The National Bathymetric Source

Glen Rice  
Office of Coast Survey  
NOAA  
Durham, USA  
glen.rice@noaa.gov

Katrina Wyllie  
Office of Coast Survey  
NOAA  
Durham, USA  
katrina.wyllie@noaa.gov

Barry Gallagher  
Office of Coast Survey  
NOAA  
Silver Spring, USA  
barry.gallagher@noaa.gov

Phuntsok Geleg  
Lynker Technologies, LLC  
Silver Spring, USA  
phuntsok.geleg@noaa.gov

Abstract—Bathymetric data are foundational for many important marine and coastal uses such as increasing the safety of navigation through updated high-resolution nautical charts and services, studying the changing of coastline features in response to climate change, creating hydrodynamic models for coastal resilience efforts, and studying marine life, among others. NOAA’s Office of Coast Survey’s National Bathymetric Source (NBS) program curates and regularly maintains a bathymetric data compilation for navigation, planning, and public needs. The traditional manual workflow of various disparate sources applied to a navigation product has been reimagined by leveraging automation in compiling sources using hydrographic quality metrics to determine the best available data for any geographic area. All bathymetric data sources acquired must be evaluated for depth, quality, and source information to make NBS data-driven workflows effective. The originating metadata informs the normalization of format, coordinate reference system, coverage, and quality assessment to prepare each source for compilation. The quality metrics used during the compilation phase to determine the best available bathymetry are carried through to the extracted products to inform effective use. The extracted bathymetry is prepared for navigation, planning, and public pipelines in customer-specific format, coordinate reference system, and resolution. All phases of the NBS workflow are discussed, from acquiring source through extracting bathymetry. The NBS program is driven to support the United States’ bathymetry needs with timely, curated, and accessible bathymetric data.

Keywords—navigation, hydrography, bathymetry, BlueTopo, automation

I. INTRODUCTION

The Office of Coast Survey (OCS) established the National Bathymetric Source (NBS) program to support its navigational and hydrographic survey mandates [1,2,3,4,5,6] with the best available bathymetry. As a maintained and metadata-rich compilation of bathymetric data, the resulting high-resolution bathymetric compilation is used by multiple sectors and stakeholders outside of navigation [7]. In addition to supporting the modernization of NOAA navigational products, the resulting NBS bathymetry is made publicly available, enabling a more complete understanding of bathymetry in numerous applications.

Navigation products are modernizing. Supporting updates to the bathymetry datasets that are mainstays of the United States’ geospatial infrastructure is a primary driver for the NBS program. Current OCS efforts focus on rebuilding the national chart suite to new geographic footprints and scales [8,9], which require recompiling bathymetry to correspond to the new product specifications. In addition, OCS plans to support Precision Marine Navigation (PMN) [10] with S-102 [11] in coordination with other associated International Hydrographic Organization S-100 standards [12].

Bathymetry is critical to marine navigation, marine geographic information systems, commerce, science, and industry. The NBS supporting processes enable rapid inclusion of new data with an emphasis on quality metrics to inform effective use. Bathymetry enhances the growth of the new blue economy by aiding in the selection of alternative energy sites and identifying geologic features and hazards. The NBS program provides critical bathymetric data for new National Ocean Service operational numerical oceanographic forecast modeling systems [13]. These real-time modeling systems provide short-term forecast guidance of water levels, currents, water temperature, and salinity to mariners. Further, the NBS products support coastal resilience by enabling simulations of sea level effects on coastal ecosystems, helping to identify and manage sensitive marine habitats.

Other bathymetry compilations exist for the United States [14,15]. And other hydrographic offices have built compilations for their jurisdictions [16,17,18,19,20,21,22]. The NBS program is unique because its compilation process evaluates hydrographic quality metrics of uncertainty, feature detection, and seafloor coverage while approximating temporal and spatial rates of change [23]. These metrics and methods for evaluating quality are standard throughout OCS’s process to support the prioritization and allocation of resources for future survey projects.

In addition to Findability, Accessibility, Interoperability, and Reuse of digital assets (FAIR) [24] principles, the NBS processes were engineered to restart from source data, producing reproducible and consistent results by removing undocumented and subjective work. The NBS process and products carry the primary pillars of hydrography, which can be distilled as depth, quality, and source lineage. The resulting bathymetry also directly supports the transition from product-driven to data-driven workflows by increasing the quality, accessibility, and timeliness of source data. With these drivers, the NBS program continues to build the National Bathymetry by region while periodically updating the compilation in established regions with new sources when available. Work is underway to distribute products of this critical bathymetric compilation in various coordinate reference system, formats, and resolutions.

The NBS program has two teams. The first team, the operations team, is responsible for building and maintaining the National Bathymetry. The second team, the development team,
is responsible for building and maintaining the infrastructure and workflow that supports the operations team. Previous versions of the NBS program focused on building a “bathymetry database” to centralize disparate bathymetric sources [25]. The current iteration is focused on building data-driven workflows to deliver the best available bathymetry for the nation in purpose-built and prompt packages to support streamlined downstream processes. Rather than a single centralized database of bathymetry, the NBS workflow incorporates several databases to support specific workflow functions. The requirements for the final extraction pipelines drive the upstream processes in terms of metadata. The following methods section provides an overview summarizing the workflow phases while the discussion covers how and why the various workflow components were implemented as summarized. This discussion begins with requirements of the final pipelines to illustrate how they drive the preceding parts of the workflow.

II. METHODS

The NBS model is built on scientific and hydrographic principles which serve to make the process adaptable to varied end uses. The phases of the NBS workflow are built for throughput by arranging modules that facilitate automation where possible and human input where required. Throughput is important for building new regions and for integrating new bathymetry, but it is also important when accommodating updates to current products if new standards, requirements, or process improvements are implemented. The workflow phases are summarized as: acquire, normalize, compile, and extract. In addition, the types of data sources, coordinate reference systems, data formats, and access for each of the five delivery pipelines currently generated are described.

A. Acquire

There are two primary approaches NBS employs to discover, acquire, and incorporate various data sources into the National Bathymetry. Significant sources of bathymetry include NOAA and U.S. Army Corps of Engineers hydrographic surveys and topographic bathymetric (topo-bathy) lidar data. For well-known sources that publish data through an application programming interface (API), such as NOAA’s National Centers for Environmental Information (NCEI) [26] or U.S. Army Corp of Engineer’s eHydro [27], new or updated data or metadata are identified and periodically pulled down using Python code. To support this operation, internally maintained databases of source information are referenced when new data are found. For other bathymetric data sources that are not well known and do not publish through an API, data is searched for and collected manually in collaboration with OCS’s External Source Data (ESD) team. Typical ESD sources include federal agencies, state and local groups, and universities. The work of the ESD team is fundamental to the NBS process for discovery, quality assessment, and defining necessary metadata for input data sources such as source data licenses [28]. Gaps in waters deeper than 200m are filled through Global Multi-Resolution Topography [16], a merged bathymetry model. A data license is assigned to each source survey to seamlessly manage which sources are accessed for the various derived products [28]. Preferentially, data are acquired directly from the source.
provider as data that travel through other distributors could be modified in undocumented ways and may come with ambiguous licensing. In the final step of the acquire phase, data are sorted into separate buckets, which are databases and storage locations, to ensure different source data feeds the desired end product.

B. Normalize

Acquired bathymetric data come in varying formats, coordinate reference systems, quality metrics, and licensing. Normalization of data and metadata into standard metrics and formats simplifies compilation. It also provides a single place where the NBS operations team can evaluate the source data metrics that support compilation. While the normalization process is largely automated in Python, humans supervise and correct coordinate reference system transformations to a compilation coordinate reference frame and perform interpolation of source surveys to represent survey coverage. Human input is recorded in databases to support end product lineage research. All source data must have the required metadata[29] to complete normalization and define a quality score to support the compilation and requirements of the end products. While data originators generally provide metadata and quality metrics, the NBS team also regularly augments this information when missing during normalization.

C. Compile

Compilation of the best available data is the heart of the NBS process. Previous phases of the process are meant to support compilation, while follow-on steps are built to ensure the data meet the end purpose. The compilation step represents a transition from survey-based workflows to location-based workflows. The upstream processes provide the sense of "best available" data through the survey score, which is decayed as a function of time and location. The compilation engine is built around evaluating the best depth estimate by evaluation locations. Evaluation locations and their horizontal footprint are based on the re-scheme chart resolution requirements provided by OCS [8,9]. The rules applied during compilation are considered in the following hierarchy to resolve a draw between source data in the same evaluation footprint with the same quality: (1) Highest decayed quality score, (2) Finest resolution, (3) Least Depth, (4) Name of the source. Compilation occurs within each data bucket type and subsequently between buckets to suit specific compilation specifications for downstream purposes. These specifications are stored in a database representing each tile with a name, geographic footprint, resolution as well as the source data buckets. Finally, compilation completes coverage by generalizing through interpolation between the provided sources.

D. Extract

Extraction takes a compiled raster dataset for each compiled tile update and derives fit-for-purpose products in the required format with the prescribed metadata and coordinate reference system. Each destination has its own "pipeline" that provides data as the NBS updates with new sources. At this time, there are five pipelines. Two are navigational: the Electronic Navigational Chart (ENC) maintained in OCS's Marine Chart Division and the S-102 high-resolution bathymetry products to be built by NBS and distributed by NOAA's PMN program. The two pipelines for disseminating NBS data for public use are called BlueTopo and "Modeling." The final pipeline is strictly for internal OCS use to support planning for future survey acquisition. Requirements for these various pipelines were either provided by the pipeline stakeholder or defined as a best estimate of need by the NBS program. To distinguish between the final products from differing pipelines the product file names are prepended with a tag that describes the target product. For the public pipelines, these files are prepended with BlueTopo and Modeling. Other products target naming conventions defined in collaboration with the pipeline stakeholders.

III. RESULTS AND DISCUSSION

The NBS program believes that bathymetry data should carry depth, quality metrics, and metadata about the source. The elevation information alone, while most meaningful to many users, does not provide the needed context for the data provided. The additional pieces of quality and source information carried forward to the products allow downstream processes to extract the needed bathymetric value from the NBS compilations.

Because data quality is vital to the NBS model, the throughput of qualified data, meaning data reviewed by OCS's Hydrographic Survey Division (HSD) for proper quality evaluation, is central to all NBS pipelines. However, to ensure that data gets back to the public promptly, NBS also adds unqualified data to non-navigational pipelines. Unqualified data have some preliminary quality metrics applied for NBS compilations, but these metrics may change once HSD comprehensively reviews a survey. Separate from qualified and unqualified data, precompiled data such as data from the ENC or GMRT are compilations where the individual source metadata may not be fully known and are thus rated at the lowest quality. Since ENC source data should be included in the NBS as an individual source, ENC soundings only contribute when the ENC is populated by an older dataset not directly referenced in the soundings. While OCS is working to resolve this issue, researching old soundings and weighing their value is manual and cumbersome. Sensitive data, designated as unsuitable for public distribution, are only used internally by OCS. These four data "buckets," qualified, unqualified, sensitive, and precompiled, constitute the NBS program's current model to sort data into specific pipelines.

The NBS makes bathymetry available to the public through BlueTopo and modeling pipelines, which carry the described metrics in GeoTIFFs and Raster Attribute Tables (RAT). These GeoTIFFs have three layers: the first is elevation, the second is vertical uncertainty, and the third is contributor. The attached RAT defines the contributor layer and contains source and quality metrics [30]. BlueTopo GeoTIFF/RAT datasets are available for download from Amazon Web Services (AWS) [31] and to increase the accessibility of this public product, the datasets can also be viewed in NOAA nowCOAST, both through the web viewer and as web map services [32]. Data dissemination through the modeling pipeline is also available on AWS. Neither BlueTopo nor modeling products are suitable for navigation as these datasets include unqualified data.
BlueTopo and modeling products are suitable inputs for NOAA’s efforts with the new blue economy and coastal resilience. The vertical coordinate reference system is the primary difference between BlueTopo, referenced to local Mean Sea Level such as NAVD88, and modeling, which refers to a low water datum such as Mean Lower Low Water [33]. The transformation to the final vertical reference system is performed at the extraction step if the target coordinate system does not match the coordinate system used at compilation.

Navigation pipelines are only derived from qualified data and are provided on a navigational vertical coordinate reference system. Data for charting are delivered upon request and directly to OCS's Marine Chart Division. This delivery currently includes three single-layer GeoTIFFs; the first of measured depths, the second with survey coverage interpolation, and the third with all data. This process ensures that all cartographic soundings are selected from actual data points while contours and depth areas are generated from contiguous grids, streamlining the automated process for each product while retaining data integrity and quality. Cartographic soundings are attributed to the source survey and survey date, as well as the uncertainty of the elevation. The quality metrics associated with each node are aggregated into areas of the same quality and provided as seamless quality polygons. Finally, a comparison to the chart is performed, and requested significant differences are identified for further research into potentially missing source.

The second navigation pipeline includes S-102 formatted high-resolution bathymetry for the next-generation navigation services in the S-100 framework. The NBS creates a Bathymetric Attributed Grid (BAG) [34] which is converted to S-102 format and available in the PMN program [35]. One of the crucial updates to the S-102 specification [11] was the addition of the metadata layer with quality metrics so that source data can be differentiated from interpolated data for a sounding selection on-the-fly.

The internal pipeline products are derived from all four data buckets and delivered to the HSD Operations Branch. The products are provided on a navigational vertical coordinate reference system in GeoTIFF format. The RAT includes targeted attributes to support survey planning and prioritization.

Providing high-resolution information is a balance between tile sizes, footprints, and the number of tiles. The NBS program tessellates the world into square tiles in geographic space. While these tiles follow the same geographic sizes as the reschemed ENC cells [8,9], only the navigation pipelines use the same naming convention and explicit tile sets as the ENCs. The naming convention for other pipelines differs from the reschemed ENC cells to allow for algorithmic and automatic generation and supersession between cells of different sizes globally making the tile set more adaptable to changing requirements and data availability. In the future the resulting resolution in various pipelines may differ depending on the pipeline requirements. In some cases, this may be the highest resolution that source data can support, adding a requirement for a variable-resolution product. This modification from a requirements-driven compilation to a data-driven compilation could have significant implications for the future development of the compilation step.

The NBS compilation methodology is designed to establish the best available data based on quality metrics and a decay model. It may be appropriate for some compilations to average or smooth data, but these methods obfuscate the lineage between the final compilation and the source information. The NBS combine process is location-based as surveys cannot be sorted as a unit as one survey might supersede in one location but not in another location due location based decay of quality. The product resolution drives the evaluation location footprint, but the NBS workflow is built on node-based logic corresponding to the BAG specification [29]. Accounting for decay in survey quality over time and by locality allows for the possibility of newer, lower quality data to supersede older.

<table>
<thead>
<tr>
<th>Compilation</th>
<th>Buckets</th>
<th>Vertical Coordinate Reference System</th>
<th>Pipeline</th>
<th>Location</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>Qualified Precompiled</td>
<td>Navigational</td>
<td>Charting</td>
<td>Charts.noaa.gov</td>
<td>ENC</td>
</tr>
<tr>
<td>Public</td>
<td>Qualified Unqualified Precompiled</td>
<td>National</td>
<td>BlueTopo</td>
<td>NBS AWS [31]</td>
<td>S-102</td>
</tr>
<tr>
<td>Internal</td>
<td>Qualified Unqualified Precompiled Sensitive</td>
<td>Navigational</td>
<td>Planning</td>
<td>Internal Storage</td>
<td>GeoTIFF and RAT</td>
</tr>
</tbody>
</table>
higher quality data in highly changeable areas. Conversely, older and higher quality data may supersede newer, lower quality data in less changeable areas. The decay time is calculated by differencing a survey's end date and a reference time that approximates new tile generation. The techniques used to merge the various sources ensure representation of the best available data. Metadata and quality metrics are carried through to the product to inform effective use. Notably, the compile step represents a transformation from a unit of work as a survey to a unit of work as a tiled area.

Because different pipelines require different data sources, NBS compilation is done in two phases. First, each source data bucket is combined to establish the highest quality data by location within each bucket. Then, depending on the target pipeline, the bucket results are combined. The second compilation phase includes generalization, which is interpolation to estimate depth where no survey explicitly has coverage. Historically and in a cartographic sense, this step was performed by cartographers drawing depth curves. For the NBS to drive automated cartography and support modeling, among other downstream uses, the same bathymetric continuity must be represented in NBS compilations. The same quality metrics that support other elevation estimates must also be made for various forms of interpolation, namely feature detection capability, uncertainty, and coverage. Interpolated data from NBS are generally denoted by a "bathymetry coverage" field in the RAT, which indicates if the area has a measured depth.

Interpolation is an important piece of the NBS workflow to represent coverage with the best estimate of depth and the associated quality metrics. Side scan sonar is a common acquisition method in the United States to detect features while also carrying a sense of survey coverage where bathymetry measurements may not be present. In the case of systematic surveys conducted without the support of side scan sonar and meant to represent a region without feature detection capability, the interpolated areas are marked as bathymetry coverage false, but survey coverage is true. NBS generalization is interpolation over areas that are not supported by any survey, so the bathymetry coverage and survey coverage fields are false. Interpolation and uncertainty estimation for interpolated elevations remains an area of active development in the NBS program, and most interpolation uncertainty estimates are biased toward a conservative, larger estimate. All interpolation performed within the NBS workflow is linear to maintain stationarity of uncertainty statistics between source data depth estimates. Ensuring interpolation covers or does not cover an area appropriately is a key goal of the interpolation step.

Another important piece of the NBS workflow is the normalization of the coordinate reference system for compilation. While limited transformations occur during the extract phase, normalization must be able to accommodate the transformation of various source data into the compilation reference frame. Input horizontal reference frames may be geographic or projected in some version of WGS84 or NAD83. Input projected systems are often Universal Transverse Mercator (UTM) or a version of a State Plane. As NOAA hydrographic surveys use NAD83 (2011) UTM for final products, NBS compilation uses this horizontal coordinate reference system. Surveys that overlap UTM zones are processed in both zones such that data is readily available for compilation. The vertical coordinate systems of the final products are prescribed by the pipeline requirements. The NBS workflow uses VDatum vertical transformation definitions where available [36].

The primary pieces of the normalization workflow include reading the source data and metadata, transforming the data to a common frame expected by the compilation process, and ensuring the sense of coverage meets the needs of the final tiled product. Surveys that do not have depth estimates where a survey is meant to represent, such as side scan sonar-supported surveys or systematic coverage surveys, are interpolated to a resolution that corresponds to the needs of the final product. In some cases this also means complete coverage source data, in an area sense, is upsampled to ensure the survey continues to have a sense of coverage. The feature size attribute in the RAT and coverage flags maintain the sense of coverage scale intended by the source data. In some cases, the readers for a particular format and source have default attributes. Once all bathymetric data sources are put into a common reference frame and all data have quality metrics, metadata, and coverage assigned, the final phase of normalization is to score the data.

The NBS operators currently run the normalization phase through a few Python scripts. The process can be run to look for new surveys, which compares stored surveys to surveys registered in the associated databases. Surveys that are not registered are incorporated into the database and then processed as far as the supporting metadata allows. Ideally, surveys process through normalization without assistance from the operations team. If human input is required or the source survey changes, the survey can be updated by setting database fields to look for new metadata or bathymetry on disk. These fields can also trigger reprocessing of data to prepare for compilation. Because it is challenging to get the best representation of interpolation coverage automatically every time, the process produces several alternate interpolation options with different calculations of coverage for each survey requiring interpolation. While the process provides a specific version of interpolation for downstream use, the operator can select an alternate or create a manual interpolation. These changes are documented and preserved.

The acquire step delivers surveys to the storage referenced by normalization to present new or updated data to the normalization process. In the case of automated acquisition, the data are ideally advertised with a spatial query function to allow for discovery by area. Previously downloaded surveys have their metadata compared to previous metadata preserved by the NBS, and ideally, a hash of the bathymetry can also be compared. If the query does not return a hash, a survey must be downloaded to compare the bathymetry against the last data downloaded. No automated acquisition is usually performed by the HSD ESD team.

While the NBS team continues to build out the National Bathymetry for the United States and territorial waters,
stakeholders in the navigation, modeling, and public sectors have found the derived products valuable thus far. Jacobson Pilots in the Ports of Los Angeles and Long Beach are navigating vessels with critical under keel clearances into port with the support of the S-102 high-resolution bathymetry supplied by NBS navigational pipelines. Reschemed ENCs in the New England region have used NBS navigation products to derive cartographic bathymetry at new chart scales. Within OCS, the NBS modeling products are used in operational storm surge and tide forecast modeling systems, such as STOFS-2D-Global and STOFS-3D-Atlantic for the U.S. East Coast, Gulf of Mexico, and Puerto Rico [37,38]. NOAA Office for Coastal Management, in conjunction with coastal zone managers of Maine, New Hampshire, and Massachusetts, have developed a regional map of seafloor characteristics based on the NBS BlueTopo products [39].

Coast Survey’s navigational mandate drives a requirement for quality and timeliness that make the NBS program’s mission distinct from other bathymetry compilations. With a workflow that is built to update and maintain the national bathymetry through targeted automation of new and different sources, the NBS program recognizes the value this compilation can add to the United States’ marine spatial infrastructure.

IV. CONCLUSION

The National Bathymetry, as realized through NOAA OCS’s NBS program, is unique because of the techniques used to combine various hydrographic data sources and generate purpose built seamless seafloor models. While initially designed to serve maritime navigation, these techniques are valuable to many other coastal health and resilience efforts, such as oceanographic modeling, coastal planning and recovery, and scientific investigation activities. Built for throughput to support updated navigational products, this effort supports the NOAA Strategic goal for Resilient Coastal Communities and Economies. Integrating bathymetric data across federal entities best serves the U.S. taxpayer and ensures that NOAA is a responsible steward of federal resources.

ACKNOWLEDGMENT

The authors thank the many past and present federal and contracted members of the NBS program and the rest of NOAA Office of Coast Survey for their support. In particular, the authors thank RDML Richard T. Brennan for championing the NBS program.

REFERENCES


