intel.
GEOSPATIAL

Intel® Geospatial Platform
Whitepaper: Geovisual Data Management

Date: 10/22/2020

Contents

Copyright

This publication is protected by copyright. No part of the publication may be used, reproduced, or translated, in any form, without the prior written consent of Intel Corporation.

1   Introduction: What is Geovisual Data?

Geovisual data is visual data that is geotagged – that is, it contains metadata that describes the exact geographic location where it was captured with varying levels of accuracy. It is collected by remote sensing devices such as drones, satellites, manned aircraft and autonomous vehicles. The output of these devices produce data that is visual - RGB imagery, thermal imagery and 3D point clouds, among many others.

When collected at scale, from multiple sources and at a regular cadence, geovisual data can drive transformation for critical business workflows that result in drastic improvements in efficiency, reliability, safety and cost. Use cases spanning smart city planning to infrastructure inspection can be optimized and even automated from a common repository of geovisual data. Traditional enterprise maintenance workflows can be transformed from manual, time-based scheduling to condition-based scheduling, ensuring that issues are addressed before they become critical failures.

The opportunities to unlock the value of geovisual data are nearly limitless. However, there are many challenges that enterprises must address to achieve the goal of delivering this value at scale.

2   Challenges Managing Geovisual Data

As the volume of visual data has increased exponentially, the inherent challenges of managing these data types have been difficult to overcome with traditional data management approaches and platforms. As a result, most of the collected data is relegated to siloed network shares or USB hard drives that few in the enterprise even have access to. There are several reasons that these disparate sources of data are difficult to pull together into a unified repository.

Some of these challenges include:

- File transportability: Geovisual data such as RGB imagery, LiDAR point clouds, geoTIFF orthomosaics and others can generally be characterized as large binary files that need to be maintained as groups of related files or datasets (i.e. a drone survey consisting of many RGB images). Using traditional approaches, moving and organizing these very large files and datasets efficiently to a centralized location while making them easily accessible for viewing/analyzing requires systems and tools that are

- Specialized Cataloguing Requirements: Geovisual data is typically stored in specialized binary formats. In order to organize, cleanse and prepare the data so that it can be easily queried and searched by users and systems requires specialized, extensible indexing metadata extraction routines and expertise, data catalogues and object models.

- Visualization and Analysis: The same characteristics that make geovisual data difficult to transport (large file sizes, file groups, specialized formats, etc.) also make them difficult to efficiently view and analyze - both by users and automated systems.

- Geospatial Contextualization and Analysis: Geovisual datasets are often collected to build intelligence around a specific asset or location in the field. Viewing and analyzing this data in geospatial context and enriching it with other geospatial information and enterprise business information is critical for extracting the maximum value from the data.

- Security and Governance: While not challenges that are unique to geospatial data, security and data governance policies and procedures are often difficult to implement for geospatial datasets as a result of the data often being siloed and residing across many different storage mediums.

3 Managing Geovisual Data with the Intel Geovisual Data Lake

The Intel Geospatial Data Lake™ is an enterprise data repository that is specifically architected to meet the unique challenges of unifying and managing disparate sources of geovisual data in a secure and scalable cloud image. The Geovisual Data Lake provides a central, managed cloud repository that consolidates previously siloed geovisual data and makes it available for analysis and interrogation to users and systems across the enterprise.

Using the Geovisual Data Lake, customers can easily consolidate, aggregate and enrich geovisual data collected from various sources including drones, manned aircraft, LiDAR, Satellite and autonomous vehicles, among others. With a seamless workflow for integrating, cataloguing and contextualizing this data, users can focus on extracting value and intelligence from these datasets using powerful built-in query, filter, search, visualization and AI capabilities.

The Geovisual Data Lake provides integration capabilities to deliver intelligence derived from the data to workflows in enterprise business systems, providing a true end-to-end geovisual intelligence pipeline. REST APIs and a robust set of connectors are provided to deliver data insights directly to the systems that need them, when they need them.

3.1 Intel Geovisual Data Lake Services

The Intel Geovisual Data Lake provides services specifically designed to unlock the value of geovisual data. Architected for high availability, reliability and security, these services are designed to meet the challenges of managing geovisual data at scale.

3.1.1 Smart Data Ingestion Services

Responsible for securely and efficiently transporting large volumes of geovisual data from the edge or local network to the Data Lake storage in the cloud. These services provide high performance,

parallelized, asynchronous dataset upload with automatic failover and restart that can be operated in a supervised or unsupervised mode. Upon landing in the Geovisual Data Lake storage, data is automatically quarantined, validated and scanned for integrity and malware before being released to production storage.

3.1.2 Data Workflow Recipes

Coupled with the Smart Data Ingestion Services, Data Workflow Recipes define how datasets will be indexed and processed by the Geovisual Data Lake. Recipes can define specific metadata extraction routines, external metadata formats, and ETL tasks to perform on the dataset to prepare it for reporting and analysis.

3.1.3 Data Cataloguing and Preparation Services

Responsible for interrogating the dataset once it has been released to production storage. Based on the dataset type or a configurable Data Workflow Recipe, specialized metadata extraction services will be executed from an extensible library of routines designed specifically to handle a wide array of geovisual data types and formats. Metadata will be organized and stored in the Data Catalog, where it is indexed and prepared to be searchable and queryable by users and systems. For all geovisual data, geospatial metadata is also extracted as part of this process and stored in a geodatabase that is tightly integrated with the Data Catalog. This enables further enrichment and contextualization of the data with other data streams connected to the Data Lake, such as geospatial datasets and enterprise business datasets.

Additional processing and preparation routines may also be executed by these services, based on dataset type or a configured Data Workflow Recipe. These tasks are extensible and range from data cleansing services to data processing routines such as tiling services.

3.1.4 Data Access Services

Provides a robust set of secure REST APIs for accessing datasets in the Geovisual Data Lake through standard programming interfaces. The APIs enable the following capabilities:

- Query, search and filter datasets on metadata attributes using simple to complex criteria.
- Update metadata attributes for any file or dataset.
- Retrieval/download of files or datasets.
- Delete files or datasets.
- Trigger or invoke processing or analytic workflows on a dataset.
- Connect third party reporting or visualization tools.

3.1.5 Geospatial Analysis Services

Provides a set of secure REST APIs and Web Services that enable geospatial query and analysis of data stored in the Geovisual Data Lake. The APIs and services enable the following capabilities:

- Search for datasets in a defined area or within a radius of another dataset, object or location of interest.
- Perform general geospatial query and search functions (contains, intersects, etc.).
- Perform general geospatial functions such as buffer analysis and proximity calculations between objects or datasets in the Geovisual Data Lake.

3.1.6 Tiling and Streaming Services

Both 2D and 3D datasets can be automatically tiled in the Geovisual Data Lake and published to standard geospatial web service protocols. Users and systems across the enterprise can securely stream large geovisual datasets across a variety of devices and bandwidth environments. Downstream analytics can also programmatically request tiles or streams enabling more efficient analysis pipelines.

3.1.7 Enterprise Integration Services

Building upon the same REST APIs that are part of the Data Access and Geospatial Analysis Services, intelligence derived through analytics and reports in the Geovisual Data Lake can be connected to enterprise workflows and systems include ERM, EAM, GIS, among others. A library of connectors is available for configuration to a customer's specific environment. These connectors can trigger actions in these downstream systems, import enterprise datasets for aggregation and enrichment in the Geovisual Data Lake (for example to surface information like the asset id or last inspection date in the context of the geovisual data), and push data or insights to systems for further analysis or reporting downstream.

3.1.8 Data Lifecycle Management

The Geovisual Data Lake provides a default set of data lifecycle management policies that can also be tuned for specific instances or use cases. These policies define how and when data should be moved to different storage tiers in the cloud repository as it ages, or newer data is collected or otherwise becomes less necessary for immediate accessibility. These policies can be configured down to the dataset level, enabling fine-tuned control over data accessibility.

3.1.9 Security Services

Security is a foundational service that underpins all other services, data and tasks within the Geovisual Data Lake. Each customer instance is private and backed by both Intel and Amazon security standards

and certifications. Data validation and interrogation routines ensure that all incoming data is quarantined and scanned before release to the Data Lake Storage and downstream processes. Enterprise grade encryption is leveraged for both data at rest (256 bit SHA-2 encryption) and in transit using HTTPS (4096 bit RSA encryption).

The Geovisual Data Lake also provides authentication, identity and fine-grained access control services. The service supports industry authentication standards including Oauth 2.0 and SAML 2.0, multi-factor authentication and federated identity management with enterprise identity providers including SAML and ActiveDirectory.

To learn more or talk to an Intel expert, please visit Intelgeospatial.com