Initial Efforts Toward Coordinated Community Data Processing to Accelerate the Growth of Publicly Available Bathymetric Data Products

Vicki Ferrini[1], John Morton[1], Lindsay Gee[2], Erin Heffron[2], Hayley Drennon[1], Nicole Raineault[2], and Suzanne Carbotte[1]

[1]Lamont-Doherty Earth Observatory of Columbia University
[2]Ocean Exploration Trust

November 21, 2022

Abstract

The Global Multi-Resolution Topography (GMRT) Synthesis is an elevation model that includes curated deep-water multibeam bathymetry data at ~100 m resolution covering more than 9% of the ocean. GMRT is built with a scalable tiled raster architecture that efficiently stores and presents high-resolution elevation data nested within low resolution data. A set of tools are available for users to access the compilation through simple user interfaces (e.g. GMRT MapTool) and web services, while also providing full attribution and access to source swath files. The availability of raw/unprocessed multibeam sonar data in the National Centers for Environmental Information (NCEI) archive has increased dramatically over the last decade, but transforming these data into high-quality integrated products suitable for use by scientists and the public alike requires significant effort. The GMRT Team has built workflows and tools for data preparation and review that are optimized for cleaning and integrating sparse globally distributed multibeam data, enabling the addition of ~60-80 research cruises per year. Once raw swath data files are cleaned and corrected, they are gridded/tiled with the GMRT Tiling tools so they can be reviewed and quality controlled in the context of other data in the GMRT Synthesis. Working with processed swath files generated by the community, we have observed that this process frequently reveals issues that are overlooked during data processing. In order to accelerate the rate of data integration and leverage the data processing efforts of the community, GMRT Tiling tools are being adapted for distributed use. Ocean Exploration Trust is an initial partner in this effort, and all processed swath files from the 2017-2019 Nautilus field seasons were prepared with GMRT Tiling tools and reviewed by the OET team. This revealed problems in processed swath data files from several cruises that were addressed prior to submission to NCEI, thereby improving the quality of data in the archive. We are now working to include the GMRT Tiling tools into at-sea standard operating procedures of the Nautilus as a testbed for broader community distribution, to ensure consistent quality of processed multibeam data, and to accelerate the production of high-quality integrated data products including GMRT and Seabed 2030.

Initial Efforts Toward Coordinated Community Data Processing to Accelerate the Growth of Publicly Available Bathymetric Data Products

Vicki Ferrini(1), John J Morton(1), Lindsay Gee(2), Erin Heffron(1,2), Hayley Drennon(1), Nicole Raineault(2) and Suzanne M Carbotte(1)

(1)Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, United States, (2)Ocean Exploration Trust, Narragansett, RI, United States

Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE

OCEAN EXPLORATION TRUST

PRESENTED AT:
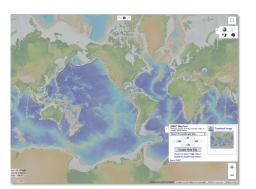
GLOBAL MULTI-RESOLUTION TOPOGRAPHY (GMRT)

Seabed bathymetry is an essential ocean observation that has a wide variety of uses. Despite its importance, only ~20% of the global ocean has been mapped with direct measurement. Several efforts are underway to acquire new data in local, regional, national, and global waters, but processing and integrating those data into useful products remains a time consuming endeavour.
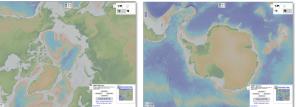
GMRT is:

- a global multi-resolution topography **data synthesis**
- an **infrastructure** for delivering elevation data as grids, images, profiles and points at user-defined locations/elevations & full access to source data
- a **tiling scheme** for efficiently storing and delivering gridded multi-resolution data for the entire planet, maintained simultaneously in 3 projections
- a **scalable methodology for QA/QC**'ing multibeam sonar data that is very well-suited for integrating multibeam data acquired during transits

GMRT: Overview

- Multi-resolution tiled synthesis
  - Topography and bathymetry
- Comprehensive metadata
  - Full attribution to sources & access to source data
- Simultaneously maintained in 3 projections
- Elevation data available in many formats
  - Grids, Images, Points, Profiles
- Accessible via:
  - GMRT MapTool Web App
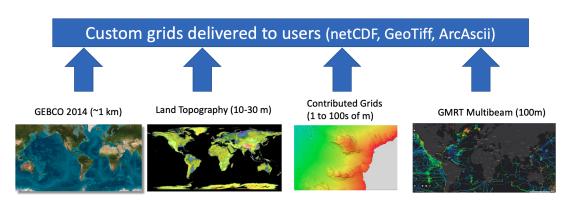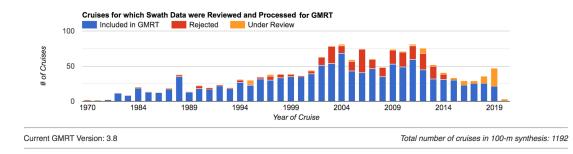 - GeoMapApp Desktop App
  - GMRT Web Services

GMRT seamlessly combines four elevation components: (1) curated multibeam bathymetry data, (2) gridded bathymetry, (3) land elevation data, (4) the GEBCO world map.

GMRT: Multiple Elevation Sources

- Four discrete tiled multi-resolution elevation components
- Maintain and updated independently and on different schedules
- Images pre-rendered, grids are generated on-the-fly

Custom grids delivered to users (netCDF, GeoTiff, ArcAscii)

GEBCO 2014 (~1 km)   Land Topography (10-30 m)   Contributed Grids (1 to 100s of m)   GMRT Multibeam (100m)
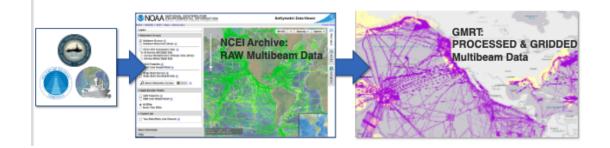
Extensive curatorial effort is focused on processing and integrating multibeam bathymetry data that is made publicly available through the NOAA NCEI Mulibeam Archive. The current version of GMRT includes data from 1,192 research cruises covering more than 9.6% of the global ocean at ~100m resolution.

Cruises for which Swath Data were Reviewed and Processed for GMRT
Included in GMRT   Rejected   Under Review

Current GMRT Version: 3.8   *Total number of cruises in 100-m synthesis: 1192*

PUBLICLY-AVAILABLE DATA

Multibeam data made publicly available in the NOAA NCEI multibeam archive are the primary source of data for the Multibeam Synthesis component of GMRT. These data have been acquired on academic, government and industry vessels operated around the global by many organizations.

Data from the US Academic Research Fleet are delivered to the NCEI archive via the Rolling Deck to Repositiry (R2R) Program. These utlibeam data are raw and unprocessed. Best practices developed by the Multibeam Advisory Committee (MAC) help to ensure that data quality at acquisition is optimized, but data processing is left to individual scientists onboard and/or others in the community who have use-cases that require them to download sonar data and process it from raw.

The raw unedited data in the NCEI archive is important to preserve and is a globla data resource. While some data were acquired during dedicated surveys, signficant amounts of these data were acquired opportunistically during transits or research cruises for which no mapping expert was aboard. Preserving these data in its raw form is an important aspect of data stewardship and ensures that data are preserved without errorsor bias induced during the data processing.

The NCEI multibeam archive also serves as the International Hydrographic Organization (IHO) Data Center for Digitial Bathymetry - the recognized international arhive for bathymetry source data.

IHO Data Centre for Digital Bathymetry (DCDB)

The IHO DCDB was established in 1990 to steward the worldwide collection of bathymetric data. The Centre archives and shares, freely and without restrictions, depth data contributed by mariners. The IHO DCDB is hosted by the U.S. National Oceanic and Atmospheric Administration (NOAA) on behalf of the IHO Member States.

IHO DCDB Data Viewer highlighting ship tracks and data availability over the Pacific Ocean and neighboring regions

The DCDB archive includes over 30 terabytes of oceanic depth soundings acquired with multibeam and singlebeam sonars by hydrographic, oceanographic and industry vessels during surveys or while on passage.

The DCDB also archives and provides access to data contributed in support of the IHO Crowdsourced Bathymetry (CSB) initiative.

The IHO DCDB Data Viewer shows the global coverage of the DCDB's bathymetric data holdings as well as the spatial extent of data archived at other repositories via web services.

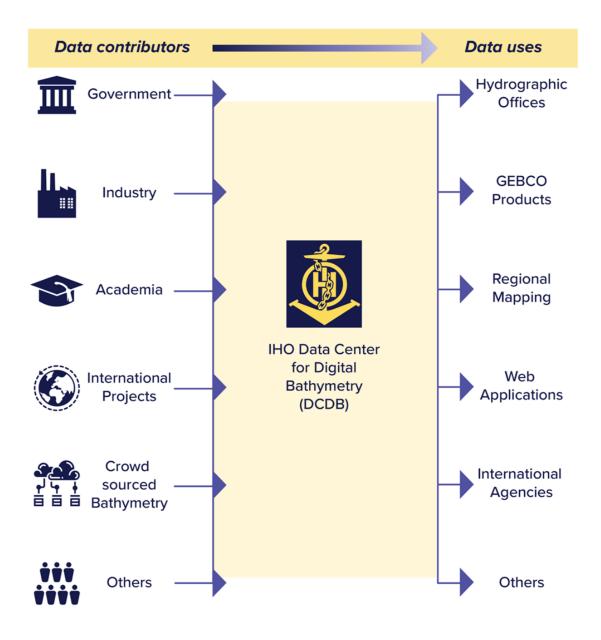**Access Data**

The World Reference for Raw Bathymetry

The IHO DCDB consists primarily of unedited single and multibeam bathymetric data contributed by industry, government, academia, and crowdsource efforts. These data are a public resource that are routinely used to produce improved, regional and global bathymetric maps and grids in support of science and exploration.

The DCDB also serves as the long-term archive for the GEBCO Ocean Mapping Programme and the Nippon Foundation-GEBCO Seabed 2030 project, which is a global initiative that aspires to create a complete global ocean map by 2030.

Heightened awareness and global focus on the ocean has resulted from a number of high profile initiatives (e.g., The Paris Agreement under the UN Framework Convention on Climate Change, The Sendai Framework for Disaster Risk Reduction 2015-2030 and the UN Decade of Ocean Science for Sustainable Development (2021-2030)). These initiatives highlight the lack of comprehensive global bathymetric coverage, a recognised fundamental element to achieve the goals of each.
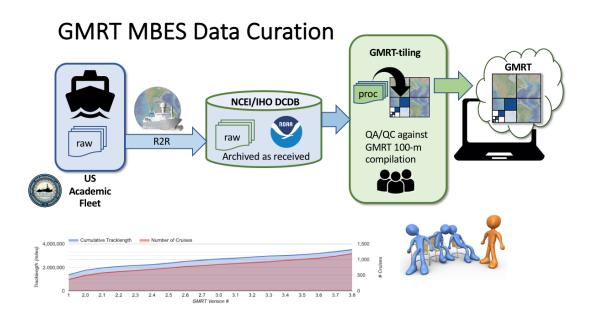

Less than 18% of the deep ocean floor has been mapped with direct measurement and approximately 50% of the world's coastal waters remain unsurveyed. (Source: GEBCO 2019)

Data contributors

Government

Industry

Academia

International Projects

Crowd sourced Bathymetry

Others

IHO Data Center for Digital Bathymetry (DCDB)

Data uses

Hydrographic Offices

GEBCO Products

Regional Mapping

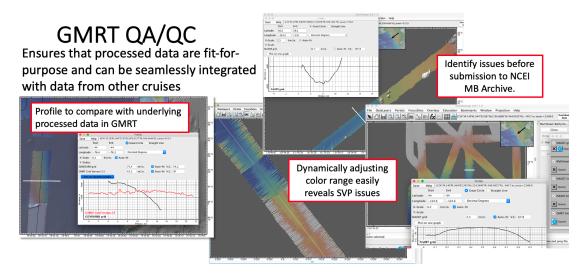Web Applications

International Agencies

Others

THE NEED FOR COORDINATED COMMUNITY DATA PROCESSING

Transforming the raw data made available in the NCEI archive into high-quality data products can require significant effort.
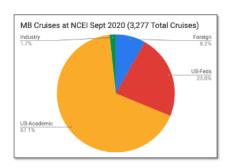

GMRT v3.8 includes processed data from 1,192 research cruises conducted world-wide. Data processing of the vast majority of these cruises was performed by the GMRT Team and focuses on addressing bad soundings, sound velocity corrections and other corrections that arise based on comparison of cruise data with underlying high-resolution data that have already been integrated.
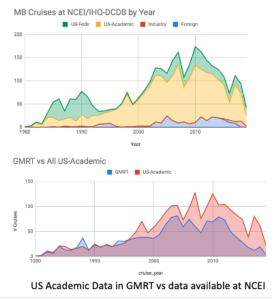

With growing volumes of data and increased efforts to build regional and global bathymetric syntheses, it is difficult for a single project to keep pace with available data. While algorithms may eventually help us programmatically work through the vast collection of global ocean mapping data, developing common tools for data QA/QC leverage the efforts of experts around the globe while establishing standards for fit-for-purpose data products is important.

Growing data volumes of raw MB data at NCEI

MB Cruises at NCEI Sept 2020 (3,277 Total Cruises)

Industry 1.7%
Foreign 8.2%
US-Feds 23.0%
US-Academic 67.1%

Data contributions by sector (based on number of cruises) in public holdings in NCEI multibeam archive Sept. 2020

MB Cruises at NCEI/IHO-DCDB by Year

US-Feds, US-Academic, Industry, Foreign

GMRT vs All US-Academic

GMRT, US-Academic

US Academic Data in GMRT vs data available at NCEI

A few vessels in the US routinely process data as part of their standard operating procedures and deliver these processed data files to the archive. These processed data are accepted by NCEI and made publicly available with no quality control or process process to ensure that they are fit for purpose. As a result, when processed data are downloaded from the NCEI archive for integration into GMRT, additional data processing is often required to ensure that the quality of data being integrated meets the standard of the GMRT Synthesis.

CRUISE	# Files Affected	Average Depth (m)	TVU	MAGNITUDE Observed Offset (M)*
EXAMPLE 1	168	848	19.504	6
	25	522	12.006	16
	150	570	13.11	30
	5	227	5.221	43
EXAMPLE 2	17	280	6.44	46
	15	816	18.768	10
	5	708	16.284	14
EXAMPLE 3	2	262	6.026	20
EXAMPLE 4	5	203	4.669	11
	4	796	18.308	6
EXAMPLE 5	4	788	18.124	27
EXAMPLE 6	11	82	1.886	5
	5	126	2.898	12
	100	94	2.162	11
EXAMPLE 7	50	92	2.116	7
EXAMPLE 8	40	146	3.358	13
EXAMPLE 9	8	77	1.771	7
	4	681	15.663	28
EXAMPLE 10	15	855	19.665	10

*A red value indicates the magnitude is outside the IHO standards for Total Vertical Uncertainty (TVU)
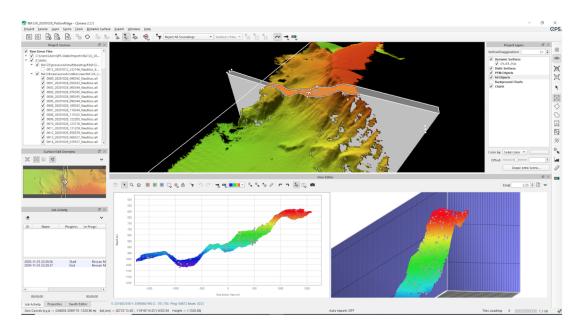
While the International Hydrographic Organization (IHO) has established strict standards for data accuracy in shallow water, these standards are not suitable for deep water data. The table above shows examples from 10 cruises of data for which processed swath files were downloaded from NCEI and the quality of those data files did not meet data integration standards of GMRT. These data require additional data processing and corrections to make them fit for purpose and ready for integration. During this particular exercise of reviewing processed swath files from 10 cruises at NCEI, more than 5,000 were reviewed, and ~17% of them required additional work before they could be integrated into GMRT, and ~10% of them had vertical uncertainty that exceeded IHO recommendations.
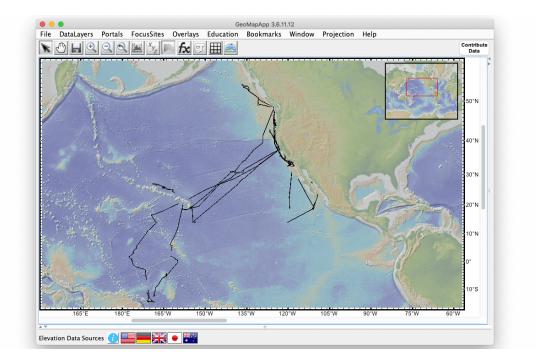
E/V NAUTILUS CONTRIBUTIONS TO GMRT

The E/V *Nautilus* is operated by the Ocean Exploration Trust and is equipped wtih a Kongsberg EM302 Multibeam Echosounder. Mapping data are acquired on all cruises as part of routine operations to support ROV dives and ocean exploration.

The multibeam bathymetry, backscatter and water column data are processed during the cruise with QPS software, and a set of data products are delivered at the conclusion of each expedition. These include edited data as GSF files, bathymetry grids, and backscatter mosaics.
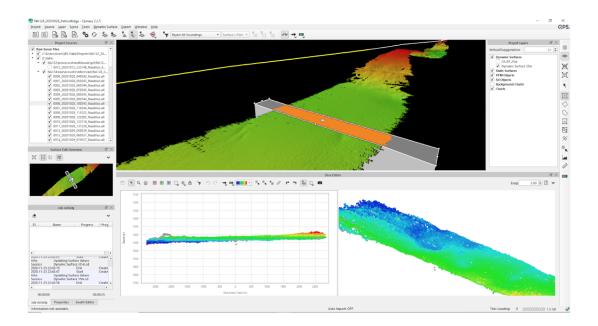

Earlier this year, processed multibeam data from 28 E/V *Nautilus* cruises were prepared for integration into GMRT before they were submitted to NCEI for long term archiving and public access. As part of this process, data from 5 cruises were found to require additional editing before they could be integrated into GMRT. Colleagues at OET performed the additional processing and the data were integrated into GMRT. Combined, these data cover more than 300,000 km2 of seafloor in the Pacific Ocean.
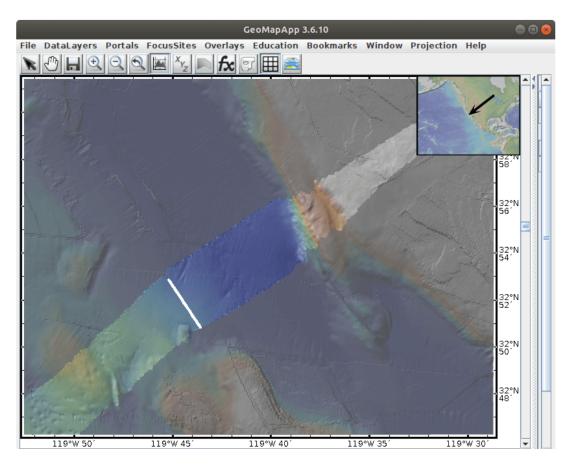
GMRT TILING AT-SEA WITH E/V NAUTILUS

The tools for GMRT tiling/gridding were recently installed aboard the E/V *Nautilus*. The use of these tools has been added to Standard Operating Procedures of the *Nautilus* to ensure that processed data meet community standards and can be readily integrated with publicly available data acquired during other cruises.
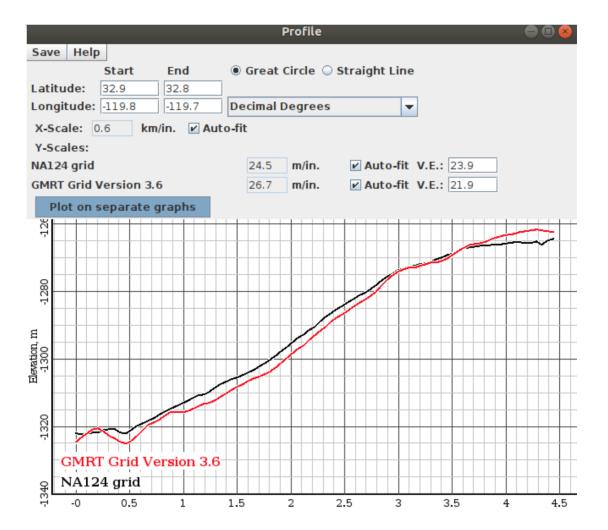

During NA124, EM302 multibeam sonar data was collected opportunistically between ROV dive sites as well as on site to prepare for ROV operations. The Kongsberg ALL files were processed by the OET team during the expedition using QPS Qimera. Once processing and normal QC were completed, GSF (Generic Sensor Format) files were exported. GSF files contain all of the soundings as well as processing flags, allowing OET to provide users with a 'clean' data set while still providing access to the full sounding set.

The resulting GSF files were then used to create GMRT tiles, which were subsequently loaded into an at-sea version of GeoMapApp (no internet access required). The tiles were scanned GeoMapApp for any missed outliers or sound speed issues.
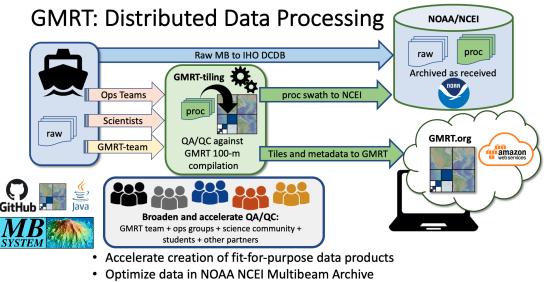
Additionally, the new GMRT tiles were compared to MBES data already integrated into the GMRT dataset; consistency with underlying multibeam data was taken as an indication of good data quality. In the image above, the white line corresponds to the profile below.

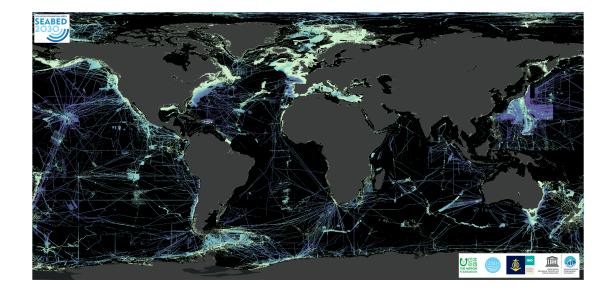# BENEFITS OF COORDINATED COMMUNITY DATA PROCESSING

The GMRT griding/tiling tools are now available in beta-form for use by members of the community interested in contributing to grow content in GMRT. We envision that this will accelerate the world toward producing a high-quality global ocean map, while distributing the effort of data processing and QA/QC. The adoption of this tool also ensures that processed deep-water multibeam data submitted to the NCEI archive are of a consistent quality that are fit for purpose. This not only improves the quality of publicly available processed data but leverages the efforts of many individuals around the world to process data. Recognizing that this effort relies heavily on the work of many individuals, full attribution to source data, as well as data processors is a core component of the GMRT metadata schema.

- Accelerate creation of fit-for-purpose data products
- Optimize data in NOAA NCEI Multibeam Archive
- Minimize need to reprocess and version data in NOAA NCEI

The GMRT tiling/gridding approach can be beneficial to regional, national and international efforts, including those related to the Presidential Memorandum on Ocean Mapping and Exploration. Our current goals are to distribute these tools to colleagues in the US research community, but they are also available for international use. In order to avoid redundancy of effort in data processing, a dashboard is currently being developed to help coordinate shore-side data processing efforts with the goal of leveraging the GMRT infrastructure and approach to data processing to help accelerate the pace of building data products. The dashboad will also be a critical component of engaging students in data processing to build a high-quality community-built data product.

Data integrated into GMRT through the use of these tools are contributed to the Nippon Foundation - GEBCO Seabed 2030 Project and have been included in the GEBCO World Ocean Map since 2014.

AUTHOR INFORMATION

Vicki Ferrini, LDEO - Technical Lead and co-PI for GMRT

John J Morton, LDEO - Applications Developer for GMRT

Lindsay Gee, OET - Mapping and Science Coordinator for OET

Erin Heffron, LDEO/OET - Data Specialist

Hayley Drennon, LDEO - Data Specialist and Data Manager

Nicole Raineault, OET - Chief Scientist for OET

Suzanne M Carbotte, LDEO - PI for GMRT and GeoMapApp

ABSTRACT

The Global Multi-Resolution Topography (GMRT) Synthesis is an elevation model that includes curated deep-water multibeam bathymetry data at ~100 m resolution covering more than 9% of the ocean. GMRT is built with a scalable tiled raster architecture that efficiently stores and presents high-resolution elevation data nested within low resolution data. A set of tools are available for users to access the compilation through simple user interfaces (e.g. GMRT MapTool) and web services, while also providing full attribution and access to source swath files.

The availability of raw/unprocessed multibeam sonar data in the National Centers for Environmental Information (NCEI) archive has increased dramatically over the last decade, but transforming these data into high-quality integrated products suitable for use by scientists and the public alike requires significant effort. The GMRT Team has built workflows and tools for data preparation and review that are optimized for cleaning and integrating sparse globally distributed multibeam data, enabling the addition of ~60-80 research cruises per year. Once raw swath data files are cleaned and corrected, they are gridded/tiled with the GMRT Tiling tools so they can be reviewed and quality controlled in the context of other data in the GMRT Synthesis. Working with processed swath files generated by the community, we have observed that this process frequently reveals issues that are overlooked during data processing.

In order to accelerate the rate of data integration and leverage the data processing efforts of the community, GMRT Tiling tools are being adapted for distributed use. Ocean Exploration Trust is an initial partner in this effort, and all processed swath files from the 2017-2019 *Nautilus* field seasons were prepared with GMRT Tiling tools and reviewed by the OET team. This revealed problems in processed swath data files from several cruises that were addressed prior to submission to NCEI, thereby improving the quality of data in the archive. We are now working to include the GMRT Tiling tools into at-sea standard operating procedures of the *Nautilus* as a testbed for broader community distribution, to ensure consistent quality of processed multibeam data, and to accelerate the production of high-quality integrated data products including GMRT and Seabed 2030.

REFERENCES

GeoMapApp, website: http://www.geomapapp.org

Global Multi-Resolution Topography (GMRT), website: https://www.gmrt.org

NOAA NCEI Multibeam Archive, website: https://www.ngdc.noaa.gov/mgg/bathymetry/multibeam.html (https://www.ngdc.noaa.gov/mgg/bathymetry/multibeam.html)

Memorandum on Ocean Mapping of the United States Exclusive Economic Zone and the Shoreline and Nearshore of Alaska, 2019: https://www.whitehouse.gov/presidential-actions/memorandum-ocean-mapping-united-states-exclusive-economic-zone-shoreline-nearshore-alaska/ (https://www.whitehouse.gov/presidential-actions/memorandum-ocean-mapping-united-states-exclusive-economic-zone-shoreline-nearshore-alaska/)

Multibeam Adivsory Committee (MAC), website: http://mac.unols.org

Rolling Deck to Repository (R2R), website: https://rvdata.us