Open Geospatial Consortium

Submission Date: 2018-11-11

Approval Date: 2018-09-13

Publication Date: 2019-01-20

External identifier of this OGC® document: http://www.opengis.net/ doc/DP/CDB-leveraging-Geopackage

Internal reference number of this OGC® document: 18-077r2

Category: OGC® Discussion Paper

Editors: Jay Freeman, Kevin Bentley, Ronald Moore, Samuel Chambers, Glen Quesenberry

OGC CDB, Leveraging GeoPackage Discussion Paper

Warning

This document is not an OGC Standard. This document is an OGC Discussion Paper and is therefore not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, an OGC Discussion Paper should not be referenced as required or mandatory technology in procurements.

.

Document type:	OGC® Discussion Paper
Document stage:	Approved
Document language:	English

Contents

i. Abstract

Creating modern geospatial terrain standards has proven to be a challenge for the modeling and simulation (M&S) industry. Multiple standards have promised to be a panacea for all; but few, if any, have delivered on the promises. Challenges facing new geospatial terrain standards have included proprietary formats, rarely used data containers, obsolescence, disconnected data silos, and heavy reliance of Government funding and management. While much effort has gone into developing new formats, the most common used formats are based on antiquated concepts with proprietary limitations. These ageing formats are a hindrance to terrain data reuse, runtime terrain database correlation, innovation, and system interoperability.

In the commercial industry, geospatial data content and use is exploding at a rate that is outpacing the innovation and utilization of the traditional M&S industry. The M&S and geospatial-intelligence (GEOINT) industries are on a path of convergence. The Open Geospatial Consortium (OGC) is a key forum influencing and facilitating this unification process. Within the OGC, there are two geospatial standards that best enable the unification of the M&S and GEOINT industries: OGC CDB and GeoPackage. OGC CDB and GeoPackage are both standards increasingly used in M&S and GEOINT industry, but they both contain weaknesses and strengths when it comes to the combined needs of both industries.

This paper offers the results of research, design, and prototype efforts to present the OGC standards working group an approach to creating “GeoCDB”—a technology mashing of GeoPackage and OGC CDB—as a deterministic repository of easily read data geospatial datasets suitable for storage, runtime access, and dissemination for live, virtual, constructive, gaming, and mission command (MC) systems.

ii. Keywords

The following are keywords to be used by search engines and document catalogues:
ogcdoc, OGC document, OGC CDB, OGC GeoPackage, modeling and simulation

iii. Preface

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

iv. Submitting Organizations

The following organizations submitted this document to the Open Geospatial Consortium (OGC): OGC CDB Standards Working Groups (SWG), supported by Army Geospatial Center (AGC), CAE USA, Leidos Inc, Cognitics, and Joint Staff J7.

v. Submitters

All questions regarding this submission should be directed to the editor or the submitters:

Name	Affiliation
Glen Quesenberry	AGC
Jay Freeman	CAE USA
Ronald Moore	Leidos, Inc
Kevin Bentley	Cognitics
Samuel Chambers	JS J7 EAD

1. INTRODUCTION

Creating modern geospatial terrain standards has proven to be a challenge for the modeling and simulation (M&S) industry. Multiple standards have promised to be a panacea for all; but few, if any, have delivered on the promises. Challenges facing new geospatial terrain standards have included proprietary formats, rarely used data containers, obsolescence, disconnected data silos, and heavy reliance of Government funding and management. While much effort has gone into developing new formats, the most common used formats are based on antiquated concepts with proprietary limitations. These ageing formats are a hindrance to terrain data reuse, runtime terrain database correlation, innovation, and system interoperability.

The Joint Staff J7 and Army Geospatial Center (AGC) share a common objective to adopt a standardized terrain format. In the past, Joint and Service simulation training capabilities adopted or developed differing standards, often in isolation, tailored to their specific training needs (Chambers and Callaway, 2017). This has resulted in costly integration efforts to link simulation-training capabilities together in federations.

An advantage to the Open Geospatial Consortium is that standards working groups actively manage and update those standards in a cooperative fashion, leveraging the synergy of the key Department of Defense (DoD) and intelligence community members, along with industry and academia.

Freeman (2017) described alternative deployment techniques in CDB that include leveraging GeoPackage.

The Joint Training Synthetic Environment (JTSE) stood up a Terrain Data Standards for Joint Training working group to address the need for a "cooperative establishment of a joint training standard or specification for the encoding, storage, access, and modification of a representation of the natural and man-made terrain for virtual and constructive simulation applications" (Chambers and Callaway, 2017). The mid-term recommendations included a common library framework and standardized attributes, while the long-term recommendations included defining a common storage format for terrain data by 2021. The long-term recommendations included the ability to modify the common standard to meet unfulfilled requirements. The working group paper also compared four existing terrain database standards, CDB, the Synthetic Environment Core (SE Core) Master Database (MDB), the Air Force Common Dataset (AFCD), and the NAVAIR Portable Source Initiative (NPSI). Although the standards have much in common, the paper identified CDB as a potential candidate due to a variety of reasons, but primarily because it was 1) an open OGC international standard and 2) it supports storage and runtime applications.

Although the Service, Agency, and Combatant Command representatives at the Joint Training Synchronization Conference approved the JTSE white paper, when the OGC CDB v1.0 was nominated for entry into the Defense Information Technology Standards Registry (DISR) as an emerging standard, the U.S. Army Modeling & Simulation Enterprise raised some concerns with the CDB standard. The key technical concern of the standard was that it incorporated two formats, Esri Shapefiles and Presagis OpenFlight, which although they are open specifications, were initially developed as proprietary formats. In addition, Shapefiles were introduced in the early 1990s, and leverage the dBase IV format to store feature attribution. The use of dBase IV limits the ability of OGC CDB v1.0 to adopt the National System for Geospatial-Intelligence (NSG) Application Schema (NAS) version 1.8 profile. The key limitation is that only 10 characters can be stored as a field name in dBase IV while many attribute names in the aforementioned profile are longer.

This paper explores potential designs to integrate OGC GeoPackage with OGC CDB to address the limitations of Esri Shapefiles within CDB.

1.1 Joint Staff J7

The mission of the Joint Staff J7 is to support the Chairman of the Joint Chiefs of Staff and the Joint Warfighter though joint force development in order to advance the operational effectiveness of the current and future joint force. The Joint Training Division has as its mission statement:

> "Develops and delivers a continuum of individual, staff, and collective joint training utilizing subject matter experts; adaptive processes; and integrated information technologies, in order to enhance the operational effectiveness of the current and future joint force."

Joint training requires terrain data modifications to meet training objectives. An onsite Terrain Services team enables trained, ready, and adaptable Joint Forces by providing realistic terrain to support M&S for training events; however, the current Joint Live Virtual Constructive (JLVC) and Service simulation requires the knowledge, skills, and processes to support around 15 proprietary formats and growing. Labor is limited, workload is extremely high, and demand from Service customers exceeds resources.

In 2012, the JS J7 began to create a roadmap for the future joint training capability. The mandates for change included DoD polices and directives that included the Joint Integrated Environment (JIE) and cloud-based computing, as well as the migration to a persistent cybersecurity scheme under the risk management framework (RMF). Since development for many of the simulations in the JLVC area began in the 80s and 90s, they do not lend themselves to updating to the new environment.

Other influencers on change include diminishing resources to send teams of system administrators and simulation instructor controllers overseas to conduct major joint training events, and fewer resources to create custom proprietary runtime databases for each scenario and simulation involved in these events. The J7 goals are increased efficiency, ability to adjust fidelity as needed, sustainability, service-oriented architecture, discovery/accessibility/usability, and the ability to run in a cloud-based architecture.

The outcome of this roadmap is the Joint Training Tools (JTT), a set of web-based tools that will enable the planning and execution of an exercise in a single user interface, with all digital data stored and available for modification and reuse in the future. A key component of JTT is the Terrain Generation Service (TGS) to drive to a centralized location for terrain and geospatial data to support the JTT. The requirements for TGS included the following:

1. Global terrain – not limited to predefined play boxes;

2. Ability to easily update and manage the global terrain and quickly add content from many sources and users (crowd sourcing);
3. Based on open standards and specifications, leveraging the billions of dollars already invested into the GEOINT and commercial geospatial industry;
4. Ability to integrate quickly with other online geospatial data services and support discovery, sharing, and reuse of data;
5. Continued support to existing legacy simulations in the JLVC using output compliers;
6. Ability to function as a dynamic runtime terrain for the future M&S as a Service (MSaaS) capability of JTT; and
7. Support for planning and command and control applications that are involved in the Joint Exercise Life Cycle (JELC) process by global terrain in TGS.

TGS leverages CDB as its terrain format. JS J7 is committed to a collaborative effort to leverage geospatial standards to improve interoperability and integration of modeling and simulation, thus saving time and money for the DoD.

1.2 Army Geospatial Center

The AGC mission is to provide timely, accurate, and relevant geospatial information, capabilities and domain expertise for Army Geospatial Enterprise implementation in support of unified land operations.

Within the AGC, the Systems Acquisition Support Directorate (SASD), Enterprise Services Branch (ESB) focuses on major Army programs and their interoperability as a holistic system. The AGC staff in this branch manages the geospatial cross-cutting capability (CCC) within the U.S. Army Common Operating Environment (COE). This effort requires the understanding and mapping of various geospatial technologies and standards. ESB staff engages industry, standards bodies, and Army programs to identify technical gaps and works to address these issues to create a more efficient Army fighting force.

AGC identified an enterprise-wide gap for disconnected data storage and dissemination in a lightweight geospatial standard and focused on the creation and specifics of OGC GeoPackage to address this gap. This technology was designed to ensure mobile / handheld devices could interoperate with systems within the larger Command Post and U.S. Army Enterprise.

After its inception, the GeoPackage specification has undergone many adaptations (both within and without OGC) to address a similar need. Current GeoPackage adaptations can natively store 3D information, routing information, various types of data ‘tiles,’ and other geographically related datasets.

As AGC and the Army assessed and studied the OGC CDB standard, AGC was wary that Shapefiles, the native vector storage format listed in OGC CDB (as a best practice), was not in alignment with the OGC or the U.S. Army Geospatial Enterprise. Furthermore, as Army systems do not utilize CDB, it was reasonable to address some of the OGC CDB implementations to ensure interoperability with CDB and the AGC. As such, an implementation of OGC GeoPackage within CDB aligns CDB vector data storage with the U.S. Army architecture.

2. CONFORMANCE

This discussion paper promotes the adoption of GeoPackage as a replacement for Shapefiles within the OGC CDB standard. GeoPackage would be added as an alternative best practice of Volume 4: OGC CDB Best Practice use of Shapefiles for Vector Data Storage.

No conformance is required.

3. REFERENCES

The following documents are referenced in this discussion paper text. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the latest edition of the document referred to applies.

LTC Charles Robinson, J. Conover, D. Hastedt. (2002). Web-Enabled, Rapid, Distributed, Database Development. The Interservice/Industry Training, Simulation & Education Conference. Orlando FL, November 2002.

Wiesner, B, Brockway, D. & Stanzione, T. (2011). Open Streaming Terrain for Modeling and Simulation. The

Interservice/Industry Training, Simulation & Education Conference. Orlando FL, November 2011.

Freeman, Jay (2017). Alternative Deployment Techniques for OGC CDB. IMAGE Conference, 2017.

Presagis (2012). Common Database Specification. Draft Version 3.2, December 2012.

Open Geospatial Consortium (2014). Retrieved April 2014, from http://www.opengeospatial.org/.

Chambers SD, Callaway, SE (2017). White Paper on Enterprise Terrain Data Standards for Joint Training. Joint Training Synthetic Environment (JTSE) Working Group (WG) of the Joint Training Synchronization Conference (JTSC). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/1040231.pdf

Geospatial Intelligence Standards Working Group. National System for Geospatial-Intelligence (NSG) Feature Data

Chambers S, Freeman J, Blount, E Ph.D. Cart Before the Horse: Leveraging Web-Based Global M&S Terrain Repositories for Scenario Development. MODSIM World 2017.

JFD Web Site – Retrieved on 11 June 2018 from http://www.jcs.mil/Directorates/J7-Joint-Force-Development/

OGC CDB Standard – Retrieved on 01 June 2018 from http://www.opengeospatial.org/standards/cdb.

OGC GeoPackage Standard – Retrieved on 01 June 2018 from http://www.opengeospatial.org/standards/geopackage.

Joint Staff, CJCSM 3500.03E Joint Training Manual For the Armed Forces of The United States. 20 Apr 2015.Dictionary (NFDD), Version 1.8, May 2007. Retrieved June 2018, from http://www.gwg.nga.mil/asfe_oversight.php.

4. LESSONS LEARNED FROM CREATING DOD STANDARDS

Multiple M&S terrain standards have promised to be a panacea for all; but few, if any, have delivered on the promises. Challenges facing new geospatial terrain standards have included proprietary formats, rarely used data containers, obsolescence, disconnected data silos, and heavy reliance of Government funding and management.

4.1 Synthetic Environment Data Representation and Interchange Specification (SEDRIS)

SEDRIS was initiated in 1994 as a co-sponsored activity by the program manager for Combined Arms Tactical Trainer (PM-CATT) at the Simulation Training and Instrumentation Command (STRICOM, now PEO STRI) and the Synthetic Theater of War (STOW) program at the Defense Advanced Research Projects Agency (DARPA). The DoD led the development of SEDRIS with representation from industry. The objective was to provide solutions to the complex problem of environmental data representation and interchange for networked heterogeneous applications. The SEDRIS architecture centered on the creation of DoD unique data representation models, environmental coding specifications, spatial reference models, file formats, and application programming interfaces (API). From the inception of SEDRIS to its final software release in July 2011, the overwhelming majority of funding for SEDRIS originated from the DoD. SEDRIS utilization for terrain distribution significantly decreased and ultimately ended when DoD resources diminished.

The SEDRIS initiative has both positive and negative lessons learned to help steer future terrain initiatives. A strength of the SEDRIS is well-developed documentation describing the various data models in use and a level of sufficiency in open source software to read and write the terrain. An area affecting the potential success with the SEDRIS initiative was its reliance on DoD unique formats with no commercial adoption.

4.2 Naval Air Systems Command (NAVAIR) Portable Source Initiative (NPSI)

NPSI is a simple concept with a simple goal: minimize the redundancy in database production across platforms without inhibiting innovation. The basic concept of NPSI is to capture value-added work performed on raw source data. This concept has resulted in significant cost savings and increased efficiency of database production to many DoD programs by minimizing the purchase and processing of new source data required for future developments. The NPSI archive stores refined source data in datasets and makes the datasets available for utilization by future programs. However, unlike the refined source data, runtime databases are not stored in the archive.

The NPSI initiative has both positive and negative lessons learned to help steer future terrain initiatives. A strength of the NPSI is the utilization of widely used commercial data formats to simplify the reading and distribution of NPSI data. An area for improvement with the NPSI initiative is its weak and unenforced conceptual data model that hinders and complicates data reuse.

4.3 Master Database (MDB)

The SE Core MDB is a folder- and file-based repository of geospatial data representing all of the geographical areas that the SE Core program has collected, prepared, and delivered to the U.S. Army. The SE Core MDB is defined for use within the SE Core production process and is designed to support the production of the runtime formats required for each U.S. Army Integrated Training Environment (ITE) systems.

The MDB elevation data is contained in GeoTIFF and exportable in DTED formats. The feature data is stored in Esri SDE™ and exportable to Esri SHAPE and File Geodatabase formats, and now OGC GeoPackage. The data model is unique to SE Core and uses the SEDRIS Environmental Data Coding Specification (EDCS) data dictionary. The MDB imagery is stored in GeoTIFF and exportable to TIFF and JPEG2000. The MDB 3D models are stored in both OpenFlight® (Presagis) and FilmBox™ (AutoDesk) for virtual and gaming, respectively. The 3D models the textures and material maps are stored in Photoshop Document (PSD) files and exported to Truevision Graphics Adapter (TGA) for processing to runtime formats. All metadata is stored in XML. The SE Core MDB contains only full government purpose rights data and is distribution-controlled by PEO-STRI.

The SE Core MDB is not a sharing standard. It is a content data store, defined in industry formats, used internally in the production of the U.S. Army ITE simulation system runtime formats. Long term, the goal of SE Core is to align the U.S. Army M&S Geospatial Data standards with Operational Mission Command (MC) Geospatial Data standards. SE Core is working with the AGC in the development of a GeoPackage extension to support M&S. SE Core is leveraging the Army's Ground-Warfighter Geospatial Data Model (GGDM) with the NSG Feature Data Dictionary (NFDD) feature codes and attribution, with the intent to replace the SE Core data model with the GGDM. Additionally, SE Core is migrating the internal formats to use OGC GeoPackage where practical.

The MDB initiative has both positive and negative lessons learned to help steer future terrain initiatives. A strength of the MDB initiative is the employment of commercial file standards as the means for terrain data storage and well-defined use case models. Improvements of the MDB initiative include using National Geospatial-Intelligence Agency (NGA) logical data model in lieu of M&S unique logical data models (e.g., EDCS) and employing a conceptual model that supports data sharing.

4.4 CDB

CDB is an open standard defining physical, logical, and conceptual models for a single, "versionable," virtual representation of the earth. CDB structured data stores provide for a geospatial content and model definition repository that is plug-and-play interoperable between database authoring workstations. Moreover, a CDB structured data store can be used as a common online (or runtime) repository from which various simulator client-devices can simultaneously retrieve and modify, in real-time, relevant information to perform their respective runtime simulation tasks (OGC CDB Standard, 2018).

CDB was developed by USSOCOM to meet the need for a rapid, large-scale production capability for worldwide simulation databases for the Regiment and other SOF simulations. The key design goals of CDB were to assure correlation between simulation subsystems by eliminating alternate storage formats of the same dirt, meet the ability to rapidly generate databases for mission rehearsal timelines, reduce the size of databases stored in a facility by eliminating the need of replication for each individual simulation subsystem, and simplify configuration by eliminating multiple representations of the same dirt (Freeman, 2017).

Beyond the key design goals of CDB, core ideas of CDB include: organizing Geographic Information Systems (GIS) data in a standardized and documented structured data repository to promote data reusability and interoperability amongst systems and vendors, prebuilt levels of detail (LOD) to enable the same correlated terrain data to be used by numerous simulation systems even though they may be operating at different altitudes and distances from the ground, and the ability to dynamically update the terrain during a simulation exercise and share the resultant terrain in near real time across all of the simulation clients participating within the exercise. Figure 1 outlines the CDB structure.

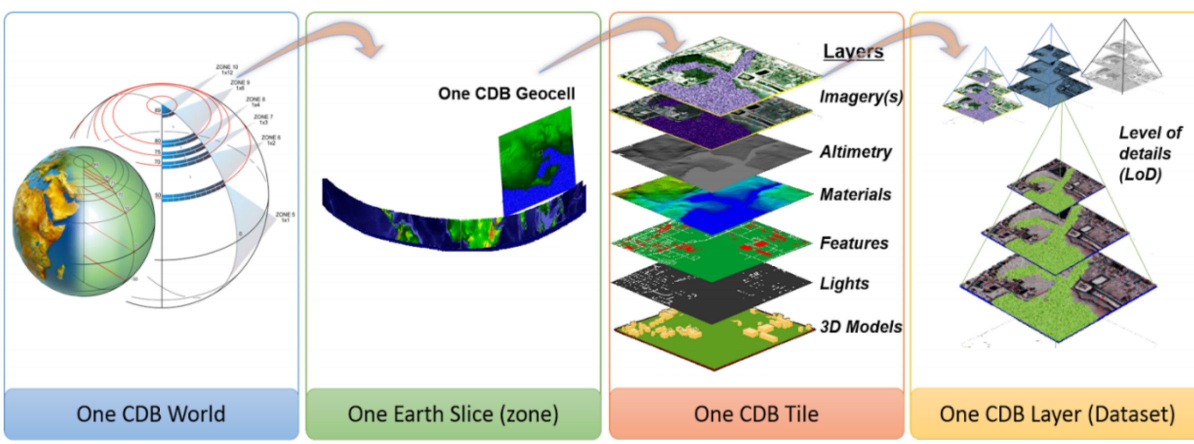

Figure 1: CDB Structure

The CDB initiative has both positive and negative lessons learned to help steer future terrain initiatives. A strength of the CDB initiative is the employment of commercial file standards as the means for terrain data storage, well-defined use conceptual models and an ability to serve as both a source repository and runtime format. Improvements to the CDB initiative include reducing the complex and voluminous file and folder system on disk.

5. GEOPACKAGE

GeoPackage is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information. The GeoPackage standard describes a set of conventions for storing the following within a SQLite database: vector features, tile matrix sets of imagery and raster maps at various scales attributes (non-spatial data), and extensions. GeoPackage is capable of storing feature geometries as Points, LineStrings, Polygons, MultiPoints, MultiLineStrings, MultiPolygons, and GeomCollection. For each feature, attributes can be stored as text, real numbers, Booleans, or raster/photos. A unique aspect to GeoPackage is the definitions that enable it to store binary blobs associated to features. The binary blobs associated to features are defined by the encoding standards as hand-held photos; however, these binary blobs could be used to store geospecific models associated to features. GeoPackage also stores multiple raster and tile pyramid data sets. “Tile pyramid” refers to the concept of pyramid structure of tiles of different spatial extent and resolution at different zoom levels, and the tile data itself. “Tile matrix” refers to rows and columns of tiles that all have the same spatial extent and resolution at a particular zoom level. “Tile matrix set” refers to the definition of a tile pyramid’s tiling structure (OGC GeoPackage Standard, 2018).

The GeoPackage initiative has both positive and negative lessons learned to help steer future terrain initiatives. A strength of the GeoPackage is the portability and utility in non-traditional simulation environments, such as hand-held tactical devices. Improvements to the GeoPackage initiative include defining a full breadth of components that are necessary to perform live, virtual, and constructive simulations.

6. THE FUTURE M&S AND GEOINT INDUSTRIES

In his presentation at the GEOINT conference in 2017, Robert Cardillo, director of the National Geospatial-Intelligence Agency, stated that “… in five years, there may be a million times more than the amount of geospatial data that we have today.” Geospatial data—and its uses—are exploding at a rate that outpaces the levels of innovation and utilization in the traditional M&S industry. Geospatial data is exploding at advanced rates of innovation for many reasons that include the advent of powerful and easily deployed sensors and drones, GPS tracking devices, crowd sourcing, and other disruptive technologies. As the quantity of data collected increases, the desire and technologies to synthesize, disseminate, and analyze the data will also increase.

The traditional M&S and GEOINT industries are on a path of unification. The GEOINT industry is rapidly advancing its abilities to store, collect, synthesize, disseminate, analyze, and render geospatial data. A key component to this advancement is the utilization and management of geospatial standards offered by the OGC. The M&S industry has lagged with geospatial standardization in comparison to the GEOINT industry for many reasons; the most notable reason being the inability of the industry to sustain a global, innovative, collaborative, hands-on engineering, and rapid prototyping program for validating and testing geospatial technologies in an open forum that is not exclusively DoD managed.

The M&S and GEOINT industries are also converging with respect to geospatial data. The OGC is a key forum influencing and facilitating this unification process. Within the

OGC, there are two geospatial standards that best enable the unification of the M&S and GEOINT industries: OGC CDB and GeoPackage. Usage of these standards is increasing in M&S and GEOINT industries, but they both contain weaknesses and strengths when it comes to the combined needs of both industries.

"GeoCDB"—a technology mashing of OGC GeoPackage and CDB—offers a deterministic repository of easily read geospatial datasets suitable for storage, runtime access, and dissemination for live, virtual, constructive, gaming, and mission command systems.

7. DESIGNS

Multiple approaches exist for leveraging GeoPackages within a CDB data store to offer an improved and multifaceted geospatial format capable of serving as a data repository and runtime format. For the scope of this paper, the approaches are limited to the vector data domain; a future paper should include approaches to leverage GeoPackage to store raster and 3D objects in a CDB data store.

The following sections consider four (4) designs for the incorporation of GeoPackage into a CDB data store where each design has a rationale, benefit with respect to improving CDB and leveraging GeoPackage capabilities, and disadvantage in terms of maintaining performance and/or fundamental design aspects of the CDB standard.

7.1 Approach #1 (Replace each Shapefile with a GeoPackage)

The easiest way to integrate a GeoPackage container into a CDB data store is to replace each Shapefile in a CDB data store with a GeoPackage. Within a CDB geocell, vector data are tiled by data layers (e.g., roads, rivers, geotypicals, etc.) and by LODs. CDB defines a level of detail required for vector features when feature count exceeds 4096. CDB defines 34 LODs based on a geotile definition where the dimensions of a geotile are 1x1 degree between 50 degrees north and south of latitude (at higher latitudes geotiles expand in longitude to account of longitudinal compression). In a fully populated CDB data store utilizing the Shapefile, best practice will have a vector file count of 33 at LOD 0, 45 at LOD 1, 93 at LOD 2, 16,413 at LOD 6 and an astounding 2.8×10^{14} at LOD 23. Given the quantity of files, transferring a highly populated CDB is untimely.

There a several advantages to this approach. Replacing each Shapefile with a GeoPackage reduces the number of files representing a CDB tile by at least 3 to 1. A Shapefile is composed of the following files: .SHP (vector geometry), .DBF (vector attribution), and .SHX (bounding information). A CDB tile may contain additional files representing instance level attribution (.DBF) and projection information (.PRJ). A GeoPackage is a single file that contains all the vector geometry, attribution, and bounding information. Employing a GeoPackage in lieu of a Shapefile for every CDB vector tile would reduce file counts to 11 at LOD 0, 15 at LOD 1, 31 at LOD 2, 5471 at LOD 6, and 9.8×10^{13} at LOD 23. Reducing the number of files in a CDB vector tile by 3:1 also improves the transfer rates of CDB. This approach has no relevant disadvantage other than it minimally exploits the capabilities available within GeoPackage.

Experimentations with approach #1 are positive. On average, the read time for a Shapefile is approximately 46 features per millisecond compared to 59 features per millisecond using GeoPackage. This represents an approximately 23% improvement in the quantity of features read per unit time. Transferring a single geotile of medium density vector data (e.g., LOD 4 with some LOD 5) in a Shapefile-based CDB from one folder to another executes in about 11,781 millisecond (total file count of 1378 files), on average. Transferring the same CDB dataset encoded in GeoPackage takes about 3734 milliseconds, a 68.3% improvement.

7.2 Approach #2 (Make each CDB tile a layer in a single GeoPackage)

Constructing each vector tile within CDB as a table within a GeoPackage for a given CDB dataset is a straightforward approach to utilize GeoPackage capabilities and significantly reduce file counts in a CDB (note that in GeoPackage a table is known as a layer). SQLite is the underlying framework for GeoPackage. The limitation on the number of tables in SQLite is 2,147,483,646 (a little over two billion). SQLite supports the CDB LOD conceptual model up to LOD 14. The concept of this approach is to exploit the high number of tables supported in SQLite to consolidate CDB tiles.

There are advantages and disadvantages of this approach. Making each Shapefile within a CDB into a table with GeoPackage reduces the number of files representing a CDB significantly. However, as the time to open a GeoPackage exponentially increases as the number of tables increase; a fatal disadvantage.

Experimentation with approach #2 is extremely negative. The open time for a fully populated GeoCDB for LOD 0 is 7 milliseconds, LOD 3 is 24 milliseconds, and LOD 6 is 5882 milliseconds, and LOD 7 is 53,114 milliseconds. Shapefiles consistently open in only a few milliseconds. The GeoCDB approach significantly degrades the performance of reading content from CDB in a rapid fashion and is a non-starter approach as a tenant of CDB is runtime utilization. The methodology to open the GeoPackages includes GDAL and SQLiteBrowser. Various versions of GDAL have different performance opening a GeoPackage suggesting a GDAL limitation; however, when using SQLiteBrowser to open the GeoPackages the same decay of open times is observed, albeit at a different magnitude. An analysis of SQLiteBrowser code suggests the increase in open time is due to each table being analyzed to deduce the schema.

7.3 Approach #3 (Store each CDB LOD as a layer in GeoPackage)

Design approach #3 incorporates the lessons learn from experimentation with approach #2 to limit the number of tables within a GeoPackage and reduce the number of files in a CDB. In this approach, the tables in the GeoPackage correspond to each LOD of CDB. The GeoPackage would contain 24 tables for each of the CDB LODs. Each CDB geotile would contain a GeoPackage to correspond to the CDB data stores (such road networks, geospecific points, etc.). CDB tiles for a data store combine into a single GeoPackage table within that given LOD where the tile definition (row and column) would be queryable attributes for each feature. In simple language, to find the features in a tile for a particular geotile's road network in LOD 3 of CDB, a consumer would open the road network GeoPackage, open the table that corresponds to LOD, and query for results where the column and row reference matches the CDB tile.

Combining CDB tiles by LODs is a significant reduction in the number of files representing a CDB. Employing approach #3 reduces file counts to only one (1)

GeoPackage for a CDB geotile per CDB dataset. A limitation of this approach is current CDB consumers will need to alter their method to finding tiles within a GeoPackage.

Experimentations with approach #3 are positive. Transferring a single geotile of medium density vector data (e.g., LOD 4 with some LOD 5) in a Shapefile-based CDB from one folder to another executes in about 11,781 millisecond (total file count of 1378 files), on average. Transferring the same CDB dataset encoded in GeoPackage with this approach executes in about 421 milliseconds, a 96.4% improvement (27.9 times faster).

7.4 Approach #4 (Store each CDB Data Store as a layer in GeoPackage)

Design approach #4 extends design approach #3 to have a single GeoPackage per geotile of CDB. In this approach, the tables in the GeoPackage correspond to each data store of CDB (such road networks, geospecific points, etc.). The GeoPackage would contain eight (8) layers representing each of the CDB data stores (GSFeature, GTFeature, GeoPolitical, VectorMaterial, RoadNetwork, RailRoadNetwork, PowerLineNetwork, and HydrographyNetwork). CDB tiles and LODs for a data store combine into a single GeoPackage table where the tile definition (row and column) and LOD would be queryable attributes for each feature. In simple language, to find the features in a location for a particular geotile's road network in LOD 3 of CDB, a consumer would open the geotile's GeoPackage, open the table that corresponds to data store, and query for results where the LOD column and row reference matches the CDB tile and LOD.

Combining CDB tiles and LODs by data stores is the maximum reduction in the number of files representing a CDB. Employing approach #4 reduces file counts to a single GeoPackage for a CDB geotile. With this approach, current CDB consumers would need to alter their method to finding CDB LODs and tiles within a GeoPackage.

Experimentations with approach #4 are positive and similar to approach # 3.

8. PATH AHEAD

CDB and GeoPackage are both OGC standards. The CDB standard is composed of 13 volumes where volume 4 describes the best practice utilization of Shapefiles within CDB. The CDB Standards Working Group (SWG) will construct a new volume defining a best practice for the employment of GeoPackage within CDB. The end state design of this best practice is a collaborate effort of the CDB SWG members that include AGC, SOCOM, Softwerx, CAE, Leidos, Cognitics, and NGA.

The CDB conceptual model outlining the storage of imagery is applicable to GeoPackage as its tile matrix. The raster tiles of GeoPackage are stored as either JPEG or PNG, which presents limitations of storing elevation data and raster materials from CDB. A worthy extension of GeoPackage would include the incorporation of storing raster data in uncompressed structures including the ability to store floating-point values. Another worthy extension to GeoPackage is the addition of 3D models to its structure.

9. SUMMARY

The Joint Staff J7 and AGC share a common objective to adopt standardized terrain formats. The (JTSE stood up a Terrain Data Standards for Joint Training working group to address the need for a “cooperative establishment of a joint training standard or specification for the encoding, storage, access, and modification of a representation of the natural and man-made terrain for virtual and constructive simulation applications” (Chambers and Callaway, 2017). The mid-term recommendations include a common library framework and standardized attributes, while the long-term recommendations include defining a common storage format for terrain data by 2021 and beginning the migration of constructive simulations and tactical simulators to the common standard as a runtime format. Additionally, the long-term recommendations include the ability to modify the common standard to meet unfulfilled requirements.

The traditional M&S and GEOINT industries are on a path of unification. The GEOINT industry is rapidly advancing its abilities to store, collect, synthesize, disseminate, analyze, and render geospatial data. A key component to this advancement is the utilization and management of geospatial standards offered by the Open Geospatial Consortium. The M&S industry has lagged with geospatial standardization in comparison to the GEOINT industry for many reasons; the most notable reason being the inability of the industry to sustain a global, innovative, collaborative, hands-on engineering and rapid prototyping program for validating and testing geospatial technologies in an open forum that is not exclusively DoD managed.

Figure 2 presents a deterministic repository of easily read data geospatial datasets suitable for storage, runtime access, and dissemination for live, virtual, constructive, gaming, and mission command systems managed by an international standards body where the technologies can evolve and grow to meet currently unfulfilled and future requirements. GeoCDB offers substantial improvements in the file transfer times and access of M&S content.

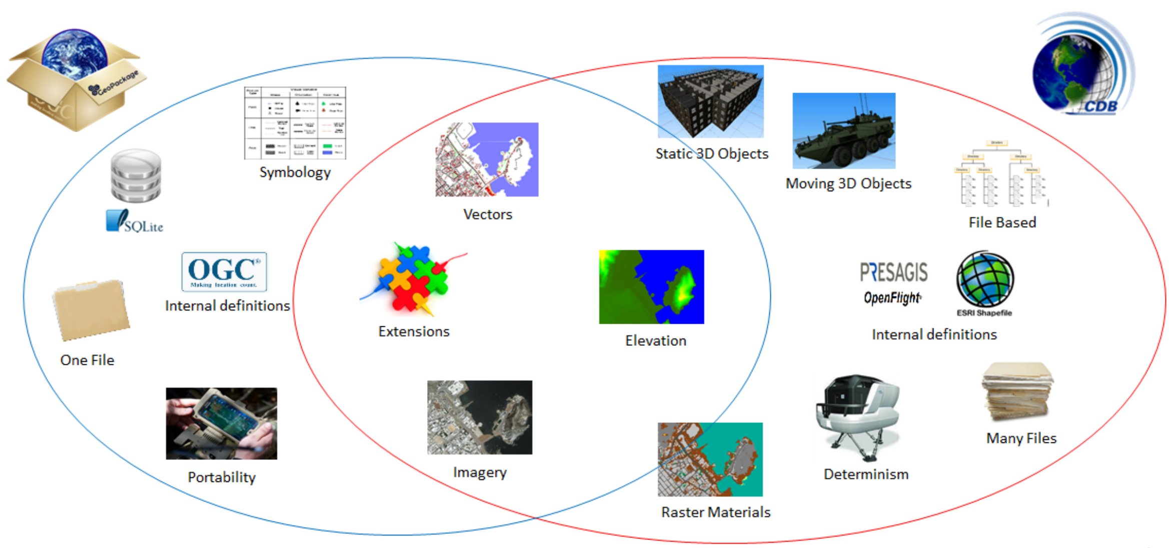

Figure 2: CDB and Geopackage Mashup

Annex A: About the Authors

Glen Quesenberry is the Architectures, Standards, and Test & Certification team lead for the U.S. Army Geospatial Center in Ft. Belvoir, Virginia. His duties include the alignment of geospatial standards with the overarching U.S. Army Geospatial Enterprise and Mission Command systems, the architectures that enable these systems, and the testing of these systems for interoperability. His specific area of expertise is the creation of modeling and simulation runtime products and the management of such representations across the DoD. As the M&S subject matter expert, Mr. Quesenberry has provided support to various Army customers, to include the Synthetic Training Environment (STE) Cross Functional Team (STE CFT) and the U.S. Army Modeling and Simulation Office (AMSO). Mr. Quesenberry has an undergraduate degree in Communications (focused on media production) and a Master's of Science in Information Systems Management.

Jay Freeman is the Synthetic Environment Technical Authority for CAE USA. He serves as the CAE's technical lead for Joint Staff J7 Environmental Development Division's development of a Terrain Generation Service and USSOCOM Geospatial Services where both capabilities leverage OGC CDB. Mr. Freeman previously served as the system and software architect for SE Core DVED. Prior to working for CAE USA, Mr. Freeman has worked at TERREX, Lockheed Martin STS (ATARS - SOFPREP), and Intergraph Services Company. Mr. Freeman attended Hobart College (Geneva, NY) for undergraduate studies and the University of Alabama in Huntsville (Huntsville, AL) for graduate studies.

Ronald Moore is currently the Chief Architect on SE Core. He has over 35 years of experience in the modeling, simulation, and training industry with expertise in software development, computer graphics, computer image generation, simulation geospatial terrain database production, sound simulation, streaming audio and video, and PC and console game development.

Kevin Bentley is the founder and president of Cognitics, Inc. He has over 20 years of experience developing software for GIS, simulation, and games. He currently is the principal investigator for a Phase II SBIR focused on research and development of technologies that improve the representation of transportation models that integrate with terrain in simulation systems. He has over 20 years of experience developing commercial and open source software applications and libraries for GIS, simulation, and video games. Mr. Bentley was the principal investigator for a Phase I, two Phase II enhancements, and a Phase III SBIR focused on research and development of technologies that improve the representation of transportation and hydrography models that integrate with terrain in simulation systems.

Samuel Chambers is the lead for Data Services at the Joint Staff J7's Joint Training, Environment Architecture Division. His DoD experience includes creating and maintaining geospatial products and databases, supporting Enterprise-level GIS, and managing the development of web-based data services to support modeling and simulation for training. He currently manages the Joint Training Data Services (JTDS), which provides Order of Battle and Terrain Data services to support Joint and Service Training. Mr. Chambers received a Bachelor's of Science in Geology from Elizabeth

City State University (Elizabeth City, NC) and a Masters in Geography from Virginia Tech (Blacksburg, VA), and served in the Marine Corps Reserves (Norfolk, VA).

Annex B: GeoCDB Experiment Statistics and Information

The test CDB data consisted of:

- 24 Geocells
- LODs populated up to LOD 12
- 237,296 Shapefile related files (.shp, .shx, .dbf, .dbt)
- 276GB size on disk
- 203GB compressed size
- 447,679 total files (Original with Shapefiles)

This experiment only placed vector data in GeoPackage files. There were four (4) options tested with each option storing the data that is normally in a Shapefile in a GeoPackage. Python code was written to convert an existing CDB for each option. This Python code is available on GitHub at: https://github.com/Cognitics/GeoCDB.

Each option was affected by the scalability of the SQLite engine and the GeoPackage implementation. The time required to open a file increased exponentially with the number of layers as shown in Table 1 and Figure 3. When testing with 50,000 layers, we abandoned the test after waiting 24 hours.

Table 1: GeoPackage Time to Open by Layer Count

Layer Count	Time in Seconds - GDAL	Time in Seconds - SQLite
1	0.014959	0.003989
10	0.010971	0.001995
50	0.054853	0.003989
100	0.10472	0.009973
500	0.890617	0.054856
1000	3.081757	0.122669
5000	118.041246	1.568804
10000	470.700899	7.179794
50000	Greater than 24 hours	343.679698

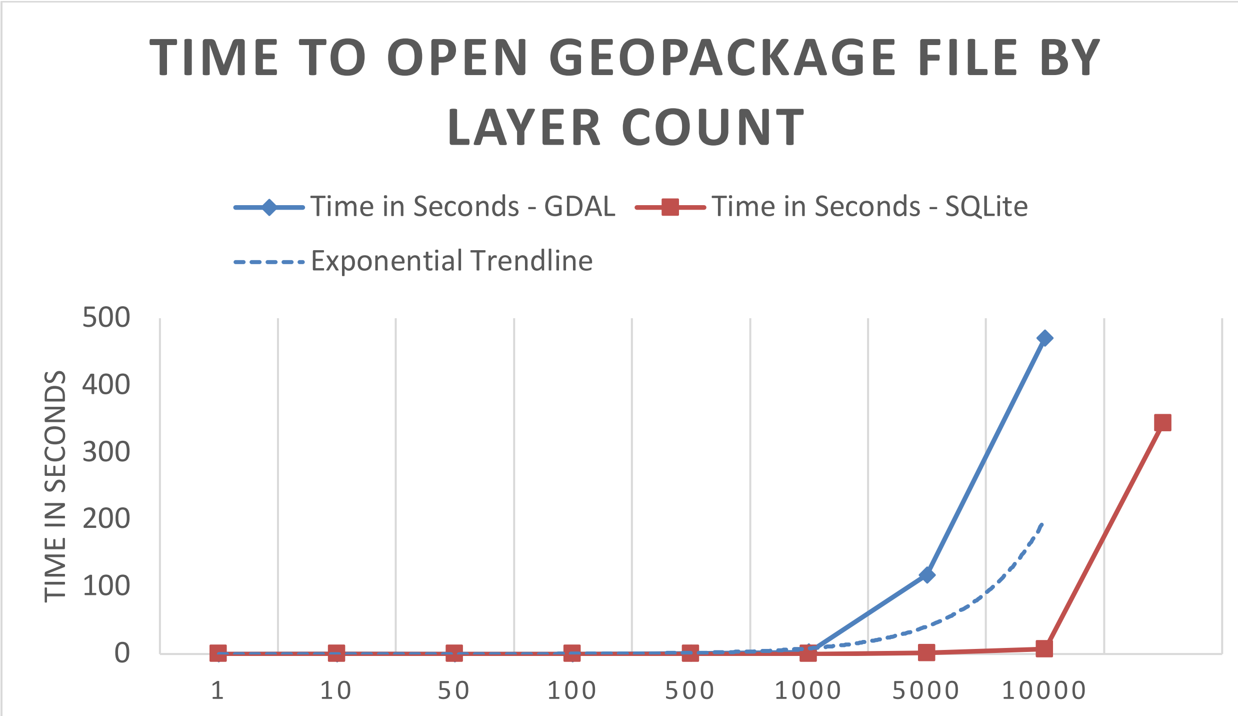

Figure 3: GeoPackage Open Performance by Layer Count

Each empty Shapefile still takes disk space and four files. An empty layer in GeoPackage has a negligible footprint.

Disk allocation units may greatly penalize large numbers of small files as is seen with Shapefiles, depending on the file system.

None of the tests took advantage of spatial indexing for queries. Spatial filters may offer additional performance benefits and functionality capabilities.

For each option, the total number of files and the reduced disk space are shown in Table 2 and Table 3.

Table 2: GeoPackage Statistics Summary

	File Count Reduction	Disk Space Reduction
Option 1	40%	21%
Option 2	53%	26%
Option 3	53%	26%
Option 4	57%	22%

Table 3: GeoPackage Statistics

Option	Total GeoPackage Files	Average Layers Per File	Total CDB Files after Conversion	Total CDB File Count Delta	Size on Disk After Conversion (GB)	Total CDB Size on Disk Delta (GB)
1	59324	1	269,707	-59,324	217	-59
2	1	59324	210,384	-237,295	205	-71
3	24	2472	210,407	-237,272	205	-71
4	163	1.8	210,546	-257,275	214	-62