Naval Research Laboratory

Washington, DC 20375-5320

NRL/7340/MR—2022/2

Real Time Management of Geospatial and Environmental Data

SETH CHAMPAGNE, PA

JOHN T. SAMPLE

Center for Geospatial Sciences Branch
Ocean Sciences Division

November 14, 2022

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
14-11-2022

2. REPORT TYPE
NRL Memorandum Report

3. DATES COVERED *(From - To)*
1/1/2017 – 1/1/2019

4. TITLE AND SUBTITLE
Real Time Management of Geospatial and Environmental Data

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Seth Champagne and John Sample

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
N2M5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
1005 Balch Boulevard
Stennis Space Center, MS 39529

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/7340/MR--2022/2

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
875 N Randolph St.
Arlington, VA 22217-1995

10. SPONSOR / MONITOR'S ACRONYM(S)
ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

NRL's Center for Geospatial Sciences conducts a broad research program in data management and data dissemination for DoD users. This report details specific research conducted to support management of real time and highly dynamic data sets. Three distinct problem areas were studied. Development was conducted in each area to address key need areas. This paper will review the problem areas and the development results for each one. Two of the problem areas research involve collection and maintenance of very large and fast changing data sets. The final problem area involves distribution of geospatial and environmental data to end users including mobile devices.

15. SUBJECT TERMS
Geospatial data Environmental data GeoPackage
OpenStreetMap METOC

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	U	9	Seth J. Champagne
					19b. TELEPHONE NUMBER *(include area code)* (288) 688-4670

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

Real Time Management of Geospatial and Environmental Data

Objective

The Naval Research Laboratory's Center for Geospatial Sciences conducts a broad research program in data management and data dissemination for DoD users. This report details specific research conducted to support management of real time and highly dynamic data sets. Three distinct problem areas were studied. Development was conducted in each area to address key need areas. This paper will review the problem areas and the development results for each one. Two of the problem areas research involve collection and maintenance of very large and fast changing data sets. The final problem area involves distribution of geospatial and environmental data to end users including mobile devices.

OpenStreetMap

OpenStreetMap (OSM) is a very large crowd sourced geospatial data set and software stack. The data and associated visualization software are open sourced and freely available. It includes streets, buildings, natural, business, government and covers the entire world. Users around the world update it thousands of times each day.

OSM data is often more accurate and more current than commercial or government data sets for many areas of the world. As such, OSM data is of high interest to the DoD community. Specifically, mission planners and analysts, though the open source tools do not fully meet the needs of this user base. First, the open-soured visualization tools do not allow for on-the-fly customization of maps. Secondly, DoD users are often working on DoD restricted networks or disconnected systems and are not able to access the Internet connected servers. Finally, the data is so voluminous and fast changing that the current software tools are often too slow for high demand services. Fast ingest of new data is an absolute requirement for successful use of the OSM data. The data changes frequently and the ingest/update cycle will determine the currency of the data provided to DoD users. DoD support for humanitarian crisis represent a use case where currency is especially important. In a disaster such as a major earthquake or tsunami, the OSM data will change frequently to reflect damage and other changes to infrastructure such as roads and bridges. This updated data is critical to relief efforts.

A technical integration plan was created and executed during this work to support OSM for DoD users. OSM distributes its data in the Protocol Buffer Format (PBF). The PBF format is compact and compressed. It is not suited for serving data for on demand queries. Typical users of the OpenStreetMap data download the PBF data and ingest it into a GIS enabled database such as the PostGIS database. For our purposes, this system would not work. It is not portable across multiple installation scenarios and does not support high performance, render on demand use cases that are needed to support custom styling demands. We developed a conversion tool to read

Manuscript approved November 2, 2022.

the data from the PDF files into NRL's vector cluster format (vcluster). Vector Clusters use a high performance binary format with a geospatial indexing scheme powered by R*Tree data structures. This design allows for large amounts of data to be used in on-the-fly map creation. These are the backing data source for NRL's Chartserver.

To produce images from the raw OSM vector data, we adapted the existing NRL Mapmaker tool. Mapmaker uses a series of XML configuration files to define how vector features are symbolized, what map scale they should appear, and in what order they are drawn to a map. Mapmaker is NRL's in house counterpart to Mapnik, the open source map-rendering tool that OSM uses. Mapnik also uses an XML configuration to describe how features are symbolized. We developed a tool to convert the Mapnik XML symbolization to Mapmaker XML files. This was a tedious process that required a lot of validation checking.

The data access methods used by OSM had to be reverse engineered to replicate the intended map functions of the OSM data. OSM has a tool to export data from the PDF files to a PostGIS database. The Mapnik XML configuration contained embedded SQL statements for retrieving the data out of the PostGIS database for map creation. We developed new methods to mimic the work done by these SQL statements by the new code that reads data from the vcluster files.

All of this work enabled for on-the-fly creation of OSM images served through a WMS and WMTS. NRL had previously produced an in house version of OGC WMS, WMTS and WFS, but these services were outdated and did not meet the standard OSM style guidelines. We created updated versions of these services. Users can access the base map containing all of the OSM data, or selectively choose feature layers to view. We also supports an "overlay" option of OSM with transparent backgrounds. None of these features are provided by the open source tools.

Results from this effort have been highly successful. Customized visualization and high performance ingest support critical DoD requirements and the system has been transitioned to operational programs.

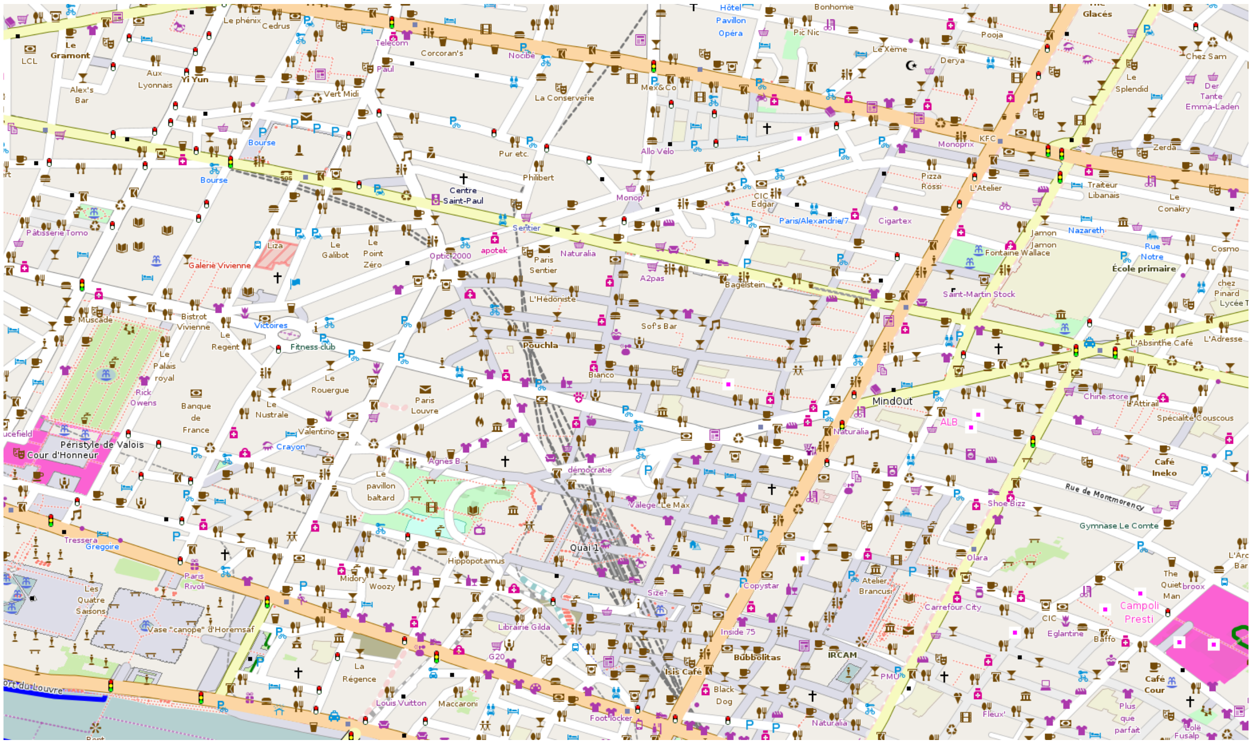

Figure 1: Paris Segment from OpenStreetMap

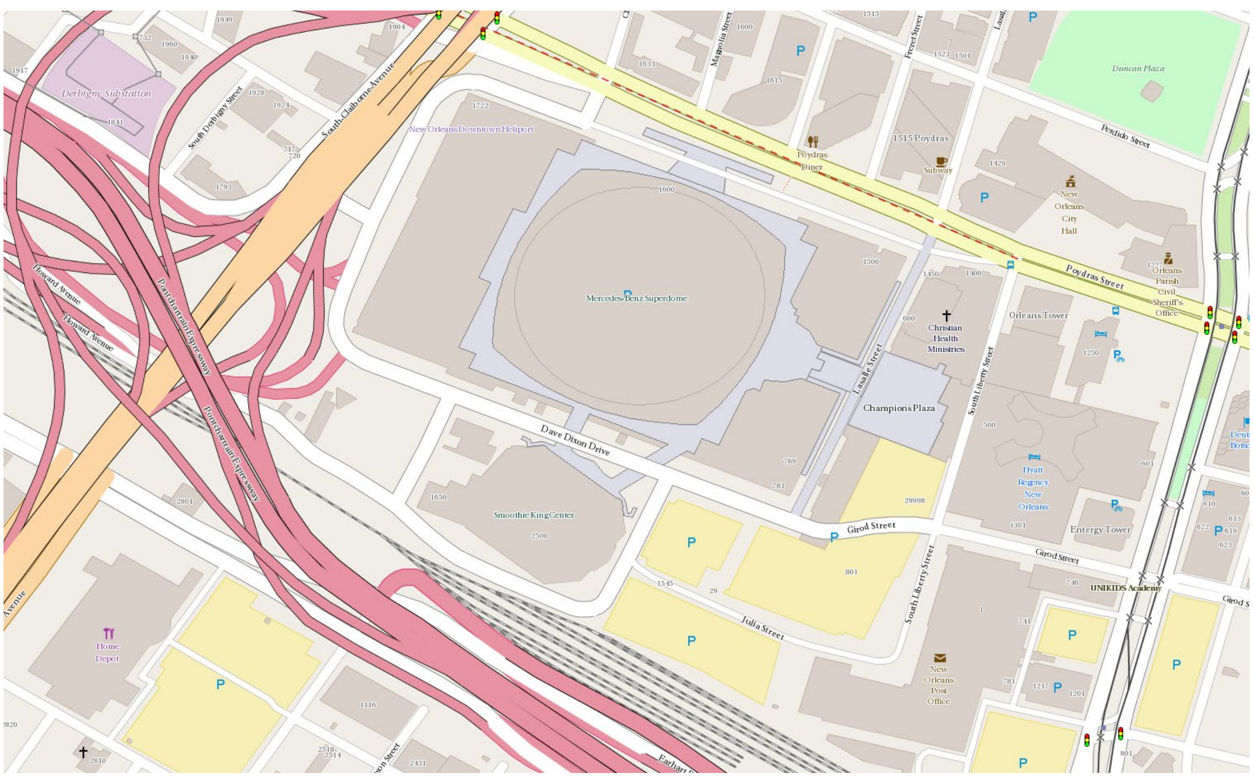

Figure 2: New Orleans Segment from OpenStreetMap

METOC Server

Distributing Meteorology and Oceanography (METOC) data is a longstanding problem in the METOC community due to a lack of standards, lack of scalable software, and the large amounts of data files produced per day by METOC models. It is common for a model to produce tens of thousands of files totaling gigabytes worth of data per analysis, with multiple analyses run per day. These models are produced by supercomputers and require large amounts of storage. Getting the data efficiently from production centers to end users and client applications is a major open requirement for the DoD.

The major task we addressed in this research was to catalog the metadata of METOC data files into a high-performance, portable binary cache. The metadata is stored in a flexible data model to fit the wide variety of file types and data products relevant to METOC models. METOC data is multi-dimensional, spanning a geographical area as well as a vertical area and a span of time (the time frame being modeled). We created an API (application programming interface) that supports queries by geospatial extant and temporal range. The API also supports contextual queries such as what variables and dimensions are modeled by the data. The results of the API are a listing of files that match the query along with the metadata of the files. The API is exposed by a RESTful web interface.

METOC data is complex due to a variety of file formats, large updating datasets, and multi-dimensional products represented across multiple files. The two most common file formats are NetCDF and GRIB, both of which support "gridded data", a matrix of data values representing a rectangular grid over a geospatial area. Less common are GeoTIFF and KMZ formats and custom text based formats for special applications. The METOC Server catalog design can handle any file format because it uses a Java interface to drive the reading of the metadata from the files.

The file metadata consists of bounding box, time, parameter, and attribute information of the file. Some files contain only a 2D slice of data for a single time and parameter, and other files can contain 3D or 4D data over multiple times and multiple parameters. All of the file types and dimensions are supported through the same query API using the metadata to filter the results and efficiently find relevant files.

The bounding box or "coverage" of the file is the geospatial location of the data. Coverages are assigned names for defined operating regions or METOC data. Some METC models are only available in specific areas of interest. The coverages allow METOC Server to support geospatial filtering of the data.

The time consists of an "analysis" timestamp representing when the data was created, and one or more "taus" representing an offset from the analysis time. The taus are normally hour intervals, but can be minutes, seconds, etc. Taking the analysis time and adding the tau gives the "valid time" of the data. The valid time is the time that the data is forecasting. For example, a file with analysis time 2000-01-01T00:00:00Z, and a tau of 06 then has a valid time of 2000-01-01T06:00:00Z. The analysis and valid times of the metadata allow METOC Server to support time based filtering.

The final component of the metadata is the parameter and attribute information. Parameters or “variables” are the different categories that the METOC data is forecasting. This could be air temperature, cloud cover, winds, sea temperature, salinity, wave height, etc. METOC Server allows the data to define its own variables so that it can adapt to other contexts beyond METOC data, such as acoustic and electromagnetic derived products. Along with the variables are the “dimensions” of the file. These typically represent the vertical component of 3D data, such as elevation, depth, isobaric level, etc., but can be extended to define other dimensionality of the data such as frequency in the acoustic data. This information defines the context of the data, what it represents or forecasts, and allows the METOC Server to filter by context.

The METOC Server’s cataloger is configurable to support many different data products in different deployment situations. The cataloger runs as a scheduled process to scan a file system or cloud storage for new data. Since METOC data is produced on a predictable schedule, a scheduled process is reliable for updating the catalog with new data as it becomes available. The interval of the scan is configured to run more often or less often depending on how often the data is updated. A multi-threaded process reads the metadata of each file and appends it to the catalog. The web services can then access those files by querying its metadata from the catalog.

Some data products supported by the METOC Server are not updated on a schedule but instead on ad hoc usage. For these products, the METOC Server supports publish-subscribe patterns through Apache Kafka or alternatively a REST interface to update the catalog. This allows for a low latency solution to updating the data in the catalog.

Our system also supports discoverability of data. Users can find what data is available and acquire the specific files relevant to their use case. The server supports forward deployed to push METOC data closer to the end users. METOC Server also OGC WMS and WCS services for data visualization and analytics support. It can also run in the cloud to scale to a large number of users. The METOC Server has had multiple success in early integration experiments and will directly support the fleet on transition.

GeoPackage

GeoPackage is an Open Geospatial Consortium (OGC) specification for a lightweight, all-in-one file format to store geospatial data. GeoPackage uses SQLite to store both vector, image and grid data. It uses a single file and is highly portable. GeoPackage is highly utilized by mobile devices. This format is needed by DoD systems to support ad hoc sharing of geospatial and environmental data.

All Android and iOS devices have SQLite installed and it is a common technology for mobile app development. GeoPackage allows for raster and vector geospatial data to be stored in a

SQLite database designed for use by a mobile app map client. SQLite is restricted to single threaded reading and writing, but it is also convenient for small applications and distribution of data. All of these considerations make GeoPackages the ideal format for exporting geospatial data for offline use by a mobile device.

To support raster data, the individual map images are stored as binary BLOBs into a SQLite table. The images can be PNG to support transparency, or JPG for better compression and smaller file size. The table follows a standard tile scheme that defines the bounding box and resolution of each "zoom level" of the map. Each image is defined by a row, column, and scale from the tile scheme. The zoom levels are then stacked on top of each other in a pyramid shape, so each higher resolution zoom level has twice the number of images as the one above it. For this project, we supported both the WGS 84 (EPSG:4326) and World Mercator (EPSG:3395) map projections.

For vector data, individual vector features are stored in table rows. Each row has a geometry column with Well-Known Binary representation of the feature. All of the other columns of the table represent the key-value pairs of the feature attributes. Exporting vector features alongside raster images is a unique feature of the GeoPackage format. Other portable formats support only vector data (such as SHP files) or only raster data (such as GeoTIFF). The vector features allow a mobile app to support detailed "feature info" views of the map data, where the feature attributes can describe areas of interest to the user.

A GeoPackage ordering tool was created with a RESTful API. Users can specify arbitrary area of interests either by bounding box or name (country or COCOM). They can also specify the desired map resolutions. The web service will then gather the requested data and bundle it into a GeoPackage. Finally, the GeoPackage is provided for download. A front-end thin client exists for the service both as a standalone web app and as a plugin to Map of the World.

The web service supports a number of different data sources. The service supports OGC Web Map Service (WMS), Web Map Tile Service (WMTS), Web Feature Service (WFS), ArcGIS MapServer and FeatureServer, GeoTIFF and GeoJSON. The RESTful API also exists in the NRL products Tileserver and Chartserver. This enabled all of NRL's large data products to be exported in GeoPackage format. The GeoPackage service is fully configurable and multi-threaded to speed up the writing process. The service can also send out alert emails when GeoPackages are finished building.

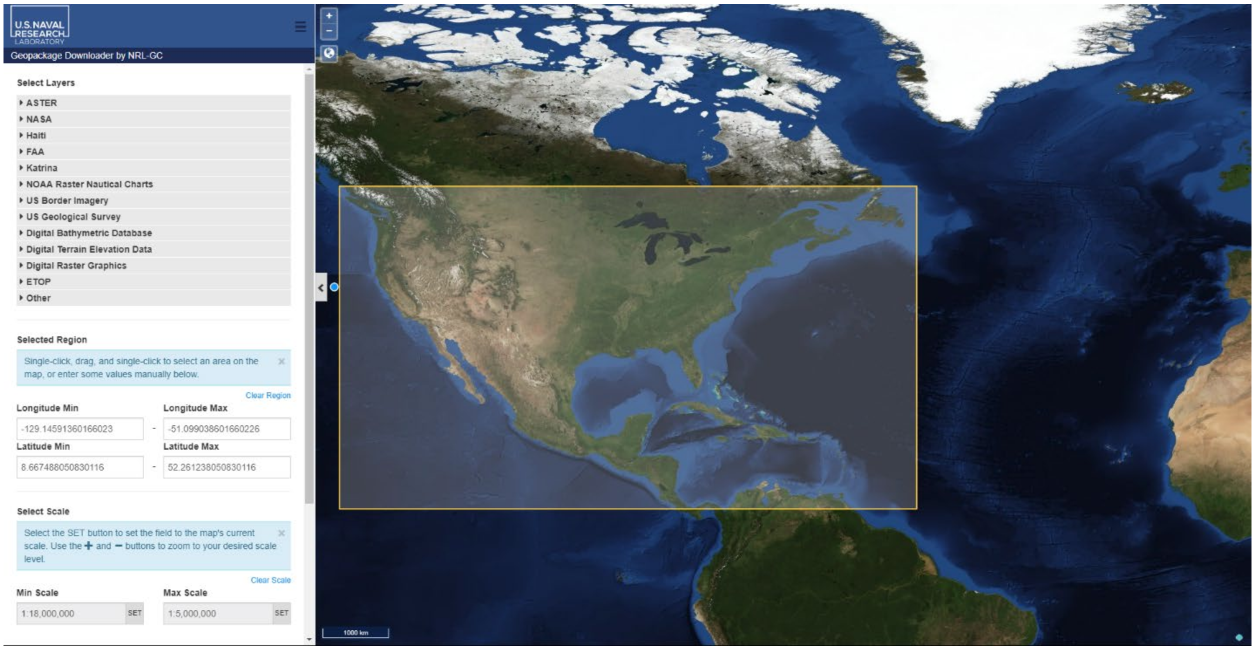

Figure 3: GeoPackage Ordering Client

References

1. OpenStreetMap contributors; OpenStreetMap database, OpenStreetMap Foundation: Cambridge, UK; 2021 www.openstreetmap.org.
2. Open Geospatial Consortium, GeoPackage Specification http://www.geopackage.org/spec version 1.3.1
3. Sample, J.T. and Ioup, E.Z.; Tile Based GIS: Principles and Practices, Springer 2010