

CSBWG Work Item H

Data Life Cycle

Submitted by Brian Calder (CCOM/JHC, University of New Hampshire)

SUMMARY

Executive Summary: This document provides details on activities conducted related to data lifecycle for volunteer bathymetry data, especially collation of suggested updates to DCDB for their data services, and merging the sub-group for Work Item F.

Action to be taken: See below

Related documents: None

Work Item Team Members: Calder, Masetti, Miles, Morissette, Solomon, Thompson, Zelenak

Work Item Background

The practice of collecting, archiving, and using volunteer data inevitably raises questions of how the data should be treated, what manipulations are required, and how the various parts of this enterprise interact. Although the CSBWG as a whole has addressed components of this through the development work on the B.12 recommendations document, as data has started to be collected the group concluded that there were a number of details that required further investigation. This work item is intended to provide input on these details, essentially providing clarification to the high-level description of the process in B.12. The work item is not intended to modify or update B.12, but rather to provide input to this process, making recommendations for required updates, and providing specific solutions for problems encountered in practice where possible.

Current Work Item Purpose

This work item aims to assess the components (theoretical, procedural, technical) for the data life cycle of volunteer data, with the ultimate goal of making recommendations for best practice for data collection, processing, and management. This includes recommendations for update to B.12 (e.g., expanded or modified metadata requirements), best practice for data processing (e.g., how to generate estimates of uncertainty or correct for unknown vertical offsets), and archive (e.g., additional services from DCDB to make use of data more efficient). The WG is pursuing these goals through a mixture of coordination meetings, directed development, and white papers.

Work Item Update

Work Item	Title	Priority <i>H-high</i> <i>M-med</i> <i>L-low</i>	Next milestone	Start Date	End Date	Status <i>P-planned O-ongoing C-completed S-Superseded</i>	Remarks
1	Recommendations for DCDB data access	H	CSBWG15 Intersessional	2024-06-01	2024-08-28	C	See below.
2	Consensus on workflow for developers	M	CSBWG16	2024-04-26	N/A	O	Proposed workflow developed; editing and consensus required.
3	Consensus on workflow for end-users	M	CSBWG16	2024-04-26	N/A	O	Proposed workflow developed as part of OpenVBI; editing and consensus required.
4	OpenVBI code review and development	H	CSBWG16	2024-04-26	N/A	O	Initial code base review required to provide stable basis for further work.
5	Vertical bias corrections in OpenVBI	L	CSBWG16	2024-04-26	N/A	P	Background research done, implementation required.
6	Uncertainty estimation in OpenVBI	L	CSBWG16	2024-04-26	N/A	P	Background research done, implementation required.
7	Merge activities of sub-group (work item) F.	M	CSBWG16	2024-08-01	N/A	O	Agreed in principle; actual adoption to be discussed at next WG meeting.

Progress Since Prior Meeting

The most significant work since CSBWG15 has been on item H.1 (recommendations to DCDB), with a tentative agreement on item H.7 (merging activities of work item F).

For work item H.1 (recommendations for DCDB data access), the team solicited opinions on modifications to the current DCDB access protocols and methods for volunteer data using a discussion topic on the OpenVBI GitHub site (<https://github.com/CCOMJHC/OpenVBI/discussions/1>). This provided a series of recommendations, including:

- Improved metadata structure for access to DCDB data through the AWS S3 bucket CSV files, rather than having to do multiple lookups through the GIS interface.
- Extended search capability in the primary API so that data can be retrieved with multiple search

criteria.

- Updates to allow for newer versions of the GeoJSON metadata schema (3.2.0)
- Use of a vertical or 3D coordinate reference system to allow for data referenced to the ellipsoid rather than an arbitrary datum (or none).

After some discussion with DCDB, many of these (potentially in slightly modified form) have been accepted as recommendations for the next development cycle, closing this work item.

For work item H.7 (merging activities of work item F), with the agreement of the work item F team lead (Brian Jensen) and the CSBWG management team, there has been preliminary discussion of dissolving work item F and merging the activities there into work item H. The rationale here is that work item F is focussed on software tools for HOs, but could be considered a sub-topic of work item H, which focussed on the wider data life cycle, including software for all users. The formal structure for merging the work items still needs to be discussed, and is likely a side-discussion for CSBWG16.

Reporting

N/A

Planned Work & Timeline

Further discussion on H.2-H.3 (recommendations for a nominal “standard” collection and processing workflow) is expected prior to CSBWG16, subject to scheduling. Progress on H.4-H.6 is expected by 2024-12-30, but certainly before CSBWG16 in 2025-03. All of these tasks, however, are likely to be ongoing work, since the community view of workflow structure and details, and new methods for integration of uncertainty and vertical bias estimation are likely to continue to develop over time. It is expected, however, that at least a first pass version should be available by mid-2025.

As mentioned above, we expect significant effort on H.7 through a side-meeting at CSBWG16, at which point further tasks may be added to the list above.

Issues/Risks/Concerns/Barriers

None

Proposed Changes to Work Item

None

Action to CSBWG

The CSBWG is requested to:

- A. **Note** the information provided;
- B. **Take** any other actions, as appropriate.