Steps Towards a True Crowd, Collecting Bathymetry via Electronic Navigation Systems

Authors

Adam Reed, LT/NOAA. Assistant Coordinator, Integrated Oceans and Coastal Mapping, NOAA. adam.reed@noaa.gov
NOAA Office of Coast Survey
1315 East West Highway
Silver Spring, MD 20910
301-713-2702 x114

Anthony Klemm, LT/NOAA. Marine Charting Division, NOAA. anthony.r.klemm@noaa.gov
NOAA Ship Thomas Jefferson
439 W. York Street
Norfolk, VA 23510
757-647-0187

Evan Robertson, National Centers for Environmental Information, NOAA. Evan.robertson@noaa.gov
NOAA National Centers for Environmental Information
325 Broadway
Boulder, CO 80305
303-497-4613

Abstract

Strained resources and budgets for hydrographic offices have resulted in increased scrutiny for the acquisition of hydrographic data and the clear need to acquire data from non-traditional sources to complete their charting mandates. Crowdsourcing bathymetric data has been listed as a priority issue by the International Hydrographic Organization (IHO) as an important source contributing to the future of nautical charting.

The IHO's Crowdsourced Bathymetry Working Group in partnership with NOAA's Office of Coast Survey and Rose Point Navigation Systems established a citizen science program to provide publicly available crowdsourced bathymetry data. This follows a successful 2015 pilot project to harvest crowdsourced bathymetry from Electronic Navigation Systems. The program incorporates efforts to minimize cost barriers to entry for contribution, expand the vessel demographics participating, and present a publicly available crowd solution. In this paper we will discuss the methods used for data collection, the location and formats in which the data is available, what quality has been observed, and the lessons learned.

Background

The Office of Coast Survey has designated over 500,000 square nautical miles of "navigationally significant" area that the United States is responsible for charting. At the current rate of survey it will take approximately 300 years to survey this area using modern survey methods, not including any areas that require re-survey due to change [Office of Coast Survey, 2016]. It is clear that additional sources of data and new methods are required to fulfil the hydrographic office's area of responsibility.

The International Hydrographic Organization (IHO) published the aggregate quality of every member nation's reported hydrographic coverage as of 2014. The United States reported that in depths shoaler than 200 meters, only 10% of US waters are surveyed adequately. In depths deeper than 200 meters, less than 1% of US waters are surveyed adequately. Cognizant of the limited resources the hydrographic programs of the world face, the IHO has listed crowdsourcing bathymetry as a priority for future solutions for the world's hydrographic offices [IHO, 2014]. In 2015, the Inter-Regional Coordination Committee (IRCC) of the IHO established a dedicated crowdsourced bathymetry working group to develop policy and guidance for the IHO's position on crowdsourced bathymetry programs [IRCC7, 2015]. The IRCC additionally named the IHO Data Centre for Digital Bathymetry (DCDB) as the designated repository for crowdsourced bathymetry data hosted by the IHO to be available for public use, whether for hydrographic or other purposes. This repository is co-located with NOAA's National Centers for Environmental Information (NCEI) in Colorado.

Current Examples Crowdsourced Bathymetry

The first formal system of crowdsourcing maritime information for marine charting in the US was established in 1963 with the Cooperative Charting Program between NOAA's Office of Coast Survey and the United States Power Squadron (USPS) [Office of Coast Survey, 2016]. This program provided members of the USPS the ability to submit point value data by postal service to cartographers for chart application. Since 1963 the Cooperative Charting Program has continuously demonstrated that volunteered geographic information from mariners is often the only way a chart may be updated, as the majority of updates from the program occur in areas that Office of Coast Survey has not surveyed since the program's inception.

The major drawback to this, and similar programs currently in existence, is that they are only able to provide singular point data. Often these data are hazards and features, not necessarily bathymetry. To update charted soundings on a large scale, it would be necessary to receive bathymetric information as the primary data contribution from the crowd. When the hydrographic community refers to "crowdsourced bathymetry," the implication is that it is bathymetry as a primary data stream, and in wholesale quantities. There are currently several efforts harvesting bathymetry around the world from crowd sources, several operated by industry and several available in a public domain.

The largest model of crowdsourcing bathymetry involves the contributor purchasing or renting a combination of equipment and/or software, commonly referred to as a data logger, from a company to install on their vessel. This data logger will then harvest position and bathymetry data and relay it to the company, which then may share data from all contributors. The major difference between these options typically comes down to the price of the equipment and the level of processing performed on the data by the company. Some manufacturers provide a completely quality controlled, regular grid of bathymetry based on uncertainty estimation, such as Olex [Olex, 2017]. Other entities provide the raw data once it passes basic quality control checks to ensure consistency, such as Sea-ID [IHO, 2017].

All of these current solutions pose a number of barriers to entry, costs that must be overcome to contribute crowdsourced bathymetry. The cost of the data logger is invariably necessary, and in the case of buying and installing added physical equipment can dissuade potential contributors. Crowdsourcing efforts can often target certain demographics of mariner, which leads to a selective crowd contributing data. For example, one can largely expect contributions to follow specific traffic trends by targeting only tug vessels. The goal of crowdsourcing is to have as much data available as possible from as many sources as possible, a true crowd.

Project Scope

In an attempt to surpass several cost barriers to entry, we explored the avenue of pre-installed equipment aboard vessels to act as data loggers. Electronic Charting Systems (ECS) are an ideal data logger, since they are systems already designed to take inputs of position, time, and depth and record them in log files. In 2016, we performed a pilot project aboard the NOAA research vessel *Bay Hydro II* to extract latitude, longitude, and time from logged GGA NMEA records and bathymetry from logged DPT NMEA records. The pilot project displayed base level bathymetric data could be acquired from Electronic Charting Systems and be used as a valid tool to provide meaningful hydrographic information [Reed, 2016].

This project begins where the pilot project left off, turning the concept study into a supply of crowdsource bathymetry. It involved a collaborative effort between Rose Point Systems, NOAA’s Office of Coast Survey, NOAA’s National Centers for Environmental Information (NCEI), and the IHO’s Crowdsourced Bathymetry Working Group.

Crowdsourced Data

Data

The four primary fields highlighted as required data for each sounding are Latitude, Longitude, depth, and time.

Latitude, Longitude, and time are supplied to Rose Point’s software via input from the vessel’s Global Navigation Satellite System (GNSS) receivers. For the majority of vessels, their GNSS system will use the Global Positioning System (GPS) as the primary source of positioning. Depth is supplied by the transducer input into the ECS.

Metadata

All metadata supplied by the contributor was considered optional. This was an intentional decision designed to promote as large and varied of a crowd as possible. By constraining the submissions to as few required inputs by a contributor as possible, we can encourage contributors who were unable or unwilling to provide specific metadata to still supply base bathymetry data. Contributors would be able to remain anonymous or supply as little identity information as desired, alleviating the concerns that users could track their specific vessel. Likewise, contributors who were unwilling to put effort into measuring offsets could still provide meaningful data.

In this model, we provide the end user to sort and discover data sets with the ability to filter by metadata supplied, if any. For example, users who want to differentiate between commercial traffic and residential traffic can discount vessels that fail to provide information on vessel type, and could choose to download only data with metadata of their interested traffic type. Users interested in charting may be interested in all vessels, regardless of type, and could choose not to filter the data based on this metadata.

It was still necessary to be able for a user to correlate contributions together if made by a single contributor. This allows users to associate several submissions as coming from the same platform and address uncertainty or errors associated with those contributions. To maintain the option of anonymity, it was necessary to develop a unique identifier that was independent from other vessel identification methods, such as IMO number. To solve this issue, Rose Point supplies a unique vessel identification generated by the software as a Universally Unique Identifier (UUID) that each installation of the software will generate separately. This unique identifier is present in all data files submitted by the vessel, so that an end user can use the identifier to correlate any vessel specific biases present. Data regarding the identifier are not submitted beyond the vessel's ECS, preventing the end user from tracing the information back to the vessel using the unique identifier alone.

Format

Rose Point's ECS products capture and log all raw sensor data NMEA strings as a part of normal operational functionality. For the pilot project aboard NOAA vessel *Bay Hydro II*, these log flies were kept in their raw NMEA format and parsed accordingly to extract the relevant information. Latitude, Longitude, and time were extracted from GGA NMEA strings, and depth was extracted from DPT NMEA strings, shown in Figure 1.

```
000000400891: $GPDPT,3.4,0.0*50
000000401000: $GPGGA,115541,3819.9158,N,07627.4480,W,2,9,1.1,4,M,,M,,*61
000000401889: $GPDPT,3.4,0.0*50
000000402030: $GPGGA,115542,3819.9158,N,07627.4480,W,2,9,1.1,4,M,,M,,*62
000000402888: $GPDPT,3.4,0.0*50
000000403012: $GPGGA,115543,3819.9158,N,07627.4480,W,2,9,1.1,4,M,,M,,*63
```

Figure 1: Raw NMEA string examples for the GPS and fathometer input.
GGA strings indicate GPS position, and DPT strings indicate fathometer depth.

There are several limitations to keeping raw NMEA strings as the product of when converting this idea to crowdsource. The two NMEA inputs used, GGA and DPT, are provided by two separate sensors that do not share a common clock. This equates to a significant time uncertainty when attempting to pair depth to a corresponding position measurement. For the pilot project, we were able to assess the relative uncertainty by using the characteristics of speed and network latency aboard the vessel. When supplied by a crowd, it would be completely unreasonable to assume that this would be able to be done equally across all contributors. Additionally, the chosen repository for crowdsourced bathymetry, the IHO DCDB, was unable to accept NMEA outputs into the current format of the database. We required the data to be submitted in vector point format.

Rose Point solved both issues by offering the ability to log sensor data in a specifically designated file designed for submission directly to the IHO DCDB. These files contain latitude, longitude, time, and depth as data values in a CSV format, shown in Figure 2. Metadata supplied by the crowdsourced contributor is contained in JavaScript Object Notation (JSON) format headers for each file, shown in Figure 3. These file format options were chosen in collaboration between Rose Point and the IHO DCDB to satisfy both size constraints for storage and transmittal of the files, and the ability for the database to store and extract data from the files for web interface and visualization. Ultimately, the database will be able to convert files to more GIS friendly options upon download request by an end.


```
47.666520,-122.098525,21.49,20161017T234638Z
47.666518,-122.098525,11.98,20161017T234739Z
47.666517,-122.098527,14.63,20161017T234839Z
47.666515,-122.098527,17.16,20161017T234935Z
47.666490,-122.098472,19.72,20161017T235044Z
47.666505,-122.098522,20.18,20161017T235141Z
47.666477,-122.098507,20.42,20161017T235241Z
47.666512,-122.098432,20.63,20161017T235342Z
47.666497,-122.098417,20.33,20161017T235443Z
47.666512,-122.098470,20.33,20161017T235548Z
47.666507,-122.098490,20.57,20161017T235644Z
47.666533,-122.098453,20.33,20161017T235832Z
47.666575,-122.098445,20.33,20161018T000042Z
47.666585,-122.098460,20.21,20161018T000236Z
```


Figure 2: Crowdsourced bathymetry file containing Latitude, Longitude, depth, and GNSS time recorded in CSV format.


```
{
    "platform":
    {
        "uniqueID":"ROSEP-e8c669f8-df38-16e5-b86d-9a79606e9478",
        "type": "Ship",
        "name": "SS Dinghy",
        "length": 65,
        "lengthUnitOfMeasure": "meters",
        "IDType": "IMO",
        "IDNumber": "1008140"
    }
}
```


Figure 3: JSON formatted metadata headers. Rose Point supplies the unique vessel identifier for all submissions, and any additional keys and values are only present when the user enters metadata in the application interface.

The crowdsourced bathymetry log files are submitted to the IHO DCDB via HTTP post. The HTTP Post occurs when the mariner initiates a chart catalog update or an ECS software update. This ensures that vessel bandwidth is utilized when the vessel is expecting higher bandwidth use and the navigational performance of the ECS is unimpeded. Log files are stored locally on the ECS machine until the mariner performs the update.

Correctors

Tidal

Tidal corrections are not applied to any data crowdsourced by Rose Point Systems. The IHO DCDB requires that all archived crowdsourced data be un-corrected for tide. The primary rationale behind this decision was to preserve the data quality. It is virtually impossible to undo improperly corrected tide if the processing information is not recorded with the data. For typical hydrographic surveys or systematic ocean mapping projects, this is reasonable to require. For crowdsourcing, it becomes a much more daunting task, as it requires extensive quality assurance on all data in the archive. It could be reasonable to assume that the amount of submissions during the early stages of a crowdsourced bathymetry project would be small enough to allow for such assurance. As a crowdsourced project grows to thousands of submissions per day, or more, this becomes a far more challenging task.

Additionally, the IHO represents a number of member states that utilize different datums for their charting products. Hosting data referenced to Mean Lower Low Water (MLLW), the datum used by NOAA’s nautical charts, would add additional complexity to entities that vertically reference in another datum.

Transducer offset

The horizontal and vertical offsets between the transducer and the GNSS location also pose a significant challenge for crowdsourced bathymetry. The offset must first be known to the mariner and must then be provided into the data logger. For several data logger manufacturers, this is a step that is solved during installation with someone assisting the mariner. For an ECS model of crowdsourcing, we must rely on the mariner to input it as metadata. As an added complication, the data logger must have heading information to correctly apply the proper X and Y coordinates for the offset. This poses a significant issue, as heading information is not necessarily an input into Rose Point. Rose Point’s software is capable of receiving heading input, but the vessel will need a digital gyrocompass in order to provide this information. This configuration is limited based on vessel’s size and bridge equipment.

Rose Point will apply an offset if input by the mariner. The value of an offset is captured in metadata fields in addition to a binary flag indicating whether offset information is present. The heading issue currently remains unresolved and will be addressed in future updates.

Time synchronization

The observations of position and depth information are supplied by two separate sensors and must be paired into point value data. The frequency at which each sensor supplies data can be set individually on the sensor. Rose Point's software logs both sensor data and can supply corresponding time stamps. However, only GNSS data has time inherently in the data. Depth data does not receive a time stamp until it is received by the ECS software. This creates uncertainty due to connection latency, time synchronization, and observation rates.

Of these, frequency of observations is the easiest issue to address. The closest observations in time are paired into points by Rose Point's software when it is logged. Fathometers and GNSS receivers are connected to the ECS computer using predominately USB or serial connections, often with adaptors. Assuming this configuration, we can expect latency between fathometer observations and data reaching the ECS software to be considerably low. These can also be assumed to be equal amounts of latency unless there is considerable discrepancy between configurations of the fathometer or GNSS unit. This issue is currently unresolved, and will need to be addressed by subsequent uncertainty evaluations.

Archival

IHO Data Centre for Digital Bathymetry

The first concept of an ideal pipeline for this project conceived by the IHO Crowdsourced Bathymetry Working Group utilized a crowd of contributors submitting observations directly to the data repository. End users would then access the repository to download bathymetry when needed, and the data would be converted to the desired format upon download. This model became known as the direct contributor model, shown in Figure 4.

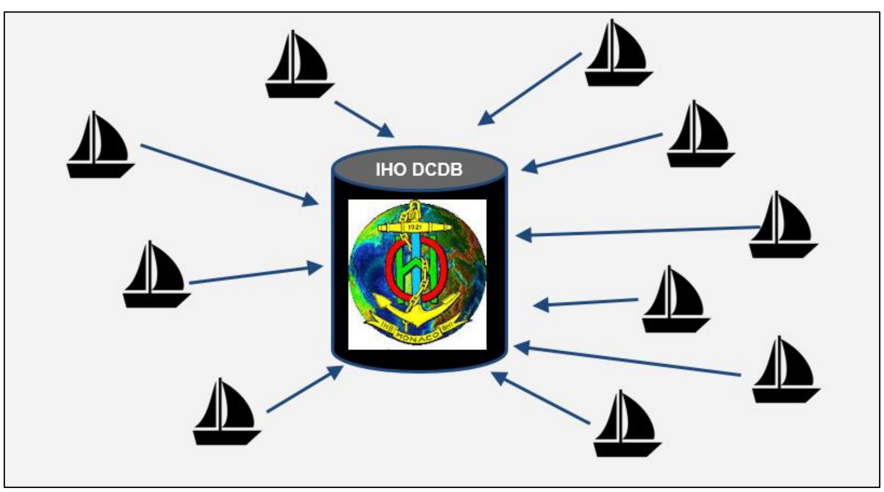


Figure 4: Individual Contributor Model of crowdsourcing bathymetry [Robertson, 2016].

The quality assessment of these data would fall on the end user in this model. This was a conscious choice based on the consideration that not all end users for these data would be expected to be a member of a hydrographic office, or that the crowdsourced bathymetry would always be used for nautical charting purposes. Users with different end products would undoubtedly have different quality control criteria. The working group also indicated that best practices would allow end users access to the primary raw data. Application of correctors that was performed incorrectly would invalidate these data from becoming a reliable source. Therefore all data should be available in as close to the raw format as practicable, and any value added products would be hosted separately.

In practice, the individual contributor model is incredibly difficult to achieve. Complications immediately began with IT management, data formatting consistency, and reliability of reliable submissions. The IHO DCDB is hosted on a NOAA IT infrastructure and security protocols follow NOAA requirements. The IT structure follows NOAA's National Multibeam Archive, which is a relational database that allows widespread access and download of hosted data but not necessarily allow widespread access to submitting files. The crowdsourced bathymetry database would be a document-oriented based database using GeoJSON, JSON with geometry and geography data embedded, documents as schema. In order to make any contributed data recognizable by the database applications it would need to appropriately follow the schema, something that was not guaranteed using the individual contributor model.

The concept of having an entity that would facilitate providing data seemed necessary to ensure that it follows the necessary format, schema, and IT protocols. The working group referred to these entities as trusted nodes. The trusted node model, shown in Figure 5, was also designed to incorporate crowdsourced data owned by a company providing data loggers to contributors. In this model, trusted nodes such as Olex and Sea-ID already had an existing contribution method to the company as a part of their business model [IHO, 2017]. For the purposes of this project, Rose Point Systems is the trusted node.

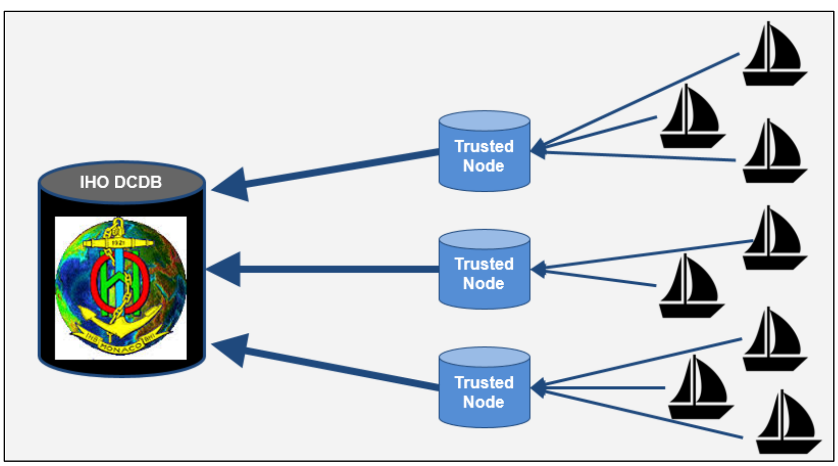


Figure 5: Individual Contributor Model of crowdsourcing bathymetry [Robertson, 2016].

It is important to note that the phrase trusted node does not carry an implication that the data supplied meets specific accuracy or precision requirements. The trusted aspect of a node is to provide data and metadata in the required formats that is ingestible by the database. Some nodes perform quality controls on the data to ensure that retained crowdsourced bathymetry data is free of erroneous data. Rose Point does not perform quality assessments of the data, and solely provides the framework for mariners to contribute data to the database.

Pipeline

The Rose Point project utilized a pipeline variation that represented a hybrid of trusted node and individual contributor models. Rose Point Systems does not store any data on offsite servers, nor receive the data before it is archived. The submission proceeds directly from the mariner to the database, as shown in Figure 6. The ECS software transmits the data via HTTP post when the mariner updates the software or chart catalog. The DCDB then checks file validity using automated scripts before storing the GeoJSON files. Ingest scripts also extract supplied metadata for the database metadata catalog. File geometries are then extracted and loaded into a spatial database.

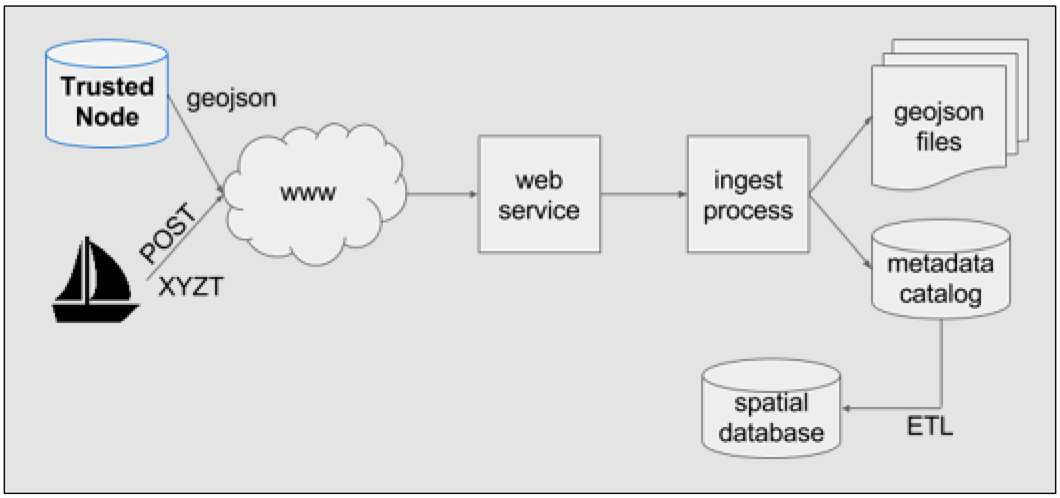


Figure 6: Pipeline for data submission from mariner to availability in the database [Robertson, 2016].

Crowdsourced Bathymetry Viewer

The IHO DCDB crowdsourced bathymetry viewer allows users to access and discover crowdsourced bathymetry data hosted by the database. NCEI developers created this viewer using existing templates of bathymetry data viewers based on ESRI GIS tools.

The first version of the viewer allowed the application to display the point observations with a generated track line connecting points based on time stamp and file similarities. This initial design contained several major issues. The track line generation algorithm generally functions as intended, but did not contain enough checks on time and file pairing. As a result, it would display track lines that were clearly in error. Figures 7 and 8 display examples of the viewer displaying vessel track lines errors over land, and across the globe. As the quantity of archived data grew, a significant memory issue caused viewer load times to increase dramatically. The viewer's algorithm of loading point values from the data files simply attempted to load too many points and began crashing or become overwhelmed. It was clear that both track line and point value mining from data files would need to be redesigned before the quantity of crowdsourced data reached anticipated levels.

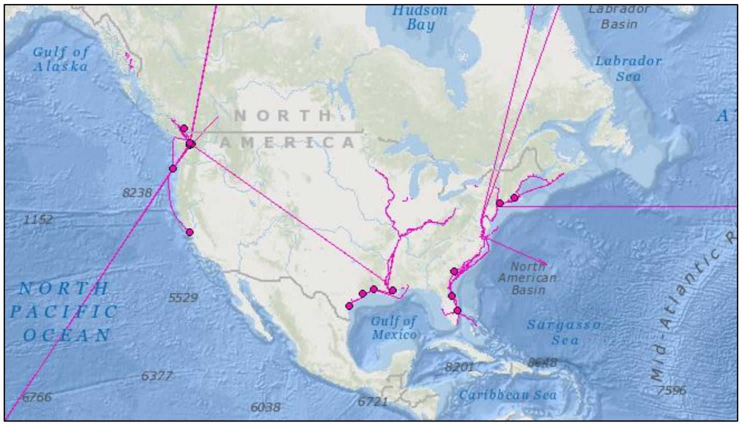


Figure 7: Track line generation issues with the initial crowdsourced bathymetry viewer.

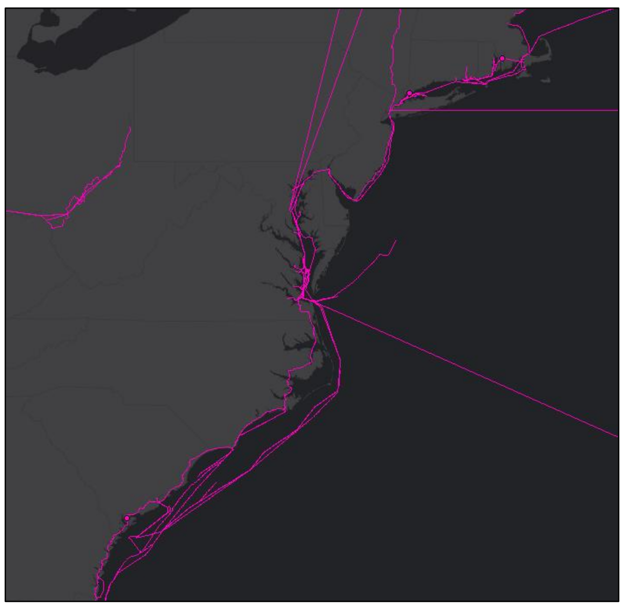


Figure 8: Track line generation through narrow waterways in the initial crowdsourced bathymetry viewer.

The next developments of the viewer will utilize an improved line segmentation algorithm, point down sampling, and a website outside of NOAA domains. Additional quality checks on line segmentation display generated track lines from the point values that consistently stay within the proper boundaries even inside of narrow passages and canals. Figure 9 shows generated track lines that accurately represent traffic transiting the C&O canal, Cape Cod Canal, and the Great Dismal Swamp Canal. Actively down sampling point value extraction from data files allow reasonable load times while still allowing improved track line generation.

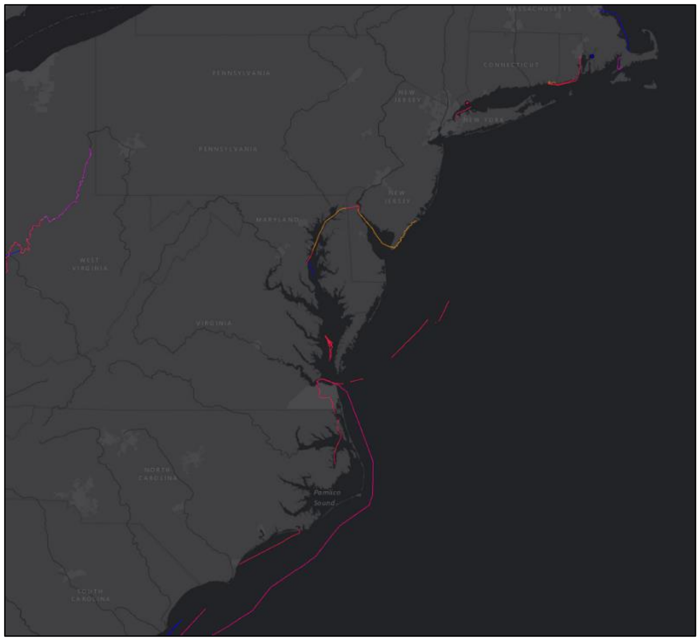
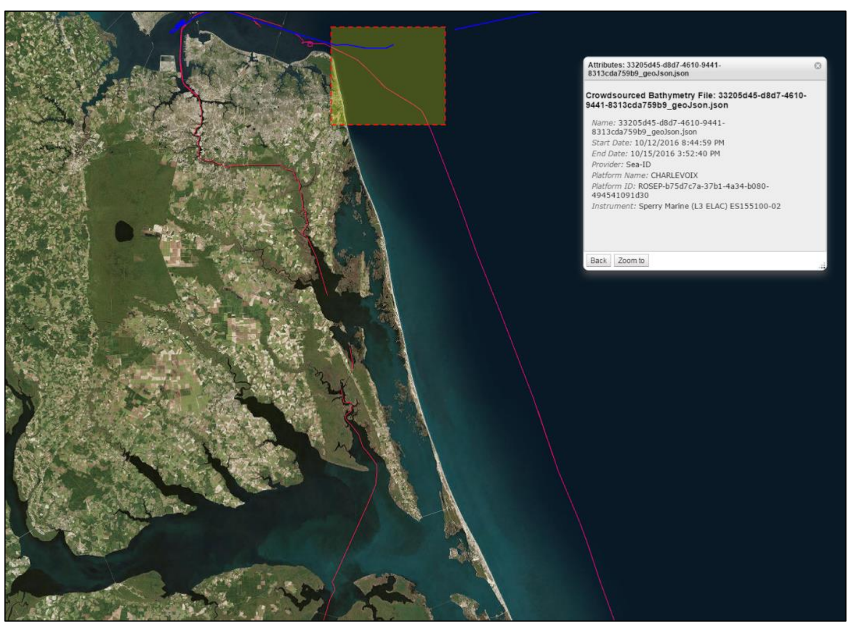


Figure 9: Crowdsourced bathymetry viewer after line segmentation fix.
(Note: only a sample of available files were displayed for this image for clarity purposes)

Crowdsourced bathymetry data are currently available to be downloaded in the native CSV format with JSON headers, which are recognizable by ArcGIS. Future developments to the viewer are planned to provide files in shapefile and XYZ. The viewer extracts metadata from the JSON tags in each file to provide a user with information regarding vessel identification and time of acquisition, shown in Figure 10.



.

Figure 10: File query and metadata fields supplied by the viewer.

Considerations for Further Work

Uncertainty analysis

The greatest strength crowdsourced bathymetry is expected to offer is a higher quantity of observations to compensate for reduced quality standards. Modeling of accuracy and uncertainty of crowdsourced bathymetry is a daunting project. Such models would require the quantity of contributions to reach appropriate levels to compare data sets adequately. The Rose Point crowdsourced supply of data has been active since May of 2016, during which time the level of contributions has only recently reached into the thousands of contributions. While it was not a large enough sample size in any location to begin an uncertainty analysis at the time of this paper, future studies should be able to include such assessments.

Expand to more ECS providers

The collaborative effort with Rose Point paved the way for future efforts. This project developed several best practices and lessons learned that will undoubtedly aid in expanding to other ECS manufacturers. These lessons learned and best practices are captured in the IHO document produced by the Crowdsourced Bathymetry Working Group. Nearly all electronic means of navigation, from ECS, chart plotters, and Electronic Chart Display Information Systems (ECDIS), are capable of acquiring this data.

Feedback

One of the most valuable aspects of crowdsourcing is the ability to develop feedback to the contributing crowd. Crowdsourced programs such as the Cooperative Charting Program provide a reward system based on number and accuracy of contributions [Office of Coast Survey, 2016]. Future iterations of crowdsourcing bathymetry programs could benefit from exploring contributor interaction providing feedback to and from contributors. Similar to an uncertainty model, this task represents a significant project in itself.

References

International Hydrographic Organization. “EIHC5 Decisions related to IRCC.” Inter-Regional Coordination Committee 7. Mexico City. 2015.

International Hydrographic Organization. “Crowdsourced Bathymetry Working Group Guidance Document.” Monaco. 2016.

International Hydrographic Organization. “Publication C-55: Status of Hydrographic Surveying and Nautical Charting Worldwide.” Monaco. 2014. http://www.iho.int/iho_pubs/CB/C-55/index.html.

NOAA Office of Coast Survey. “A Brief History of the National Ocean Service – United States Power Squadrons Cooperative Charting Program.” National Oceanic and Atmospheric Administration. 2016.

NOAA Office of Coast Survey. *NOAA Hydrographic Survey Priorities, 2012 Edition*. National Oceanic and Atmospheric Administration. 2016. http://www.nauticalcharts.noaa.gov/hsd/NHSP.htm.

Reed, A., Klemm, A. “Crowd Sourced Bathymetry Concept Study Using Data Logged with Electronic Navigation Systems.” Hydro 2016. Halifax. 2016.

Robertson, E., Reed, A., Jencks, J. “Crowdsourced Bathymetry Data via Electronic Charting Systems.” ESRI Ocean Forum. 2016.