

# Data Life Cycle (Work Item H)

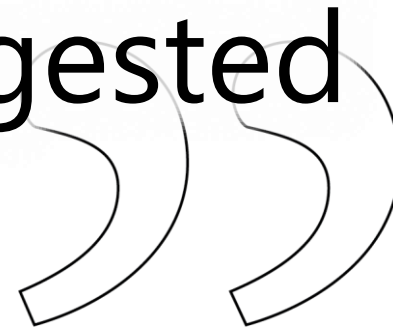
## IHO Crowdsourced Bathymetry Working Group



Brian Calder, Giuseppe Masetti, Brian Miles, Guillaume Morissette, Shaul Solomon,  
Colin Thompson, Georgie Zelenak



Clarify all aspects of the CSB data cycle and capture known issues, requirements and suggested enhancements.



1. Capture Processing
2. Metadata
3. Data Format Extension
4. Data Access
5. Use Cases
6. Availability

1. Data Access
2. Metadata
3. Capture Processing
4. Data Format Extension
5. Use Cases
6. Availability

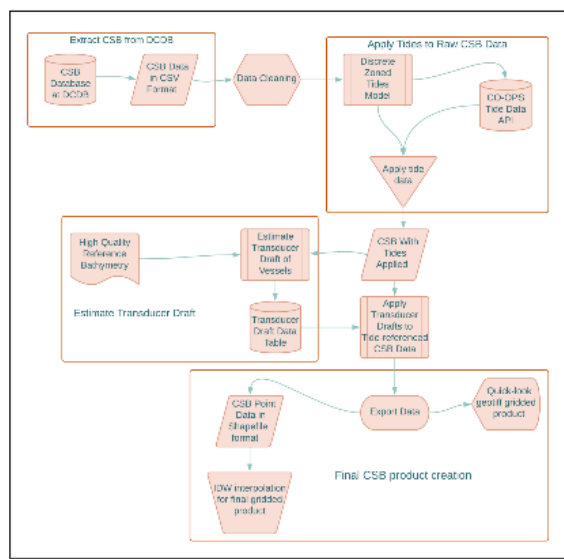


Figure 2: General workflow diagram for CSB processing script

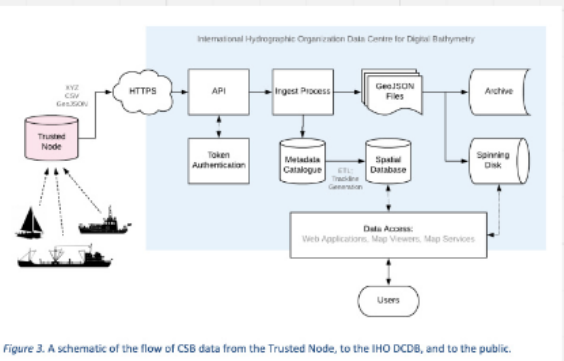
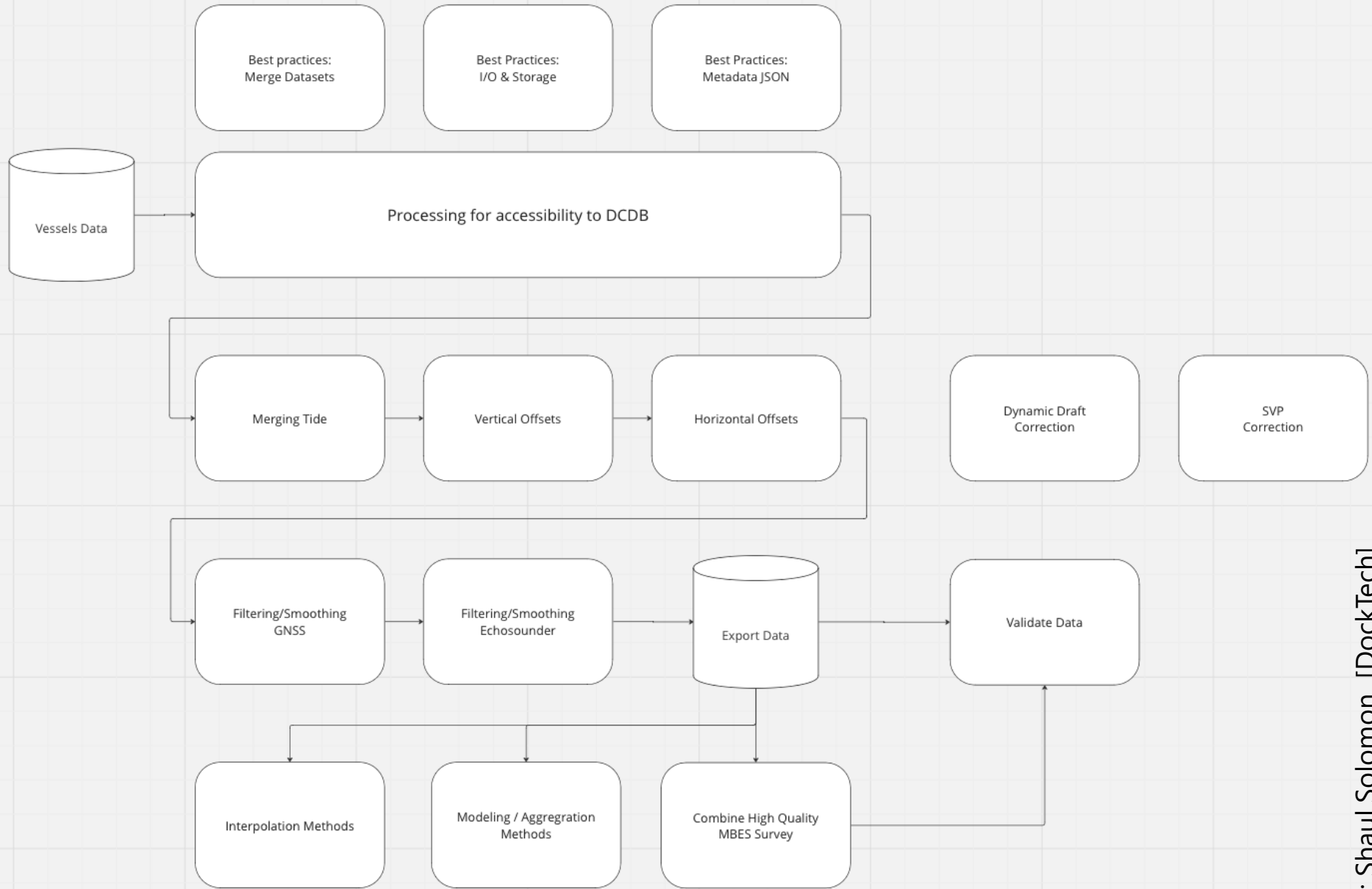


Figure 3. A schematic of the flow of CSB data from the Trusted Node, to the IHO DCDB, and to the public.



# Metadata

Review the current **metadata** structure for CSB data for completeness, *encoding* methods, *validation mechanisms*, and support for end-user database mapping (e.g., for groups of collectors and other aggregations), and recommend modifications.

```
noaa_b12_v3_0_0_suggested.json — NOAA (git: main)
1 {
2   "type": "FeatureCollection",
3   "crs": {
4     "horizontal": {
5       "type": "EPSG",
6       "value": 8252
7     },
8     "vertical": "Transducer"
9   },
10  "properties": {
11    "providerContactPoint": {
12      "orgName": "Example Cruises Inc",
13      "email": "support@example.com",
14      "logger": "Rose Point ECS",
15      "loggerVersion": "1.0"
16    },
17    "convention": "GeoJSON CSB 3.0",
18    "dataLicense": "CC0 1.0",
19    "platform": {
20      "uniqueID": "EXAMPLE-f8c469f8-df38-11e5-b86d-9a79f06e9478",
21      "type": "Ship",
22      "name": "USCGC Healy",
23      "length": 128,
24      "IDType": "MMSI",
25      "IDNumber": 908338000,
26      "sensors": [],
27      "correctors": {
28        "positionReferencePoint": "GNSS",
29        "soundSpeedDocumented": true,
30        "positionOffsetsDocumented": true,
31        "dataProcessed": true,
32        "motionOffsetsApplied": true,
33        "draftApplied": true
34      }
35    },
36    "algorithms": [
37      {
38        "name": "deduplicate",
39        "Params": ""
40      }
41    ],
42    "lineage": [
43      {
44        "type": "CRSTransformation",
45        "timestamp": "2021-11-22T16:10:09.346821",
46        "detail": {
47          "origin": "EPSG:4326",
48          "destination": "EPSG:8252",
49          "method": "GeoTrans"
50        }
51      },
52      {
53        "type": "TimeInterpolation",
54        "timestamp": "2021-11-22T16:10:09.346821".
55      }
56    ]
57  }
58 }
```

```
b12_v3_1_0_example.json — IHO (git: main)
1 {
2   "type": "FeatureCollection",
3   "crs": {
4     "type": "name",
5     "properties": {
6       "name": "EPSG:4326"
7     }
8   },
9   "properties": {
10    "trustedNode": {
11      "providerOrganizationName": "Sea-ID",
12      "providerEmail": "support@sea-id.org",
13      "uniqueVesselID": "SEAID-e8c469f8-df38-11e5-b86d-9a79f06e9478",
14      "convention": "GeoJSON CSB 3.1",
15      "dataLicense": "CC0 1.0",
16      "providerLogger": "Rose Point ECS",
17      "providerLoggerVersion": "1.0",
18      "navigationCRS": "EPSG:4326",
19      "verticalReferenceOfDepth": "Transducer",
20      "vesselPositionReferencePoint": "GNSS"
21    },
22    "platform": {
23      "uniqueID": "SEAID-e8c469f8-df38-11e5-b86d-9a79f06e9478",
24      "type": "Private vessel",
25      "name": "White Rose of Drachs",
26      "length": 65,
27      "IDType": "MMSI",
28      "IDNumber": "369958000",
29      "sensors": [],
30      "soundSpeedDocumented": true,
31      "positionOffsetsDocumented": true,
32      "dataProcessed": true,
33      "contributorComments": "On 2022-03-08, at 20:30 UTC, the echo sounder lost bottom tracking
34    },
35    "processing": [
36      {
37        "type": "CRSChange",
38        "timestamp": "2023-02-14T02:00:00.0000Z",
39        "original": "EPSG:4326",
40        "destination": "EPSG:8252",
41        "method": "GeoTrans"
42      },
43      {
44        "type": "VerticalReduction",
45        "timestamp": "2023-02-14T03:00:00.0000Z",
46        "reference": "ChartDatum",
47        "datum": "CANNORTH2016v1HyVSEP_NAD83v6_CD",
48        "method": "Predicted WaterLevel",
49        "model": "CANNORTH2016v1HyVSEP_NAD83v6_CD"
50      },
51      {
52        "type": "GNSS",
53        "timestamp": "2023-02-14T04:00:00.0000Z",
54        "algorithm": "RTKLib",
55        "version": "1.2.0"
56      }
57    ]
58 }
```

```
b12_v3_1_0_example.json — IHO (git: main)
1 {
2   "type": "FeatureCollection",
3   "crs": {
4     "type": "name",
5     "properties": {
6       "name": "EPSG:4326"
7     }
8   },
9   "properties": {
10    "trustedNode": {
11      "providerOrganizationName": "Sea-ID",
12      "providerEmail": "support@sea-id.org",
13      "uniqueVesselID": "SEAID-e8c469f8-df38-11e5-b86d-9a79f06e9478",
14      "convention": "GeoJSON CSB 3.1",
15      "dataLicense": "CC0 1.0",
16      "providerLogger": "Rose Point ECS",
17      "providerLoggerVersion": "1.0",
18      "navigationCRS": "EPSG:4326",
19      "verticalReferenceOfDepth": "Transducer",
20      "vesselPositionReferencePoint": "GNSS"
21    },
22    "platform": {
23      "uniqueID": "SEAID-e8c469f8-df38-11e5-b86d-9a79f06e9478",
24      "type": "Private vessel",
25      "name": "White Rose of Drachs",
26      "length": 65,
27      "IDType": "MMSI",
28      "IDNumber": "369958000",
29      "sensors": [{}],
30      "soundSpeedDocumented": true,
31      "positionOffsetsDocumented": true,
32      "dataProcessed": true,
33      "contributorComments": "On 2022-03-08, at 20:30 UTC, the echo sounder lost bottom tracking"
34    },
35    "processing": [
36      {
37        "type": "CRSChange",
38        "timestamp": "2023-02-14T02:00:00.0000Z",
39        "original": "EPSG:4326",
40        "destination": "EPSG:8252",
41        "method": "GeoTrans"
42      },
43      {
44        "type": "VerticalReduction",
45        "timestamp": "2023-02-14T03:00:00.0000Z",
46        "reference": "ChartDatum",
47        "datum": "CANNORTH2016v1HyVSEP_NAD83v6_CD",
48        "method": "Predicted Waterlevel",
49        "model": "CANNORTH2016v1HyVSEP_NAD83v6_CD"
50      },
51      {
52        "type": "GNSS",
53        "timestamp": "2023-02-14T04:00:00.0000Z",
54        "algorithm": "RTKLib",
55        "version": "1.2.0"
56      }
57    ]
58  }
59 }
Line: 1 | JSON | Tab Size: 4 |
```

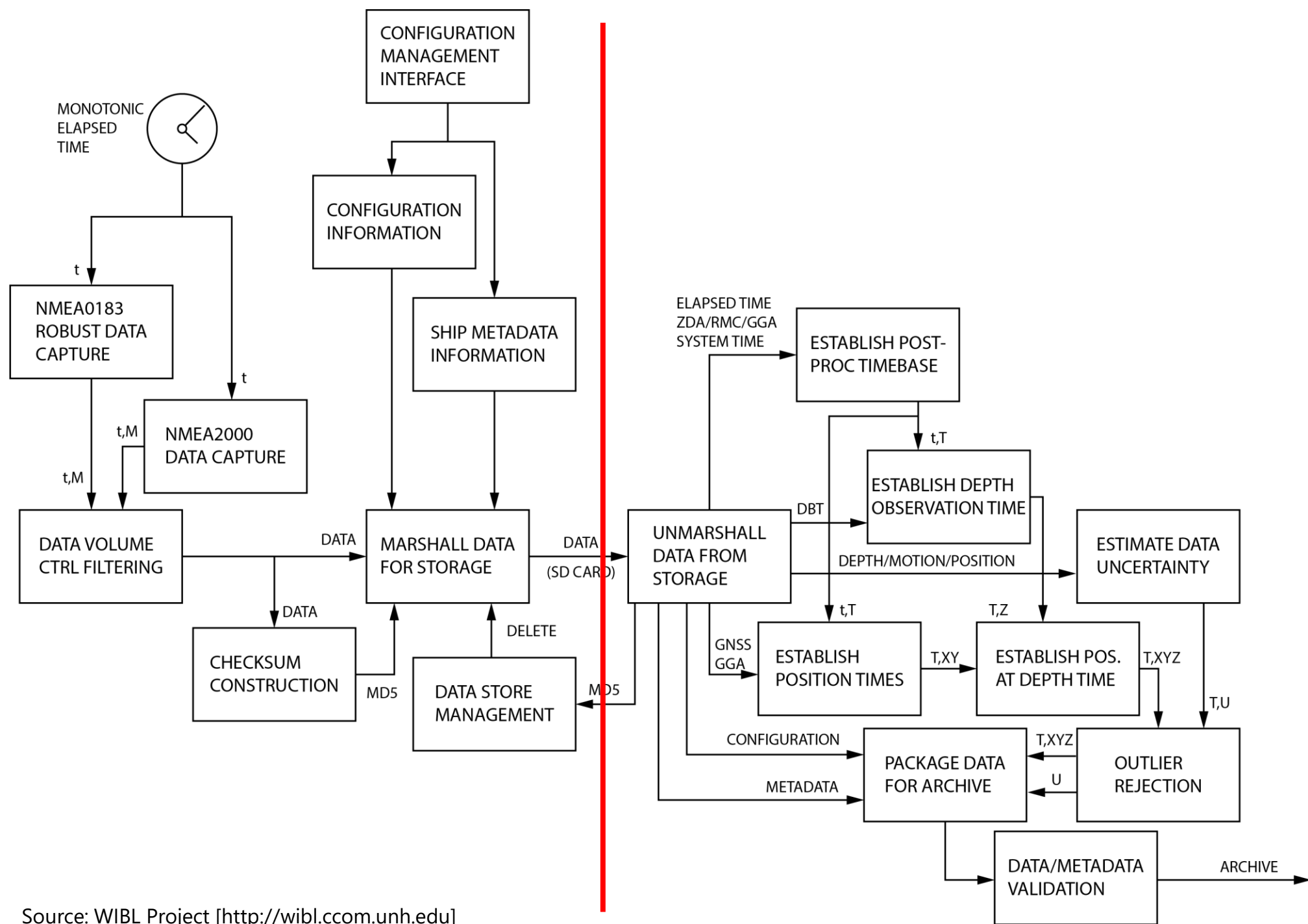
```
b12_v3_2_0-BETA_example.json — IHO (git: main)
1 {
2   "type": "FeatureCollection",
3   "crs": {
4     "type": "name",
5     "properties": {
6       "name": "EPSG:4326"
7     }
8   },
9   "properties": {
10    "trustedNodePlatform": {
11      "providerOrganizationName": "Sea-ID",
12      "providerEmail": "support@sea-id.org",
13      "uniqueVesselID": "SEAID-e8c469f8-df38-11e5-b86d-9a79f06e9478",
14      "type": "Private vessel",
15      "name": "White Rose of Drachs",
16      "IDType": "MMSI",
17      "IDNumber": "369958000"
18    },
19    "observationCollection": {
20      "trustedNode": {
21        "convention": "GeoJSON CSB 3.2",
22        "dataLicense": "CC0 1.0",
23        "providerLogger": "Rose Point ECS",
24        "providerLoggerVersion": "1.0",
25        "navigationCRS": "EPSG:4326",
26        "verticalReferenceOfDepth": "Transducer",
27        "vesselPositionReferencePoint": "GNSS"
28      },
29      "platform": {
30        "length": 65,
31        "sensorDescriptions": [{}],
32        "soundSpeedDocumented": true,
33        "positionOffsetsDocumented": true,
34        "dataProcessed": true,
35        "contributorComments": "On 2022-03-08, at 20:30 UTC, the echo sounder lost bottom tracking"
36      },
37      "processing": [
38        {
39          "type": "CRSChange",
40          "timestamp": "2023-02-14T02:00:00.0000Z",
41          "original": "EPSG:4326",
42          "destination": "EPSG:8252",
43          "method": "GeoTrans"
44        },
45        {
46          "type": "VerticalReduction",
47          "timestamp": "2023-02-14T03:00:00.0000Z",
48          "reference": "ChartDatum",
49          "datum": "CANNORTH2016v1HyVSEP_NAD83v6_CD",
50          "method": "Predicted Waterlevel",
51          "model": "CANNORTH2016v1HyVSEP_NAD83v6_CD"
52        },
53        {
54          "type": "GNSS",
55          "timestamp": "2023-02-14T04:00:00.0000Z"
56        }
57      ]
58    }
59  }
60 }
Line: 1 | JSON | Tab Size: 4 |
```

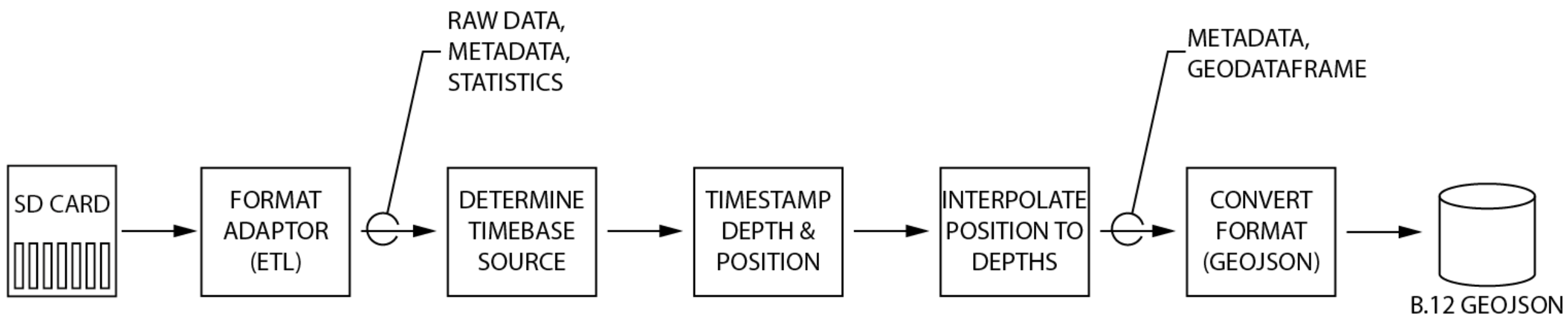


1. Collect suggested modifications in a draft specification
2. Test the implementation with example data
3. Iterate as required to beat out any deficiencies
4. Bring proposal to CSBWG/DCDB for adoption

# Capture Processing

Review, recommend, and document (in conjunction with the available open source code development projects) the data flow of standard processing stages for **data capture**, including extensions for uncertainty estimation and associated **metadata**/data format extensions.





The image shows a code editor window with a file explorer on the left and a code editor on the right. The file explorer shows a project structure for 'OPENVBI' with subfolders like 'adaptors', 'core', 'corrections', 'examples', and 'filters'. The 'examples' folder is expanded, showing 'basic\_processing.py' selected. The code editor displays the content of 'basic\_processing.py', which is a Python script for processing YachtDevices data. The script includes imports for 'uuid', 'load\_data', 'md', 'write\_geojson', and 'write\_csv\_json'. It defines constants for 'dcdb\_id', 'provider\_id', and 'provider\_email'. The main logic involves loading data from a file, populating metadata, and writing the output to a GeoJSON file and a CSV file. The script is 40 lines long.

```
openvbi > examples > basic_processing.py > ...
1 import uuid
2 from openvbi.adaptors.ydvr import load_data
3 import openvbi.core.metadata as md
4 from openvbi.adaptors.dcdb import write_geojson, write_csv_json
5
6 # In order to fill out the metadata, we need a DCDB-style Trusted Node identification string,
7 # and some core data for the provider and e-mail.
8 dcdb_id = 'OPNVBI'
9 provider_id = 'OpenVBI'
10 provider_email = 'hello@openvbi.org'
11
12 # In practice, you'd want to have a fixed unique identifier for the particular logger that you're
13 # working on. Since this isn't available in the file being loaded, you'd need to look up the
14 # filename (or other information) in a database to translate to the unique identifier. For example
15 # purposes here, we just make up a random identifier
16 unique_id = dcdb_id + '-' + str(uuid.uuid4())
17
18 # Load from YachtDevices, and convert into a dataframe
19 data = load_data('/Users/brc/Projects-Extras/OpenVBI/ExampleData/00030095.DAT')
20
21 # Since there isn't a lot of information on the logger from the datafile, we need to populate the
22 # metadata separately. We focus here on the mandatory metadata. In practice, you would probably want
23 # to look up the unique identifier in a database of loggers, and pick out the appropriate metadata (or
24 # potentially a pre-formatted object).
25 data.meta = md.Metadata()
26 data.meta.setProviderID(provider_id, provider_email)
27 data.meta.setIdentifiers(unique_id, 'YDVR', '1.0')
28 data.meta.setReferencing(md.VerticalReference.TRANSDUCER, md.VerticalReferencePosition.GNSS)
29 data.meta.setVesselID(md.VesselIdentifier.MMSI, "00000000")
30
31 # YachtDevices is a NMEA2000 device, so we convert 'Depth' messages for depth information
32 data.generate_observations('Depth')
33
34 # Finally, write the output to a DCDB-compatible, IHO CSBWG B.12 GeoJSON file. Note that the
35 # keyword parameters here are passed on to json.dump() directly, so you can control the output
36 # a little more directly. Using "indent" here gives something that's more verbose but easier
37 # to read.
38 write_geojson(data, '00030095.json', indent=2)
39 write_csv_json(data, '00030095_copy', indent=2)
40
```

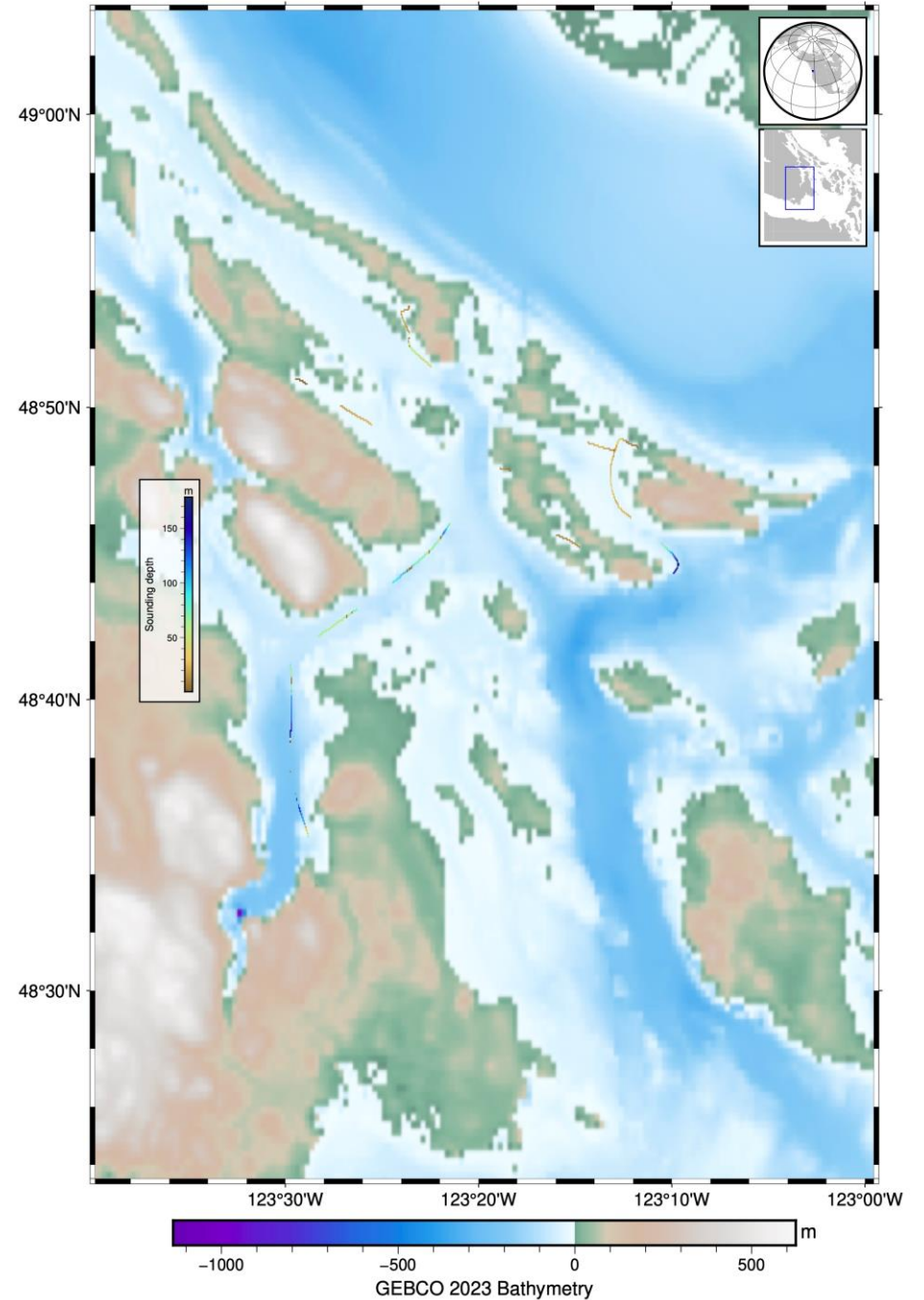
# Use Cases

Establish recommendations for one or more use cases of CSB, and associated guidelines for data products (e.g., interpretation of quality, known problems, required processing, etc.). Collaborate with other task groups preparing specialized products (e.g., for HO use).



Wireless  
Inexpensive  
Bathymetry  
Logger

Soundings from 'Test observer'



# Availability

Promote and contribute to an open source reference implementation of available algorithms for all the processing steps, preferably with a permissive license such as MIT or CC0.

## Adapters

TeamSurv  
YachtDevices  
Generic ASCII  
DCDB [I/O]

## Core

Dataset Management  
Time-based Interpolation  
Timestamping/Depth Construction  
Metadata Management  
Packet & General Statistics  
Uncertainty

## Corrections

NOAA Waterlevels  
CHS Waterlevels  
Vertical Bias

## Filters

Depth [Shoal/Deep]  
Time [Early/Late]  
Deduplication  
Outliers

## Examples

ELT → GeoJSON  
Metadata Addition  
Algorithmic Filtering

OpenVBI

EXPLORER

- OPENVBI
  - openvbi
    - core
      - \_\_init\_\_.py
      - interpolation.py
      - metadata.py
      - observations.py
      - statistics.py
      - timebase.py
      - types.py
    - corrections
      - \_\_pycache\_\_
    - waterlevel
      - \_\_pycache\_\_
    - noaa
      - \_\_pycache\_\_
      - \_\_init\_\_.py
      - \_\_init\_\_.py
      - \_\_init\_\_.py
    - examples
      - basic\_processing.py
      - dedup.py
      - metadata.py
      - prep.py
    - filters
      - \_\_pycache\_\_
      - \_\_init\_\_.py
      - deduplicate.py
      - thresholding.py**
      - timeslot.py
      - \_\_init\_\_.py
    - openvbi.egg-info
    - .gitignore
    - Dockerfile.build
    - LICENSE
    - pyproject.toml
  - OUTLINE
  - TIMELINE

thresholding.py

```
openvbi > filters > thresholding.py > shoaler_than > _metadata
27 from openvbi.core.observations import Dataset
28 import openvbi.core.metadata as md
29 from openvbi.filters import Filter
30 from openvbi import version
31
32 # Remove any points that are shoaler than the threshold specified
33 class shoaler_than(Filter):
34     def __init__(self, threshold: float) -> None:
35         self._threshold = threshold
36         super().__init__()
37
38     def _execute(self, dataset: geopandas.GeoDataFrame) -> geopandas.GeoDataFrame:
39         self.n_inputs = len(dataset)
40         dataset = dataset[dataset['z'] > self._threshold]
41         self.n_outputs = len(dataset)
42         return dataset
43
44     def _metadata(self, meta: md.Metadata) -> None:
45         meta.addProcessingAction(md.ProcessingType.ALGORITHM, None,
46                                 name='ShoalDepth Filter',
47                                 source='OpenVBI',
48                                 version=version(),
49                                 comment=f'After filtering, total {self.n_outputs} points selected from {self.n_inputs}.')
50
51 # Remove any points that are deeper than the threshold specified
52 class deeper_than(Filter):
53     def __init__(self, threshold: float) -> None:
54         self._threshold = threshold
55         super().__init__()
56
57     def _execute(self, dataset: geopandas.GeoDataFrame) -> geopandas.GeoDataFrame:
58         self.n_inputs = len(dataset)
59         dataset = dataset[dataset['z'] < self._threshold]
60         self.n_outputs = len(dataset)
61         return dataset
62
63     def _metadata(self, meta: md.Metadata) -> None:
64         meta.addProcessingAction(md.ProcessingType.ALGORITHM, None,
65                                 name='DeepDepth Filter',
66                                 source='OpenVBI',
67                                 version=version(),
68                                 comment=f'After filtering, total {self.n_outputs} points selected from {self.n_outputs}.')
```

Ln 46, Col 29 Spaces: 4 UTF-8 LF Python 3.12.2 ('openvbi': conda)

The screenshot shows a code editor with a file explorer on the left, a main code editor in the center, and a preview pane on the right. The file explorer shows a project named 'CIDCO-UNCERT...' with subfolders 'doc' and 'src'. The 'src' folder contains 'datalogger\_examples.py', 'georeference.py', 'test.py', and 'uncertainty.py'. The main code editor displays the contents of 'datalogger\_examples.py', which is a Python script for calculating uncertainty. The script includes comments and code for calculating uncertainty from GNSS readings and IMU/sonar data. The preview pane shows a README.md file with a section titled 'Uncertainty' and example values for the script's parameters.

```
src > datalogger_examples.py > ...
39
40 print("Hydroblock Uncertainty")
41 u = uncertainty.TotalPropagatedUncertainty()
42
43 tpu = u.compute(latitude, longitude, ellipsoidal_height, heading, pitch, roll,
44
45 print(tpu)
46 error_x=math.sqrt(tpu[0][0])
47 error_y=math.sqrt(tpu[1][1])
48 error_z=math.sqrt(tpu[2][2])
49
50 print(f"Approximate x error: {error_x:.3f} m")
51 print(f"Approximate y error: {error_y:.3f} m")
52 print(f"Approximate z error: {error_z:.3f} m")
53
54 print("-----")
55
56 # -----
57 # Low-cost WIBL datalogger
58
59 #10m-ish error at equator for uncorrected GNSS readings
60 latitude_sigma = 0.0001
61 longitude_sigma = 0.0001
62 ellipsoidal_height_sigma = 1
63
64 #FIXME: what kind of IMU figures?
65 #IMU accuracy: the BNO055 datasheet gives +-1 degree on roll/pitch, and +
66 heading_sigma = math.radians(2/2)
67 pitch_sigma = math.radians(1/2)
68 roll_sigma = math.radians(1/2)
69
70 #FIXME: what kind of sonar ?
71 #Imagenex 852 datasheet advertises a 1mm range resolution...which is prob
72 depth_sigma = 0.001
73
74 #FIXME: those are purely fictional
75 #Offsets are unknown
76 offset_x_sigma = 1 #Length of the boat
77 offset_y_sigma = 1 #Half the width
78 offset_z_sigma = 1 #Twice the depth of the boat assuming transducer fixed
79
80
81 print("WIBL Uncertainty")
82 tpu = u.compute(latitude, longitude, ellipsoidal_height, heading, pitch, roll,
83
84 print(tpu)
85 error_x=math.sqrt(tpu[0][0])
86 error_y=math.sqrt(tpu[1][1])
87 error_z=math.sqrt(tpu[2][2])
88
89 print(f"Approximate x error: {error_x:.3f} m")
```

```
# offset_z = 3 //meters

import georeference
g = georeference.Georeference()
result =
g.compute(latitude, longitude, ellipsoidal_height, heading, pitch, r
```

## Uncertainty

Get the uncertainty associated with an XYZ position (in the ECEF frame) of a depth sounding (in the WGS84 frame)

```
# Example values
# latitude = math.radians(48.4525) //decimal degrees
# longitude = math.radians(-68.5232) //decimal degrees
# ellipsoidal_height = -28 //meters
# heading = math.radians(350)
# pitch = math.radians(0.5)
# roll = math.radians(1.5)
# depth = 5 //meters
# offset_x = 1 //meters
# offset_y = 2 //meters
# offset_z = 3 //meters
# latitude_sigma = 0.00001
# longitude_sigma = 0.0001
# ellipsoidal_height_sigma = 1
# heading_sigma = 1
# pitch_sigma = 1
# roll_sigma = 1
# depth_sigma = 0.5
# offset_x_sigma = 0.0000001
# offset_y_sigma = 0.0000001
# offset_z_sigma = 0.0000001

import uncertainty
u = uncertainty.TotalPropagatedUncertainty()
result =
u.compute(latitude, longitude, ellipsoidal_height, heading, pitch, r
```

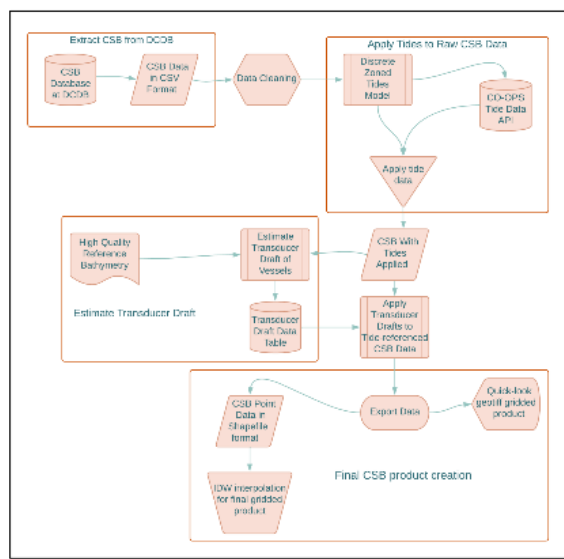


Figure 2: General workflow diagram for CSB processing script

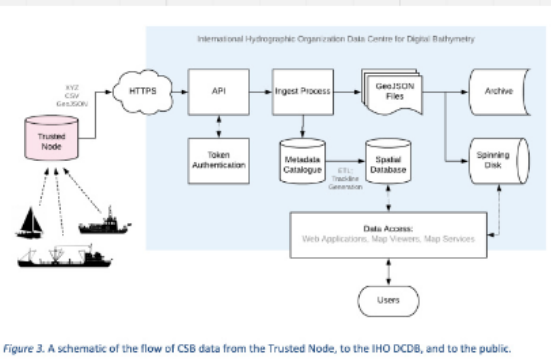
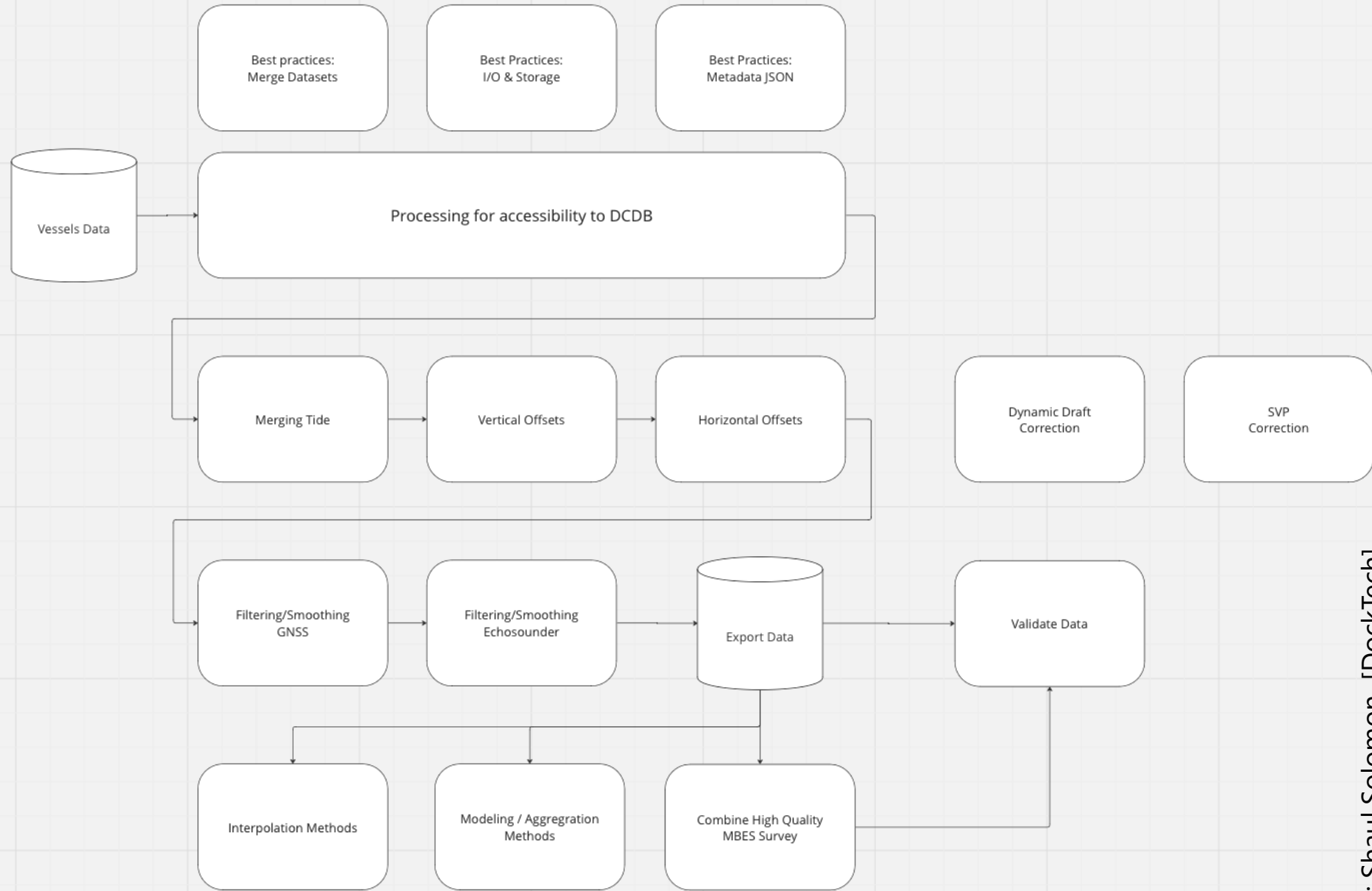


Figure 3. A schematic of the flow of CSB data from the Trusted Node, to the IHO DCDB, and to the public.



```
openvbi > examples > prep-simple.py > ...
1  import copy
2  from openvbi.adaptors.ydvr import load_data
3  from openvbi.filters.thresholding import shoaler_than, deeper_than
4  from openvbi.filters.timeslot import before_time, after_time
5  from openvbi.corrections.waterlevel.noaa import SingleStation, ZoneTides
6  from openvbi.adaptors.dcdb import write_geojson
7
8  # Pull in data from YachtDevices raw binary file, and convert to depths assuming NMEA2000 data
9  data = load_data('00030095.DAT')
10 data.generate_observations('Depth')
11
12 # Calibrate acceptable depth and time windows for data (note that these are simply for
13 # demonstration purposes: this cuts off a lot of valid data!)
14 min_depth = data.depths['z'].min()
15 max_depth = data.depths['z'].max()
16 depth_range = max_depth - min_depth
17 shoal_threshold = min_depth + depth_range/3.0
18 deep_threshold = max_depth - depth_range/3.0
19
20 min_time = data.depths['t'].min() + 10.0*60.0 # Remove first ten minutes
21 max_time = data.depths['t'].max() - 10.0*60.0 # Remove last ten minutes
22
23 # Generate filters for shoal/deep depth, and early/late time
24 shoal = shoaler_than(shoal_threshold)
25 deep = deeper_than(deep_threshold)
26 early = before_time(min_time)
27 late = after_time(max_time)
28
29 # Filter for depth window, and time window
30 data = early.Execute(late.Execute(deep.Execute(shoal.Execute(data))))
31
32 # Correct for waterlevel using NOAA zoned tides and live API for waterlevels
33 zone_tide_wl = ZoneTides('tide_zone_polygons_new_WGS84_merge.shp')
34 zone_tide_wl.preload(data)
35 zone_tide_wl.correct(data)
36
37 # Generate B.12-format GeoJSON output
38 write_geojson(data, '00030095.json', indent=2)
39
```

main Sign in to Jira No active issue Bitbucket: Brian Calder: 0 0 0 Ln 39, Col 1 Spaces: 4 UTF-8 LF Python 3.12.2 ('openvbi': conda)

# Upcoming Work

- Recommendations for DCDB data access
- Consensus on workflows for:
  - Developers (capture, conversion)
  - End-users (data discovery, understanding, re-use)
- OpenVBI code review & development
- Vertical bias correction in OpenVBI
- Uncertainty estimation in OpenVBI

