OPEN NAVIGATION SURFACE WORKING GROUP

MEETING SUMMARY

2011-04-27
US Hydro 2011
Tampa FL

DRAFT VERSION
2011-05-09

1 Introduction

This document details the result of the meeting held at the US Hydro 2011 conference in Tampa, FL on 2011-04-27 to discuss future directions for the Open Navigation Surface project, and in particular additions to the ONS structure to support hydrographic survey quality control, and compression of the HDF base files. The summary of all meetings and teleconferences of the Open Navigation Surface Working Group (ONSWG) can be obtained from the project's website, http://www.opennavsurf.org. For a list of participants, see section 0.

In the following, names people with action items are shown in **BOLD SMALL CAPS**; expected deadline release dates are shown in red. Sizes of variables are indicated by 'U' for unsigned, 'S' for signed, 'F' for floating-point, and a size in bits (e.g., U8 is an eight bit unsigned integer, F64 is a 64-bit (double precision) floating-point number). Data sizes are given in bytes (B) with the usual convention that the SI multipliers are taken to mean multiples of 2^{10}B (i.e., 1kB = 2^{10}B = 1024B). The acronym 'CR' means 'Candidate Release' (i.e., a release of the library for comments) and 'FR' means 'Full Release'.

2 Summary of Discussion

2.1 Compression of HDF5 base files

Compression in the underlying HDF5 V1.8 files that provide the supporting structure for BAG had been addressed previously, but without experience of the performance costs of applying compression on the fly had not been adopted for the last release. McDonald reported that they have now conducted preliminary experiments in the performance costs associated with compression of the HDF5 files, and find that although there is a performance cost in compression, the decompression appears to be affected only slightly, and therefore this might be a feasible option. The compression is variable, with levels from 1-9. The early experiments appear to show significant improvement in file size with compression on (compressing down to approximately 10% of the original size), and only modest increases in computational time for compression levels up to 5 (perhaps 30% extra time). After level 5, little gain was observed in the test set but there were significantly increased computational costs in compression.

It therefore appears that compression of the underlying files would be a significant improvement, and there was general agreement that we should move ahead with compression for the next release (1.4.0). The group discussed the implementation of the API change required for compression. Due to the variable level of compression that is allowed by HDF5, it seems likely that we should allow the user to control the compression so that they can adjust the run-time cost as required for their application, but that we should provide a default setting, with compression on, that we would recommend for most users, and have the API use this unless otherwise configured. (The read-back of compressed data is transparent.) The default settings have to be confirmed, and McDonald and Byrne recommended that this is something that would probably require further investigation before the release. (**ACTION: MCDONALD**.)

The group then discussed the concern that the code might have to decompress a large area of the file in order to get an idea of which areas are active (e.g., when the file is first opened). The option to add an 'activity' layer was discussed, meaning a simple binary indicator or scalar value at low resolution that indicates where the data is valid; the goal would be to provide low-resolution data that indicates which areas are active and therefore should be included in any queries. Whether we add this depends on how the HDF5 library decompresses the data; since the code provides a chunking mechanism to break the entire data area into smaller blocks, it might be decompressing only the chunks required, in which case this might not be as strongly required. This, however, requires confirmation. **(ACTION: MCDONALD.)**

Finally, Ducet brought up the question of whether a uniformly structured grid is appropriate for the data. That is, it would probably be more efficient (particularly in a sparse data area) if the whole grid were split into tiles with some organizing data structure (e.g., a quadtree) to provide structuring information on which areas are active, and their size, etc. This would potentially provide another level of compression by allowing sparse grids, and might speed up look-ups (at some computational cost). Since this would require a fairly major change to the structure of the files and possibly more API changes, the idea was deferred to a later release, most likely 1.6 or later.

2.2 Addition of Optional 'Hydrographic Quality Assurance/Control' Layers

Quintal provided further detail on a previous request to add an optional layer to support quality control and assurance metrics for multibeam echosounder surveys. These are mainly intended to support intermediate processing stages in the lifecycle of survey data, particularly the hand-off from field surveyor to office processing, where the extra metrics and information should better support inspection of the data and survey verification. There is no expectation that these layers will be part of an archive product both because they would generally be superfluous at that stage, and because they are expected to be voluminous.

The group generally agreed to the idea in principle (since it is an optional layer, and therefore low cost), but debated what structure the change should take. The main decision was in how to balance complexity of the options allowed against the volume of data and ease of access that this would imply. After some discussion it was agreed that the most likely structure would be to group all of the parameters that describe a node into one structure, and all of the parameters that describe a depth hypothesis into another, and make the optional layer an HDF5 group that could include an array of one or both of these entities. With this design, we avoid the difficulty of having some, but not all, of the parameters required for a particular purpose which might readily occur if each parameter was given its own layer, but allow the user to store summary statistics for the nodes without being concerned about hypotheses, and vice versa, to avoid data bloat or missing entries. The primary question is which parameters should be in each group. Quintal's previous proposal had included:

- Shoal depth at the node (with question of whether the shoal depth should be the shoalest sounding in any hypothesis at the node, or the shoalest depth in the selected hypothesis).
- Sounding density (with question of whether the density should include all the soundings that contributed to any hypothesis at the node, or just those that contributed to the selected hypothesis).
- Standard deviation of soundings (with question of whether this should be the standard deviation of all soundings that contributed to any hypothesis, or just those that contributed to the selected hypothesis; and whether a simple or scaled standard deviation should be recorded).
- Number of hypothesis on reconstruction depth reported by the processing algorithm (typically CUBE).
- Hypothesis strength for the selected hypothesis reported by the processing algorithm (typically CUBE).

so a logical choice would be to have the first three as a structure in one layer, and the second two as a structure in a second layer; this needs to be confirmed as acceptable for purpose, however. (**ACTION: QUINTAL.**) Determining the answers to the auxiliary questions posed above also needs to be done before the release, which will take some discussion. (**ACTION: CALDER.**)

2.3 Release of New Library Version

The group finally discussed the potential for new releases, and the approximate schedule for these releases. Byrne suggested that the modifications for addition of compression appeared to be relatively simple, but could have a large impact on performance and utility of the format, and therefore should be considered sooner rather than later. A proposed date of 2011-07-01 for a candidate release of 1.4.0 was agreed, with the usual two-week comment period before the release was finalized. Release 1.4.0 is expected to also incorporate modifications submitted since the last release that address compilation issues on certain platforms, and possibly some metadata modifications to support the on-going IHO S-102 development process.

The group agreed that the addition of optional new layers would likely take a little more work, and therefore suggested that a release date later in the year, most likely around 2011-10/11 would be appropriate, with the final date to be confirmed after release 1.4.0. Since the changes are relatively significant, this was expected to be release 1.5.0 rather than 1.4.1.

2.4 Updates on International Hydrographic Organization S-102 Efforts

Although unable to attend the meeting, Ladner communicated separately that the effort to adopt BAG for the core of the IHO S-102 standard is progressing. Some recommendations were made from the initial review, and a revised version will be submitted with a target date of 2011-06-01. A particular request was for a tiling scheme of some kind for storage efficiency (the arguments are likely similar to those in Sec. 2.1)

Ladner agreed to circulate a copy of the revised submission, based on BAG 1.3 (or the 1.4 prototype is possible), before the submission. (**ACTION: LADNER**).

3 Summary of Action Items and Dates

The following actions and dependencies were agreed:

Action	Person	Actions(s)	Section	Release
1	McDonald	Confirm compression performance; recommend default settings for release candidate; propose API for compression control on construction.	2.1	1.4.0
2	Quintal	Update definition of optional layers to include node-related and hypothesis-related information.	2.2	1.5.0
3	Calder	Facilitate discussion of QA/QC layer parameter definitions.	2.2	1.4.0
4	Ladner	Circulate draft of S-102 proposal with updates for BAG 1.3 compatibility and comments from country reviewers.	2.4	N/A

The release dates agreed were:

- Release Candidate 1 (BAG 1.4.0): 2011-07-01
- Full 1.4.0: 2011-07-15
- Release Candidate 1 (BAG 1.5.0): 2011-10 [Tentative Date]
- Full 1.5.0: 2011-11 [Tentative Date]

4 Participants

Shannon Byrne (SAIC Newport)
Webb McDonald (SAIC Newport)
Rebecca Quintal (SAIC Newport)
Brian Calder (CCOM/JHC)
Susan Sebastian (NAVO)
Jack Riley (NOAA)
Barry Gallagher (NOAA)
Moe Ducet (IVS3D[QPS])