies can lead to highly-successful data sharing where each program participant enjoys a large return on their investment.

It is also worthwhile to note that, to date, the NGA has been very inclusive regarding the quality standards maturation and GAIT development processes. The NGA is intent on providing high-quality data in response to customer needs. As such, interested Government representatives are typically welcome to participate in workshops and meetings where future developments for standards and GAIT are discussed and vetted. In fact, some of the RDG geospatial data producers have participated in these deliberations. A larger presence implies a larger voice representing the simulation community in influencing the data that typically serves as the source data from which simulation environments are constructed.

The basic ingredients for a more pervasive geospatial data sharing program are already available for adoption and modification as necessary. The missing piece of the puzzle is the identification or formation of a controlling body that could utilize and adapt the existing quality standards, provide an assessment capability, and enact an enforcement capability. Data sharing programs, mandated or not, are most likely to succeed when all parties can expect a favorable return on their investments. The MGCP program offers an existence proof that successful data sharing can become a reality when investments are small compared to the returns.

REFERENCES

Dellagnello, Marzio. (2012). *Multinational Geospatial Co-Production Program (MGCP).* Presentation at the Joint Agency Commercial Imagery Evaluation (JACIE) conference, St Louis, MO, April 17 – 19, 2012.

Doyle, Sir Arthur Conan. (1982). “The Adventure of the Copper Beaches” *in The Adventures of Sherlock Holmes.* London, England: George Newnes Ltd.

Miller, E. and Blank, H. (2013). *Creating Synthetic Environmental Databases for DoD Mission Driven Simulations*. Tutorial 1316, Proceedings of the Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL, December, 2013.

Nichols, K. W. (2004). *The Portable Source Initiative*. Paper 1569, Proceedings of the Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL, December 2004.

Scrudder, Roy, et al. (2011). *Rapid Data Generation for Modeling and Simulation*. Paper 11163, Proceedings of the Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL, December 2011.