Work Item H

Data Life Cycle

Report to CSBWG16 NIWA, Wellington, NZ 27 March 2025

Brian Calder CCOM/JHC, University of New Hampshire





Work Item Team Members

- Brian Calder (CCOM/JHC)
- Clint Campbell (NCEI)
- Knut Hartmann (EOMAP)
- Brian Jensen (Raymarine)
- Giuseppe Masetti (GST)
- Brian Miles (CCOM/JHC)

- Guillaume Morissette (CIDCO)
- Mathieu Rondeau (CHS)
- Thierry Schmitt (SHOM)
- Shaul Solomon
- Colin Thomson (OFM)
- Emma Wise (Teledyne Marine)



Work Item Background & Purpose

International Hydrographic Organization

This work item is intended to provide input on [details of data treatment, processing, and management], 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.



IHO Progress Since CSBWG15

International Hydrographic Organization	Work Item	Title	Priority	Next milestone	Start Date	End Date	Status
	1	Recommendations for DCDB data access	Н	CSBWG15 Intersessional	2024-06-01	2024-08-28	С
	2	Consensus on workflow for developers	Μ	CSBWG16	2024-04-26	N/A	0
	3	Consensus on workflow for end-users	Μ	CSBWG16	2024-04-26	2025-06-30	0
	4	OpenVBI code review and development	Н	CSBWG16	2024-04-26	2025-03-30	0
	5	Vertical bias corrections in OpenVBI	L	CSBWG16	2024-04-26	2025-06-30	0
	6	Uncertainty estimation in OpenVBI	L	CSBWG16	2024-04-26	N/A	0
	7	Merge activities of sub-group (work item) F.	Μ	CSBWG16	2024-08-01	N/A	0

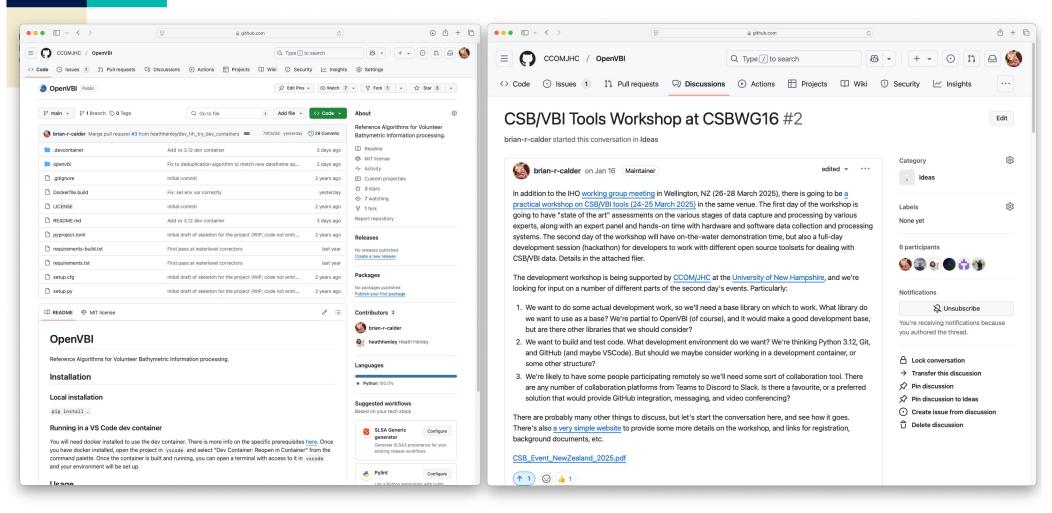


1HO 2&3: Consensus on Workflow

nization	ndori novem eventur eventur	1		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	Q		nthonyklemm / csb_processing_pipeline D Issues 휴 Pull requests 다 Discussio	ins 💿 Actions 🖽 Projects 💿 Security	Q Type (7) to search	
Volun	issues e 30(1)	100 - 100 -		- Second		Cs	b_processing_pipeline Public		③ Watch	1 ▼ V Fork 0 ▼ ☆ Star 1 ▼
Volun	e 29(2) e 29(1) ues →		Constan 2 1			۶º m	ain 👻 🕈 1 Branch 💿 0 Tags	Q Go to file	Add file 👻 <> Code 👻	About
Colle	ctions	1 Prove	SK 2 W	a star		🕲 a	nthonyklemm Update 2-csb_processing.py	6594cc1	3 days ago 🕚 20 Commits	No description, website, or topics provided.
Bathyme	ining	IL I E	Contraction of			C 1	-csb_scraper.py	Update and rename 1-csb_scraper_mod1.py to 1-c	sb_scr 4 days ago	🛱 Readme
Research and Tech All collect	nology			X		D 1	0-csb_differencing_visualizations.py	Update 10-csb_differencing_visualizations.py	4 days ago	4₫ CC0-1.0 license
All conect			175 HAS			L 2	-csb_processing.py	Update 2-csb_processing.py	3 days ago	☆ 1 star
Aims 8	About	XX	as The			L 3	-load_csb_to_duckdb.py	initial commit	last week	 1 watching
Editoria	Board		Carola Caro			L 4	-histograms_and_calibration_points.py	initial commit	last week	V 0 forks Report repository
Publication Pol De	Fig. 12 Map	of tide gauge stations and their respective are tations for which predictions are available. Pir				D 5	-apply_best_offsets_duckdb.py	Update and rename 5-apply_best_offsets_duckdb	2.py to 4 days ago	Bulance
Peer Review I Visit IHO			X			L 6	i-export_transits_to_gpkg_and_tiff_2.py	initial commit	last week	Releases No releases published
		1 martin		23		D 6	i-export_transits_to_gpkg_and_tiff_speed.py	Update 6-export_transits_to_gpkg_and_tiff_speed	l.py 4 days ago	
Subm Submit an	2 mars	384 4.79 0.4	8348			C 7	-Outlier_model_PMM_Imputation.py	Update and rename 7-Outlier_model_PMM_Imputa	tion_it 4 days ago	Packages
Author Gui		Contra -				L 8	-insert_outlier_flags_in_duckdb.py	Update and rename 10-insert_outlier_flags_in_duc	kdb.py 4 days ago	No packages published
	es	La start		<u> </u>		D 9	-csb_export_all_points_create_geotiff.py	Update and rename 11-csb_export_all_points_crea	te_geo 4 days ago	Languages
		BE	ST.			DL	ICENSE	Initial commit	last week	• Python 100.0%
90			Ring Ala	1 m		Ľ R	EADME.md	Update README.md	3 days ago	
- L			8200	<u> </u>		۵Þ	uild_leaderboard_and_tracklines	Update and rename 7-build_leaderboard_and_trac	klines 4 days ago	
	\backslash	BELO -	Contraction of the second seco	5 1 5		C •	ount number outliers duckdb.py	initial commit	last week	
58440/ihr	$\langle \rangle$		A Real	8150		C c	sb_differencing_visualizations_work.py	initial commit	last week	
22			And a second	A195		C c	sb_export_all_points.py	initial commit	last week	
org/10.58440/				- T		C d	lashboard 1.py	Update and rename 8-dashboard 1.py to dashboar	d 1.py 4 days ago	
						₽ R	EADME 5 CC0-1.0 license		0	
، ttps://doi.o	Fig. 13 Bounding	Connection of the second	posed on the global influence may e dataset to be reduced.	 D. Highlighted are the Voronoi EPI < E 		an alg	alysis/correction, uncertainty estimation, v gorithms, and exporting data as vessel-tra			
	out IHR Volume 30(1)	From volunteer ping to Hydrography Program		ity Cite Share Down	load Previous article Next article	le Th	ere are other helper scripts, and some scr	pts for dashboard/leaderboard creation as well.		



4: OpenVBI Code Review and Development





IHO 5: Vertical Bias Corrections in OpenVBI

International
Hydrographic
Organization

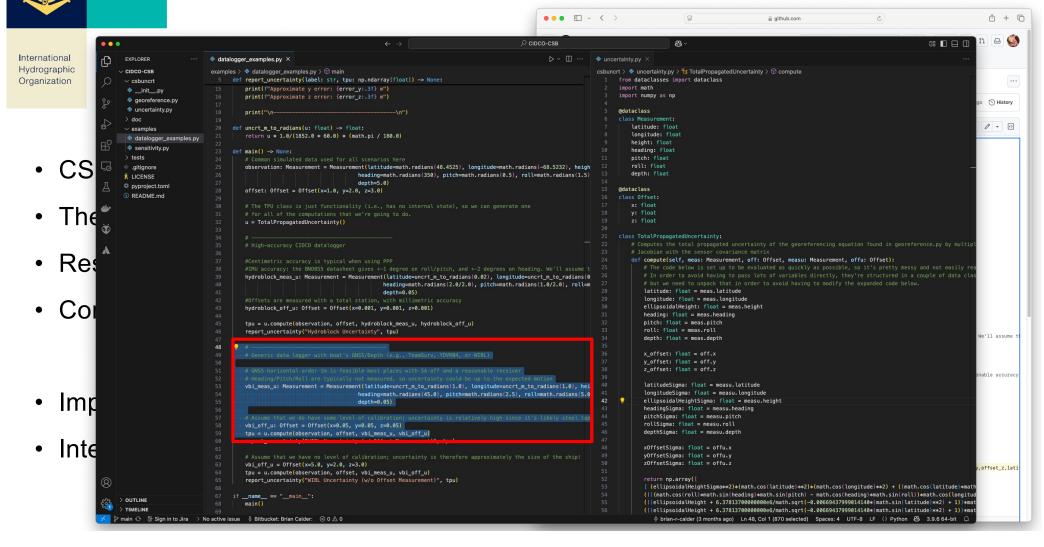
•••

	OPENVBI	ope	nvbi > examples > 🆆 prep-simple.py >	openvbi	> corrections > waterlevel > noaa > 🌳initpy >
)	\sim openvbi		from openvbi.adaptors.ydvr import load_data	108	class ZoneTides(Waterlevel):
	\sim adaptors		from openvbi.filters.thresholding import shoaler_than, deeper_than		<pre>definit(self, zone_shapefile: str) -> None:</pre>
	v leansurv.py		from openvbi.filters.timeslot import before_time, after_time	110	<pre>selfzones = geopandas.read_file(zone_shapefile)</pre>
	🍦 ydvr.py		from openvbi.corrections.waterlevel.noaa import ZoneTides	111	<pre>super()init()</pre>
	✓ core		from openvbi.adaptors.dcdb import write_geojson	112	
				113	def preload(self, dataset: Dataset) -> None:
	>pycache		# Pull in data from YachtDevices raw binary file, and convert to depths assuming NMEA2000		# Spatial join to determine which polygon each observation is in (and hence which
	🔷initpy		data = load_data('00030095.DAT')		annotated_pts = geopandas.sjoin(dataset.depths, selfzones, how='inner', predica
	🕏 interpolation.py				# List of all required stations
	🝨 metadata.py	10			<pre>selfstations = annotated_pts['ControlStn'].unique()</pre>
	observations.py	11			<pre>selftides = dict()</pre>
		12			# For each station, we need to determine the time bounds of the observations affe
	statistics.py	13			<pre># call the CO-OPS API to get the waterlevel corrections; these are stored until i</pre>
	🍨 timebase.py	14	· · · · · · · · · · · · · · · · · · ·		<pre># do the corrections for some/all of the observations</pre>
	🅏 types.py				for station in selfstations:
	\sim corrections	16	그는 이번 방법에 두 사람이 없는 것이 가지 두 것 같아요. 그는 것은 것이 두 것이 같아요.		<pre>station_times = annotated_pts[annotated_pts['ControlStn'] == station]['t']</pre>
	>pycache	17			min_time = station_times.min() - 10*60
	<pre>> waterlevel</pre>	18 19			<pre>max_time = station_times.max() + 10*60</pre>
		20			<pre>raw_levels = get_noaa_station(station, min_time, max_time)</pre>
	>pycache	20			<pre>corrections = InterpTable(['dz',])</pre>
	\checkmark noaa	21		128	<pre>for n in range(len(raw_levels)):</pre>
	>pycache	23			<pre>corrections.add_point(raw_levels['t'][n].timestamp(), 'dz', raw_levels['v</pre>
	<initpy< td=""><td>24</td><td></td><td></td><td><pre>selftides[station] = { 'min': min_time, 'max': max_time, 'raw': raw_levels,</pre></td></initpy<>	24			<pre>selftides[station] = { 'min': min_time, 'max': max_time, 'raw': raw_levels,</pre>
	<pre></pre>	25			
		26		132 133	<pre>def _execute(self, observations: geopandas.GeoDataFrame) -> geopandas.GeoDataFrame:</pre>
	🕏initpy	27		133	# Spatial join to determine which polygon each observation is in (and hence which
	\sim examples			134	<pre>annotated_pts = geopandas.sjoin(observations, selfzones, how='inner', predicate for station, data in selftides.items():</pre>
	🕏 basic_processing.py	29		135	station_points = annotated_pts[annotated_pts['ControlStn'] == station]
	🕏 dedup.py	30		130	<pre>lut_times = station_points['t'] - station_points['ATCorr']*60</pre>
	🗬 metadata.py		# Correct for waterlevel using NOAA zoned tides and live API for waterlevels	138	<pre>wl_corr = data['table'].interpolate(['dz',], lut_times)[0]</pre>
			<pre>zone_tide_wl = ZoneTides('NOAA_tide_zones/tide_zone_polygons_new_WGS84_merge.shp')</pre>	139	<pre>station_points['z'] -= station_points['RR']*wl_corr</pre>
	prep-instrumented.py				observations.loc[station_points.index, 'z'] = station_points['z']
	🅏 prep-simple.py		data = zone_tide_wl.correct(data)	140	return observations
	\checkmark filters	35		142	
	>pycache		# Generate B.12-format GeoJSON output	143	def _metadata(self, meta: md.Metadata) -> None:
	🕏initpy				meta.addProcessingAction(md.ProcessingType.VERTREDUCTION, None,
	deduplicate.py				reference='ChartDatum',
					datum='MLLW',
	🅏 thresholding.py				<pre>method='Observed Waterlevel',</pre>
	🍨 timeslot.py				algorithm='OpenVBI',
	🕏initpy				version=version(),
					<pre>model = f'NOAA Zoned Tides with stations {selfstations}')</pre>
	OUTLINE				
	TIMELINE				

83

+ 0







IHO CSB Tools Workshop Development Day

International		OpenVBI Workflow Tool	OpenVBI Workflow Tool
Hydrographic	Inputs	Results	_ trustedNode
Organization	Input Directory Extras/OpenVBI/ExampleData	Successful Files	providerOrganizationName UNHJHC
	Output Directory /Users/brc/temp		providerEmail wibl@ccom.unh.edu
	Workflow		uniqueVesselID
	Loader WIBL		convention GeoJSON CSB 3.1
	Writer DCDB GeoJson		dataLicense CC0 1.0
	Depth Mesage Depth (NMEA2000)		providerLogger WIBL
	Metadata File tDay/rukuwai2-metadata.json Create		providerLoggerVersion 1.3/1.1.0/1.0.1
	Run Workflow		navigationCRS
			verticalReferenceOfDepth Transducer
			vesselPositionReferencePoint GNSS
			platform
			type Research vessel
			name Rukuwai II
		Failed Files	length
			IDType MMSI
			IDNumber 512000771
			sensors
			soundSpeedDocumented
			positionOffsetsDocumented
			dataProcessed
			contributorComments
			uniquelD
			Export
			Filename velopmentDay/rukuwai2.json Choose
			Actions
			Validate Export Quit

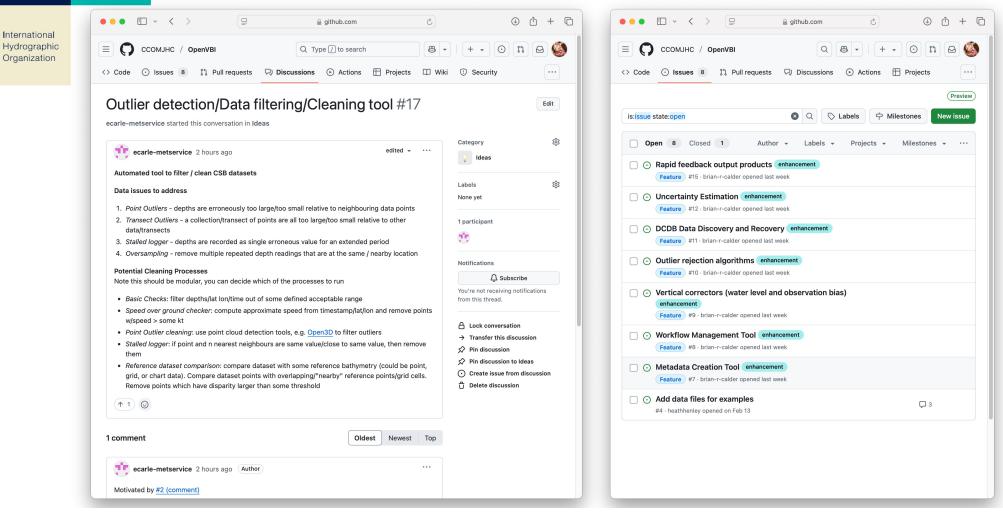


CSB Tools Workshop Development Day

openvl	bi > adaptors > 🗇 csb_rest_api_adapter.py > 😤 CSBRestApil.oader > /workspaces/OpenVBl/tests/unit/test_processing.py
	import os
	from pathlib import Path
	import requests
	from openvbi.adaptors.dcdb import CSVLoader
	from openvbi.core.observations import Dataset
	from openvbi.adaptors import Loader
	<pre>class CSBRestApiLoader(Loader):</pre>
10 11	definit(self, output_path: Path, query_result_loader: Loader = CSVLoader(), default_bucket_path: str = ' <u>https://noaa-dcdb-bathymetry-pds.s3.amazonaws.com/csb/csv</u> ') -> None:
11	<pre>self.query_result_loader = query_result_loader cold_subtry ext = other active to the set of th</pre>
12	<pre>selfoutput_path = output_path selfbucket_path = resolve_bucket_path(default_bucket_path)</pre>
	settouckec_path = resource_oucket_path(uerautc_oucket_path)
	def suffix(self) \rightarrow str:
	def load(self, filename: str) -> Dataset:
19	object_path = get_object_path(selfbucket_path, filename)
	file_path = download_object(object_path, filename, selfoutput_path)
	return selfquery_result_loader.load(file_path)
	def resolve_bucket_path(default_bucket_path: str) -> str:
	<pre>bucket_path = os.environ.get('0PENVBI_S3_BUCKET')</pre>
	if bucket_path is None:
	return default_bucket_path
	return bucket_path
28 29	def get_object_path(bucket_path: str, filename: str) -> str:
30	file me_without_extension: st, r = filemem.replace(".tar.gz", "")
	time_code = file_mae_without_extension.split("", 2][0]
	year = time_code(0:4)
	month = time_code[4:6]
	day = time_code[6:8]
	return f"{bucket_path}/{year}/{month}/{day}/{file_name_without_extension}_pointData.csv"
	def download_object(object_path: str, filename: str, output_path: Path) -> Path:
	<pre>if output_path.exists() and output_path.is_dir():</pre>
	<pre>print (f"{output_path} is valid")</pre>
	<pre>print (f"{output_path} is not a valid path")</pre>
	with requests.get(object_path, stream=True) as response:
	status_code = response.status_code
	if status de le 200:
	raise RuntimeError(f"object request status did not return 200, returned {status_code}")
	with open(output_path.joinpath(filename), 'w') as output_file:
	for line in response.iter_lines():
	if line:
	output_file.write(line)
	return output_path
55 56	



IHO CSB Tools Workshop Development Day







- ✓✓✓
- Updates to code in OpenVBI as consequence of CSB Tools Workshop development day (see separate report)
 - Architecture review of OpenVBI base
 - Capture of development ideas from CSB Tools Workshop
 - Discussion with (former) Work Item F members to update task list for next year (at CSBWG16)



IHO "Official" Work Item H Tasks

	H-1	Review, recommend, and document the data flow of standard processing stages for data capture	М	Present to CSBWG for discussion	2023	Dec 2024	0	CCOM/B Calder	
✓	H-2	Review current metadata structure for CSB data for completeness, encoding methods, validation mechanisms, and support for end-user database mapping	М	Recommend modifications	2023	March 2025	Р	CCOM/B Calder	
	Н-3	Consider potential extensions of the current GeoJSON	М	Recommend modifications	2023	Dec 2024	0	CCOM/B Calder	

\checkmark	H-4	Investigate possible extensions to the DCDB S3 data store and OGC services	Н	Consolidate suggestions for future development at DCDB.	2023	May 2024	С	CCOM/B Calder	
	H-5	Establish recommendations for one or more use cases of CSB, and associated guidelines for data products	М	Collaborate with other working groups preparing specialized products (e.g., for HO use)	2023	Ongoing	0	CCOM/B Calder	



- Note the information provided in the written report & here
- Provide feedback on tasks, development direction as necessary