Thursday, 15th September

Session 1: Products, Data Handling and Standards

0900: The Open Navigation Surface Project: A Grid File Format for Hydrography

John Shannon Byrne[3]

Richard T. Brennan[1,2]**, Brian Calder**[2]**, James D. Case**[2]**, David Fabre**[4]**, Barry Gallagher**[1]**, R. Wade Ladner**[4]**, Bill Lamey**[5]**, Friedhelm Moggert**[6]**, Mark Paton**[7] **, Jack Riley**[1]**, and the Open Navigation Surface Working Group**

[1] National Ocean Service, National Oceanic and Atmospheric Administration, Silver Spring MD, USA.
[2] Center for Coastal and Ocean Mapping & NOAA/UNH Joint Hydrographic Center, University of New Hampshire, Durham NH, USA.
[3] Science Applications International Corporation, Marine Science and Technology Division, Newport RI, USA.
[4] Naval Oceanographic Office, Stennis Space Center MS, USA.
[5] CARIS Ltd., New Brunswick, Canada.
[6] Seven Cs AG & Co., Hamburg, Germany.
[7] IVS3D Ltd., New Brunswick, Canada.

Abstract

Many hydrographic and oceanographic agencies have moved or are moving towards gridded bathymetric products. Grids provide an efficient and powerful way to represent geospatial data. A processing pipeline that produces data products at the full spatial resolution warranted by the survey systems and parameters offers clear advantages in the preservation and dissemination of information over the traditional approach of archiving "smooth-sheet" type products from hydrographic surveys. However, there is no accepted standard format to allow such gridded data products to be exchanged between various organizations while still maintaining necessary data and metadata integrity. The Open Navigation Surface (ONS) Project is an open-source software project designed to provide a freely available, machine-independent, software library to encapsulate gridded bathymetric surfaces and associated uncertainty information. The data file format, called a Bathymetric Attributed Grid (BAG), contains at least two co-located gridded layers, these being the bathymetry data and the bathymetric uncertainty. A list of changes resulting from hydrographer specified (i.e. "golden") soundings is maintained within the file for audit purposes. Metadata detailing all pertinent attributes, ownership, parameters, and processing history are saved within the BAG to help ensure that these details remain physically associated with the data. A mechanism is provided to digitally 'sign' the BAG, allowing a data user to verify that the contents of the BAG have not been modified since it was certified. The BAG specification optionally allows for additional layers of gridded data, but does require that the contents of such optional layers be defined. The prototype implementation of the BAG software access libraries is described herein, and the software access library functionality is demonstrated via a prototype BAG constructed from a sample of the 2005 common data set. Visualization of the data contained within the prototype BAG is accomplished with commercial software packages.

Background

The Navigation Surface [1, 2], developed at the University of New Hampshire (UNH), defines an approach to hydrographic data processing that results in a gridded database of bathymetry data and corresponding bathymetric uncertainty. A fundamental concept in this approach is the replacement of the standard archival data product of sounding points and contours, with a gridded database product. The standard archival data product is traditionally generated at a scale (data density) appropriate for the intended navigational chart to be produced from the survey data, and includes a (potentially) significant level of data decimation and corresponding loss of information detail. Whereas, the new archival data product is a gridded database whose data density is consistent with the highest spatial resolution appropriate for the survey data. The Navigation Surface approach describes a defocusing process (to account for horizontal error effects) and a generalization processes (to account for chart scale) that allow for one data product to readily contribute to one or more chart data products, potentially at different scales.

With increasing acceptance of the Navigation Surface concepts comes the need for a gridded data format that provides for encapsulation of the bathymetry data, uncertainty data, survey acquisition and post-processing related metadata, a node change-list (a.k.a. 'Golden Sounding'), and data file signature and certification/authentication needs. The archival nature of the Navigation Surface product suggests that a suitable data format should provide for straightforward interoperability among the effected hydrographic and oceanographic offices, agencies, and institutions.

The Open Navigation Surface Working Group (ONSWG) was formed based on responses to an email message soliciting interest from the community. Initial members included representatives from the National Oceanic and Atmospheric Administration (NOAA), the US Naval Oceanographic Office (NAVOCEANO), Science Applications International Corporation (SAIC), Interactive Visualization Systems (IVS), CARIS Ltd., 7Cs GmbH, and the Center for Coastal and Ocean Mapping and NOAA/UNH Joint Hydrographic Center, (CCOM/JHC). The primary goals of the ONSWG are to define an open, platform independent, grid database file format suitable for access, archival, and interchange of Navigation Surface results, and to develop an open source software access library to operate on this format. The fundamental idea behind the ONSWG is to develop **a data format for the community, by the community**. The ONSWG has convened two separate working sessions hosted at CCOM/JHC, each of one-week duration. Member communications are facilitated via an email distribution listing, and a Concurrent Versions System (CVS) server provides file revision control. The first meeting, held in January 2004, completed with consensus on the major elements of the fundamental approach, established smaller working groups, and assigned areas of responsibility. Efforts continued following the first working meeting with significant contributions in the areas of signature/authentication and metadata XML schema definition and parsing. The second meeting, held in July 2005, completed with establishment of baseline versions of the dependant libraries, a refinement of the overall call sequence for opening, creating, and accessing BAGs, and a significant percentage of the required integration code completed. The individual members of the ONSWG, working on a voluntary basis over the last 18 months as their schedules allow, have completed contributions to their assigned areas of responsibility such that a prototype version of the access libraries now exists, and a sample BAG file has been created.

Approach

Through an extensive exchange of concepts and ideas in advance of the first ONSWG working session, many basic requirements and objectives became solidified. During the first working session, the following fundamental approach decisions were made. 1) The data format and software access library will be freely available to parties that express an interest. 2) The data format itself does not specify or infer any processing policies, procedures or algorithms. 3) The data format is defined to support elevations both above and below sea level. 4) The name Bathymetric Attributed Grid (BAG) was selected in part to avoid being limited to just the concepts of the Navigation Surface, even though encapsulation of Navigation Surfaces is a primary objective. 5) The mandatory elements of a BAG include metadata, co-located elevation and uncertainty grids, a change-list to support hydrographer specification of the depth in a certain area, and file signature and certification/authentication. 6) The ISO19115 [3] standard for geospatial metadata, and the XML encoding of ISO19139 [4] were selected as the metadata standards for the BAG format. 7) HDF-5 [5] was selected to provide overall data encoding and platform independence. 8) All data values are encoded in metric SI units. 9) The format supports both projected and geographic coordinates. 10) A limited number of coordinate system projections, horizontal datums, and vertical datums are supported. 11) In addition to the mandatory elements of a BAG, optional extensions can be included, but these must be open and defined, and comply with BAG conventions. 12) The NGA Geotrans [6] software package provides the services for all coordinate system transformations. 13) Mechanisms are provided within the format definition and software libraries to digitally sign a data file, and to authenticate a previously signed data file to ensure data reliability and original content have not been compromised since the file was signed. 14) The access library is written in ANSI C, but with intent to support both C and C++ applications. 15) For simplicity, the initial focus is to support a single spatial resolution within a BAG. Growth potential to support multiple spatial resolutions within a BAG does exist with HDF-5. 16) Platform independence is a basic requirement, with intended support for at least the following operating systems: Windows, various versions of Linux, and Macintosh. Towards this end, the group agreed on the basic need to be able to compile and link from source code all modules on which the BAG depends.

By the end of the first working session, smaller working teams were formed to complete more detailed design and concept refinement. Teams were formed around the following areas: metadata definition, HDF-5 encapsulation, coordinate system transformations, signature/authentication, application program interface (API) definition, library build structure, documentation, unit testing, and sample programs. A CVS site hosted at CCOM/JHC maintains file revision control with updates checked into CVS as milestones have been reached. A web site (http://www.opennavsurf.org) exists to provide a basic forum for information access. The web site contains contact information and describes how to sign up for the supported e-mail lists. In contrast to a strictly standards based up front process, the activities of the ONSWG aim to establish a working version of a gridded data file format through community involvement. However, this approach does not preclude a standards process at some point in the future.

File Structure

Figure 1 provides a conceptual layout of a BAG data file. The metadata, elevation grid, uncertainty grid, change-list, and certification elements of the file are all mandatory. The extensions section of the file is optional, and must be explicitly defined. The metadata, elevation, uncertainty, change-list, and extensions are each HDF-5 datasets with a dataset substructure appropriate for the information being stored. The signature section is simply a byte string appended after the end of the HDF-5 file using standard C file access mechanisms. The rationale for adding the signature byte stream after the end of the HDF-5 file is to ensure that the contents of the signature do not modify the file that it is trying to protect.

The metadata section contains all of the descriptive information about the overall dataset including the extents, resolution, coordinate system, all parameters necessary to convert between geographic and projected coordinate systems, datum information, and the exact location of the south-western most node. The metadata also contains extensive information about who acquired the dataset, when the dataset was acquired, how the dataset was acquired, chain of custody information, and security information. The metadata section of the BAG is encoded as an HDF-5 dataset that contains a one-dimensional character array with a defined maximum length. The character array is loaded with the XML encoded string containing all metadata parameters.

The elevation section of the BAG contains a two-dimensional matrix of elevation values, organized in row major order, and starting from the south-western most data point, where each value is defined to be the elevation at an exactly specified geographic point (node). An important distinction to note is the contrast between this node-based geo-referencing as opposed to cell oriented grid-based geo-referencing whereby a grid value may be defined to have the spatial extents of a grid cell, or a sounding footprint. The units of the elevation values are meters, and the sign convention is for z to be positive for values above the vertical datum. In release 1.0.0, the elevation values are encoded as four byte floats within an HDF-5 dataset. The x and y location of each elevation value is not explicitly saved within the dataset. Rather, the X, Y location of the south-western most point is saved within the metadata section, along with the X node spacing and the Y node spacing. Regular spacing between nodes is required.

The uncertainty section of the BAG contains a two-dimensional matrix of values that specify the vertical uncertainty at each node. The elevation grid and the uncertainty grid are explicitly co-aligned. The values are expressed as positive quantities in units of meters. The exact definition of the elevation uncertainty has not been defined herein. This was deemed necessary to allow flexibility in the use of the uncertainty values. A BAG at the stage of final survey data processing should contain uncertainty information germane to the survey data itself and intended to be used for information compilation. A BAG intended for navigational purposes would need to specify the overall uncertainty to the mariner – these two may be quite different. In release 1.0.0, the uncertainty values are encoded as four byte floats within an HDF-5 dataset. The ONSWG plans to investigate disk space and file access efficiency alternatives in a future release that may result in additional options for the type used by HDF-5 to encode the elevation and the uncertainty.

As of the time of this writing the encoding approach for the change-list section of the BAG has not been completely defined, and is not supported in the currently available prototype version of the library and sample BAG files. The intent is to complete this definition and encoding before releasing version 1.0.0. It may be deemed necessary to enforce Hydrographer modifications, or designated soundings, when other sources of information offer a better understanding of the true depth at some location. In this situation, the designated (i.e. "golden") sounding and its corresponding vertical uncertainty will be used to replace the original values in the existing elevation and uncertainty surfaces. The displaced elevation and uncertainty values will be saved to the change-list section of the BAG along with additional information to specify the rationale for the change, thereby providing an audit record. Sufficient information will be maintained to ensure that this process can be reversed (iteratively if required) if deemed appropriate at some point in the future.

The signature section of the BAG exists to provide a mechanism for appropriate authorities to digitally sign the data file as a means to specify the file's authenticity and intended use. Through the use of an external (private) key, and a number computed from the entire contents of the BAG file, the signature section provides a robust mechanism for all subsequent users to verify the authenticity of the singing authority, and to verify that no changes have occurred to the file's content since the file was signed [7]. Any user modifications to a file's contents post-signature, or any transfer errors that occur post signature, will result in a validation failure. The signature section is not intended to prevent use of the data. Rather, the signature section is intended to ensure that once a BAG has been validated by an appropriate authority for some intended use, then any subsequent users of this BAG can verify that they recognize the authorizing agency and the specified intended use, and they can be certain that the contents of the BAG have not been modified since being signed. The BeeCrypt library [8] is used to implement the digital signature and encryption mechanisms used within a BAG.

Library Overview

To the extent possible, the design decisions made by the ONSWG were to adopt existing technologies, and use commonly accepted standards available within the public domain. This approach has been taken to simplify the efforts associated with developing and supporting the BAG concepts, and hopefully will promote community acceptance and use of the BAG format. Figure 2 provides a graphic overview of the libraries on which the BAG depends, showing the version currently used for each dependant library. Release 1.0.0 will include a sample program showing how to create a new BAG and a sample program showing how to read an existing BAG. For release 1.0.0 all libraries will be handled as static objects, conversion to .dlls and shared objects is an objective for a future release.

All data structures required for basic BAG file input/output operations, and all prototypes for functions intended to be exported from the BAG library are available in a single common header file: bag.h. The BAG APIs have been designed to isolate the application programmer from the intricacies of the data encapsulation mechanisms, metadata data encapsulation, and the mechanics of the digital security mechanisms. These high level APIs have been designed to expose only the parameters that are required to exchange the minimal necessary information between the application program and the library. Data values and parameters required for file encapsulation mechanics are maintained privately within the library and not exposed to the application program. This approach maintains the power of the technologies on which the BAG libraries depend, but through this level of isolation, simplifies the learning curve for the application programmer. The high level APIs are provided to exchange the information to and from internal form in standard engineering (SI) units. Wrapper code is planned to provide a unit and regression test capability. Several sample public domain gridded datasets are available to support the library unit test effort. A standard build process and directory structure has been established to enable a full build of all of the BAG components. At the time of this writing, the build process has been verified under both RedHat Enterprise Linux 3.0 (using gcc) and Windows XP (using Microsoft Visual Studio).

Prototype Status

A sample of the Reson 8125 data from the shallow survey 2005 common dataset was processed through CUBE [9] by UNH/CCOM to a node spacing of 0.5 meters, resulting in a grid of 3152 rows by 5686 columns on a UTM (zone 30) surface. The CUBE bathymetry and uncertainty surfaces were exported as ASCII X, Y, Z, U, and this dataset was used to create a prototype BAG. A sample program was written to read the ASCII X, Y, Z, U files and write these values to a prototype BAG file. Figure 3 shows the elevation (bathymetry) data read from this prototype BAG file and displayed as a 2-D sun shaded surface with the survey platform track lines superimposed. Figure 4 shows the vertical uncertainty values read from this prototype BAG file and displayed as a 2-D sun shaded surface. In figure 4, the sun shading is based on the bathymetric surface while the color assignment is based on the uncertainty values. Figures 3 and 4 have been produced using SAIC's SABER software updated to read and display BAGs.

The prototype version of the BAG library code is available in beta form, but several tasks remain to be finished before version 1.0.0 will be complete. The specific encoding implementation for the change-list needs to be completed. Library entry points need to be developed to support geographic and projection transformations in order to facilitate more flexible spatial file query operations. The library currently supports file access as specified in row and column coordinates. Updates are planned within these row/column operations to include support for conversion between row, column and the grid's X, Y coordinate system within the library. In addition, it is reasonable to expect that as the first several application programs are updated to support I/O operations on BAG files, additional BAG library needs will be identified.

Summary

The ONSWG is working towards completion of the first release of a gridded data format to support hydrographic data. The data format, referred to as the Bathymetric Attributed Grid (BAG) is a self contained, open data format for which both source code and documentation will be freely distributed. The format includes metadata, a grid of elevation data, a co-aligned grid of elevation uncertainty data, a change-list, and a digital signature/authentication block. The primary goal of the BAG format is to provide a data format with suitable capabilities to be used for access, distribution, and archive of gridded hydrographic information.

A prototype version of the library with basic overall functionality exists, and a prototype BAG file has been created. Commercial software packages are being updated to support the BAG format.

The members of the ONSWG are representatives from US NOAA, US NAVOCEANO, academia and industry. Membership and participation is cooperative and voluntary, but increasing community acceptance is an objective and community participation is actively encouraged. Contact and status information is available at: (http://www.opennavsurf.org).

References

[1] Smith, S. M. (2003). *The Navigation Surface: A Multipurpose Bathymetric Database*, Masters Thesis, Center for Coastal and Ocean Mapping & Joint Hydrographic Center, University of New Hampshire.

[2] Smith, S. M., Alexander, L., & Armstrong, A. A. (2002). '*The Navigation Surface: A New Database Approach to Creating Multiple Products from High-Density Surveys*', Int. Hydro. Review, Vol. 3, No. 2, pp. 12-26.

[3] ISO Standard 19115:2003 Geographic information – Meta-data (http://www.iso.org)

[4] ISO Proposed Standard 19139: Geographic information -- Metadata -- XML schema implementation (http://www.iso.org)

[5] HDF-V Documentation (http://hdf.ncsa.uiuc.edu/HDF5/doc/)

[6] NGA Geographic Translator documentation (http://earth-info.nga.mil/GandG/geotrans)

[7] Brennan, R.T. (and co-authors) "The Open Navigation Surface Project", Proceedings from Hydro 2005 San Diego Ca, March 2005

[8] Deblier, R. (2004) BeeCrypt API Documentation, http://users.telnet.be/robert.deblier/beecrypt.

[9] Calder, B.R. "Automated Statistical Processing of Multibeam Echosounder Data", Int. Hydro Review, January 2003.

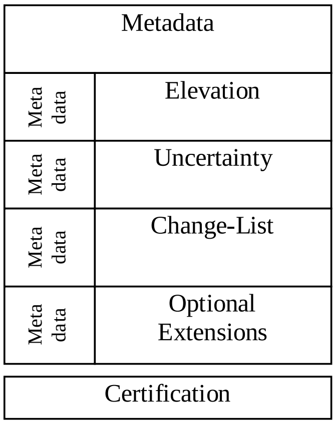

Fig. 1 – Logical structure of the BAG file format, with optional certification block appended.

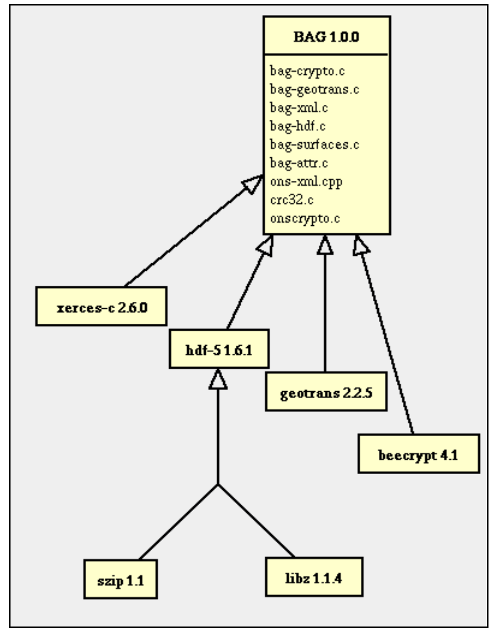

Fig. 2 – Library dependencies of the BAG library source files.

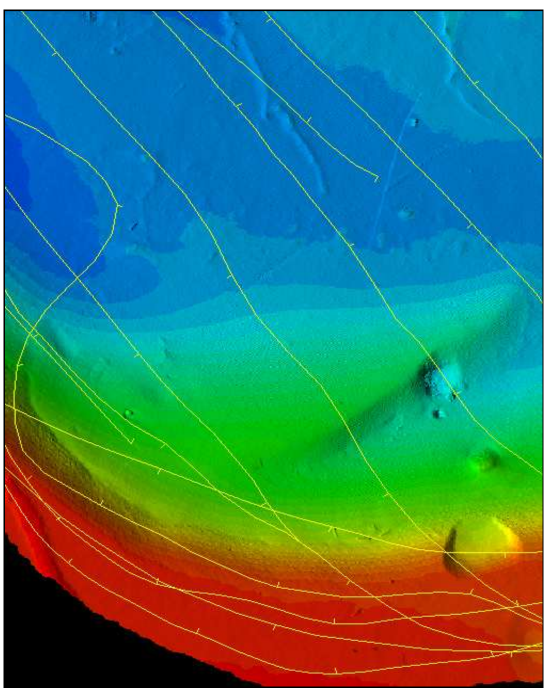

Fig. 3 – Sample BAG Elevation Surface of 8125 data from the Common Dataset. Color coded by elevation and shaded by slope with track lines overlaid.

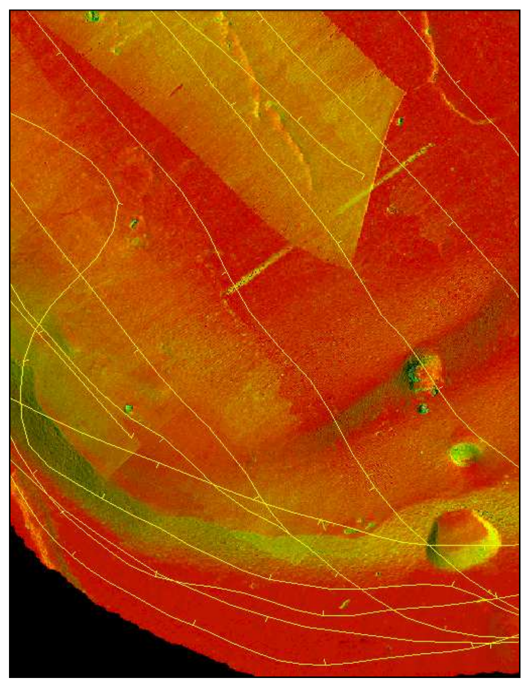

Fig. 4 – Sample BAG Uncertainty Surface of 8125 data from the Common Dataset. Color coded by uncertainty and shaded by bathymetric slope with track lines overlaid.