

# Comprehensive Inventory of Seafloor Mapping, Ground-truthing, and Predictive Habitat Modeling Datasets to Support *Deepwater Horizon* Mesophotic and Deep Benthic Communities

## Data Report

### Authors

Avery B. Paxton<sup>1</sup>, Jennifer Herting<sup>2</sup>, Arliss J. Winship<sup>3</sup>, Erik Ebert<sup>3</sup>, Will Sautter<sup>3</sup>, Jennifer Kraus<sup>1</sup>, Sarah D. Hile<sup>3</sup>, Jacob Howell<sup>3</sup>, Daniel Dorfman<sup>3</sup>, Samantha Flounders<sup>4</sup>, Matthew Poti<sup>1</sup>, Michael Coyne<sup>3</sup>, Stephen Formel<sup>5</sup>, Asmita Shukla<sup>3</sup>, J. Christopher Taylor<sup>1</sup>, Kristopher Benson<sup>6</sup>, Stacey L. Harter<sup>4</sup>, Tim Battista<sup>1</sup>, and Charles Menza<sup>1</sup>

<sup>1</sup> NOAA, National Ocean Service, National Centers for Coastal Ocean Science

<sup>2</sup> Tech Global, Inc, Under contract to NOAA, National Marine Fisheries Service, Southeast Fisheries Science Center

<sup>3</sup> CSS, Inc., Under contract to NOAA, National Ocean Service, National Centers for Coastal Ocean Science

<sup>4</sup> NOAA, National Marine Fisheries Service, Southeast Fisheries Science Center

<sup>5</sup> NOAA, National Environmental Satellite, Data, and Information Service, National Centers for Environmental Information

<sup>6</sup> NOAA, National Marine Fisheries Service, Office of Habitat Conservation

March 2023

DWH MDBC Data Report DR-23-02



## Suggested Citation

Paxton, A.B., J. Herting, A.J. Winship, E. Ebert, W. Sautter, J. Kraus, S.D. Hile, J. Howell, D. Dorfman, S. Flounders, M. Poti, M. Coyne, S. Formel, A. Shukla, J.C. Taylor, K. Benson, S.L. Harter, T. Battista, and C. Menza. 2023. Comprehensive Inventory of Seafloor Mapping, Ground-truthing, and Predictive Habitat Modeling Datasets to Support *Deepwater Horizon* Mesophotic and Deep Benthic Communities. National Oceanic and Atmospheric Administration. DWH MDBC Data Report DR-23-02. Silver Spring, MD. 73 pp. doi: <https://doi.org/10.25923/ptv8-vb15>

## Acknowledgements

Multiple individuals helped facilitate inventory compilation. David Sallis, Kate Rose, Amber Batts, Jessica Morgan, and Kirsten Larsen contributed to data management. Mark Mueller, Alicia Caporaso, Bill Shedd, Amanda Demopoulos, Jason Chaytor, Megan Cromwell, Marissa Nuttall, and Jenn Jencks served as technical advisors. Special thank you to Gabe Patterson-King for his support with the NOAA Geoplatform and more. Multiple other individuals across NOAA and other institutions provided data, technical expertise, and more. We thank the MDBC Technical Team, Steering Committee, and Leadership Team for their guidance throughout the inventory process. The report was reviewed by MDBC Technical Team, Steering Committee and Open Ocean Trustee Implementation Group (TIG). CSS, Inc. employees were supported under NOAA Contract No. EA133C17BA0062 and EA133C17BA0049.

For more information on MDBC Restoration, please visit:

<https://coastalscience.noaa.gov/project/scientific-support-for-mesophotic-and-deep-benthic-community-restoration-in-the-gulf-of-mexico/>

## Disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## ***Deepwater Horizon* Mesophotic and Deep Benthic Communities Restoration**

This report is part of the NOAA Mesophotic and Deep Benthic Communities (MDBC) Series of publications that share the results of work conducted by the *Deepwater Horizon* MDBC restoration projects.

The 2010 *Deepwater Horizon* oil spill was an unprecedented event. Approximately 3.2 million barrels of oil were released into the deep ocean over nearly three months. The plume of oil moved throughout the water column, formed surface slicks that cumulatively covered an area the size of Virginia, and washed oil onto at least 1,300 miles of shoreline habitats. More than 770 square miles (2,000 square kilometers) of deep benthic habitat surrounding the *Deepwater Horizon* wellhead and 4 square miles of the Pinnacles mesophotic reef complex, located at the edge of the continental shelf, were injured by the oil spill.

Under the Oil Pollution Act, state and federal natural resource trustees conducted a Natural Resource Damage Assessment (NRDA). The Trustees assessed damages, quantifying the unprecedented injuries to natural resources and lost services. They also developed a programmatic restoration plan to restore injured resources and compensate the public for lost services.

In April 2016, a settlement was finalized that included up to \$8.8 billion in funding for the Deepwater Horizon Trustees to restore the natural resource injuries caused by the oil spill as described in their programmatic restoration plan, Final Programmatic Damage Assessment and Restoration Plan and Final Programmatic Environmental Impact Statement. The Deepwater Horizon Open Ocean Trustee Implementation Group is responsible for restoring natural resources and their services within the Open Ocean Restoration Area that were injured by the oil spill. The Open Ocean Trustees include NOAA, U.S. Department of the Interior (DOI), U.S. Environmental Protection Agency, and U.S. Department of Agriculture.

In 2019, the Open Ocean Trustee Implementation Group committed more than \$126 million to implement four restoration projects to address the injury to MDBC. The MDBC projects are: Mapping, Ground-truthing, and Predictive Habitat Modeling; Habitat Assessment and Evaluation; Coral Propagation Technique Development; and Active Management and Protection. NOAA and DOI are implementing the projects, in cooperation with a range of partners, over eight years.

Together, the projects take a phased approach to meet the challenges involved in restoring deep-sea habitats. Challenges to restoration include a limited scientific understanding of these communities, limited experience with restoration at the depths at which these communities occur, and remote locations that limit accessibility.

More information about *Deepwater Horizon* restoration and the MDBC restoration projects is available at: [www.gulfspillrestoration.noaa.gov](http://www.gulfspillrestoration.noaa.gov).

# Table of Contents

Executive Summary.....	1
1. Background.....	1
2. Inventory Scope .....	3
2.1. Broad Inclusion Criteria .....	3
2.1.1. Geographic Extent .....	3
2.1.2. Temporal Extent .....	4
2.2. Fine Scale Inclusion Criteria for Seafloor Mapping.....	4
2.2.1. Data Types .....	4
2.2.2. Resolution.....	6
2.3. Fine Scale Inclusion Criteria for Ground-truthing.....	6
2.3.1. Data Types .....	6
2.4. Fine Scale Inclusion Criteria for Predictive Habitat Modeling .....	8
2.4.1. Data Types .....	8
3. Data Discovery.....	9
3.1. Broad Sources.....	9
3.1.1. Seafloor Mapping.....	9
3.1.2. Ground-truthing.....	11
3.1.3. Predictive Habitat Modeling .....	12
3.2. Literature Search.....	12
3.3. Partner Engagement.....	14
3.4. External Engagement .....	15
3.5. Portals.....	15
4. Data Tracking.....	16
4.1. Data Catalog.....	16
4.2. Unique Data Identifier.....	16
4.3. Data Parent and Child Structure.....	16
5. Data Screening and Acquisition.....	17
6. Data Processing.....	19
6.1. Data Processing for Seafloor Mapping .....	19
6.1.1. Goal.....	19
6.1.2. Dataset Attributes .....	19
6.1.3. File Naming Conventions.....	19
6.1.4. Folder Structure.....	20
6.1.5. ArcGIS Pro Projects.....	21

6.1.6. Custom Python and ArcPro Tools.....	21
6.1.7. Raw Multibeam Data Processing in Qimera .....	23
6.1.8. Fishery Multibeam Data Processing .....	23
6.2. Data Processing for Ground-truthing .....	24
6.2.1. Goal.....	24
6.2.2. Dataset Attributes .....	24
6.2.3. File Naming Conventions and Folder Structure.....	24
6.2.4. Dataset Reformatting.....	24
6.2.5. Footprints .....	24
6.3. Data Processing for Predictive Habitat Modeling.....	25
6.3.1. Goal.....	25
6.3.2. Dataset Attributes .....	25
6.3.3. File Naming Conventions and Folder Structure.....	25
6.3.4. Data Reformatting.....	25
6.3.5. Footprints .....	26
7. Inventory products.....	27
7.1. Data.....	27
7.2. Data Catalog.....	27
7.3. Literature Library .....	27
7.4. Footprints.....	27
7.5. Mapping Imagery .....	27
7.6. Ground-truthing Annotations.....	28
7.7. Habitat Modeling Predictions .....	28
7.8. Geodatabases.....	28
7.8.1. Base Layer Geodatabase.....	28
7.8.2. Seafloor Mapping Geodatabases.....	28
7.8.3. Ground-truthing Geodatabase.....	29
7.8.4. Predictive Habitat Modeling Geodatabase .....	29
7.9 Inventory Access .....	29
7.10. Archival Products.....	30
8. Results.....	31
8.1. Seafloor Mapping Results .....	31
8.1.1. Overall.....	31
8.1.2. Data Source .....	32
8.1.3. Sensor.....	36

8.1.4. Collection Decade .....	39
8.1.5. Resolution .....	43
8.1.6. Depth.....	47
8.1.7. Derived Datasets .....	48
8.1.8. Imagery.....	51
8.1.9. Mapping Datasets in Inventory Queue .....	53
8.2. Ground-truthing Results .....	53
8.2.1. Overall.....	53
8.2.2. Data Collection Type .....	54
8.2.3. Data Source .....	56
8.2.4. Data Collection per Decade.....	60
8.2.5. Data Collection by Depth .....	62
8.2.6. Ground-truthing Datasets in Inventory Queue .....	65
8.3. Predictive Habitat Modeling Results.....	65
9. Discussion and Conclusion .....	69
9.1. Overall Patterns .....	69
9.2. Lessons Learned and Best Practices .....	69
9.3. Updating the Inventory .....	70
9.4. Planned Uses of Data Inventory.....	70
References .....	71
Appendix.....	73

## List of Figures

Figure 1. The geographic extent of the inventory was the northern Gulf of Mexico north of 27°N latitude and generally deeper than 50 m. ....	4
Figure 2. Schematic of predictive habitat modeling process. The example shown is a predictive model of the spatial distribution of a biological taxon, but predictive models can also be developed for less specific biological variables (e.g., biological cover) or physical habitat characteristics (e.g., to predict the spatial distribution of a habitat class).....	8
Figure 3. Data storage workflow. Source data were “ingested” into a shared drive system for later review and processing.....	17
Figure 4. Footprints and tracklines for seafloor mapping data included in the inventory. These footprints and tracklines do not include derived products (Section 8.1.7). ....	32
Figure 5. Logarithmic cumulative area (km <sup>2</sup> ) of seafloor mapping footprints (A) and tracklines (B) by source. Numbers above bars represent the number of mapping survey footprints or trackline segments that comprise the cumulative area value.....	33
Figure 6. Footprints and tracklines for seafloor mapping data acquired from other sources.....	34
Figure 7. Footprints for seafloor mapping data acquired from NOAA sources. ....	34
Figure 8. Footprints and tracklines for seafloor mapping data acquired from USGS sources. ....	35
Figure 9. Footprints and tracklines for seafloor mapping data acquired from academic sources.....	35
Figure 10. Footprints for seafloor mapping data acquired from industry sources. ....	36
Figure 11. Logarithmic cumulative area (km <sup>2</sup> ) of seafloor mapping footprints (A) and tracklines (B) by sensor. Numbers above bars represent the number of mapping survey footprints or trackline segments that comprise the cumulative area value.....	37
Figure 12. Footprints and tracklines for seafloor mapping data collected with multibeam echosounders. ....	37
Figure 13. Footprints for seafloor mapping data collected with side-scan sonar. ....	38
Figure 14. Footprints for seafloor mapping data collected with fishery multibeam echosounders. ....	38
Figure 15. Tracklines for seafloor mapping data collected with seismic profilers. ....	39
Figure 16. Cumulative area (km <sup>2</sup> ) of seafloor mapping footprints (A) and tracklines (B) by data collection decade. Numbers above bars represent the number of mapping survey footprints or trackline segments that comprise the cumulative area value.....	40
Figure 17. Footprints and tracklines for seafloor mapping data from the 1980s. This includes data whose acquisition ended in the 1980s.....	41
Figure 18. Footprints and tracklines for seafloor mapping data from the 1990s. This includes data whose acquisition ended in the 1990s.....	41
Figure 19. Footprints and tracklines for seafloor mapping data from the 2000s. This includes data whose acquisition ended in the 2000s.....	42
Figure 20. Footprints and tracklines for seafloor mapping data from the 2010s. This includes data whose acquisition ended in the 2010s.....	42
Figure 21. Footprints for seafloor mapping data from the 2020s. This includes data whose acquisition ended in the 2020s.....	43
Figure 22. Cumulative area (km <sup>2</sup> ) of seafloor mapping footprints by resolution (m). Numbers above bars represent the number of mapping survey footprints that comprise the cumulative area value.....	44

Figure 23. Footprints for seafloor mapping data with resolution from <1 m.....	44
Figure 24. Footprints for seafloor mapping data with resolution from 1 m to 10 m.....	45
Figure 25. Footprints for seafloor mapping data with resolution from 10 m to 32 m.....	45
Figure 26. Footprints for seafloor mapping data with resolution from 32 m to 100 m.....	46
Figure 27. Footprints for seafloor mapping data with resolution >100 m.....	46
Figure 28. Footprints and tracklines for seafloor mapping data with coarse resolution (e.g., point or raw data).....	47
Figure 29. Footprints and tracklines for seafloor mapping data by depth bins. Depth contours represent 30 m, 300 m, 800 m, and 2500 m.....	47
Figure 30. Histogram of average depths for seafloor mapping footprints. Depth bins are 200 m each. Dotted black line represents the mean depth (760 m).....	48
Figure 31. BOEM derived data products: A) polygon of high-resolution geophysical surveys; B) interpretations of bottom anomalies from seismic surveys; and C) polygons corresponding to the extent of the BOEM Bathymetry Grid from 3D Seismic.....	49
Figure 32. USGS interpretations of seafloor features derived from side-scan sonar: A) salt diapirs (also includes terrestrial assessments with other sensors), B) mass wasting locations, C) geomorphology, D) hydrate locations, and E) hydrocarbon seeps.....	50
Figure 33. NOAA derived products: A) locations of fishery multibeam surveys conducted from the NOAA ship <i>Pisces</i> ; and B) locations of side-scan sonar surveys conducted by the NOAA NMFS Panama City Lab.....	50
Figure 34. Locations of side-scan sonar surveys conducted by Florida Fish and Wildlife Research Institute (FWC/FWRI).....	51
Figure 35. Polygons corresponding to existing A) Fugro bathymetry data and B) PGS seismic survey data.....	51
Figure 36. Representative seafloor mapping bathymetric imagery collected with multibeam echosounders from fine to coarse resolutions: A) multibeam collected from an autonomous underwater vehicle (Marine Geoscience Digital Library); B) shipboard multibeam of Elvers Bank in the Flower Gardens National Marine Sanctuary (NOAA); and C) shipboard multibeam of Green and Mississippi Canyons (NOAA).....	52
Figure 37. Representative seafloor mapping mosaic imagery collected with side-scan sonar from fine to coarse resolutions collected by: A) FWC/FWRI, B) NOAA NMFS Panama City Lab, and C) USGS GLORIA survey.....	52
Figure 38. Ground-truthing surveys represented by biological or geological collection surveys, bottom video sites, or submersible tracklines.....	54
Figure 39. Biological collections conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.....	54
Figure 40. Bottom video sites conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.....	55
Figure 41. Geological collections conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.....	55
Figure 42. Submersible vehicle and towed camera system tracklines conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.....	56



Figure 43. Bar plots identifying the total number of ground-truthing datasets that were acquired from various sources used for the purpose of this inventory (A), and the number of datasets acquired per source and the ground-truthing data type (B).....	57
Figure 44. Ground-truthing surveys that were acquired from NOAA sources as biological collection surveys, bottom video surveys, and submersible surveys conducted in the northern Gulf of Mexico from 1980 to present.....	58
Figure 45. Ground-truthing surveys that were acquired from USGS sources as biological and geological collection surveys and submersible surveys conducted in the northern Gulf of Mexico from 1980 to present.....	58
Figure 46. Ground-truthing surveys that were acquired from academic sources as biological collection surveys, bottom video surveys, and submersible surveys conducted in the northern Gulf of Mexico from 1980 to present.....	59
Figure 47. Ground-truthing surveys that were acquired from the Smithsonian (SMITH), as biological collection surveys conducted in the northern Gulf of Mexico from 1980 to present.....	59
Figure 48. Ground-truthing surveys that were acquired from BOEM, as bottom video surveys conducted in the northern Gulf of Mexico from 1980 to present.....	60
Figure 49. Ground-truthing surveys that were acquired from other sources outside of external MDBC partners and proprietary efforts, as biological and geological collection surveys, bottom video surveys, and submersible surveys in the northern Gulf of Mexico from 1980 to present.....	60
Figure 50. A bar plot showing the types of ground-truthing data collected per decade between 1980 to present. The Multi_decade datasets are those surveys that spanned over multiple decades.....	61
Figure 51. Ground-truthing surveys that were conducted over multiple decades (Multi_Decade) as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.....	61
Figure 52. Ground-truthing surveys that were conducted from the year 2000 to 2009 as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.....	62
Figure 53. Ground-truthing surveys that were conducted from the year 2010 to 2019 as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.....	62
Figure 54. A bar plot showing different ground-truthing dataset surveys conducted over mesophotic (50m to 300m), deep (300m and greater), and both mesophotic and deep (Meso_Deep) depth ranges.....	63
Figure 55. Ground-truthing surveys that were conducted in mesophotic depth range between 50m and 300m as biological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.....	63
Figure 56. Ground-truthing surveys that were conducted in a depth range greater than 300m as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.....	64
Figure 57. Ground-truthing surveys that were conducted in both mesophotic and greater depth range, 50m and greater as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.....	64
Figure 58. Ground-truthing surveys that were conducted in depths greater than 50m as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.....	65
Figure 59. Footprints for predictive habitat modeling data included in the inventory, excluding datasets with global coverage.....	66
Figure 60. Footprints for relatively local predictive habitat modeling data included in the inventory.....	66

Figure 61. Number of predictive habitat modeling datasets by A) model type, B) predicted response type, C) method, and D) spatial resolution. BRT refers to boosted regression tree modeling, NNL refers to a nearest neighbor learning method, and RF refers to random forest modeling.....67

## List of Tables

Table 1. Broad and fine scale inclusion criteria for data to be included in the inventory.....	3
Table 2. Data sensors targeted for inclusion in the seafloor mapping component of the inventory .....	5
Table 3. Platforms from which seafloor mapping data included in the inventory may be collected.....	5
Table 4. Ancillary data types that may accompany seafloor mapping data and were included in the inventory on ad hoc bases.....	6
Table 5. The targeted gear types and/or sensors for the ground-truthing data inventory. Also included are short descriptions and relevance to the MDBC ground-truthing data inventory.....	7
Table 6. Platforms where data types in Table 5 can be collected that were included in the inventory.....	7
Table 7. Broad sources from which data were discovered for the inventory. ....	9
Table 8. BOEM data sources and their status for inclusion or exclusion in seafloor mapping inventory. Description provides accompanying details.....	10
Table 9. Number of articles returned in the literature search (initial search) and included or excluded in either the title and abstract screening, full text screening, or dataset screening.....	14
Table 10. Attributes for datasets (e.g., sensor, source) and products with accompanying descriptions and codes; these codes were used for file naming.....	20
Table 11. Description of tools developed to process seafloor mapping data and incorporate them into the inventory.....	21
Table 12. Groups and bins used to display seafloor mapping feature layers within geodatabases. ....	29
Table 13. Description of the subgroups of layers in the ground-truthing geodatabase. ....	29
Table 14. Number of seafloor mapping datasets screened according to coarse and fine scale inclusion criteria.....	31
Table 15. Seafloor mapping summary statistics for footprints and tracklines (type) by source from which data were acquired.....	33
Table 16. Seafloor mapping summary statistics for footprints and tracklines (type) by sensor.....	36
Table 17. Seafloor mapping summary statistics for footprints and tracklines (type) by data collection decade.....	40
Table 18. Seafloor mapping summary statistics for footprints and tracklines (type) by resolution .....	43
Table 19. Number of ground-truthing datasets screened according to coarse and fine scale inclusion criteria. Numbers of datasets excluded and included according to the inclusion criteria are also provided..	53
Table 20. The total count as well as percentages were calculated per each source of surveys inventoried for fish ecology datasets.....	56
Table 21. Methods employed by predictive habitat modeling datasets in the inventory.....	68

## **Commonly Used Acronyms**

AUV	autonomous underwater vehicle
BOEM	Bureau of Ocean Energy Management
DWH	Deepwater Horizon
ESPIS	BOEM's Environmental Studies Program Information System
FWC	Florida Fish and Wildlife Conservation Commission
FWRI	FWC's Fish and Wildlife Research Institute
GLORIA	Geological Long-Range Inclined Asdic
GOM	Gulf of Mexico
MBES	multibeam echosounder
MDBC	mesophotic and deep benthic communities
NCEI	NOAA's National Centers for Environmental Information
NMFS	NOAA's National Marine Fisheries Service
NOAA	National Oceanic and Atmospheric Administration
NRDA	Natural Resource Damage Assessment
NSF	National Science Foundation
ROV	remotely operated vehicle
USGS	U.S. Geological Survey

## **Key Definitions**

### *Data extraction*

The process of retrieving data and associated metadata from datasets that pass the data screening process based on their adherence to defined inclusion criteria.

### *Data Inventory and Analysis Implementation Plan*

A systematic and reproducible plan for compiling the comprehensive inventory. Also referred to as “Inventory Plan.”

### *Data screening*

The process of reviewing datasets for inclusion in the inventory based on whether the data meet defined inclusion criteria. Data screening occurs in multiple phases, where data are reviewed for adherence to broad or coarse level inclusion criteria and then reviewed again at a narrower level for adherence to finer-scale inclusion criteria. Datasets that pass the screening steps are those from which data will be extracted and/or compiled.

### *Deep or deepwater benthic habitats*

Areas on the seafloor that exist in the depth zone where light is not present and are inhabited by an interacting group of benthic species. For the purpose of this MDBC project, work in deep benthic habitats will generally focus on depths deeper than 980 feet (300 m) deep. The dominant communities providing structural habitat among deepwater benthic habitats comprise corals and sponges.

### *Ground-truthing*

The process of collecting in-situ observations to check the accuracy of MDBC habitat model predictions and information documenting the abundance and distribution of MDBC habitats (OO TIG 2019 Section 4.4.6.1). The specific collection methods and types of data to collect will be determined during the planning phase of this project.

### *Inclusion criteria*

Objective standards or criteria used to evaluate datasets for inclusion in the inventory. Inclusion criteria can be broad or fine-level, and they set the coarse and fine filters for data compilation, respectively.

### *Inventory*

For the purposes of this report, refers to a collection of seafloor mapping, ground-truthing and predictive habitat modeling data, literature, metadata, and derived products.

### *Mapping Requirements*

Defines the intended use or application of the seafloor mapping and ground-truthing data. Requirements are provided by the data users and includes information on the estimated size or extent of features to be mapped, or expected level of uncertainty in measurements, qualities or quantities of features on the seabed. Requirements often have a dimension, scale or extent for reference (e.g., defining hard vs. soft bottom, delineating habitat classes at defined taxonomic resolution, identifying and locating individuals of key taxa, etc.)

### *Mesophotic and Deep Benthic Communities (MDBC)*

MDBC are an interacting group of benthic species, including fish, mobile and sessile invertebrates and benthic infauna, that grow within mesophotic and deepwater benthic habitats. Some taxa, such as mesophotic corals, create biogenic habitats and are both a taxonomic unit of the community and define the habitat. In the context of the DWH Programmatic Damage Assessment and Restoration Plan, MDBC are defined as a restoration type.

### *MDBC habitats*

Areas on the seafloor inhabited and/or associated with mesophotic and deep benthic communities. MDBC habitats include mesophotic reefs, deep-sea corals, soft-bottom habitats and non-natural substrates (e.g., oil and gas structures, artificial reefs) located on the seafloor. MDBC habitats may be characterized by geomorphological forms, substrates or biological cover depending on data availability.

### *Mesophotic benthic habitats*

Areas on the seafloor that exist in the depth zone where light levels are low and are inhabited by an interacting group of benthic species. Mesophotic habitats are typically found at depths ranging from approximately 100 feet (30 m) and extending to over 500 feet (approximately 150 m) in tropical and subtropical regions (Puglise et al. 2009, Hourigan et al. 2015, Sulak and Dixon 2015, DWH NRDA Trustees 2016). For the purpose of these MDBC projects, work in mesophotic habitats will generally focus on biological communities existing between approximately 165 feet (50 m) and 980 feet (300 m) deep. The use of 980 feet (300 m) as the lower limit of this zone is consistent with BOEM and the Bureau of Safety and Environmental Enforcement's current management practices for benthic resources (BOEM 2021). This depth range has also been referred to as the "twilight zone" (Pyle 1996, Kahng et al. 2010). The dominant communities providing structural habitat among mesophotic benthic habitats comprise coral, sponge and algal species.

### *Model validation*

The process of comparing predictive habitat model predictions to independent data and quantifying how well the predictions match those data through established statistical performance measures and tests. Ideally, data used for validation should be new data collected in a statistically robust manner designed specifically for the purpose of validating a model's predictions. Model validation provides an assessment of model reliability.

### *Predictive Habitat Modeling*

The process of predicting MDBC habitats from remotely sensed information, including seafloor mapping data, and in situ MDBC observations. For this project, predictive habitat modeling will focus on producing refined northern Gulf of Mexico regional-scale models of habitat for mesophotic and deepwater coral species (OO TIG 2019 Section 4.4.6.1). MDBC habitats may be defined by geomorphological forms, substrates or biological cover depending on data availability.

### *Prioritization of Mapping Requirements*

The process of engaging data users across MDBC projects as well as subject matter experts, resource management agencies, and other organizations (both internal and external to the DWH NRDA) to develop prioritization criteria, and identify and prioritize spatial gaps in derived products that characterize the seafloor for the MDBC portfolio. The goal of the prioritization will be to recommend

locations where derived products are needed and to inform plans for future collaborative field work during the implementation phase of this project. This will support DWH NRDA and the MDBC portfolio as approved in Open Ocean RP2.

### *Seafloor Mapping*

Seafloor mapping refers to the process of using remote sensing to survey and derive depth, shape, texture, reflectivity, and backscatter of the seafloor. In the context of this project, mapping is generally used to refer to seafloor mapping, including in the project title. Data from seafloor mapping form the basis in predictive habitat modeling to interpret, classify, and characterize the geomorphological forms, substrates, and benthic biological cover used to define MDBC habitats. We do not include the bottom layers of the water column in the definition of seafloor mapping. For this project, seafloor mapping activities would be undertaken using surface (i.e., ship-based), subsurface (i.e., ROV, AUV, HOV, technical diving), and towed sonar operations, and could also include electromagnetic operations and/or laser operations. Mapping operations will collect high-resolution data to document the distribution and abundance of MDBC habitats and improve existing predictive habitat models (OO TIG 2019 Section 4.4.6.1).

### *Specifications or Survey Specifications*

Describes what the data should look like that allows for delineating or detecting features. The data specifications are determined by mapping and ground-truthing experts and include considerations of resolution and precision to achieve requirements. The specifications help guide the experts to sensors, platforms and optimal survey designs to achieve the desired resolution. A survey specification may achieve many requirements.

# Executive Summary

Successful restoration of MDBC impacted by the DWH oil spill requires knowledge of the distribution, abundance, and status of these communities and associated habitats. Accurate high-resolution seafloor maps and data on abundance and distribution of MDBC represent key requirements for restoration efforts. Partial information about MDBC can be gleaned from existing datasets. As such, one component of the DWH Natural Resources Damage Assessment (NRDA) called for the creation of an inventory of existing seafloor mapping and ground-truthing data and of existing predictive habitat models.

A comprehensive inventory of seafloor mapping, ground-truthing, and predictive habitat modeling data was compiled for a large part of the northern GOM for years 1980-2021 to meet the DWH restoration goals and support restoration activities. Thorough and reproducible data discovery protocols were applied to search a range of data sources including project partners, external contacts, scientific literature, and online data portals. Reproducible protocols were also applied to screen, acquire, standardize, store, and document inventoried datasets. Summary products, such as spatial footprints, were derived from the original datasets to provide simplified, standardized views of the inventoried data across datasets and data types.

The resulting inventory incorporated 693 seafloor mapping footprints, 13,722 seafloor mapping tracklines, 20 seafloor mapping derived products, 737 seafloor mapping raster images, 57 ground-truthing datasets comprising 43 feature layers of ground-truthing surveys, and 12 predictive habitat modeling datasets, which included 198 model prediction layers. Despite the quantity of inventoried datasets, there are large parts of the northern GOM with few or no seafloor mapping or ground-truthing data, and much of the study area also lacks habitat model predictions with a spatial resolution  $\leq 10$  m. Locations where all three data types exist provide the most information about MDBC, yet in many locations one or more data types are lacking.

Footprints of the datasets in the inventory are archived and available online through NOAA's National Centers for Environmental Information (Accession 0268765). While the data inventory will be updated as new data are collected, it currently serves and will continue to support DWH restoration activities for MDBC. The data inventory has already directly facilitated a prioritization exercise with stakeholders and subject matter experts to refine data gaps and prioritize future surveys, and is currently supporting a gap analysis to quantify existing data gaps and inform data collection needs. These data will also support operations planning to meet mapping, ground-truthing, and predictive habitat modeling requirements, for field work to be carried out by the MDBC project teams to document the abundance and distribution of MDBC for restoration where such information is lacking.

# 1. Background

On April 20, 2010, the *Deepwater Horizon* (DWH) mobile drilling unit exploded, resulting in a massive release of oil from the BP Exploration and Production Inc. Macondo well, causing loss of life and extensive natural resource injuries in the Gulf of Mexico (GOM). Oil spread from the deep ocean to the surface and nearshore environment from Texas to Florida. Extensive response actions were undertaken to try to reduce harm to people and the environment. However, many of these response actions had collateral impacts on the environment and on natural resource services.

After the Deepwater Horizon oil spill, federal and state agencies came together to form the DWH Natural Resource Damage Assessment (NRDA) Trustee Council. The DWH Trustees studied the effects of the oil spill and continues to implement projects to restore for injuries to the GOM ecosystem that resulted from the spill. The DWH Trustees documented extensive injuries to deep-sea benthic habitats surrounding the wellhead and extending up the continental slope. Injuries in the deep sea affected mesophotic and deep benthic communities (MDBC). MDBC are largely composed of corals, sponges, infauna, and fish and mobile invertebrates, and are distributed across hard and soft bottom habitats deeper than 50 m. MDBC are vast and complex ecosystems on the ocean floor that are a foundation of GOM food webs and provide food, refuge and reproductive opportunities for multiple species of fish and invertebrates.

The full extent of DWH oil spill impacts to MDBC is unknown. Only small fractions of the mesophotic and deepwater habitats in this region of the GOM have been previously surveyed, and the distributions of MDBC in the impacted area have not been fully determined. For example, coral injuries were documented along the rocky reefs of the Pinnacles Trend (Etnoyer et al. 2016; Silva et al. 2016) but this region has not been surveyed comprehensively since it was surveyed by the U.S. Geological Survey (USGS) in the 2000s (Gardner et al. 2000; CSA and TAMU 2001). Such knowledge gaps regarding MDBC distribution pose a significant challenge for restoring MDBC communities and achieving restoration goals prescribed by the Final Deepwater Horizon Programmatic Damage Assessment and Restoration Plan/Programmatic Environmental Impact Statement (DWH PDARP/PEIS).

Successful restoration of MDBC impacted by the DWH oil spill requires knowledge of the distribution, abundance, and status of these communities and associated habitats. Accurate high-resolution bathymetric and habitat maps, as well as data on the abundance and distribution of MDBC are nearly universal requirements for efforts to restore these communities.

Information on MDBC extent and habitats can be gleaned from existing datasets, such as seafloor mapping and ground-truthing data, as well as predictive habitat models. As such, one component of the DWH NRDA was to create an inventory of existing seafloor mapping data, ground-truthing data, and predictive habitat models. The broad goal of the inventory was to help determine the extent of MDBC communities and support restoration activities. More specifically, the resulting inventory will be used to 1) prioritize locations and key information requirements for habitat maps needed for research planning, restoration activities, or conservation management actions; 2) conduct a gap analysis to define focal regions to conduct new seafloor mapping and ground-truthing surveys or predictive modeling; 3) inform participatory workshops with subject matter experts to define best practices for habitat mapping, ground-truthing, and modeling; and 4) inform operations to newly document the extent of MDBC.



The MDBC Mapping, Ground-truthing, and Predictive Habitat Modeling (MGM) project created a comprehensive inventory by developing a reproducible, systematic approach to discovering existing data and incorporating those data into a standardized inventory framework. This report first details the inventory scope. It then describes the methods used to create the inventory, including data discovery, screening, acquisition, and extraction, before presenting inventory results and spatially explicit summaries specific to seafloor mapping, ground-truthing, and predictive habitat modeling. The report concludes by outlining a plan to update the inventory as new data are collected and discussing lessons learned during inventory compilation, as well as anticipated uses of the inventory.

## 2. Inventory Scope

The inventory aimed to include MDBC data from three broad categories: 1) seafloor mapping, 2) ground-truthing, and 3) predictive habitat modeling. These data categories were selected to provide insight into the abundance and distribution of MDBC and thereby support subsequent restoration activities. Briefly, seafloor mapping data are remote sensing observations used to survey and derive depth, shape, and texture of the seafloor and characterize the geomorphological forms, substrates, and biota used to define MDBC habitats. Ground-truthing data are direct observations or samples that provide visual, physical (e.g., benthic grabs, sediment samples or cores) and analytical characterization of information documenting the abundance and distribution of MDBC, substrate, and geomorphology. Predictive habitat models analyze habitat-environment or species-environment relationships from seafloor mapping data, in situ observations, and other environmental data to predict the distribution, extent, or abundance of MDBC habitats.

Prior to inventory compilation, objective standards or criteria (i.e., inclusion criteria) were defined for evaluating datasets for inclusion in the inventory. Inclusion criteria were broad and fine-level, setting the coarse and fine filters for data compilation, respectively (Table 1). These inclusion criteria are further detailed in the subsequent Sections.

Table 1. Broad and fine scale inclusion criteria for data to be included in the inventory. Category indicates whether the inclusion criteria are broad or fine scale. The criteria type provides a high-level description of the type of inclusion criteria (e.g., data type, geographic extent).

Category	Criteria Type	Inclusion Criteria
broad	geographic extent	Gulf of Mexico waters $\geq 50$ m depth AND north of 27° N latitude (includes intersection with these boundaries)
broad	temporal extent	1980–present
broad	spatial positioning	all data collected need to have accurate spatial positioning or location information
fine: seafloor mapping	data types	seafloor mapping data from identified sensors and platforms that allow identification of seafloor features
fine: seafloor mapping	resolution	data spatial resolution of 400 m or finer
fine: ground-truthing	data types	ground-truthing data from identified gear types and platforms that allow for characterization of seafloor features and their associated biota
fine: predictive habitat modeling	data types	predictions of MDBC habitat from relationships defined or estimated between environmental predictor variables and direct MDBC observations

### 2.1. Broad Inclusion Criteria

Broad inclusion criteria were established to set the coarse-scale filter or scope for the inventory search. These criteria helped constrain the data search by geographic extent and temporal extent.

#### 2.1.1. Geographic Extent

The geographic extent for the inventory was a defined area in the northern GOM that corresponds to the area of interest for MDBC restoration following the DWH oil spill (Figure 1). Specifically, the southern boundary of the inventory was bounded by the 27°N parallel and the northern extent was bounded by the 50 m isobath. If a dataset was partially inside and partially outside of the inventory geographic extent, then the dataset was included in the inventory.

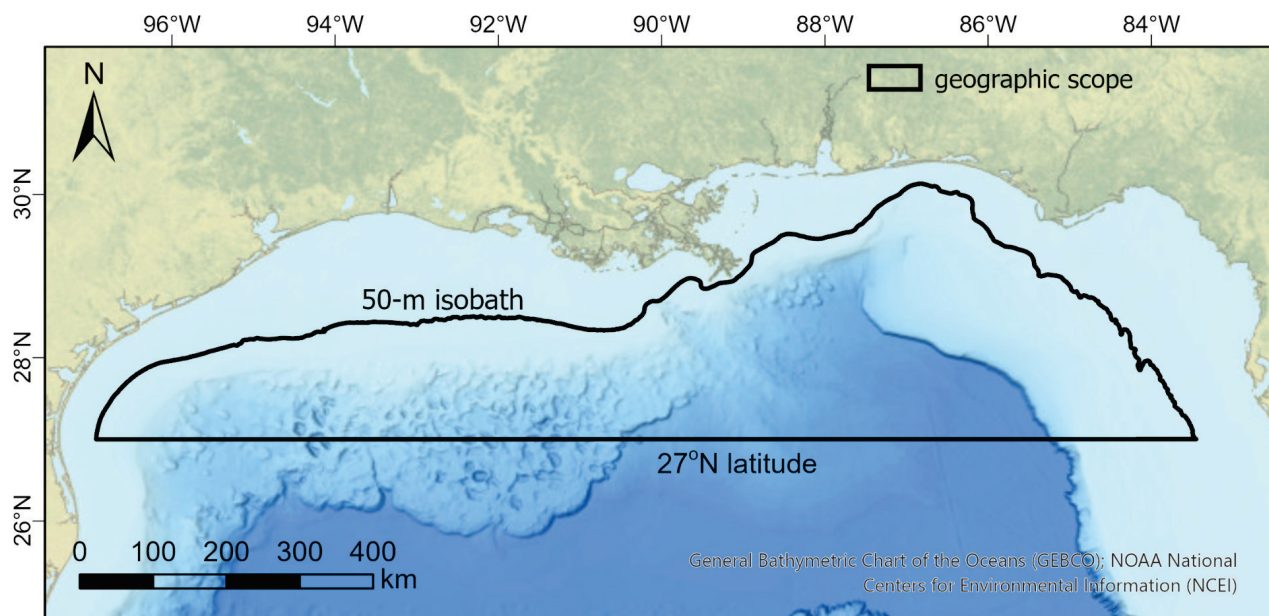


Figure 1. The geographic extent of the inventory was the northern Gulf of Mexico north of 27°N latitude and generally deeper than 50 m.

### **2.1.2. Temporal Extent**

The inventory temporal scope was defined to include data from 1980 through 2021. This was based largely upon technological considerations for seafloor mapping and ground-truthing data accuracy and reliability. For example, the collection date of seafloor mapping data relates to positional accuracy because of technological advancements permitting higher spatial accuracy over time, as well as a shift from using LORAN (long range navigation) to GPS (global positioning system) for georeferencing, allowing identification of seafloor features at resolutions relevant to MDBC. Ground-truthing data reliability also relates to the time period because of factors, such as physical forcing, which may bury or expose MDBC over time. Predictive habitat models are built upon data collected from time periods that may differ from when the model itself was constructed. These considerations helped narrow the temporal scope of the inventory from 1980 through 2021, representing nearly three decades prior to and over one decade following the DWH oil spill.

## **2.2. Fine Scale Inclusion Criteria for Seafloor Mapping**

Seafloor mapping is the process of using remote sensing to survey and derive depth, shape, and texture of the seafloor. Data from seafloor mapping guide ground-truthing and assessments and form the basis of predictive habitat modeling to interpret, classify, and characterize the geomorphological forms, substrates, and benthic biological cover used to define MDBC habitats. Fine scale inclusion criteria for seafloor mapping define target data types and data resolution.

### **2.2.1. Data Types**

Raw and processed seafloor mapping data collected using identified sensors (Table 2) deployed across a suite of platforms (Table 3) were targeted for inclusion in the inventory. The sensors ranged from

multibeam echosounders (MBES) and side-scan sonar to seismic profilers and laser-line scanners (Table 2). These identified sensors collect data capable of mapping or helping to identify the distribution of MDBC and can be operated from a variety of platforms, including vessels and remotely-operated vehicles (Table 3).

Table 2. Data sensors targeted for inclusion in the seafloor mapping component of the inventory. Description of data types, as well as their relevance for mesophotic deep benthic communities (MDBC), are also included.

Sensor Type	Collected Data	Description	MDBC Relevance
multibeam echosounder (MBES)	bathymetry, backscatter	3-dimensional maps or surfaces of seafloor features; interpretations yield morphometrics (slope, curvature, rugosity, etc.)	Maps and surfaces can support modeling of MDBC distribution
side-scan sonar	intensity or reflectance	Wide swath object detection with detailed imagery and texture of seafloor features	Seafloor texture can indicate MDBC
seismic profiler	depth, intensity	Geophysical profile mapping for seafloor features or shallow buried features	Seismic anomalies can help indicate suitable substrates for MDBC
sub-bottom profiler	depth, intensity	Geophysical 2-dimensional profile mapping for seafloor features or shallow and deeply buried features	Sub-bottom profiles can help indicate MDBC
interferometric or phase differential bathymetric sonar (PDBS)	bathymetry, backscatter	Extremely wide swath 3-dimensional bathymetry and backscatter imagery	Maps and surfaces can identify and map distribution of MDBC
synthetic aperture sonar (SAS)	intensity or reflectance	Extremely wide swath deep sea imagery, eliminates shadows behind objects	Ultra-high resolution seafloor texture images can delineate MDBC
laser-line scanning (LLS)	point cloud	Sub millimeter scale 3-dimensional point clouds or digital elevation models of seafloor features	High-resolution models over small areas can detect MDBC

Table 3. Platforms from which seafloor mapping data included in the inventory may be collected.

Platform	Example
vessel	research vessel or small boat outfitted with sonar transducer
autonomous underwater vehicle (AUV)	AUV outfitted with sonar or laser line scanner
autonomous surface vehicle (ASV)	ASV outfitted with multibeam echosounder
human occupied submersible (HOV)	HOV outfitted with laser-line scanner
remotely operated vehicle (ROV)	ROV outfitted with multibeam echosounder
towed body	side-scan sonar towfish

Seafloor mapping data are often accompanied by ancillary data, such as navigation, tide, sound velocities, and temperature (Table 4). These ancillary data can be essential to producing accurate high-resolution seafloor maps and facilitate mapping data interpretation. In cases where raw data are available but have not been processed, ancillary data can assist in advanced processing. Ancillary data were inventoried if they were required for mapping data interpretation or additional processing.

Table 4. Ancillary data types that may accompany seafloor mapping data and were included in the inventory on ad hoc bases.

Sensor Type	Collected data	Description	MDBC Relevance
motion and attitude	attitude; navigation; true heave	smooth out navigation artifacts and correct multibeam to GPS tides to ellipsoidal reference	Ancillary data that can support advanced processing and MDBC detection for multibeam echosounder data
Conductivity-Temperature-Depth (CTD), Expendable Bathythermograph (XBT), Thermosalinograph (TSG)	sound velocity	sound velocity profile	Ancillary data that can support advanced processing and MDBC detection for multibeam echosounder data
Acoustic Doppler Current Profiler (ADCP), Acoustic WAve and Current Meter (AWAC)	current velocities; echo intensity; wave measurements	2D and 3D time series/profiles	Ancillary data that can be used in models
navigation tracking (LBL/USBL/GPS tracking)	GPS position or offset	navigation data or navigational offsets	Ancillary data that can be used to log position of sensor if mounted to mobile platform, like ROV or AUV

Derived data products, which are products derived from raw or processed seafloor mapping data, were also included in the inventory. These derived products included habitat maps, such as those derived from backscatter, or synthetic data layers composed of multiple fine-scale seafloor maps or fine-scale datasets.

### 2.2.2. Resolution

In addition to the broad scale inclusion criteria of geographic, temporal, and depth extents, seafloor mapping data were included if they included data with resolutions finer than 400 m. The 400 m threshold was selected to focus efforts on data which would be most suitable for characterizing MDBC habitats and build off the unified General Bathymetric Chart of the Oceans (GEBCO; <https://gebco.net/>), which is used as the basemap for the inventory and accompanying maps. GEBCO incorporates data at an approximately 400 m resolution (i.e., 15-arc second resolution).

## 2.3. Fine Scale Inclusion Criteria for Ground-truthing

Ground-truthing is the process of collecting in-situ observations to facilitate interpretation of seafloor mapping imagery and to check the accuracy of predictive habitat models. Ground-truthing data are collected using a variety of sampling methodologies. Fine-scale inclusion criteria for ground-truthing data were defined by data collection sensors and associated gear types (Table 5). Ancillary data, as well as existing metadata, were also captured as part of the inventory.

### 2.3.1. Data Types

Ground-truthing data were collected primarily from in-situ bottom imagery, biological samples, and geological samples. Ground-truthing data targeted for inventory included a range of sensors, such as various types of cameras and collection gear types for biological and geological samples (Table 5), that were deployed from or operated from a variety of platforms (Table 6). Note that the availability of original video or photo imagery and biological samples varies by dataset and source.

Table 5. The targeted gear types and/or sensors for the ground-truthing data inventory. Also included are short descriptions and relevance to the MDBC ground-truthing data inventory.

Sensor/Gear Type	Collected Data	Description	MDBC Relevance
video camera	live and/or post-cruise annotations	.csv or Excel files with location and scientist descriptions of the bottom and associated biota; can also describe collections occurring, submersible ops, etc.	Give insight to bottom types and associated sessile organisms to identify MDBC
digital still camera	live and/or post-cruise annotations	.csv or Excel files with location and scientist descriptions of the bottom and associated biota; can also describe collections occurring, submersible ops, etc.	Give insight to bottom types and associated sessile organisms to identify MDBC
stereo camera	live and/or post-cruise annotations	.csv or Excel files with location and scientist descriptions of the bottom and associated biota; can also describe collections occurring, submersible ops, etc.	Give insight to bottom types and associated sessile organisms to identify MDBC
manipulator arm, suction sampler	biological samples	.csv or Excel files with identification of collected organisms and associated information	Collected sessile or bottom organisms can help identify MDBC
trawls	biological samples	.csv or Excel files with identification of collected organisms and associated information	Collected sessile or bottom organisms can help identify MDBC
sediment cores	geologic & biological samples	.csv or Excel files with box core, multicore, gravity core, push core information	Collected substrates types (sediment, rock, infauna) can help identify MDBC
sediment grabs	geologic & biological samples	.csv or Excel files with sediment grab information	Collected surface substrate types (sediment, rock, infauna) can help identify MDBC

Table 6. Platforms where data types in Table 5 can be collected that were included in the inventory.

Platform	Example
vessel	research vessel from which a drop camera, sediment grab, multicore, or dart core is deployed
Autonomous underwater vehicle (AUV)	AUV outfitted with a digital still or video camera
Remotely operated vehicle (ROV)	ROV outfitted with a digital still, video camera, vibracore, manipulator arm, or push core
Human occupied vehicle (HOV)	HOV outfitted with a digital still, video camera, manipulator arm, or push core
towed sled	towed sled outfitted with a digital still or video camera
lander or stationary camera array (e.g. Stereo Baited Underwater Video Array (S-BRUV))	camera array outfitted with one or more digital still imagery and video cameras

The collected ground-truthing data were often accompanied by ancillary data, such as navigation from the data collection platforms (Table 6). For example, remotely operated vehicle (ROV) tracking data can be used to estimate area covered, thereby allowing density estimates of important MDBC species. These ancillary data were identified and incorporated into the inventory if they were available.

## 2.4. Fine Scale Inclusion Criteria for Predictive Habitat Modeling

Predictive habitat modeling is the process of estimating the spatial distribution of MDBC habitat through analysis of habitat-environment or species-environment relationships using seafloor bathymetric variables, water column environmental data, and direct MDBC observations. Predictive habitat modeling synthesizes seafloor mapping data and ground-truthing data to interpret, classify, and characterize the geomorphological forms, substrates, and benthic biological cover used to define MDBC habitats. Fine scale inclusion criteria for predictive habitat modeling were related to model data types.

### 2.4.1. Data Types

The predictive modeling data category included predictions of MDBC habitat types from relationships defined or estimated between environmental predictor variables (seafloor mapping data and derived bathymetric variables and water column environmental data) and direct MDBC observations (Figure 2). MDBC habitats defined by geomorphological forms, substrates, or biota were all considered as were MDBC habitat types described in strictly physical or biological terms with any specificity (e.g., detailed habitat classes versus soft/hard substrate or biological cover versus individual species' habitat suitability). Datasets with numerical or categorical predicted quantities with associated spatial coordinates were considered. Predicted quantities considered included: habitat suitability, binary/multinomial presence/absence, probability of occurrence, abundance/density, area/percent cover, and size. Datasets were typically in the form of spatial grids/rasters or shapefiles. Predictions from a range of model types were considered including mathematical and statistical models as well as non-statistical functional or mechanistic models. Model predictions that were simply interpolations, extrapolations, or unsupervised classifications of direct MDBC observations were not included in this data category but rather were included as derived products in the Seafloor Mapping data category. Model prediction surfaces were included in the predictive habitat modeling part of the inventory. Response and predictor datasets input to models were not included, although attributes of those datasets were documented in the inventory (see Section 6.3.2).

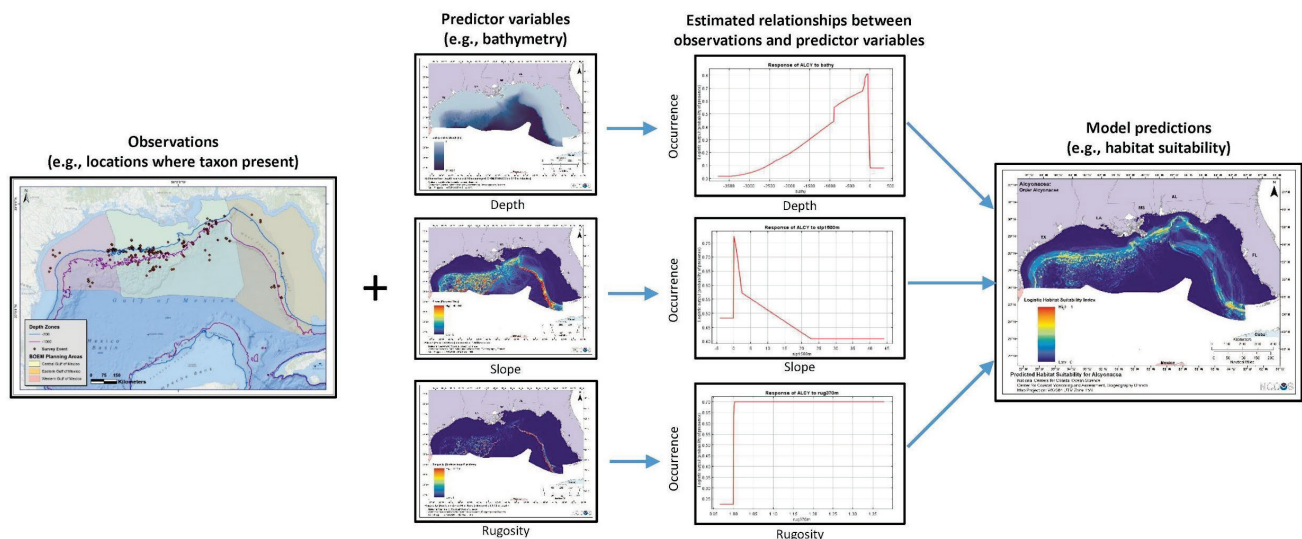


Figure 2. Schematic of predictive habitat modeling process. The example shown is a predictive model of the spatial distribution of a biological taxon, but predictive models can also be developed for less specific biological variables (e.g., biological cover) or physical habitat characteristics (e.g., to predict the spatial distribution of a habitat class). These model results are from Kinlan et al. (2013).

## 3. Data Discovery

### 3.1. Broad Sources

Seafloor mapping, ground-truthing, and predictive habitat modeling data were requested and acquired from multiple broad sources, representing federal agencies, state agencies, academic institutions, and industry organizations (Table 7). Broad sources from which data were acquired are detailed below for seafloor mapping, ground-truthing, and predictive habitat modeling.

Table 7. Broad sources from which data were discovered for the inventory.

Source Type	Source	Seafloor Mapping	Ground-truthing	Predictive Habitat Modeling
federal agency	National Oceanic and Atmospheric Administration (NOAA)	X	X	X
federal agency	Bureau of Ocean Energy Management (BOEM)	X	X	
federal agency	U.S. Geological Survey (USGS)	X	X	X
federal agency	National Science Foundation (NSF)	X		
federal agency	U.S. Navy	X		X
state agency	multiple	X	X	
academia	multiple	X	X	X
industry	multiple	X		
other	multiple	X	X	X

#### 3.1.1. Seafloor Mapping

##### National Oceanic and Atmospheric Administration (NOAA)

NOAA maintains archives and repositories of seafloor mapping data. These data were specifically requested and acquired from the National Centers for Environmental Information (NCEI) and the National Ocean Service (NOS) Hydrographic Survey Data database (<https://ngdc.noaa.gov/>). These requests largely included multibeam echosounder data, as well as data from several other mapping sensors. Some data were not yet available through NCEI or the NOS Hydrographic Database, so these data were requested directly from the entity who acquired the data. For example, a request for data from Flower Garden Banks National Marine Sanctuary yielded five multibeam surveys that were not yet available through main NOAA data repositories because their archiving was underway.

##### Bureau of Ocean Energy Management (BOEM)

Seafloor mapping data from BOEM are stored within multiple portals, databases, and other data sources. Eight main BOEM online data sources were evaluated for inclusion in the inventory (Table 8).



Table 8. BOEM data sources and their status for inclusion or exclusion in seafloor mapping inventory. Description provides accompanying details.

Data Source	Status	Description
<a href="#">BOEM Northern Gulf of Mexico Deepwater Bathymetry Grid from 3D Seismic</a>	include	interpolated grid based on XYZ data from 3D seismic surveys included in 'BOEM Seismic Water Bottom Anomalies Dataset'; shapefiles detailing extent of proprietary input data are available
<a href="#">BOEM Seismic Water Bottom Anomalies Dataset</a>	include	shapefile coverage but not actual data
<a href="#">BOEM Environmental Studies Program Information System</a>	include	required study level screening and subsequent data search and acquisition (see Section below)
<a href="#">BOEM High Resolution Geophysical Survey</a>	include	shapefile coverage indicates where industry data exist; actual data not available without formal request
<a href="#">BOEM Data Center</a>	exclude	redundant, mostly boundary data
<a href="#">BOEM Marine Minerals Information System</a>	exclude	too shallow
<a href="#">BOEM Gulf of Mexico Archaeological Information</a>	exclude	data not applicable
<a href="#">Marine Cadastre</a>	exclude	redundant with other BOEM data sources
<a href="#">BOEM Technical Information Management System (TIMS)</a>	exclude	secure web-based program that provides oil and gas industry users location for electronic submittal of data

While multiple BOEM datasets were directly available from BOEM websites, BOEM also maintains a database of final reports, technical memoranda, and other documents associated with BOEM-funded research studies. Studies from this database, called the BOEM Environmental Studies Program Information System (ESPIS), likely had data associated with them that were applicable to the seafloor mapping inventory. We performed the following steps to determine whether data referenced in these reports met the inventory inclusion criteria:

**Step 1:** ESPIS search—ESPIS data manager searched the ESPIS database using the following query on 5 May 2021 [mapping AND (sea?floor OR geomorph\* OR bathymetr\* OR sonar OR multi?beam OR side?scan OR seismic OR sub?bottom OR interferometric OR differencing)]. This search yielded 561 unique ESPIS entries that required screening.

**Step 2:** Inventory team screened each of the 561 entries to determine whether they might meet inclusion criteria for seafloor mapping inventory, including the geographic scope and temporal scope. During the screening process, ESPIS entries were also reviewed for whether they included seafloor mapping data, and if so, which types of data. This search was performed by screening documents associated with ESPIS entries. This search yielded 116 entries that may contain seafloor mapping data that meet MDBC inventory inclusion criteria. Of these, 84 entries were flagged for likely inclusion and 32 entries were flagged for possible inclusion, as we were unsure whether these should be included without further review.

**Step 3:** Data acquisition—Data associated with most ESPIS entries are not available online. Instead, data must be acquired by BOEM colleagues from the original researchers who conducted the study. BOEM colleagues facilitated this data search and upon receipt of datasets, the data were shared with the MDBC team for further screening and processing as part of the established workflow detailed below. Because of the high time investment required to acquire a dataset from the original

principal investigator(s), ESPIS entries were prioritized. Data from the past decade (2020–2010) were assigned top priority, followed by data from the previous decade (2010–2000) and then the 1990s and 1980s. Data from later decades will be added during the data accretion phase of the inventory as they are received by BOEM and shared with the MDBC inventory team.

### U.S. Geological Survey (USGS)

USGS seafloor mapping data are spread across multiple physical and digital locations. The centralized Coastal Marine Geosciences Data System (CMGDS; <https://cmgds.marine.usgs.gov/>) includes recent data, but often does not include older or legacy data. The data request to USGS included information on the data types, inventory scope, and metadata of interest. Using these bounding and metadata criteria, the USGS Data Management Group executed a search of records across their data holdings, which returned 33 seafloor mapping datasets from multibeam bathymetry or side-scan sonar and 30 from seismic or sub-bottom. USGS also provided six additional multibeam bathymetry datasets that are stored elsewhere, such as the USGS ScienceBase Catalog.

### National Science Foundation (NSF)

Data from the National Science Foundation were requested and compiled from the Marine Geosciences Data System (MGDS; <https://www.marine-geo.org/index.php>). These data included several high-resolution bathymetry grids from organizations, such as National Deep Submergence Facility (NDSF), as well as seismic survey data.

### U.S. Navy

Data from the Navy were requested, and those datasets received included footprints of survey locations. Additional data requests are in progress.

## **3.1.2. Ground-truthing**

### NOAA

Data and metadata were requested and received from NCEI, Flower Garden Banks National Marine Sanctuary, Index to Marine and Lacustrine Geological Samples (IMGLS), and various National Marine Fisheries Service (NMFS) Southeast Fisheries Science Center surveys, such as Panama City Lab Reef Fish Survey. The data requests included broad- and fine-scale inclusion criteria detailing target data types, as well as the geographic scope of the inventory. Those datasets that did not pass these criteria were not included in the inventory and were not ingested.

### BOEM

BOEM ground-truthing data are stored within multiple portals and databases. After reviewing BOEM data holdings, we decided to use ESPIS as our primary BOEM data source because the ESPIS database includes information on final reports, technical memoranda, and other documents associated with BOEM-funded research studies. Studies from the GOM region may have data associated with them that are applicable for the ground-truthing inventory. To determine whether data that meet the inventory inclusion criteria exist, we performed the following steps:

**Step 1:** ESPIS search - ESPIS data manager searched the ESPIS database using the following query on 15 April 2021 [coral OR mesophotic OR deepwater OR ROV OR AUV OR benthic OR bottom OR (benthic community\*) OR hardbottom OR softbottom OR sediment core OR macrofauna OR infauna

OR imagery OR camera OR (swiftia exserta) OR (muricea pundula) or thesea OR plexauridae OR placogordia]. This search yielded 404 unique ESPIS entries that required screening.

**Step 2:** Initial screening - Inventory team used screening parameters to decide whether to include or exclude the paper. Inclusion criteria included the temporal scope (published after 1980) and the geographic scope (north of the 27th parallel and >50m depth). First papers were sorted by the temporal and geographic scope and those that did not meet those criteria were excluded. Next papers were sorted by relevance to search string query and were reviewed for whether they included ground-truthing data, and if so, which types of data were presented in the reports. Full documents were screened and yielded 27 papers that may contain ground-truthing data that meet MDBC inventory inclusion criteria.

**Step 3:** Data acquisition - Data associated with most ESPIS entries are not available online. Instead, data must be acquired by BOEM colleagues from the original researchers who conducted the study. BOEM colleagues facilitated this data search and upon receipt of datasets, the data were shared with the MDBC team for further screening and processing as part of the established workflow detailed below. Because of the high time investment required to acquire a dataset from the original principal investigator(s), ESPIS entries were prioritized. Data from the past decade (2020–2010) were assigned top priority, followed by data from the previous decade (2010–2000) and then the 1990s and 1980s. Data from later decades will be added during the data accretion phase of the inventory as they are received by BOEM and shared with the MDBC inventory team.

## USGS

Data and metadata were requested and received from USGS repositories, such as usSEABED. The data requests included broad and fine inclusion criteria on which data types the inventory encompasses, as well as the temporal and geographic scopes of the inventory. Once the dataset list was compiled, another quality assurance/quality control (or QA/QC) of the datasets occurred, to ensure the criteria were met. Those datasets that did not pass these criteria were not included in the inventory.

### **3.1.3. Predictive Habitat Modeling**

Several key broad sources of predictive habitat modeling data were already known or discovered (Section 3.2), including NOAA, USGS, U.S. Navy, and academic institutions. There was a smaller number of predictive modeling datasets available than there were mapping and ground-truthing datasets. Individual modeling datasets were acquired directly from existing online portals and repositories (Section 3.5) or from individual dataset holders including project partners (Section 3.3) and external contacts (Section 3.4).

## **3.2. Literature Search**

Some data pertinent to the MDBC inventory were located in, referenced in, or accompanied peer-reviewed publications. We conducted a systematic literature search using the database Web of Science (Clarivate™) to search for data sources. Searches were performed separately for mapping, ground-truthing, and predictive habitat modeling using the Web of Science advanced search function; however, the syntax of each search string was consistent among the groups to include concatenated topic searches for geography and relevant topics or instruments.

The literature search string for seafloor mapping was a combination of a topic search for geography, mapping, and specific mapping instruments, as follows:

```
TS= ("Gulf of Mexico") AND TS = (geomorph* OR seafloor OR mapping OR bathy* OR backscatter)
AND TS = ("sonar" OR "multibeam" OR "multi-beam" OR "multi beam" OR "single-beam" OR "single
beam" OR "sidescan" OR "side scan" OR "side-scan" OR "synthetic aperture" OR "seismic" OR "sub-
bottom" OR "sub bottom" OR "subbottom" OR "laser-line" OR "laser line" OR "laser scan" OR
"bathymetric" OR "interferometric" OR "phase differencing" OR "phase-differencing")
```

The search string for ground-truthing was a combination of a topic search for geography, sampling gear type, and bottom habitat as follows:

```
TS=("Gulf of Mexico") AND (benthic OR mesophotic) AND (image OR camera OR video OR
Sediment OR sample* OR photogram* OR "point cloud" OR "laser-line scanner" OR "aperture sonar"
OR core OR cores OR "multicores" OR "monocores" OR "vibracores" OR "hard bottom" OR "soft
bottom" OR "epibenthic" OR "endobenthic" OR shelf OR slope OR substrate OR ROV or AUV))
```

The search string for predictive habitat modeling was a combination of a topic search for geography, habitat, biota, and methodology, as follows:

```
TS = ((marine OR ocean OR sea OR seafloor) AND (global OR "Gulf of Mexico" OR "west Atlantic" OR
"north Atlantic" OR "northwest Atlantic") AND ("aragonite saturation state" OR benthic OR
benthos OR "calcite saturation state" OR seep OR "continental shelf" OR "continental slope" OR
"deep-sea" OR "deep sea" OR geomorphology OR mesophotic OR ridge* OR "salt diapir*" OR "salt
dome*" OR seafloor OR sediment* OR "vulnerable marine ecosystem*") OR (bivalve* OR
chemosynthetic OR "cold-water coral*" OR "cold water coral*" OR "coral ecosystem" OR "deep-sea
coral*" OR "deep sea coral*" OR "deepwater coral*" OR infauna OR macrofauna OR "mesophotic
coral*" OR mussel* OR sponge* OR tubeworm*)) AND ("boosted regression tree*" OR "generalized
linear model*" OR "generalized additive model*" OR "habitat distribution" OR "habitat map*" OR
"habitat suitability model*" OR "machine learning" OR maxent OR "maximum entropy" OR "neural
network" OR "niche model*" OR "occupancy" OR "predictive habitat model*" OR "predictive
model*" OR predictor* OR "random forest*" OR "regression kriging" OR "species distribution
model*"))
```

The mapping, ground-truthing, and predictive habitat modeling searches were conducted for publications written in the English language and published between 1980 and 2021. The start date for the modeling literature search was left at the Web of Science default of 1970, but given that predictive habitat modeling studies did not happen until much later, this start date did not affect the number of discovered modeling datasets. All document types were searched, and the citation indices queried were those available via the NOAA Central Library subscription. Each Web of Science search yielded a reference list containing titles and abstracts. For example, the mapping search yielded 587 potentially relevant publications, whereas the ground-truthing search yielded 770 and the modeling search 1,450 (Table 9).

The reference lists returned from the Web of Science searches were imported into a reference screener program called Swift ActiveScreener (Howard et al. 2020). Swift ActiveScreener uses a machine learning algorithm to rank publications in order of relevance so that relevant publications can be screened earlier rather than later. The software updates the order and relevance of publications based

on completed reference screening actions. The software also presents a running estimate of the percentage of relevant references that have been screened from the initial set, referred to here as the ‘recall rate’, and a running estimate of the number of remaining relevant references that have not been screened. Those estimates allow a user to define a target recall rate at which point the screening is terminated, saving the user time by not having to screen every reference in the initial set. We chose a target recall rate of 95% at which point the screening was terminated. Keywords were manually input into Swift ActiveScreener so that these words would be highlighted in titles and abstracts to help facilitate screening.

Using the Swift ActiveScreener interface, publications were screened at the level of title and abstract by multiple people. During the title and abstract screening, the publications were assessed for inclusion or exclusion in the inventory based on geographic extent and temporal scope, as well as fine-scale inclusion criteria outlined above. Screening results were then reconciled among the multiple screeners, making decisions about whether to include a reference or not when results differed. Following the title and abstract screening, 158 of 587 seafloor mapping articles were designated as likely relevant for the inventory effort (Table 9). Likewise, 146 of 770 ground-truthing and 39 of 1,450 modeling publications also passed the title and abstract screening (Table 9).

Articles that passed title and abstract screening were then screened at the full text level. The NOAA Library assisted in finding and downloading full texts of the publications. Full texts were populated into the MDBC literature library within EndNote (The EndNote Team 2020), where they were further screened based on inclusion criteria. Articles that passed the full text screening were then screened at the dataset level to determine if the dataset was already included in the inventory or whether the dataset could be downloaded directly, required external contact, or other actions. The literature search for seafloor mapping, for example, yielded 45 new datasets and found 22 datasets that had already been discovered through other search processes and included in the inventory (Table 9). The literature search was the primary means of discovery of predictive habitat modeling datasets. The modeling literature search identified all previously known datasets and yielded several previously unknown datasets.

Table 9. Number of articles returned in the literature search (initial search) and included or excluded in either the title and abstract screening, full text screening, or dataset screening.

Category	Initial Search Results	Title and Abstract Screening		Full Text Screening			Dataset Screening		
		Include	Exclude	Include	Exclude	TBD	Include (new)	Already Included (other sources)	Exclude
seafloor mapping	587	158	399	89	70	NA	45	22	22
ground-truthing	770	146	624	104	42	NA	10	31	63
predictive habitat modeling	1450	39	1411	13	16	10	NA	NA	NA

### 3.3. Partner Engagement

Many of the broad sources of seafloor mapping, ground-truthing, and predictive habitat modeling data in the inventory (Section 3.1) were project partners within the MDBC portfolio. Discovery and requests for datasets from partners were generally conducted through direct communications using e-mail and

telephone. Because these communications were internal to the project, they were not formally documented beyond standard practices for records management, including recording call notes, dates, and times.

### **3.4. External Engagement**

Other broad sources of seafloor mapping, ground-truthing, and predictive habitat modeling data in the inventory (Section 3.1) were external to the MDBC portfolio. Discovery and requests for datasets from external sources were formally documented in a set of spreadsheets used as a simple Customer Relationship Management database. At the time of publication, the project team conducted 63 individual engagements with 81 individuals in departments or offices at 27 organizations, including 15 academic institutions and four federal agencies. For each engagement activity, contact information, status, purpose, and content of communications were documented. When an external engagement activity was relevant to more than one MDBC project, personnel from the relevant projects were contacted to coordinate that external engagement. Communications with external sources were conducted according to the MDBC external engagement policy to standardize the manner of interactions and the language describing the MDBC portfolio. Many of the predictive habitat modeling datasets were acquired from external sources.

### **3.5. Portals**

Multiple existing online data portals or repositories store and catalog data that were potentially relevant for the inventory. These portals often display footprints of sampling or modeling efforts using interactive map data viewers. We initially compiled a list of data portals with data relevant to seafloor mapping, ground-truthing, or predictive habitat modeling. Information about portals relevant to each data type was documented in the MDBC Data Portal Catalog, including access details and host organization (more information about the Data Portal Catalog can be found in DIVER [see Appendix]). Data from portals were discovered and screened for broad- and fine-scale inclusion criteria. For seafloor mapping, portals were initially assessed, but most data were acquired through other sources as part of the search process. Some mapping datasets, however, were acquired directly from portals, such as GRIIDC. Following the search process, the portals were rescreened to ensure that all data had been incorporated into the inventory effort. Using broad- and fine-screening criteria, data from portals were also searched for the ground-truthing inventory. These portals were revisited to ensure that all ground-truthing datasets were acquired during the initial search. For predictive habitat modeling, several datasets were acquired through portals and repositories (Dryad, GRIIDC, and Zenodo).

## 4. Data Tracking

### 4.1. Data Catalog

Datasets were discovered from multiple sources, including federal agencies, external contacts, peer-reviewed literature, and portals. Once datasets were identified as potentially relevant to the inventory, they were added to a dataset tracking catalog called the MDBC Data Catalog. This data catalog served as a comprehensive tracking system that provided a unified location to record and track individual datasets throughout the inventory workflow. The data catalog also documented important attributes of each dataset. The list of attributes, sources, and corresponding definitions associated with each dataset in the MDBC Data Catalog can be found on DIVER (refer to the Appendix for access details).

The metadata framework used to describe dataset attributes was guided by the traditional approach and schema for describing datasets. However, non-traditional metadata attributes were also included to facilitate data tracking across the multi-institutional and multi-disciplinary MDBC project. The metadata attributes were categorized as: 1) Core, 2) Universal, and 3) Working group specific (mapping, ground-truthing, and predictive habitat modeling working groups). Core attributes (e.g., unique data identifier, date of record creation, period of record for the dataset, etc.) are mandatory for record description and are applicable to all data types. Core attributes are used for indexing the datasets thus making them easily searchable by database search engines. Unlike core attributes, there are data characteristics which are only applicable to specific data types. Known as universal attributes, such attributes were defined as those which could be utilized to characterize one or more data types, as necessary. For example, seafloor mapping data included universal attributes for spatial resolution and survey ID, since these were key parameters used to describe mapping data. However, the said parameters could also be utilized to characterize modeling datasets. Working group specific attributes are those which are used to set apart datasets belonging to a specific group. For example, modeling software is specific to the predictive habitat modeling working group and not utilized by others.

### 4.2. Unique Data Identifier

When a dataset was added to the MDBC Data Catalog, it was immediately assigned a unique identifier for the context of the MDBC project. The unique identifier, referred to as the “MDBC\_ID” is a two-part code. The first part of the code, or prefix, is “MAP,” “GTR,” or “PHM” which designates that the dataset was relevant for the seafloor MAPping, Ground-TRuthing, or Predictive Habitat Modeling portion of the inventory, respectively. The second part of the code is a six-digit numerical suffix designating the dataset number; this number was assigned sequentially as datasets were added to the data catalog. For example, the first MAP dataset was MAP000001; the second was MAP000002, and so forth (e.g., MAP##### generically).

### 4.3. Data Parent and Child Structure

When a dataset represented data collected with a single instrument or sensor, then the dataset was assigned to one row or entry in the MDBC Data Catalog, as for example mapping dataset unique ID MAP000004. Some datasets, however, included data collected with multiple instruments or sensors. For example, a mapping dataset may include data from both a multibeam echosounder and side-scan sonar. In these cases, the dataset was assigned an overall or “parent” ID, as well as an instrument-specific dataset “child” ID. The parent ID was the unique identifier (e.g., MAP000005), whereas the child IDs were variations of the parent ID reflecting data stemming from a particular instrument or sensor (e.g., MAP000005C01 for multibeam, MAP000005C02 for side-scan sonar). The numerical code for a child ID (e.g., C##) was assigned sequentially in the order the child dataset was added to the data catalog.

## 5. Data Screening and Acquisition

All discovered datasets were screened to determine whether they met both coarse- and fine- scale inclusion criteria. If datasets did not adhere to the inclusion criteria, then they were excluded from further steps in the inventory compilation effort. If datasets adhered to the inclusion criteria or if the datasets required additional inspection before determining if they met the inclusion criteria, then the source data were acquired and stored (Figure 3). For seafloor mapping data, all pertinent available source data were obtained, ranging from raw data to processed data. The goal with source data acquisition was to obtain all data needed to create products, such as footprints or imagery, for the inventory. The ground-truthing data acquisition followed the same protocols as seafloor mapping, in that raw to processed data were collected to create spatial footprints for the inventory. For modeling, data acquisition focused on model predictions, typically raster data, but sometimes also included related modeling data that were available; e.g., output files from modeling software. Response and predictor data input to models were generally not acquired, although sometimes these data were part of the related modeling data that were available. Source data retained their original naming conventions, as they were received by the inventory team, and were not renamed. This helped preserve the provenance of the source data. Specific download and organization steps were used for mapping, ground-truthing, and modeling.

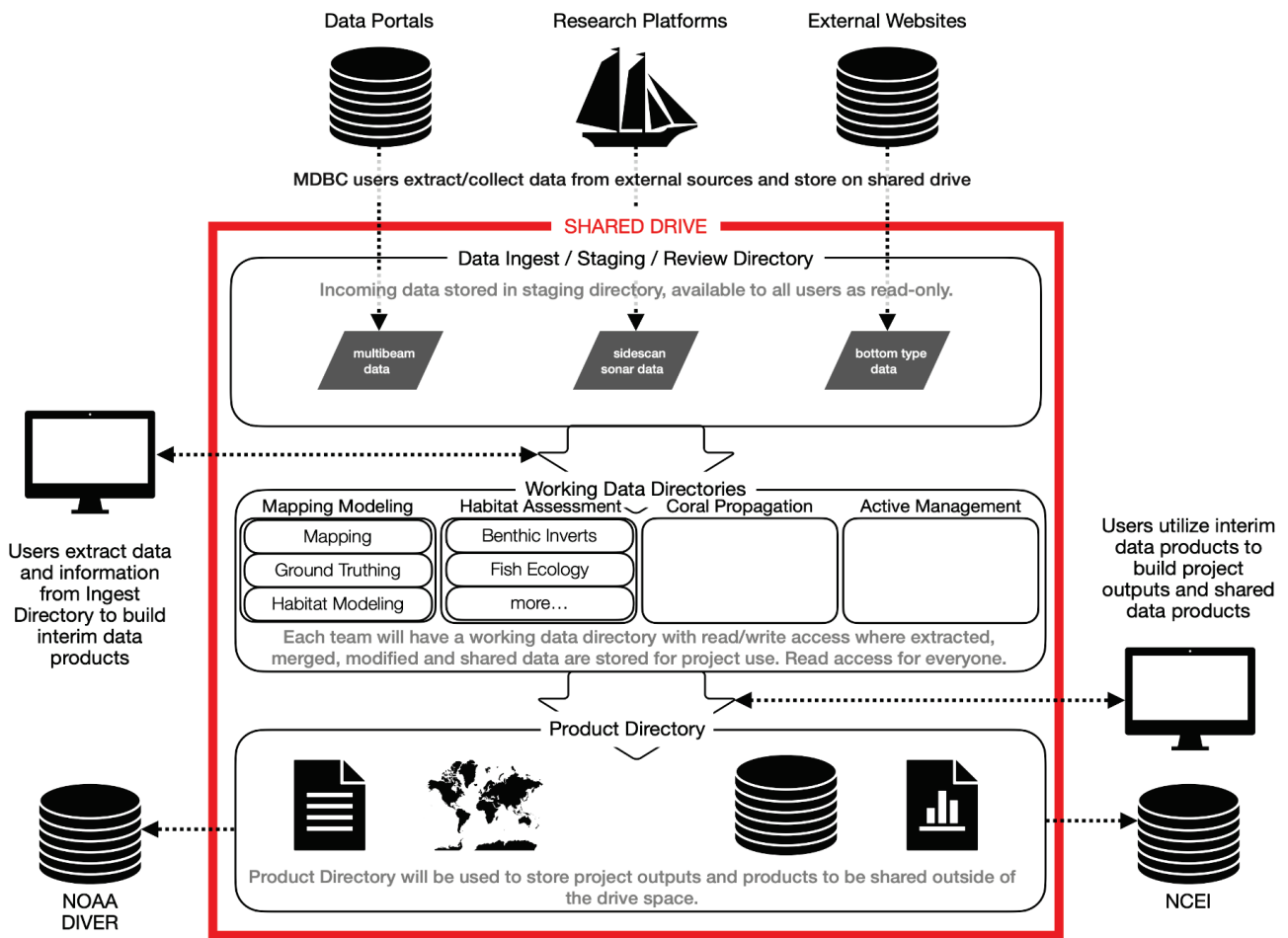


Figure 3. Data storage workflow. Source data were “ingested” into a shared drive system for later review and processing. A “working” directory was set up for each team to provide a space to share processed data and interim products.



Once a dataset was received from an archive, original source, or data provider, the data were decompressed and sorted into a predefined folder structure. A description of the process for data download and organization for seafloor mapping and ground-truthing can be found on DIVER (refer to the Appendix for access details). This process occurred on a local workspace as follows: 1) decompressed dataset locally using Winzip 25.0 (64 bit), 2) assigned unique dataset identifier (e.g., MDBC\_ID), 3) locally stored decompressed data within a folder named for the unique dataset identifier, 4) organized source data into subfolders. For seafloor mapping and ground-truthing data, the source data were organized into raw source data, processed source data, and source metadata, making sure to retain native file naming conventions to preserve dataset provenance tracking. For modeling data, datasets were placed within a sub-folder of that folder named 'original\_files' if data reformatting was conducted, and original file names and sub-folder names and structures were retained to preserve dataset provenance tracking. Background information about the original files was documented in a readme text file if necessary.

## 6. Data Processing

After data were discovered and source data were obtained, data were processed. The processing steps were customized for mapping, ground-truthing, and modeling, and those individual processes are described below.

### 6.1. Data Processing for Seafloor Mapping

Seafloor mapping data were processed using a custom workflow that reproducibly ingested source data of multiple types into the inventory and generated spatial footprints, processed imagery, and associated attributes. Here, we detail the folder structure, file naming conventions, and other required processing steps.

#### 6.1.1. Goal

The goal of data processing was to generate a spatial footprint and associated geometric and other descriptive attributes for each dataset, as well as processed seafloor imagery (e.g., .tif). Final footprints and imagery adhered to the spatial reference for geographic NAD 1983 (2011).

#### 6.1.2. Dataset Attributes

Attributes associated with each dataset were extracted and recorded into the data catalog during processing. These attributes describe each dataset and help facilitate dataset tracking, as well as future queries for particular datasets. The required attributes stemmed from multiple sources. Some required manual entry, whereas others were filled in automatically. Some attribute values were extracted from metadata (xml) files, whereas others were calculated in and exported from custom ArcGIS Pro (ESRI 2020) workflows. The list of attributes, sources, and corresponding definitions associated with each dataset in the MDBC Data Catalog can be found on DIVER (see Appendix). Briefly, the attributes include inventory level information on the unique dataset ID, data storage location, inclusion criteria, and more. Other attributes reflect dataset descriptors, including the title, abstract, survey dates, and survey ID, whereas others provide spatial context, including bounding coordinates, spatial reference, and resolution. Additional attributes were created to track a dataset through the inventory workflow; these include information on the data format, processing level, and workflow status.

#### 6.1.3. File Naming Conventions

Resulting footprints and processed seafloor imagery files were named to reflect combinations of dataset attributes, such as the sensor and data source. Standardized codes corresponding to these attributes were used to name files (Table 10). Codes were also developed to indicate whether a product was a footprint or processed seafloor imagery (Table 10).

Table 10. Attributes for datasets (e.g., sensor, source) and products with accompanying descriptions and codes; these codes were used for file naming.

Attribute	Description	Code
sensor	multibeam echosounder	MBES
sensor	side-scan sonar	SSS
sensor	seismic profiler	SEIS
sensor	sub-bottom profiler	SUBB
sensor	interferometric or phase differential bathymetric sonar	PDBS
sensor	laser-line scanner	LLS
source	National Oceanic and Atmospheric Administration	NOAA
source	Bureau of Ocean Energy Management	BOEM
source	United States Geological Survey	USGS
source	industry	INDUSTRY
source	academia	ACADEMIA
source	other sources	OTHER
spatial reference	geographic NAD 1983 (2011)	GCSNAD1983
product	spatial footprint of dataset coverage	FP
product	processed seafloor imagery	IMG
product	Table of footprint spatial attributes	FP [.csv extension]

Using these file naming codes, footprints were named to reflect key attributes as *MAP#####\_SENSOR\_SOURCE\_SPATIAL-REFERENCE\_FP.shp*. Footprints were shapefiles so included .shp files, as well as accessory files (e.g., .shx). Files were named using all capital letters; underscores separate file naming strings. Spatial reference was always GCSNAD1983 for footprints. Examples of named footprints are *MAP000001\_MBES\_NOAA\_GCSNAD1983\_FP.shp* and *MAP000003C01\_SSS\_USGS\_GCSNAD1983\_FP.shp*.

Processed seafloor imagery products were named similarly to reflect key attributes and allow easy cross-referencing with footprint products. Imagery files were geotiffs (.tif) and named as *MAP#####\_SPATIAL-REFERENCE\_ORIGINAL-FILE-NAME.tif*. Capital letters were used for all portions of the string except for the original file name, which was the original source file name as it was received (e.g., lowercase, uppercase, dash). Spatial reference was always GCSNAD1983 for imagery. An example of a named seafloor image is *MAP000001\_GCSNAD1983\_H12059\_MB\_2m\_MLLW\_1of6.tif*, where the string beginning “H12059” through “.tif” was the original file name as it was received during data acquisition.

A Table for each dataset was also created that included spatial and geographic information calculated in ArcGIS Pro. These Tables were named as *MAP#####\_SENSOR\_SOURCE\_GCSNAD1983\_FP.csv*, similarly to footprints.

#### 6.1.4. Folder Structure

Consistent folder structure and associated file naming conventions were required to support the seafloor mapping inventory. The folder structure for the inventory compilation effort facilitated: 1) use of multiple software packages and languages (e.g., ArcGIS Pro, Python), 2) hybrid local and network

workflow, 3) multiple workflow processing contributors, and 4) MDBC level project data management. More details about the folder structure can be found on DIVER (refer to the Appendix for access details).

### 6.1.5. ArcGIS Pro Projects

The majority of data processing occurred in ArcGIS Pro using files stored and written locally to optimize processing speed. This was facilitated by the creation of temporary ArcGIS Pro Projects for each mapping data processor to use on their local workspace that were developed from the same template. Steps detailing the setting up of the ArcGIS Pro template can be found on DIVER (see Appendix for details). The template file included the geographic scope and depth contours, as well as custom Python tools. Following the completion of work in ArcGIS Pro and the generation of footprints and processed data products, the final products were uploaded to the data storage system (Figure 3).

### 6.1.6. Custom Python and ArcPro Tools

Custom tools were written as Python (Van Rossum and Drake Jr. 2020) scripts and then set up as ArcGIS Pro tools with a point and click interface. They include tools to process seafloor mapping raster images to export footprint shapefiles, obtain metadata, export final seafloor imagery files, and extract attributes that describe the dataset (e.g., resolution, survey area; Table 11). The Python scripts can be used as standalone scripts or converted to ArcGIS Pro tools. Each tool/script is described in detail below. The scripts evolved over time and continue to do so as new datasets are discovered and added to the inventory. The Python scripts are available upon request, as standalone scripts as well as in the form of ArcGIS Pro tools including the instructions on how to add them to the ArcGIS Pro toolbox.

Table 11. Description of tools developed to process seafloor mapping data and incorporate them into the inventory. The interface for all tools is ArcPro custom tool; the language for all tools is Python.

Tool Name	Description	Input	Output	Common Data Types
raster processing tool	processes rasters to produce footprints and imagery	raster	footprint (polygons, .shp); imagery (.tif); attribute Table(.csv)	gridded multibeam, side-scan sonar
point data processing tool	converts point data (xyz) into coarse grid to produce footprints and imagery	points (xyz)	footprint (polygons, .shp); coarse-gridded imagery (.tif); attribute Table (.csv)	ungridded multibeam (xyz)
line data processing tool	converts positioning line or point data without depth values into shapefile	lines (.shp) or point data for lines (xy but no depth)	footprint (lines, .shp); attribute Table (.csv)	seismic survey lines (xy)
derived product - shapefile processing tool	appends inventory-required attributes to derived data product shapefile	shapefile	shapefile with appended inventory attributes (.shp); attribute Table (.csv)	derived product shapefile requiring attribute additions
derived product - imagery processing tool	transforms derived data product raster into inventory geographic reference system	raster	raster (.tif); attribute Table (.csv)	derived product raster requiring minimal processing (.tif)
metadata scraping tool	extracts metadata from xml files and writes to Table	metadata file (xml)	Table (.csv)	datasets with metadata files
NCEI imagery tool	processes NCEI coarse-gridded rasters to produce footprints and imagery	raster	footprint (polygons, .shp); imagery (.tif); attribute Table (.csv)	coarse-gridded multibeam

### Raster processing tool

The raster processing tool processes raster images to generate a footprint, imagery files, and a Table with descriptive attributes. Input rasters can be either .tif or .bag files. The tool calculates raster statistics, such as the cell size or resolution and depth (minimum, maximum, mean, standard deviation). When the raster contains an uncertainty band, such as in a .bag file, the tool also calculates uncertainty (minimum, maximum, mean, standard deviation). The tool then records the original spatial reference and projects the raster into the standardized inventory spatial reference (GCS NAD 1983, 2011 version), recording the implemented transformation. From the projected raster, the script then extracts the raster bounding coordinates and creates a detailed footprint of the survey extent and calculates the area (km<sup>2</sup>) of the footprint. To determine the data coverage in the surveyed footprint, we created a second footprint using the eliminated gaps in the detailed footprint. For example, if there were gaps in the data collection, then these voids were eliminated in the second more generalized footprint. We then compared the area covered by the detailed footprint versus the footprint without gaps to calculate the percent of surveyed area actually covered with data and the area of gaps in the footprint. If a dataset contained more than one input raster, then the steps above were performed on each raster individually so that these individual calculations and results were associated with each raster. The script then merged the multiple footprint polygons and associated attributes into one shapefile, where each polygon was retained separately as reflected in the attribute Table as separate rows. The resulting final footprint was exported as a shapefile. Then, one imagery file (.tif) per source raster was exported; we did not merge the imagery together because sometimes different raster images from the same dataset were at different resolutions, and we wanted to preserve imagery provenance as closely as possible. Last, we exported one Table (.csv) per dataset that contained attributes that we needed to record in the data catalog to help with dataset tracking.

### Point data processing tool

The point data processing tool was used to generate a footprint, coarse gridded imagery files, and a Table with descriptive attributes. The tool accepts input point files in .xyz format, where each point contains longitude, latitude, and depth values. The tool converted the .xyz data from their native tabular format into geospatial points before converting these points into a raster grid. The resulting raster grid was a coarse grid with a 32-m resolution. We then applied the same steps used in the raster processing script to calculate statistics, including cell size or resolution and depth, before converting to the standardized inventory spatial reference, extracting bounding coordinates, creating footprints, and calculating area values. Similar to the raster processing tool, the script exported a final footprint, final imagery, and a final Table with data catalog attributes.

### Line data processing tool

The line data processing tool was used to generate a set of lines as a shapefile and one .csv Table per dataset that contains attributes needed for the data catalog. This tool does not produce final imagery since the input was line data rather than rasters. The tool mainly applied to seismic survey data, where survey lines were associated with discrete point values describing the latitude, longitude, time, and date when survey activities (e.g., shots) were performed. The line data tool worked on line data that were points or already in line files; in the case of the former, the points were converted to lines at the beginning of the script. The lines were then converted to the standardized spatial reference for the inventory, and the bounding extent coordinates were calculated. The line footprints were then exported as a shapefile. A Table with data catalog attributes was also exported.

### Derived product - shapefile processing tool

Some datasets included in the inventory were derived products. Examples include: 1) BOEM High-Resolution Geophysical Survey, which contained shapefiles but no underlying data for hundreds of industry-conducted seafloor mapping surveys, 2) reef features delineated as shapefiles from NOAA NMFS side-scan sonar data, 3) salt diapirs delineated as shapefiles from USGS side-scan sonar data. These derived features included polygons, points, and lines. The shapefile processing tool for derived products transformed these geospatial data into the inventory standardized spatial reference. The script then appended new attributes to the original attribute Tables for these features. The new attributes were those required for the inventory tracking process, such as the unique ID of the dataset, the name of the source file, geographic extent, and more. The tool exported the data as a shapefile and also exported a Table with attributes needed for the data catalog.

### Derived product - raster processing tool

Several datasets included in the inventory were raster products that only required light processing. Examples include the BOEM Bathymetry Grid from 3D Seismic data. In these cases, the derived product raster processing tool transformed the rasters to the standardized inventory spatial reference, calculated extent coordinates, and then exported the transformed image along with a Table containing attributes needed for the data catalog.

### Metadata extraction tool

The metadata extraction tool extracts metadata values from .xml files and writes them into a Table. The tool requires an .xml file for input. The tool works on metadata from the following dialects: ISO-19115, FGDC, Pydro DR, and Pydro DRM. The tool extracts particular attributes from the metadata that are required for inventory dataset tracking. It then writes these attributes into a .csv file from which the values can be input into the data catalog.

### NCEI coarse imagery tool

The NCEI archive included some multibeam bathymetry data that were in point or .xyz format that had been coarse-gridded for NCEI display purposes. We obtained these coarse-gridded rasters and applied a modified version of the raster processing tool to them. The modifications of the raster processing tool (Section 6.1.6. Raster Processing Tool) allowed for the NCEI imagery ID to be matched up with the survey ID so that we could perform the workflow.

## **6.1.7. Raw Multibeam Data Processing in Qimera**

When raw multibeam data were available but processed multibeam surfaces were not, we conducted preliminary processing in QPS Qimera (<https://qps.nl/qimera/>) to generate a 32 m resolution uncleaned bathymetric grid. The resulting coarse bathymetric grid was input to the raster processing tool to generate the final footprint, imagery, and Table. An overview of QPS Qimera raster processing can be found on DIVER (refer to the Appendix for access details).

## **6.1.8. Fishery Multibeam Data Processing**

Fishery echosounder data collected using an ME70 multibeam system were included in the inventory. These data included processed surfaces, as well as footprints of surveys corresponding to collected data that had not yet been processed. These data were added to the inventory as they were received and will be updated in the future as NMFS continues data processing. In general, ME70 data can

provide CUBE bathymetric surfaces compliant with IHO and OCS charting standards when properly calibrated, yet these data can contain a higher uncertainty to depth ratio than other multibeam systems. An overview of the ME70 multibeam processing can be found on DIVER (refer to the Appendix for access details).

## **6.2. Data Processing for Ground-truthing**

Ground-truthing data were processed using a custom workflow that reproducibly ingested source data of multiple types into the inventory and generated spatial footprints and their associated attributes. Here, we detail the folder structure, file naming conventions, and other required processing steps.

### **6.2.1. Goal**

The goal of ground-truthing data processing was to generate a spatial footprint and associated descriptive attributes for each dataset. Final footprints adhere to the spatial reference for geographic NAD 1983 (2011).

### **6.2.2. Dataset Attributes**

During data processing, we extracted and recorded attributes associated with each dataset into the data catalog. These attributes describe each dataset and help facilitate dataset tracking, as well as future queries for particular datasets. Attributes were entered manually after examination of the datasets and their supporting documentation. The list of attributes, sources, and corresponding definitions associated with each dataset in the MDBC Data Catalog can be found on DIVER (refer to the Appendix for access details). Briefly, the attributes include inventory level information on the unique dataset ID, data storage location, inclusion criteria, and more. Other attributes reflect dataset descriptors, including the title, abstract, survey dates, and survey ID.

### **6.2.3. File Naming Conventions and Folder Structure**

Consistent folder structure and associated file naming conventions were required to support the ground-truthing inventory. A description of the process developed for downloading and organizing ground-truthing data can be found on DIVER (refer to the Appendix for access details)

### **6.2.4. Dataset Reformatting**

After datasets were downloaded, the files were reformatted so they were able to be presented as footprints. The majority of this effort was conducted in Microsoft Excel or ArcGIS Pro. Generally, datasets were received as tables and opened in Microsoft Excel for reformatting so they could be correctly imported into ArcGIS Pro and display the correct attributes for presentation purposes. Examples of reformatting included editing the headers to be accepted in ArcGIS Pro (ESRI 2020), changing the coordinates from decimal minutes to decimal degrees, adding missing content for the attribute Table, and also removing other attributes not suitable for display.

### **6.2.5. Footprints**

All processed Microsoft Excel files were brought into the ArcGIS Pro template using the XY Table to Point tool. Steps detailing the setting up of the ArcGIS Pro template can be found on DIVER (refer to the Appendix for access details). In a few cases, if the Table was a submersible transect, these were

translated from XY Point Table to polyline surveys by using the Points to Line tool. Datasets received in a shapefile format were added to the ArcGIS Pro template using the Add Data tool.

Once datasets were added, the attribute Tables were edited to make sure they reflected the inventory needs. This included changing or adding attribute aliases for display and adding missing content to the Tables. Next, layers were added to four subgroups depending on the type of sampling conducted during the survey: ground-truthing biological collections, ground-truthing geological collections, ground-truthing bottom-video sites, and ground-truthing submersible surveys. Lastly, the metadata provided by the dataset authors were added to each feature layer.

## **6.3. Data Processing for Predictive Habitat Modeling**

### **6.3.1. Goal**

The goal of predictive habitat modeling data processing was to generate a spatial footprint and associated geometric and other descriptive attributes for each dataset, as well as reformatted model predictions if necessary (e.g., TIFF).

### **6.3.2. Dataset Attributes**

During data processing, attributes associated with each dataset were recorded in the data catalog. These attributes describe each dataset and help facilitate dataset tracking, as well as future queries for particular datasets. Attributes were entered manually after examining datasets and their supporting documentation. The list of attributes, sources, and corresponding definitions associated with each dataset in the MDBC Data Catalog can be found on DIVER (refer to the Appendix for access details). Briefly, the attributes include inventory level information on the unique dataset ID, data storage location, inclusion criteria, and more. Other attributes reflect dataset descriptors, including the title, abstract, and study date, whereas others provide spatial context, including bounding coordinates, spatial reference, and resolution. Additional attributes were created to document details about modeling including things like the quantity modeled, modeling framework, and software. Information about predictor variables used in modeling were documented in a separate Google spreadsheet for each dataset whose link was included in the corresponding data catalog entry.

### **6.3.3. File Naming Conventions and Folder Structure**

Original data files were stored in folders named according to the unique MDBC dataset identifiers (e.g., PHM000001). Reformatted model predictions, when necessary, were stored in a sub-folder named 'GeoTIFFs'. Other than the file extension, original file names were retained to preserve dataset and file provenance tracking.

Footprints and associated processing files were stored locally within folders named for the unique dataset identifiers with the suffix '\_NAD1983\_FP'; e.g., *PHM000001\_NAD1983\_FP*. Footprint shapefiles within those folders were named similarly.

### **6.3.4. Data Reformatting**

Modeling datasets were acquired in multiple file formats. GeoTIFF was chosen as a standard format, and datasets not originally in that format were reformatted. Custom R scripts (R Core Team 2020) were developed to reformat datasets, and each script was stored in its associated dataset's folder named for its unique dataset identifier. Scripts employed the R packages 'ncdf4' (Pierce 2019), 'raster'



(Hijmans 2020), 'sf' (Pebesma 2018), and 'terra' (Hijmans 2021). Original coordinate reference systems were retained in reformatted datasets.

### **6.3.5. Footprints**

Footprints describing the spatial extent of each modeling dataset were derived using custom R scripts that employed the R packages 'raster', 'sf', and 'terra'. The scripts were stored in the associated footprint folders and named '*[MDBC ID]\_footprint.R*'. All footprints were created with the geographic coordinate reference system NAD 1983 (2011), which typically required transformations of the original coordinate systems of the modeling datasets. Footprints were stored in shapefile and Well Known Text (WKT) formats. WKT representations of the full footprint and a simplified footprint were created, the latter of which was input to the data catalog for each dataset. The bounding box coordinates for each footprint were also derived and input to the data catalog. A subset of the dataset attributes from the data catalog were included as attributes in the footprint shapefiles. The R and R package versions used to create the footprints are documented in text files named '*[MDBC ID]\_footprint\_software\_versions.txt*' in the footprint folders.

## 7. Inventory products

Inventory products included: source data, the data catalog, footprints, seafloor mapping imagery, ground-truthing annotations, modeling predictions, geodatabases, and digital inventories. These products are detailed below.

### 7.1. Data

During the inventory process, source data were downloaded and stored. We ensured that data necessary to create inventory products or incorporate the data into the inventory were acquired. In some cases, these included only processed data products from the source. In other cases, these included raw datasets that required intensive processing to incorporate into the inventory. These acquired data represent a major product associated with the inventory. The data are maintained in long-term storage associated with the inventory (Figure 3). All data are organized according to the unique inventory ID of the dataset (MDBC\_ID) and therefore are easy to relate to the more derived products, including footprints, imagery, annotations, and model predictions, which are described below.

### 7.2. Data Catalog

The data catalog represents the comprehensive dataset tracking system for the inventory. It includes descriptive information about each dataset evaluated for and included in the inventory. The data catalog provides descriptive attributes, geospatial attributes, and workflow tracking attributes for each dataset.

### 7.3. Literature Library

A library of peer-reviewed publications screened for the literature review portion of the inventory compilation was compiled. The library resides within EndNote. It includes text of all articles that passed title and abstract screening and were reviewed during or passed the full text screening portion of the literature review.

### 7.4. Footprints

For each dataset, a spatial representation of the “footprint” of the dataset was created. For mapping, these were footprints of polygons or lines where mapping data existed. For ground-truthing, these were footprints representing where video surveys, biological and sediment sampling occurred. For predictive modeling, the footprints were polygons representing the areas for which model predictions were made. For all data types, the footprints indicated actual coverage and any data gaps within study areas. The benefit of these footprints is that they provide a small size data source without having to drill into detailed data, and they are easy to open because of their smaller size. Footprints were produced during the data processing workflows described above for seafloor mapping (Section 6.1), ground-truthing (Section 6.2), and predictive habitat modeling (Section 6.3). Footprints were archived as a package at NCEI (Accession 0268765). Attributes associated with footprints along with corresponding definitions were also included in the data package.

### 7.5. Mapping Imagery

In addition to creating footprints for each dataset, we compiled imagery for the seafloor mapping component of the inventory. These images were the underlying imagery used to generate footprints.

They included imagery, such as multibeam echosounder bathymetry and backscatter, as well as side-scan sonar mosaics. Each of these images was produced as a .tif raster during the data processing workflow described above (Section 6.1) but was not included in the archival package containing the footprints.

## **7.6. Ground-truthing Annotations**

Annotations associated with ground-truthing surveys, generally submersible surveys, were downloaded from the source and organized. Steps detailing the data download and organization can be found on DIVER (refer to the Appendix for access details). Depending on the source, annotations were those created during the survey or post-cruise annotations. As with the ground-truthing footprints, these are maintained in long-term storage associated with the inventory.

## **7.7. Habitat Modeling Predictions**

Predictions from habitat models were received as or reformatted to GeoTIFF rasters as described in Section 6.3.4 and are maintained in long-term storage associated with the inventory.

## **7.8. Geodatabases**

Geodatabases are collections of spatial data stored in a common format. Geodatabases were created to store data products for seafloor mapping, ground-truthing, and predictive habitat modeling, as detailed below.

### **7.8.1. Base Layer Geodatabase**

A geodatabase with base layers was compiled. This included a feature layer defining the geographic scope of the inventory, as well as dividing the inventory scope into three regions (eastern, central, western). Depth contours were also included as a base layer.

### **7.8.2. Seafloor Mapping Geodatabases**

Three seafloor mapping geodatabases were created: 1) footprint geodatabase, 2) imagery geodatabase, and 3) derived products geodatabase.

To create the footprint geodatabase, we wrote a Python script that we ran manually within the Python console in ArcGIS Pro. The script merges seafloor mapping footprints from multiple shapefiles into one master feature layer. The master feature layer contains all seafloor mapping footprints. To ensure that the master footprint feature layer includes all necessary attributes, the script then joins the footprint attributes with attributes from the data catalog. This merge was conducted using the unique dataset ID (MDBC\_ID). Following the attribute join, the attributes were organized and standardized naming conventions were applied.

To help display footprints intuitively, we then created subsets of the master feature layer. The subsets were made for resolution, collection decade, instrument, and source (Table 12). For collection decade, for example, we created separate feature layers for data from the 1980s, 1990s, 2000s, 2010s, and 2020s. We completed these steps for footprints that are polygons, as well as for footprints that were line data, and exported the subset layers.

Derived data layers, such as the BOEM Seismic Anomalies and the Otos survey extent, that were originally shapefiles without underlying imagery were not combined using the scripting approach. During processing, MDBC attributes were appended to native attributes, which we intentionally retained because they often had classification or other important descriptors.

Table 12. Groups and bins used to display seafloor mapping feature layers within geodatabases.

resolution	<1 m, 1-10 m, 10-32 m, 32-100 m, >100 m, coarse
decade	1980s, 1990s, 2000s, 2010s, 2020s
instrument	multibeam echosounder, side-scan sonar, multibeam fishery echosounder, seismic profiler
source	NOAA, USGS, BOEM, academia, industry, other
derived products	derived data products retained as individual feature layers

### 7.8.3. Ground-truthing Geodatabase

A single ground-truthing geodatabase was created that contained all datasets and their derived footprints. To display the footprints in an intuitive manner, there were four subgroups of layers created to keep the appropriate footprints together (Table 13).

Table 13. Description of the subgroups of layers in the ground-truthing geodatabase.

Subgroup	Description
biological collections	display sites where various gears were used to collect biological samples during a survey
bottom video sites	display sites where bottom imagery was collected using a variety of camera types and arrays, for example, a stationary baited underwater video array and deployment sites of submersible vehicles
geological collections	display sites where various gears were used to collect geological samples during a survey.
submersible surveys	display track lines as polylines of various submersible vehicles, such as ROV, AUV, and HOV's, to conduct video transects and/or collect biological and/or geological samples.

### 7.8.4. Predictive Habitat Modeling Geodatabase

A single predictive habitat modeling geodatabase was created that contained all modeling datasets and derived footprints. Model prediction layers were named with a combination of the dataset identifier and the quantity predicted. Footprint layer names were identical to the names of the footprint files.

## 7.9 Inventory Access

Access to the inventory was configured differently for internal and external (to NOAA) MDBC team members, and general public. Inventory of the seafloor imagery, ground-truthing data and predictive habitat modeling datasets, and associated footprints was made available to the internal team members through the Geoplatform (NCCOS' ArcGIS Online platform) as well as via shared network drives maintained by NCCOS. The footprints inventory was also archived with the NCEI and is now available online at <https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:MDBC-MGM-Inventory> for public access.

## **7.10. Archival Products**

The collection of inventory spatial footprints and their associated attributes for seafloor mapping, ground-truthing, and predictive habitat modeling are archived at NCEI (Accession 0268765).

## 8. Results

In total, 693 seafloor mapping footprints, 13,722 seafloor mapping tracklines, 20 seafloor mapping derived products, 737 seafloor mapping raster images, 57 ground-truthing datasets, including 43 feature layers of various ground-truthing surveys, and 12 predictive habitat modeling datasets, including 198 model prediction layers were included in the inventory. The seafloor mapping data included footprints delineating survey extent, as well as actual seafloor mapping imagery. Ground-truthing data included multi-year, multi-gear sampling surveys to determine various bottom types in the areas of interest for this inventory. Modeling data included model footprints delineating the prediction area, as well as the model predictions. These datasets represent multiple sources, data types, and collection time periods. Below, we describe the general patterns associated with these datasets at the survey or sampling-event level for mapping and ground-truthing and at the level of particular models for predictive habitat modeling. Analyses beyond the presented high-level and spatially explicit summaries will be conducted as part of an upcoming formal gap analysis conducted on inventory footprint products (Section 9.4).

### 8.1. Seafloor Mapping Results

#### 8.1.1. Overall

During the seafloor mapping inventory process, 422 datasets were considered for inclusion. Of these, 404 datasets passed coarse-scale inclusion criteria and were further screened based on fine-scale inclusion criteria (Table 14). In total, 336 datasets passed fine-scale inclusion criteria and were either incorporated into the inventory or are in the process of being compiled into the inventory, as data acquisition or processing remain underway (Table 14).

Table 14. Number of seafloor mapping datasets screened according to coarse and fine scale inclusion criteria. Numbers of datasets excluded and included according to the inclusion criteria are also provided.

Screening Level	Datasets Screened	Datasets Excluded	Datasets Included
coarse scale screening	422	18	404
fine scale screening	404	68	336

As of January 2022, the seafloor mapping inventory encompassed the following (Figure 4):

- 693 spatial footprints of survey extent representing 221 datasets
- 13,722 tracklines (e.g., linear footprints) representing 13 datasets, and largely corresponding to seismic surveys
- 20 derived datasets, representing seafloor feature interpretation, survey polygons with underlying data are unavailable, or synthetic data
- 737 rasters of seafloor mapping imagery corresponding to the spatial footprints
- 257 GB (7,869 files) of data acquired directly from sources, including raw data, processed data, metadata, and relevant ancillary files for most datasets

The inventoried data were acquired from multiple sources and span a variety of sensors, resolutions, and collection decades, which are detailed below (Sections 8.1.2–8.1.6). Derived datasets and underlying imagery are detailed separately (Sections 8.1.7–8.1.8).

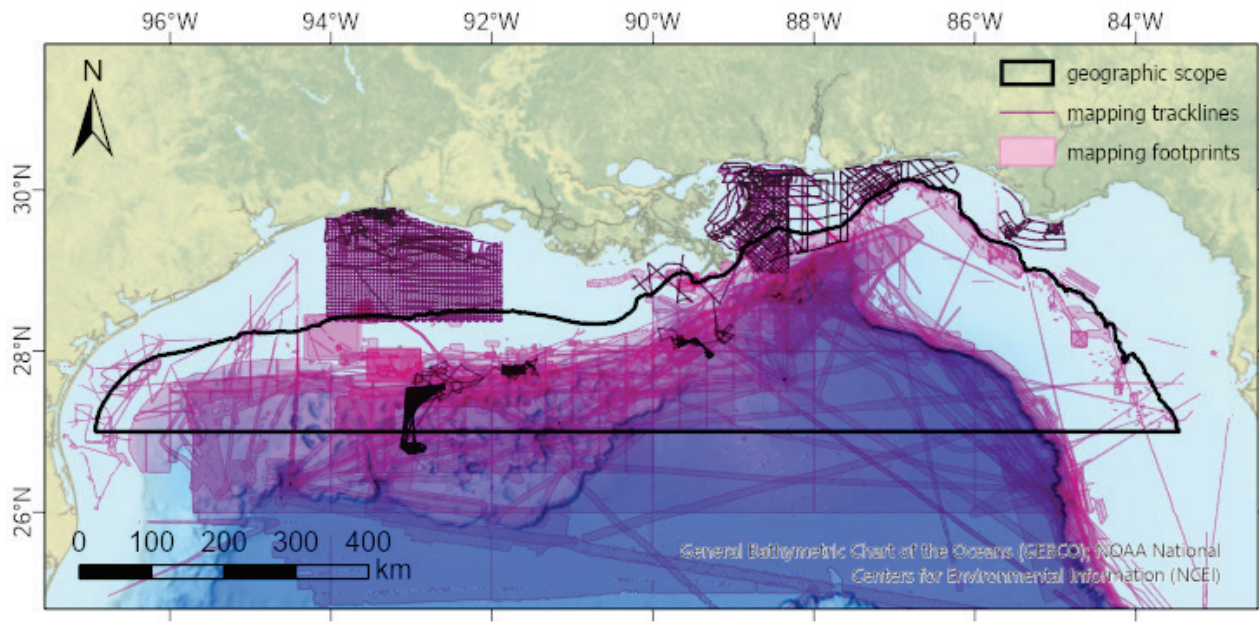


Figure 4. Footprints and tracklines for sea floor mapping data included in the inventory. These footprints and tracklines do not include derived products (Section 8.1.7).

### 8.1.2. Data Source

Inventoried data were acquired from multiple sources, predominantly NOAA and USGS (Table 16). NOAA and Florida Fish and Wildlife Conservation Commission’s (FWC) Fish and Wildlife Research Institute (FWRI; e.g., “OTHER” category) contributed the most footprints (44.1%, 41.9%, respectively; Table 16; Figure 5; Figure 6). Because the footprints contributed by FL, however, were high-resolution side-scan sonar over small areas of the seafloor, they covered a relatively small total area (approximately 1,800 km<sup>2</sup>). NOAA footprints, on the other hand, covered the largest area of the seafloor (approximately 658,000 km<sup>2</sup>; Figure 7), followed by USGS (approximately 655,000 km<sup>2</sup>; Figure 8). Despite the large seafloor coverage, USGS footprints comprised 5.2% of inventoried footprints because the few footprints covered large areas. USGS tracklines covered approximately 1,900,000 linear km (Figure 8). Academic and industry sources contributed fewer and less extensive footprints and tracklines than federal sources (Figures 9 and 10; Table 15). BOEM contributed many datasets, some of which are described in derived data below (Section 8.1.7), and others which were conducted in collaboration with BOEM and thus are associated with sources, including academia, industry, or other federal partners. Several other sources provided derived data, which are described below, as well.

Table 15. Seafloor mapping summary statistics for footprints and tracklines (type) by source from which data were acquired. Count is the total number of footprints or trackline segments. Count (%) represents the percentage of footprints or tracklines by source. Area (km<sup>2</sup>) and length (km) are both cumulative values or sums across all footprints or tracklines, respectively.

Type	Source	Count	Count (%)	Area (km <sup>2</sup> )	Length (km)
footprint	NOAA	306	44.1	657,981.6	NA
footprint	USGS	36	5.2	655,301.1	NA
footprint	ACADEMIA	59	8.5	16,141.9	NA
footprint	INDUSTRY	2	0.3	170.8	NA
footprint	OTHER	291	41.9	1,754.9	NA
trackline	USGS	13,716	>99.9	NA	1,891,102.0
trackline	ACADEMIA	6	<0.1%	NA	162.4

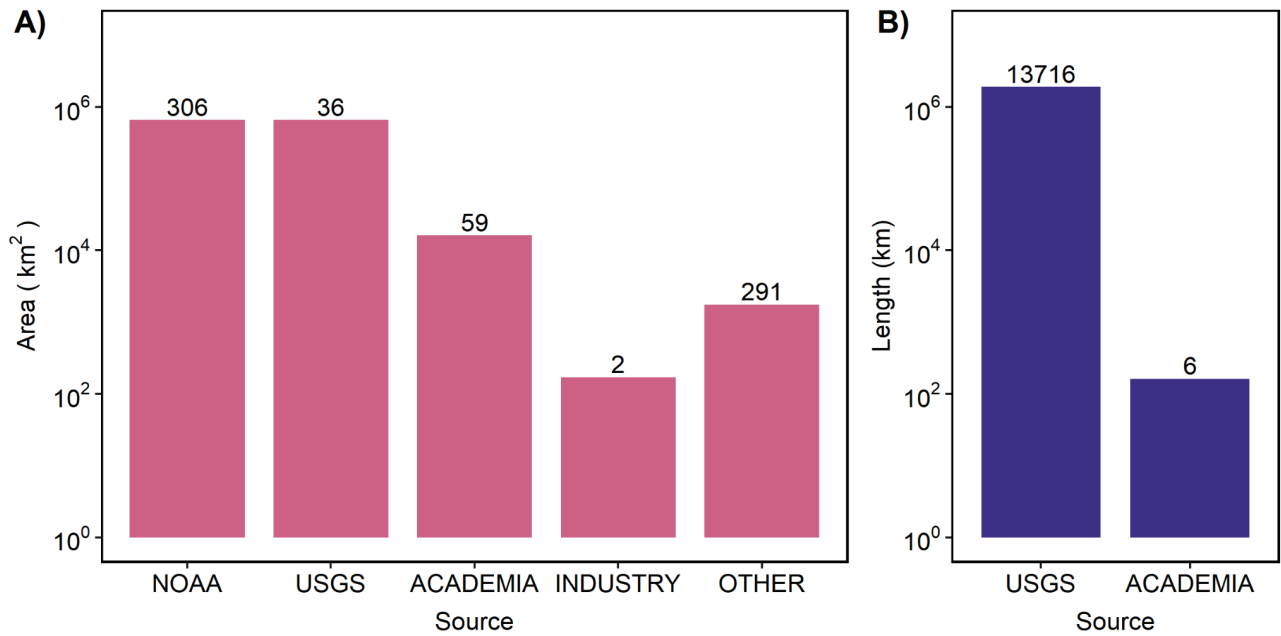


Figure 5. Logarithmic cumulative area (km<sup>2</sup>) of seafloor mapping footprints (A) and tracklines (B) by source. Numbers above bars represent the number of mapping survey footprints or trackline segments that comprise the cumulative area value.



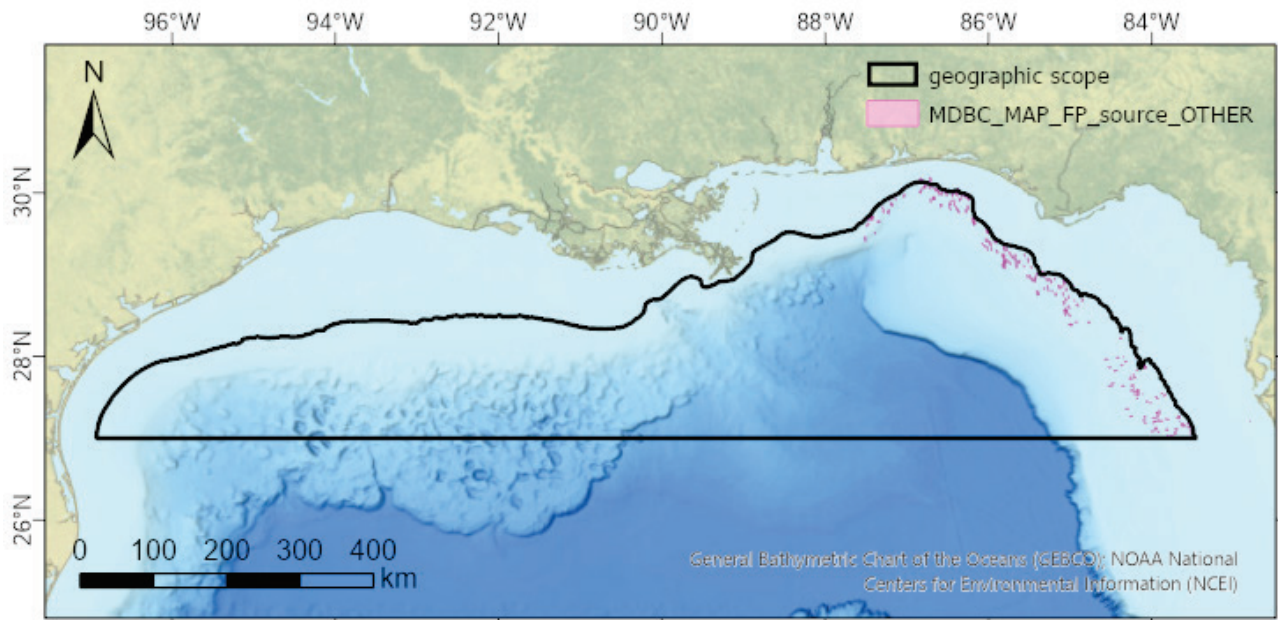


Figure 6. Footprints and tracklines for seafloor mapping data acquired from other sources.

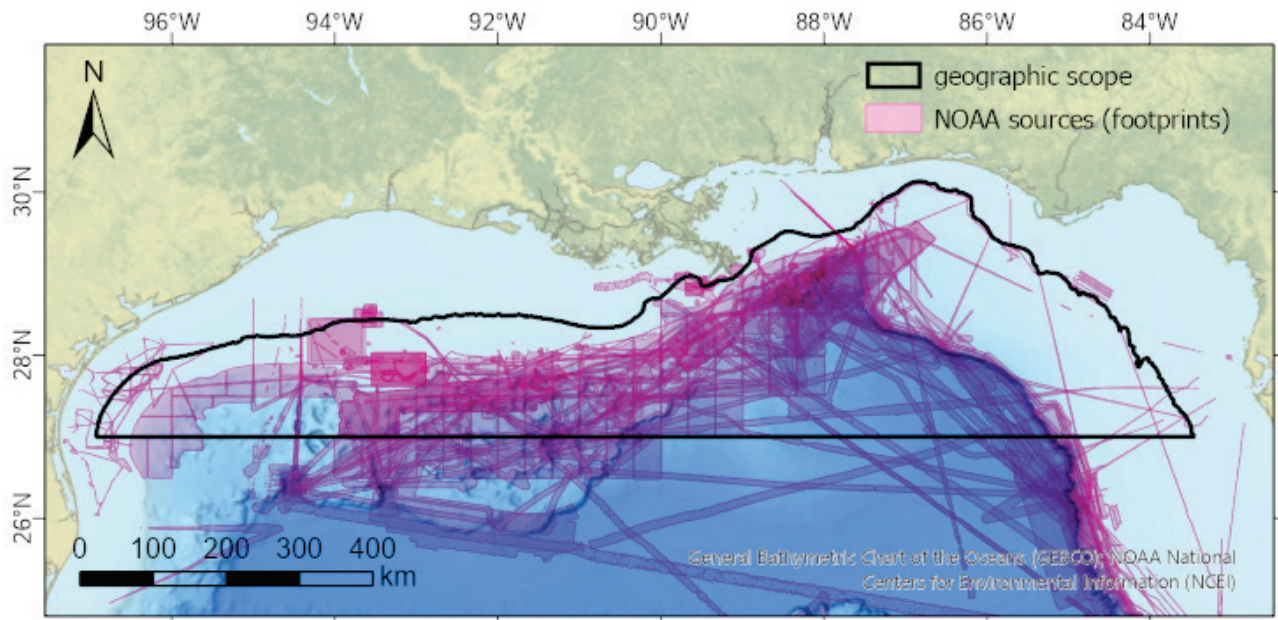


Figure 7. Footprints for seafloor mapping data acquired from NOAA sources.

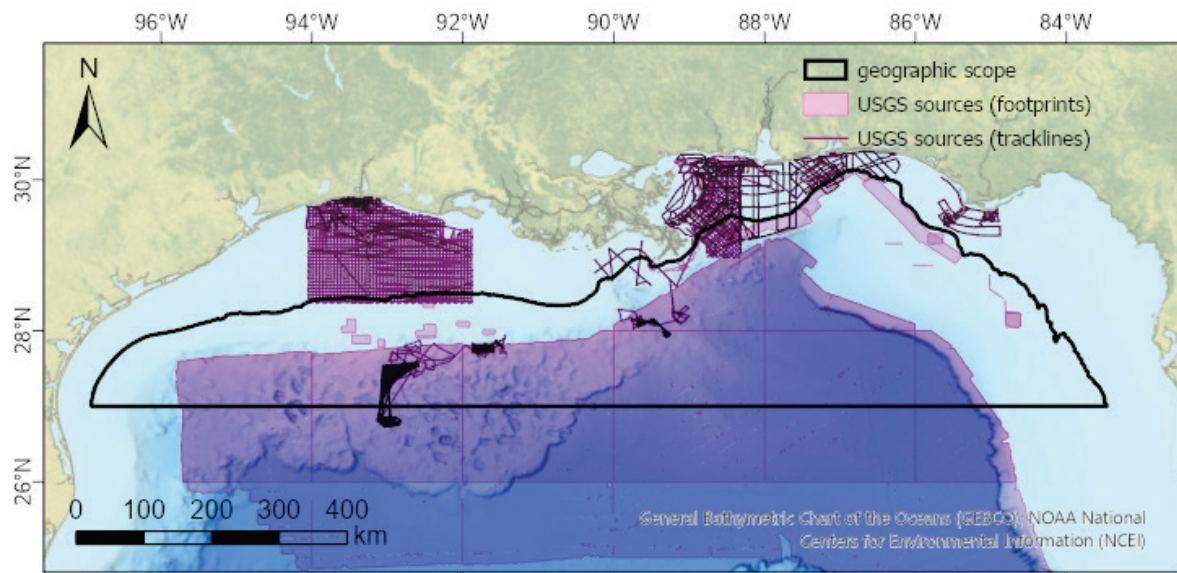


Figure 8. Footprints and tracklines for seafloor mapping data acquired from USGS sources.

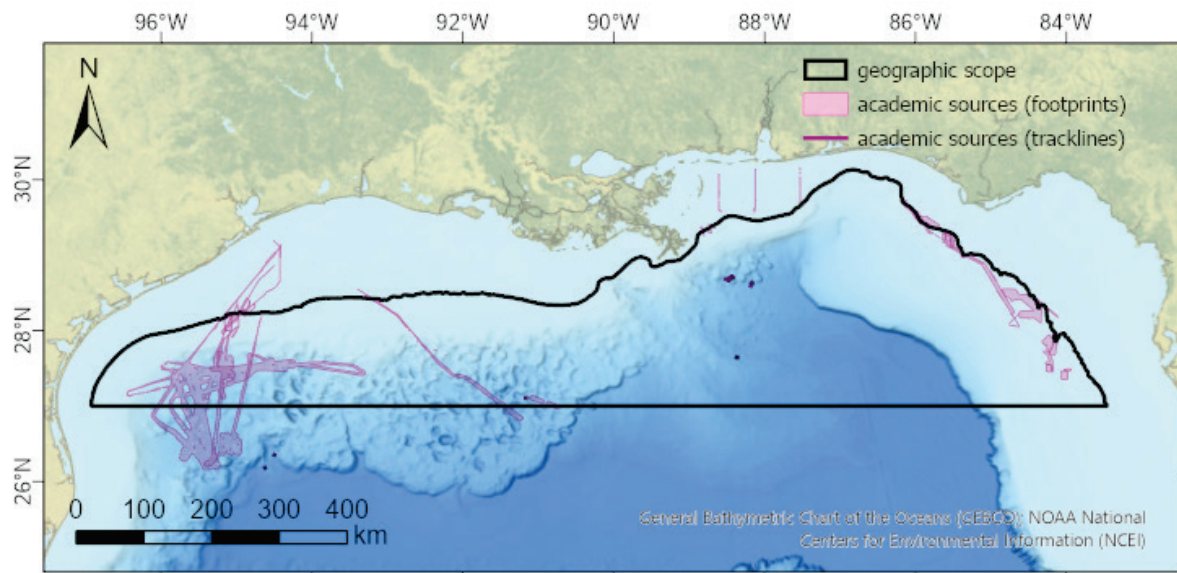


Figure 9. Footprints and tracklines for seafloor mapping data acquired from academic sources.

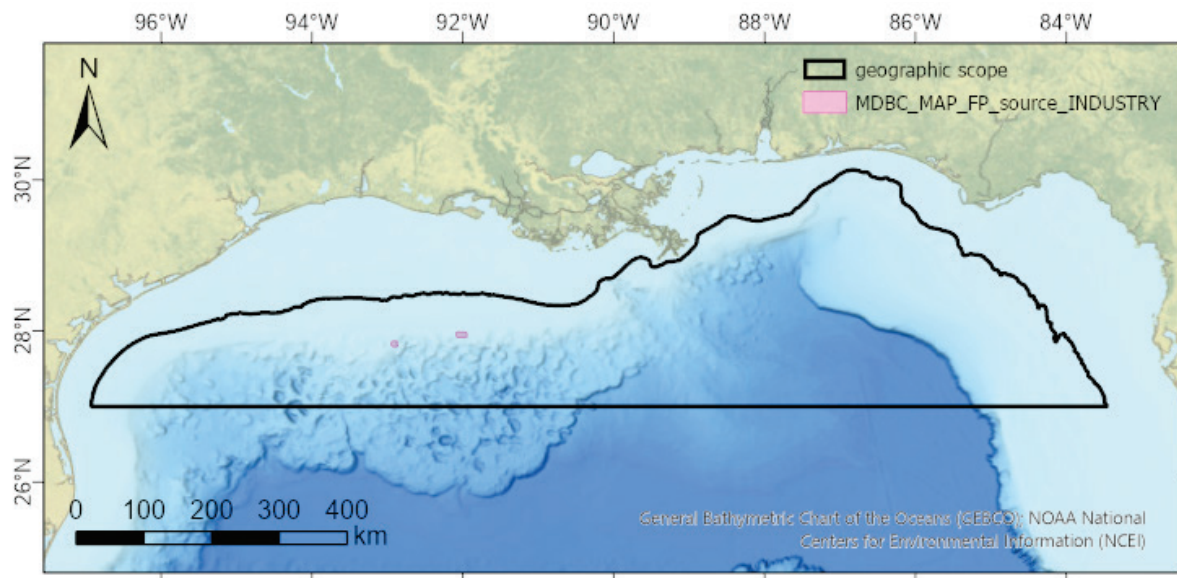


Figure 10. Footprints for seafloor mapping data acquired from industry sources.

### 8.1.3. Sensor

Seafloor mapping data included in the inventory were collected with multibeam echosounders, multibeam fishery echosounders, side-scan sonar, and seismic profilers (Table 16; Figure 11). The highest number and coverage of data were from multibeam echosounders (45.7% of footprints, 674,000 km<sup>2</sup>; Figure 12), which provided bathymetry and/or backscatter, or side-scan sonar mosaics (53.2% of footprints, 656,000.1 km<sup>2</sup>; Figure 13), whereas multibeam fishery echosounders surveyed only 555.7 km<sup>2</sup> (Figure 14). Over 99% of tracklines were associated with seismic profilers, covering a linear area of nearly 2 million km (Figure 15); <1% of tracklines were from multibeam echosounders, as well (Figure 12).

Table 16. Seafloor mapping summary statistics for footprints and tracklines (type) by sensor. Count is the total number of footprints or trackline segments. Count (%) represents the percentage of footprints or tracklines by sensor. Area (km<sup>2</sup>) and length (km) are both cumulative values or sums across all footprints or tracklines, respectively.

Type	Sensor	Count	Count (%)	Area (km <sup>2</sup> )	Length (km)
footprint	MBES	317	45.7	674,394.4	NA
footprint	MBES_FISHERY	8	1.2	555.7	NA
footprint	SSS	369	53.2	656,400.1	NA
trackline	MBES	6	<1	NA	162.4
trackline	SEIS	13,716	100	NA	1,891,102.0

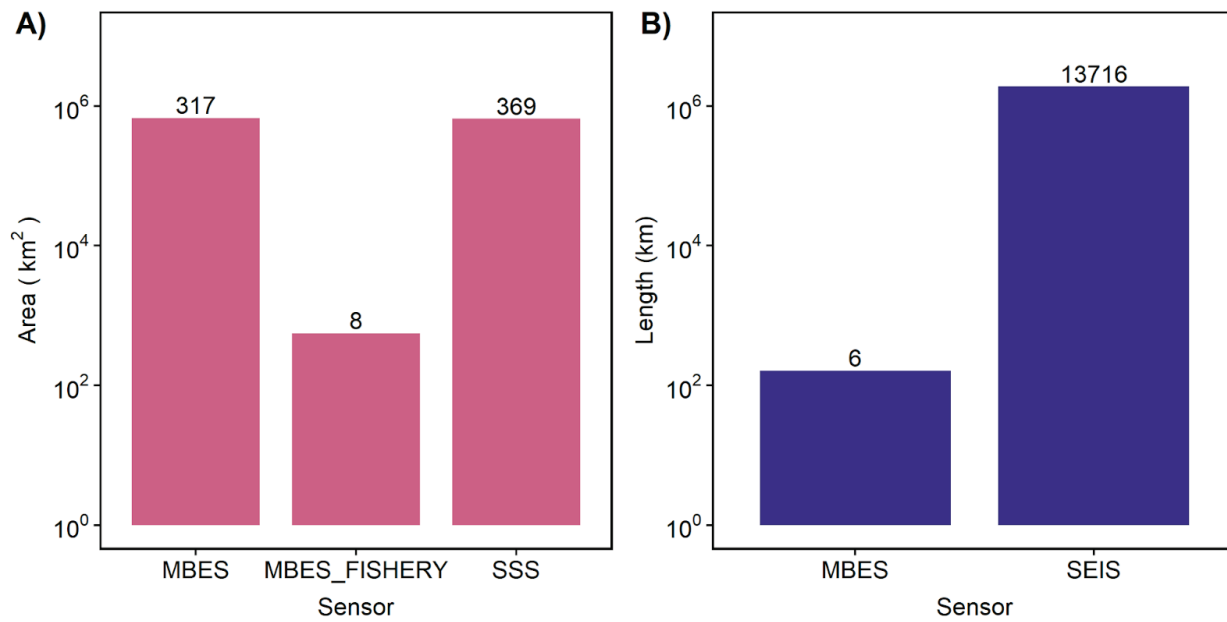


Figure 11. Logarithmic cumulative area (km<sup>2</sup>) of seafloor mapping footprints (A) and tracklines (B) by sensor. Numbers above bars represent the number of mapping survey footprints or trackline segments that comprise the cumulative area value.

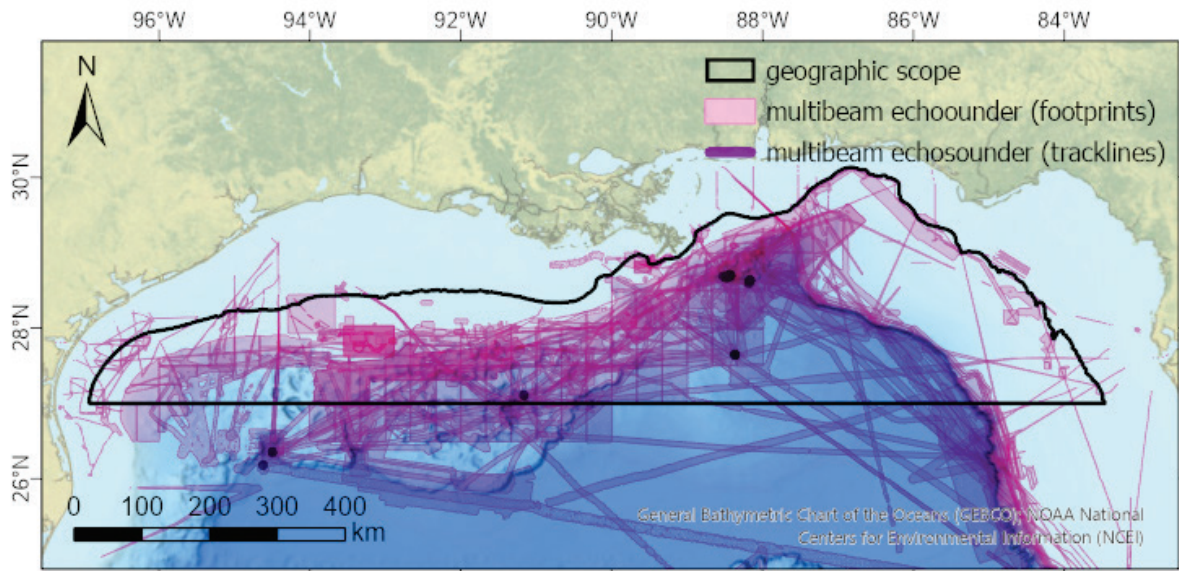


Figure 12. Footprints and tracklines for seafloor mapping data collected with multibeam echosounders.

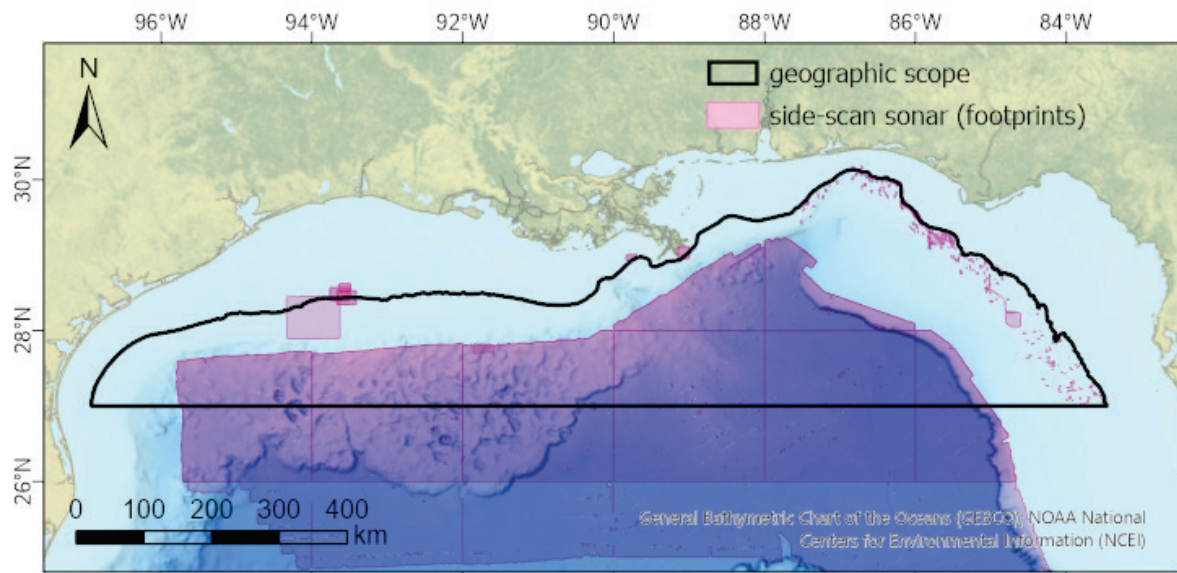


Figure 13. Footprints for seafloor mapping data collected with side-scan sonar.

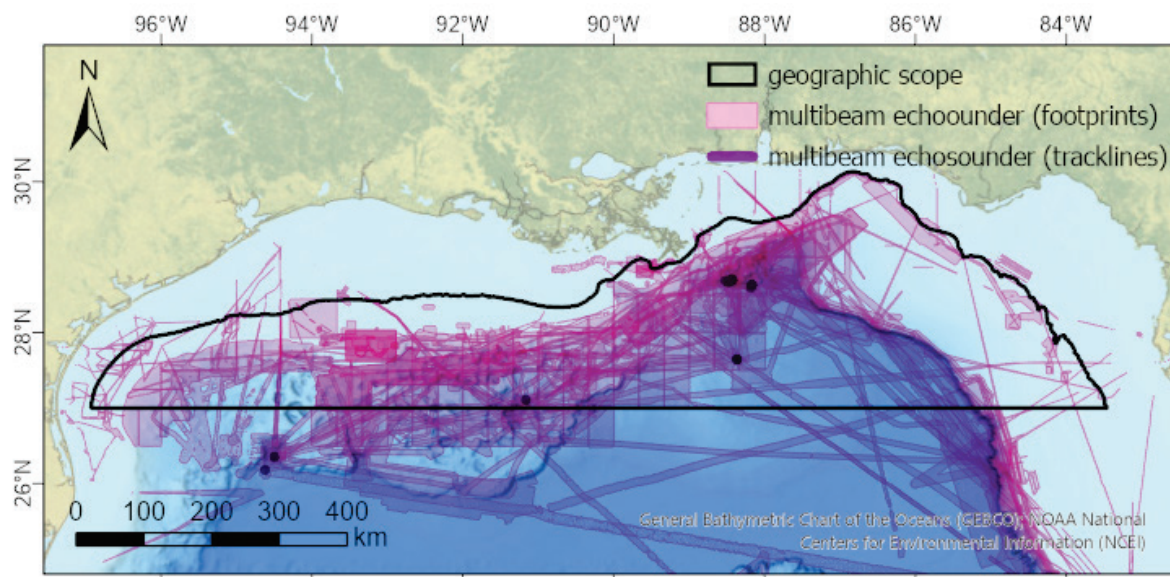


Figure 14. Footprints for seafloor mapping data collected with fishery multibeam echosounders.

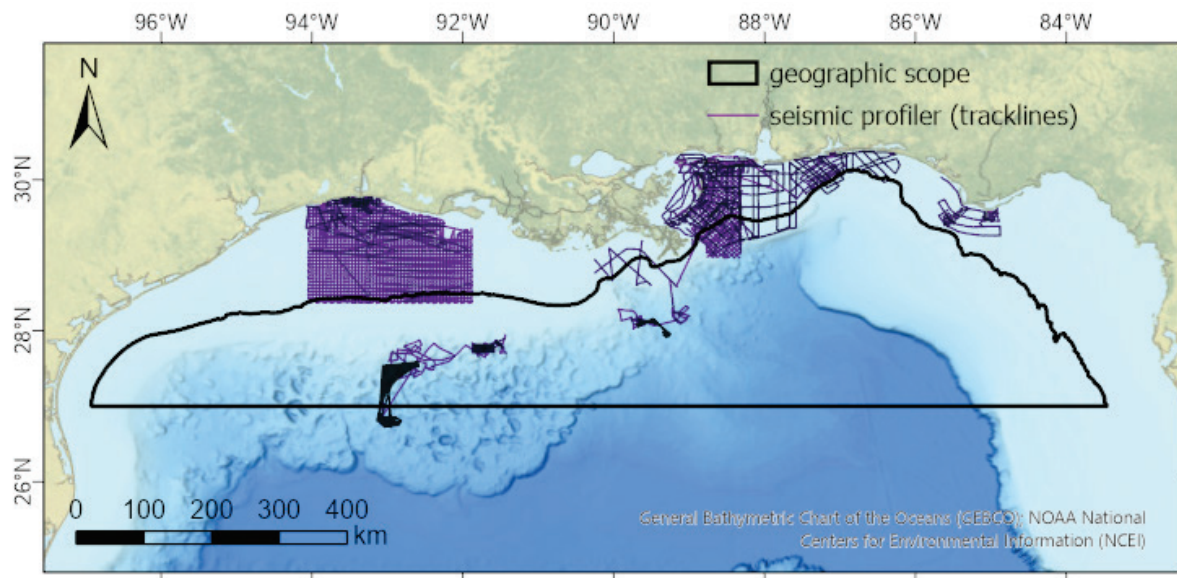


Figure 15. Tracklines for seafloor mapping data collected with seismic profilers.

#### 8.1.4. Collection Decade

The inventory included seafloor mapping data from 1980 through 2021 (Table 17; Figure 16). Interestingly, footprints from the 1980s spanned the largest area (approximately 700,000 km<sup>2</sup>) yet only accounted for 6.6% of footprints because there were few broad surveys (Figure 17). For example, the USGS-led side-scan survey (GLORIA, or Geological Long-Range Inclined ASDIC) in the 1980s mapped a large portion of the GOM at coarse resolution. 1990s data covered less area (approximately 75,000 km<sup>2</sup>; Figure 18), as did 2000s data (approximately 190,000 km<sup>2</sup>; Figure 19). 2010 footprints covered the second largest area (approximately 360,000 km<sup>2</sup>) and represented over 70% of all footprints (Figure 20). Fourteen footprints from the 2020s were included, covering approximately 7,000 km<sup>2</sup>, yet these only accounted for available data through 2021 (Figure 21). Additional data from the 2020s likely exist that have not yet been archived or published and will be added to the inventory during the accretion phase. Trackline extent was highest in the 2000s (approximately 1,900,000 km) followed by the 1990s, 1980s, and 2010s (Figure 16B).

Table 17. Seafloor mapping summary statistics for footprints and tracklines (type) by data collection decade. Count is the total number of footprints or trackline segments. Count (%) represents the percentage of footprints or tracklines by decade. Area (km<sup>2</sup>) and length (km) are both cumulative values or sums across all footprints or tracklines, respectively.

Type	Decade	Count	Count (%)	Area (km <sup>2</sup> )	Length (km)
footprint	1980s	46	6.6	699,276.1	NA
footprint	1990s	51	7.3	75,904.1	NA
footprint	2000s	93	13.4	186,375.8	NA
footprint	2010s	490	70.6	362,184.0	NA
footprint	2020s	14	2	7,610.2	NA
trackline	1980s	178	1.3	NA	21,038.0
trackline	1990s	572	4.2	NA	21,442.7
trackline	2000s	12,966	94.5	NA	1,848,622.0
trackline	2010s	6	0	NA	162.4

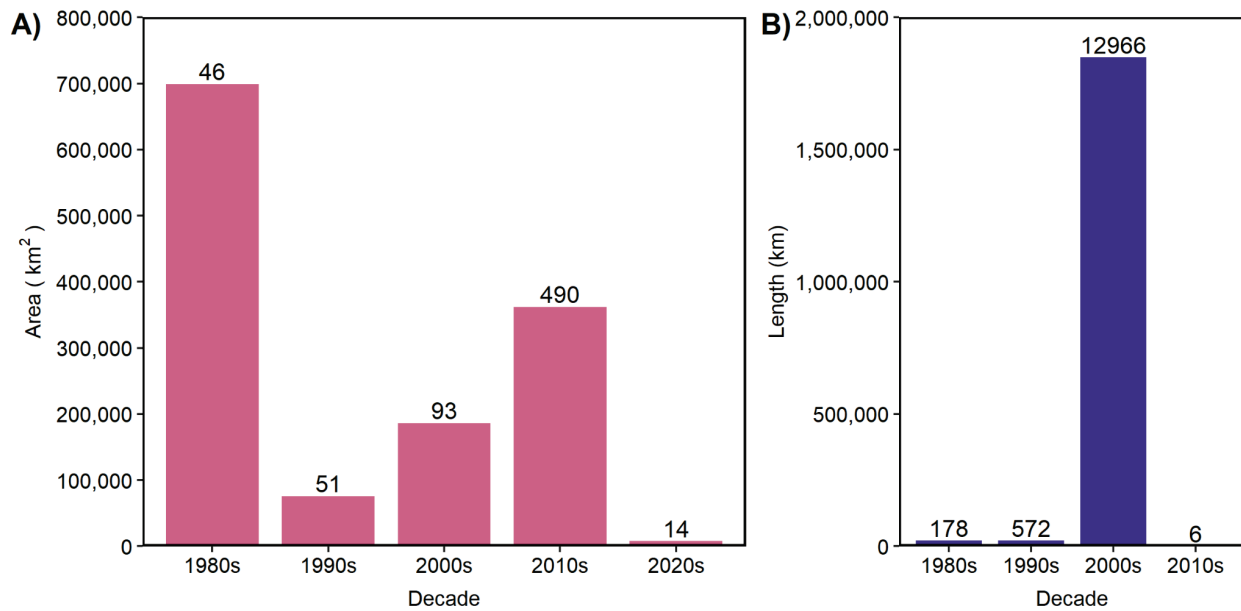


Figure 16. Cumulative area (km<sup>2</sup>) of seafloor mapping footprints (A) and tracklines (B) by data collection decade. Numbers above bars represent the number of mapping survey footprints or trackline segments that comprise the cumulative area value.

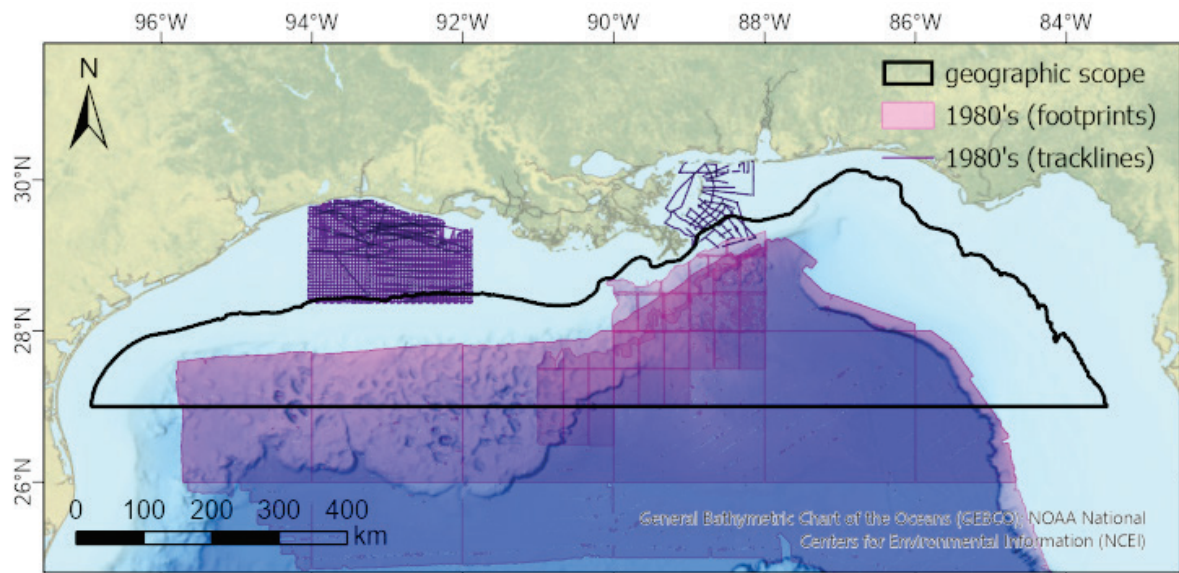


Figure 17. Footprints and tracklines for seafloor mapping data from the 1980s. This includes data whose acquisition ended in the 1980s.

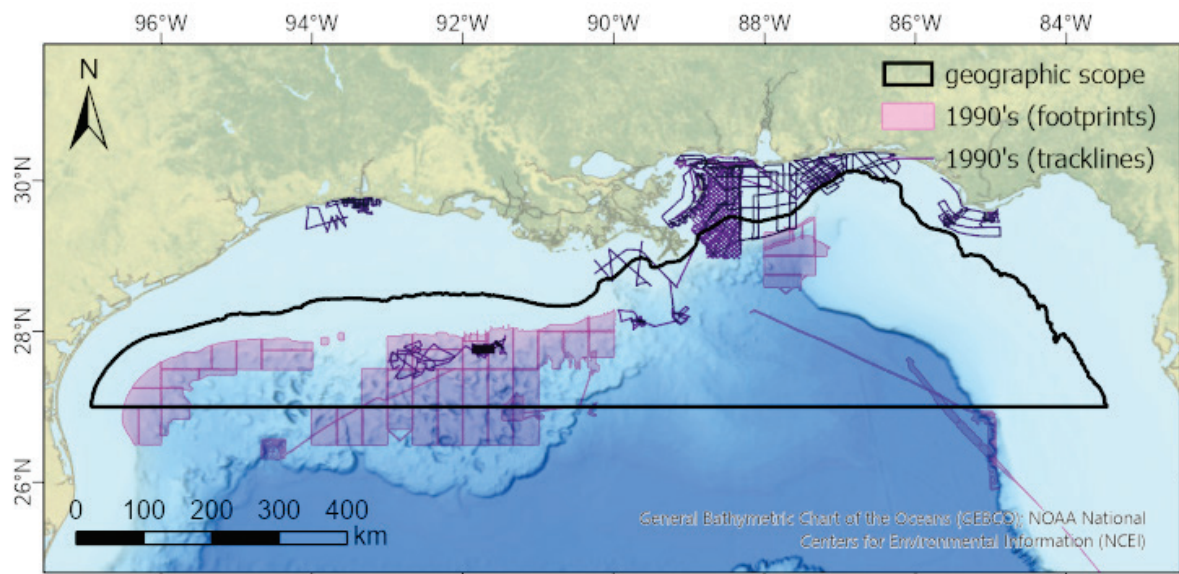


Figure 18. Footprints and tracklines for seafloor mapping data from the 1990s. This includes data whose acquisition ended in the 1990s.



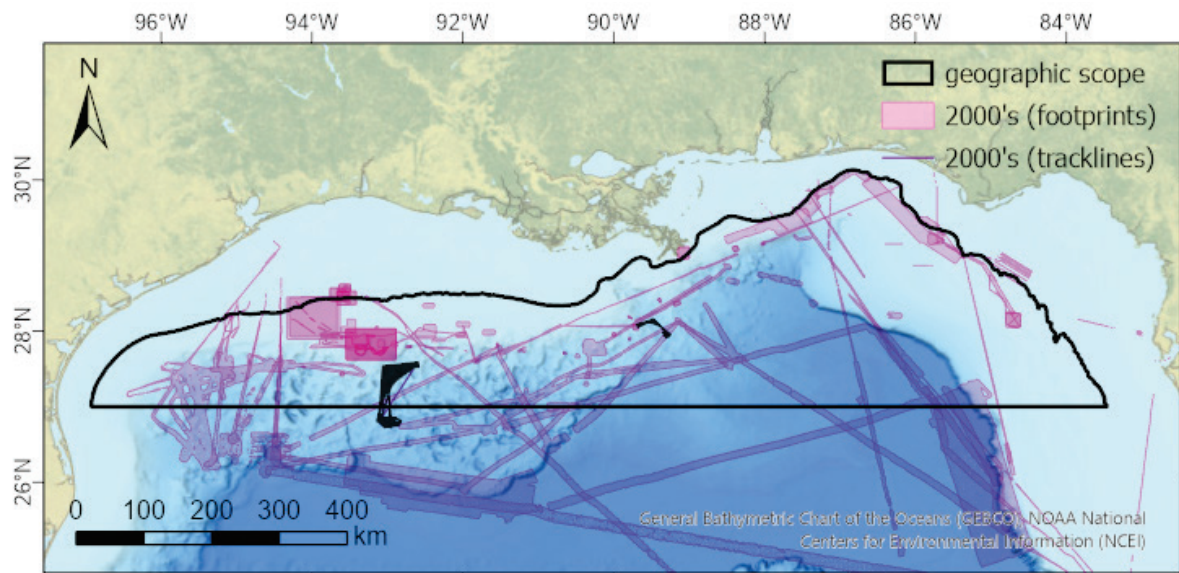


Figure 19. Footprints and tracklines for seafloor mapping data from the 2000s. This includes data whose acquisition ended in the 2000s.

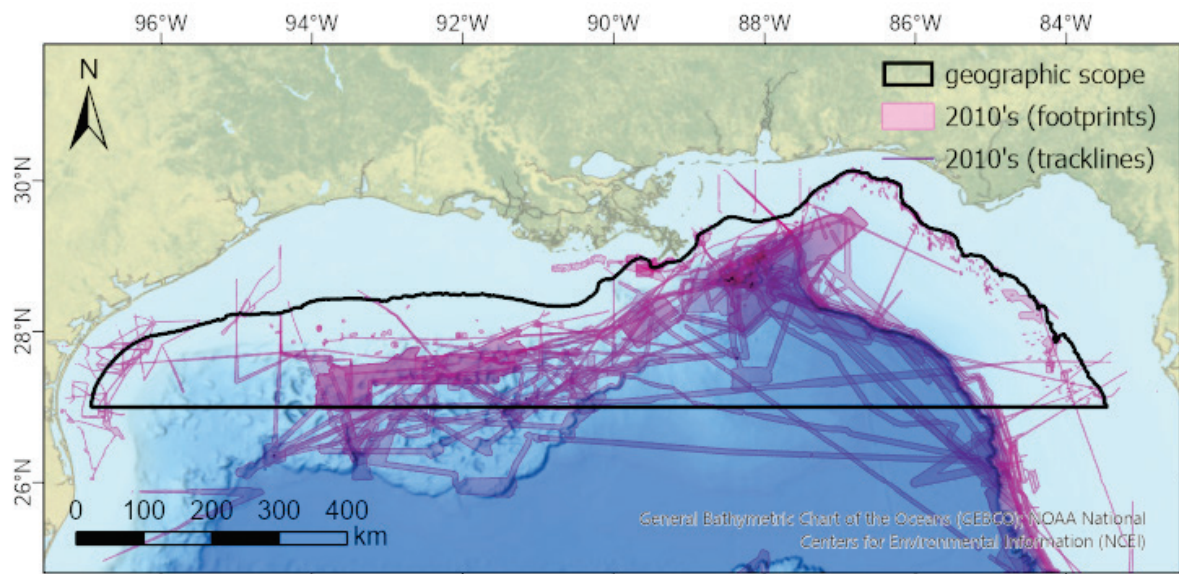


Figure 20. Footprints and tracklines for seafloor mapping data from the 2010s. This includes data whose acquisition ended in the 2010s.

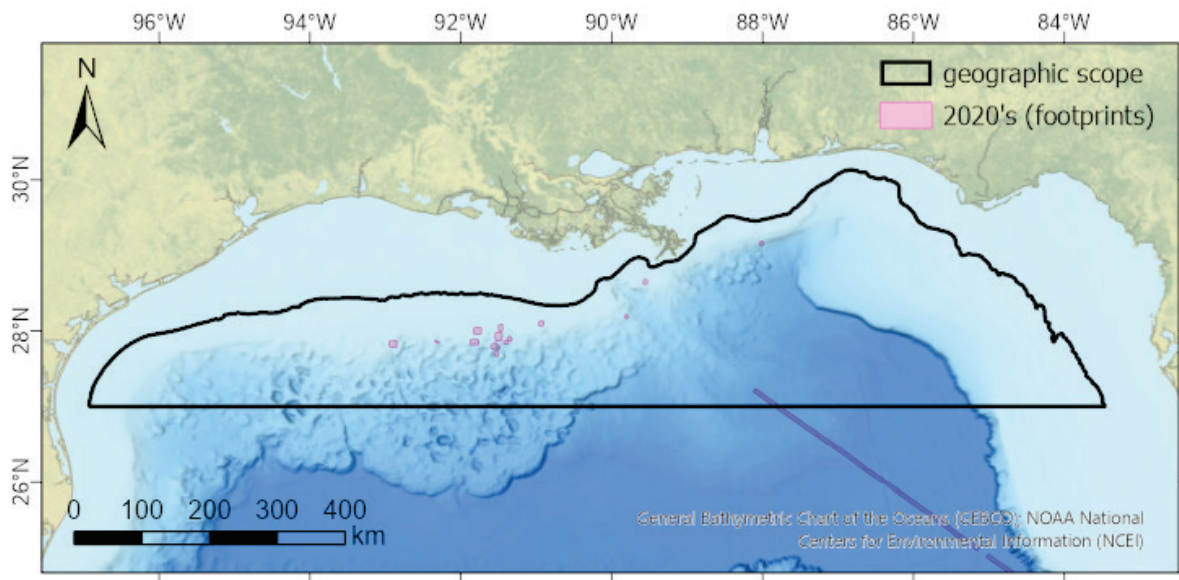


Figure 21. Footprints for seafloor mapping data from the 2020s. This includes data whose acquisition ended in the 2020s.

### 8.1.5. Resolution

Seafloor mapping data spanned from fine resolution (<1 m) to coarse resolutions (>100 m or unprocessed data; Table 19). Footprints with finer resolution covered less area of the seafloor than footprints with coarser resolutions (Figure 22), and finer resolution data were generally collected in shallow waters of the geographic scope, whereas coarser resolution data were distributed in deeper waters. The majority of seafloor mapping footprints were <1 m resolution (54.5% of footprints), yet these fine resolution footprints covered the smallest amount of seafloor (approximately 12,000 km<sup>2</sup>) compared to other resolution footprints (Figure 23). Footprints with resolutions 1–10 m (Figure 24), 10–32 m (Figure 25), and 32–100 m (Figure 26) covered progressively broader areas. Footprints of >100 m resolution, while only constituting 4.8% of the total number of footprints, encompassed the largest amount of seafloor area (approximately 478,000 km<sup>2</sup>; Figure 27). Trackline data, which were quantified by length (km) instead of area, totaled nearly 2,000,000 linear km (Table 18). Footprints and tracklines corresponding to raw or point data that were not fully processed were categorized as “coarse” and covered broad extents of the seafloor (Figure 28).

Table 18. Seafloor mapping summary statistics for footprints and tracklines (type) by resolution. Count is the total number of footprints or trackline segments. Count (%) represents the percentage of footprints or tracklines by resolution. Area (km<sup>2</sup>) and length (km) are both cumulative values or sums across all footprints or tracklines, respectively.

Type	Resolution	Count	Count (%)	Area (km <sup>2</sup> )	Length (km)
footprint	0–1 m	378	54.5	11,616.9	NA
footprint	1–10 m	122	17.6	24,964.6	NA
footprint	10–32 m	34	4.9	84,120.6	NA
footprint	32–100 m	33	4.8	382,731.3	NA
footprint	>100 m	6	0.9	478,053.3	NA
footprint	coarse	121	17.4	349,863.5	NA
trackline	coarse	13,722	100	NA	1,891,265.0

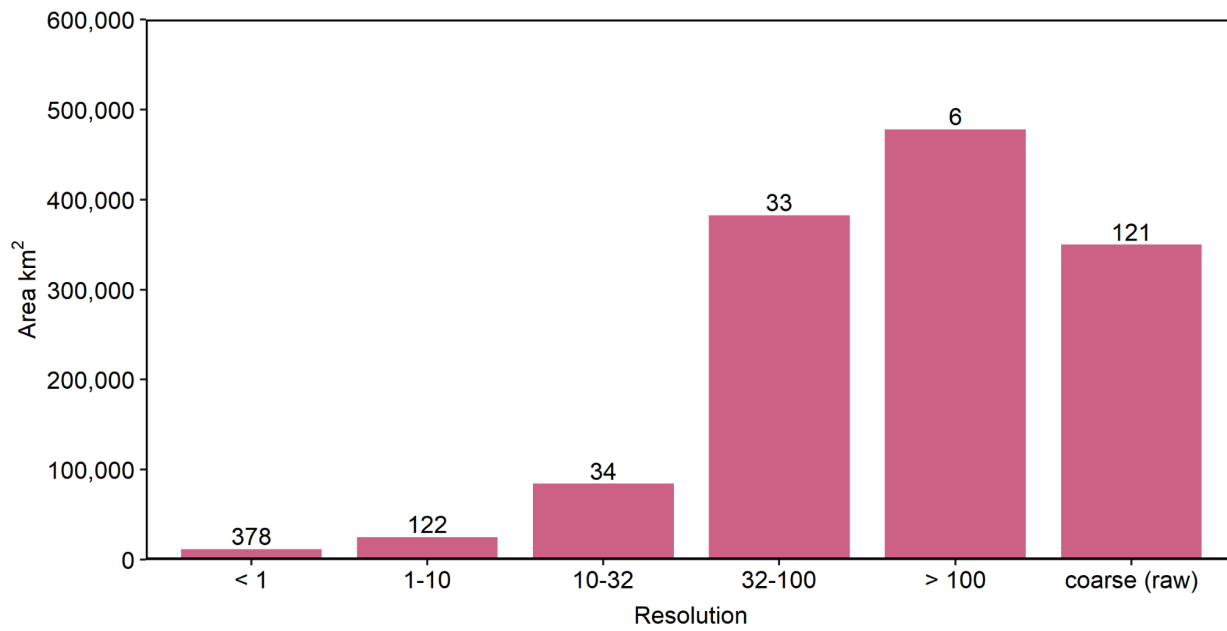


Figure 22. Cumulative area (km<sup>2</sup>) of seafloor mapping footprints by resolution (m). Numbers above bars represent the number of mapping survey footprints that comprise the cumulative area value. Coarse resolution corresponds to footprints from raw or point data that have not been fully processed.

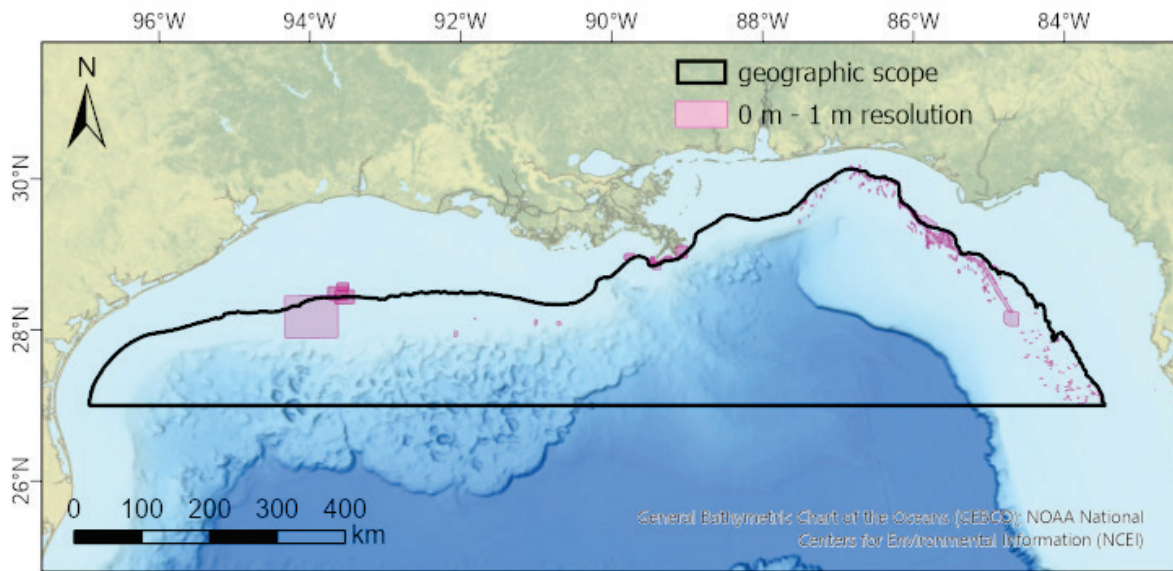


Figure 23. Footprints for seafloor mapping data with resolution from <1 m.

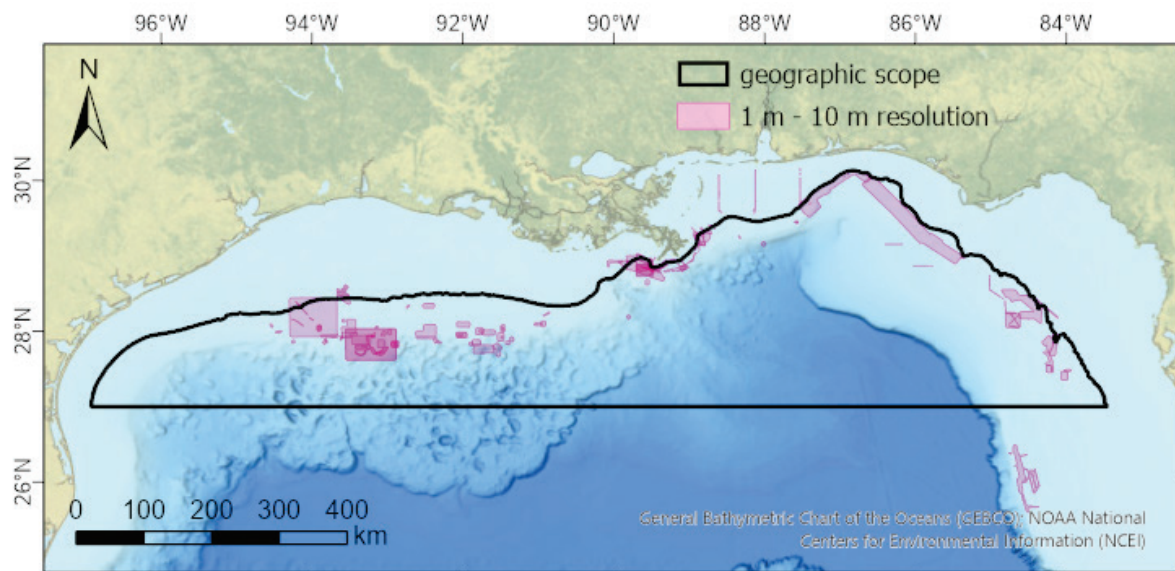


Figure 24. Footprints for seafloor mapping data with resolution from 1 m to 10 m.

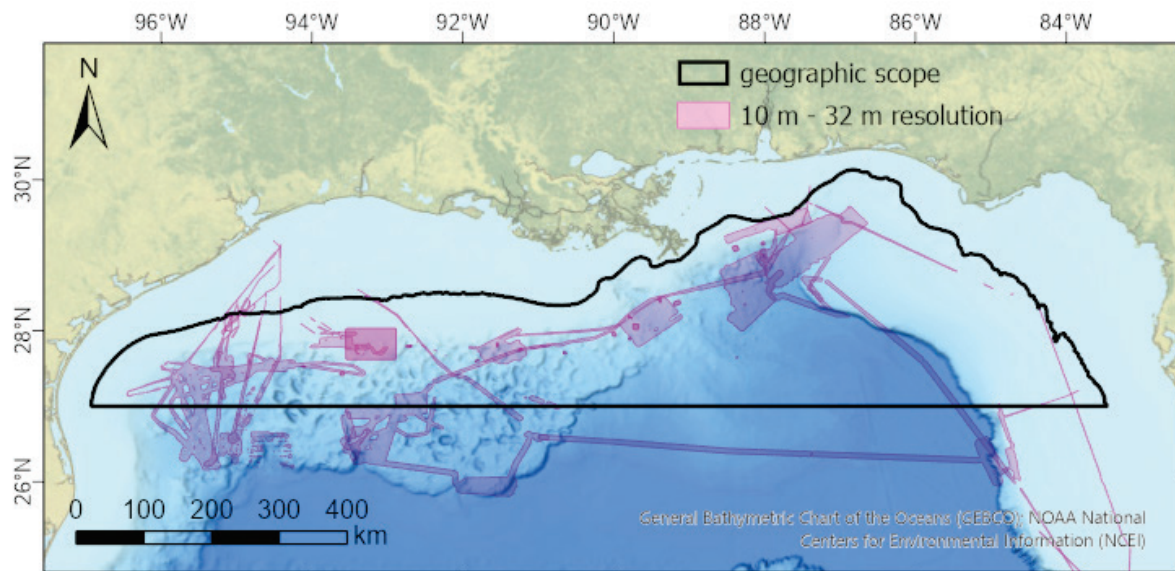


Figure 25. Footprints for seafloor mapping data with resolution from 10 m to 32 m.

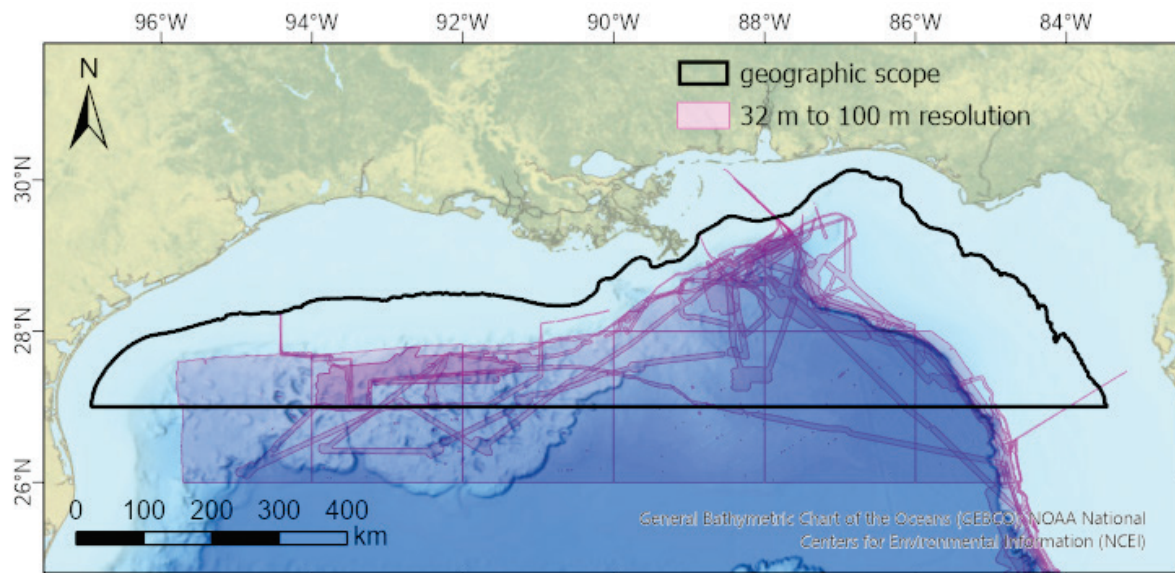


Figure 26. Footprints for seafloor mapping data with resolution from 32 m to 100 m.

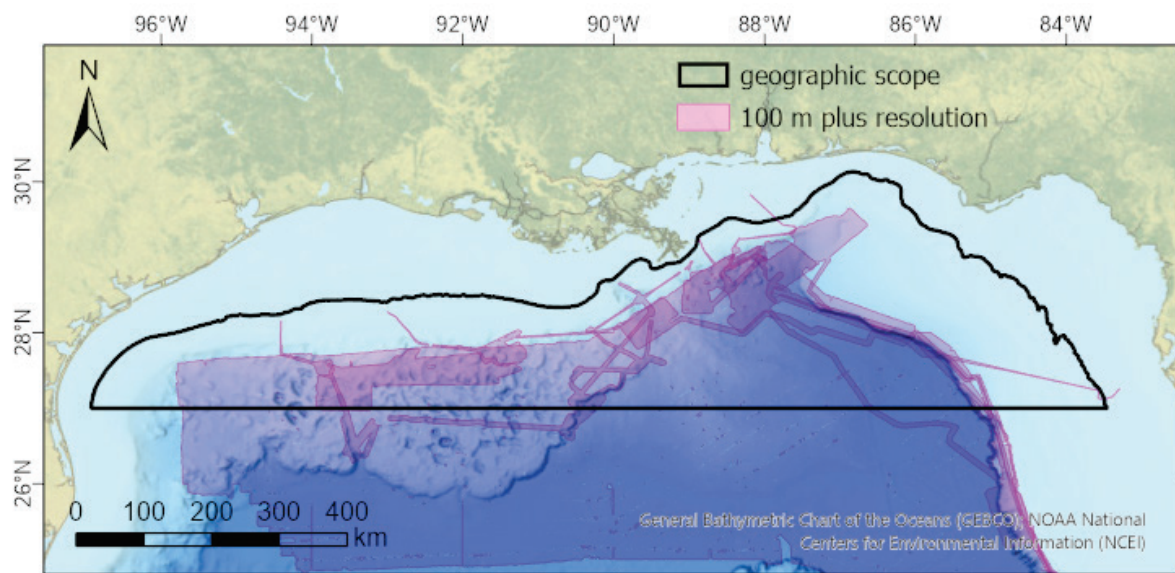


Figure 27. Footprints for seafloor mapping data with resolution >100 m.

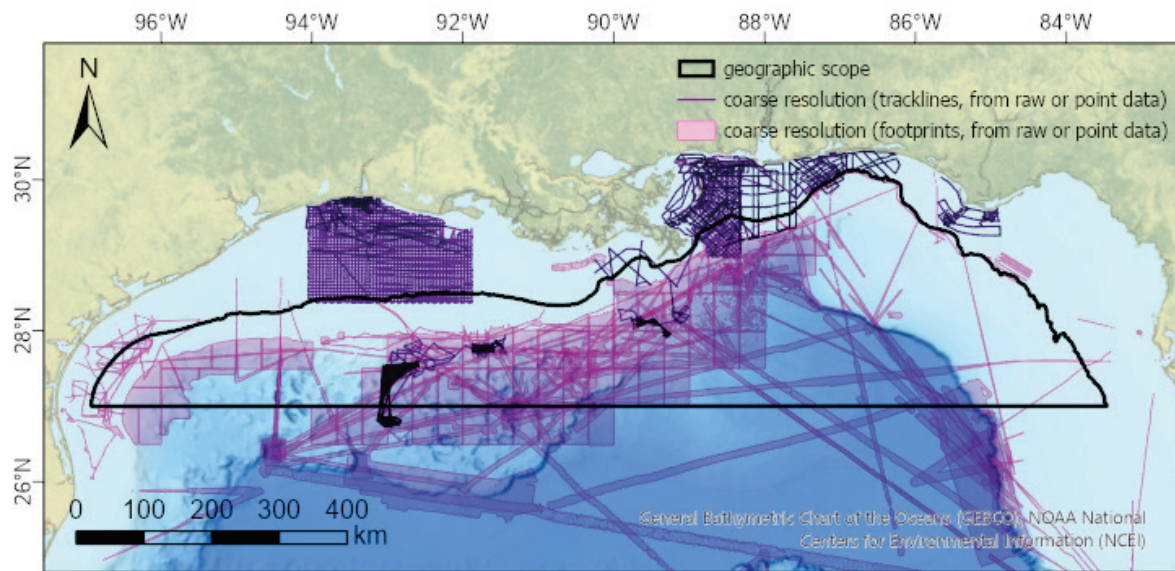


Figure 28. Footprints and tracklines for seafloor mapping data with coarse resolution (e.g., point or raw data).

### 8.1.6. Depth

The inventory geographic scope spanned from the 50 m isobath to the 27°N parallel, reaching depths greater than 2,500 m. Seafloor mapping data covered a variety of depths (Figure 29; Figure 30). The grand mean depth of all surveys included in the inventory was 760.0 m, and the median was 320.9 m. The shallowest average survey depth was 0.5 m, whereas the deepest was 3,389.6 m. These shallow and deep extents were included because data that intersected the geographic scope and extended outside of the scope to shallower or deeper depths were collected in the inventory.

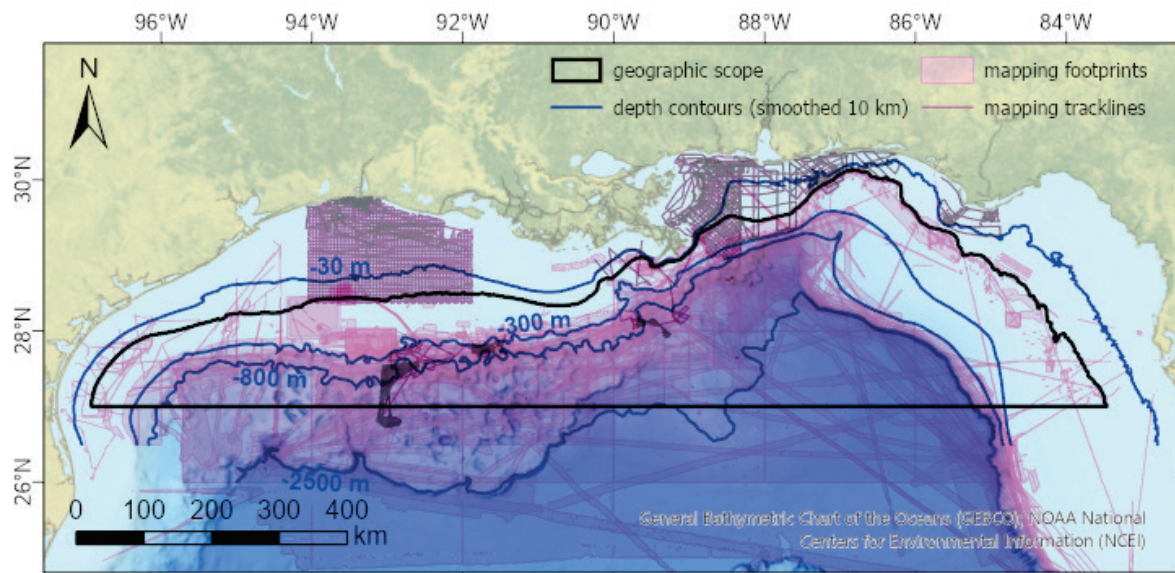


Figure 29. Footprints and tracklines for seafloor mapping data by depth bins. Depth contours represent 30 m, 300 m, 800 m, and 2500 m.

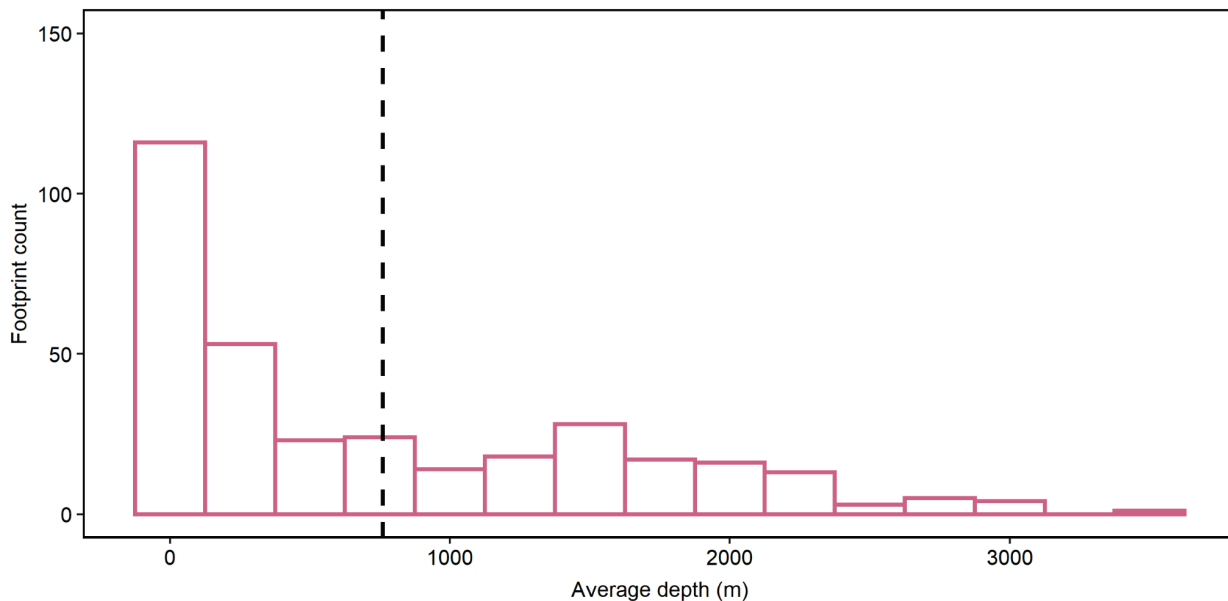


Figure 30. Histogram of average depths for seafloor mapping footprints. Depth bins are 200 m each. Dotted black line represents the mean depth (760 m).

### 8.1.7. Derived Datasets

Products derived from raw or processed seafloor mapping footprints, such as habitat maps derived from backscatter, synthetic data layers, or extents of surveys for which data exist but are not yet available, were included in the inventory. These derived datasets were handled individually, separately from the combined footprint files described above (Sections 8.1.1–8.1.6). Twenty derived datasets were included in the inventory, representing polygon, point, and line files, several of which are highlighted below.

Federally derived products from BOEM, USGS, and NOAA were included in the inventory. BOEM derived products included polygons of high-resolution geophysical surveys, for which data exist but have not yet been made available for inclusion in the inventory (Figure 31A). BOEM also contributed locations of bottom anomalies interpreted from seismic surveys (Figure 31B), as well as the polygons corresponding to 3D seismic data used to generate the BOEM Bathymetry Grid from 3D Seismic over a broad extent of the GOM (Figure 31C). USGS derived datasets include interpretations of seafloor features from side-scan sonar datasets (e.g., GLORIA survey) for salt diapirs (Figure 32A), mass wasting locations (Figure 32B), geomorphological features (Figure 32C), hydrate locations (Figure 32D), and hydrocarbon seeps (Figure 32E). NOAA derived products included locations of multibeam fishery echosounder surveys conducted by the NOAA Ship *Pisces* (Figure 33A), as well as sampling locations and reef delineations for NOAA NMFS Panama City side-scan sonar surveys (Figure 33B).

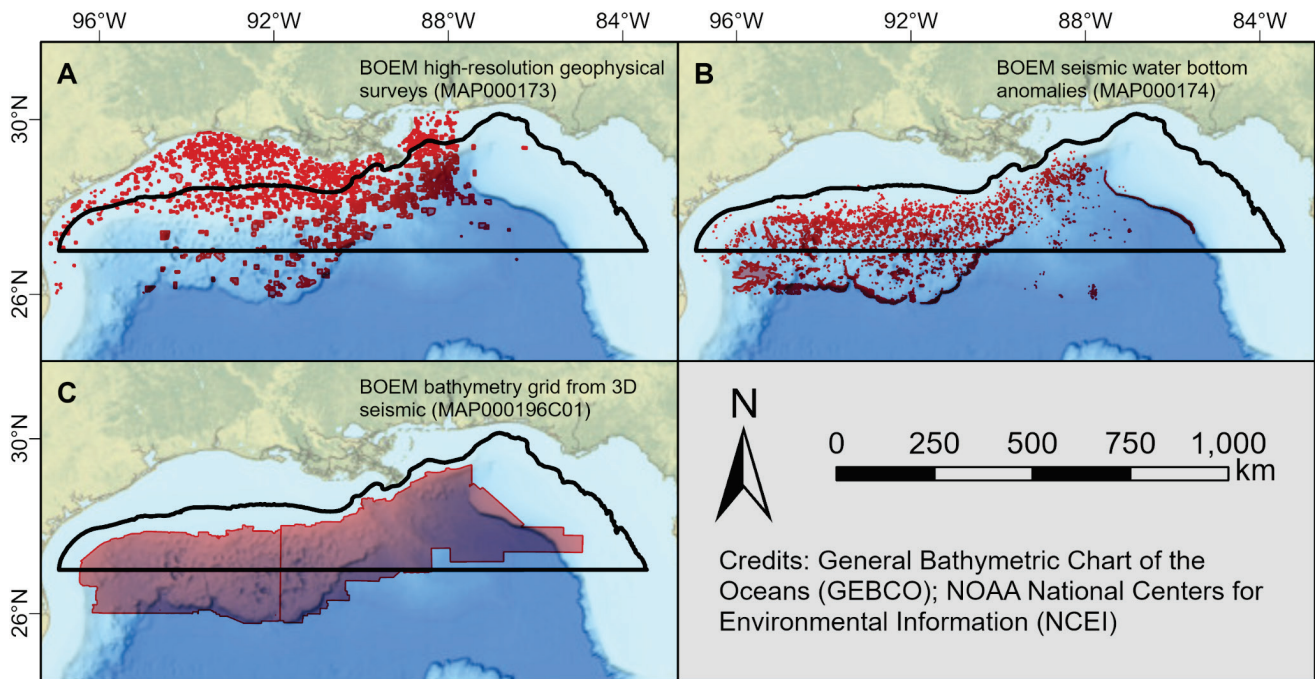


Figure 31. BOEM derived data products: A) polygon of high-resolution geophysical surveys; B) interpretations of bottom anomalies from seismic surveys; and C) polygons corresponding to the extent of the BOEM Bathymetry Grid from 3D Seismic. Black polygons represent the inventory geographic scope.



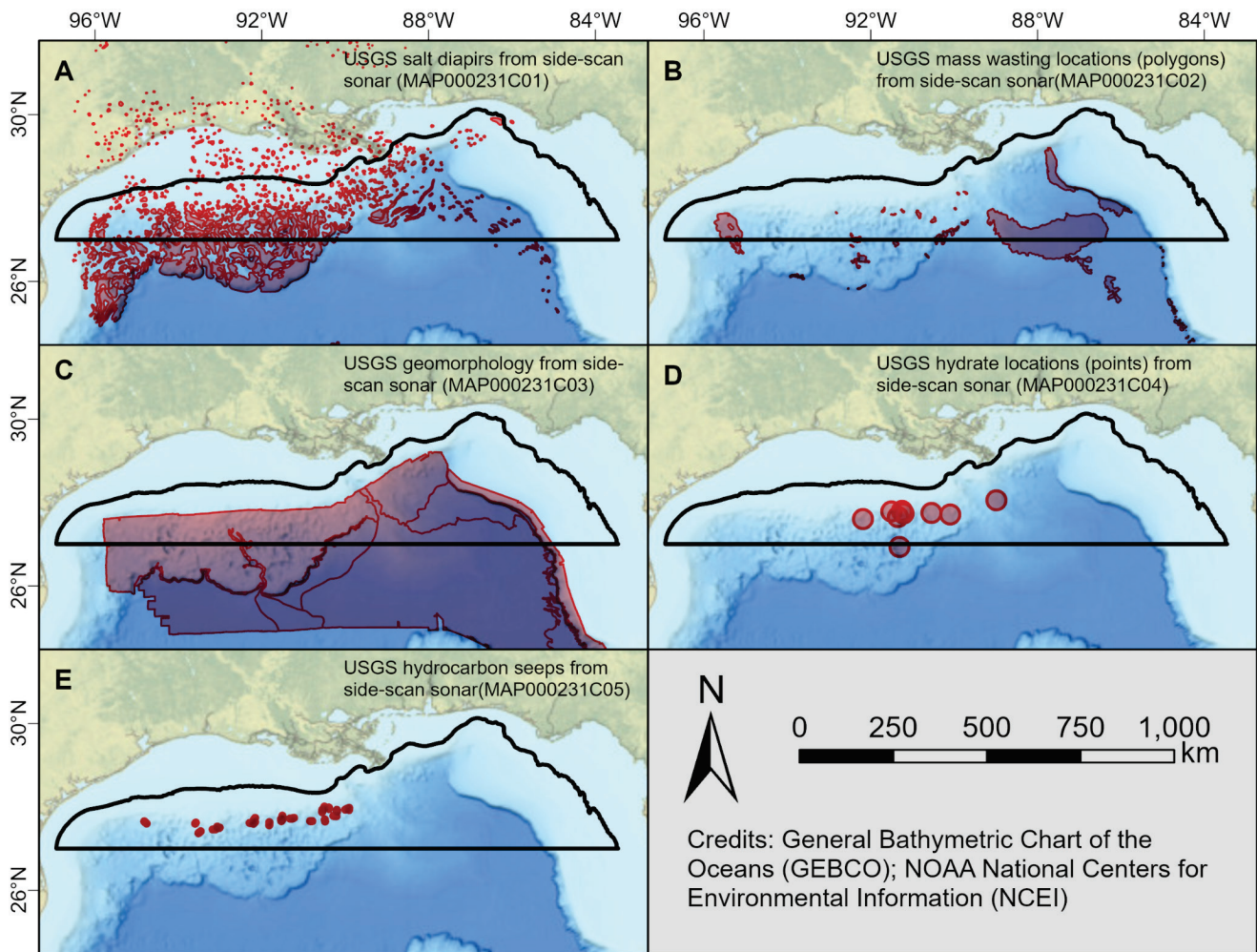


Figure 32. USGS interpretations of seafloor features derived from side-scan sonar: A) salt diapirs (also includes terrestrial assessments with other sensors), B) mass wasting locations, C) geomorphology, D) hydrate locations, and E) hydrocarbon seeps. Black polygons represent the inventory geographic scope.

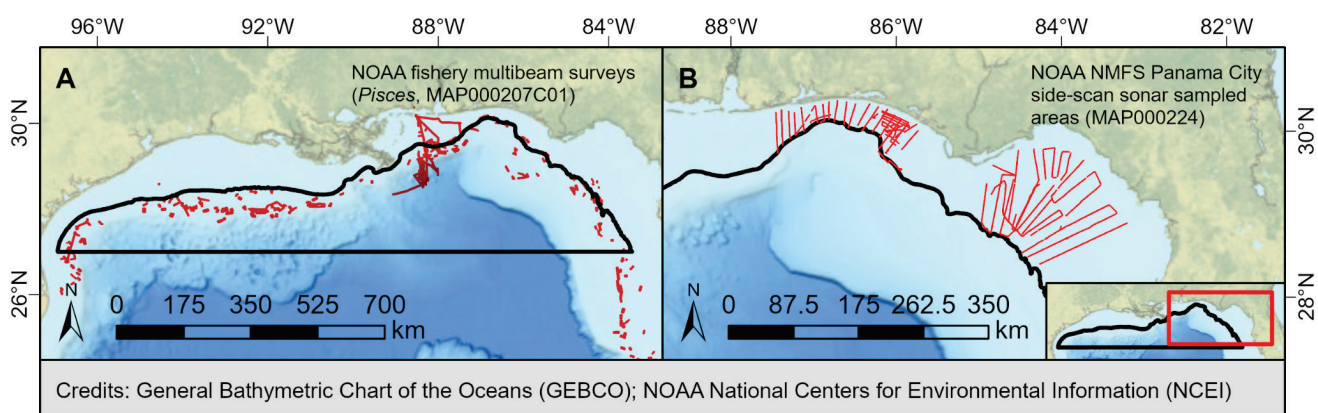


Figure 33. NOAA derived products: A) locations of fishery multibeam surveys conducted from the NOAA ship *Pisces*; and B) locations of side-scan sonar surveys conducted by the NOAA NMFS Panama City Lab. Black polygons represent the inventory geographic scope.

Non-federal sources also contributed derived products to the inventory. FWC/FWRI provided locations of side-scan sonar surveys conducted along the Gulf coast of Florida (Figure 34). These surveys were conducted across the shelf and portions extended into the geographic scope of the inventory. Polygons corresponding to Fugro and PGS (Petroleum Geo-Services) surveys, for which data exist but were unavailable for the inventory were also contributed (Figure 35).

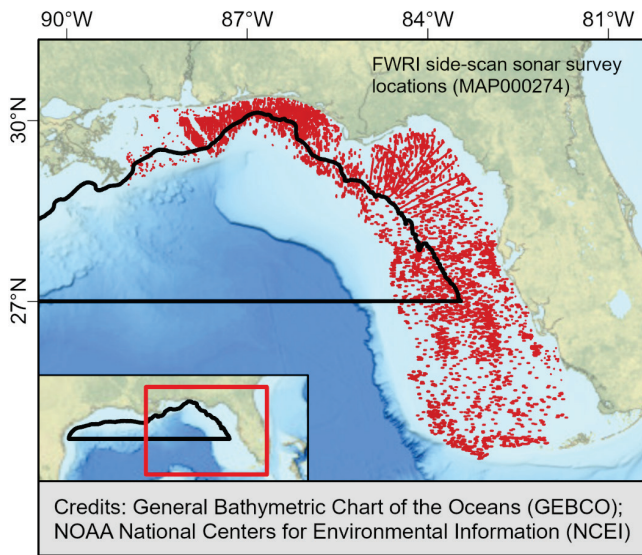


Figure 34. Locations of side-scan sonar surveys conducted by Florida Fish and Wildlife Research Institute (FWC/FWRI). Black polygon represents the inventory geographic scope.

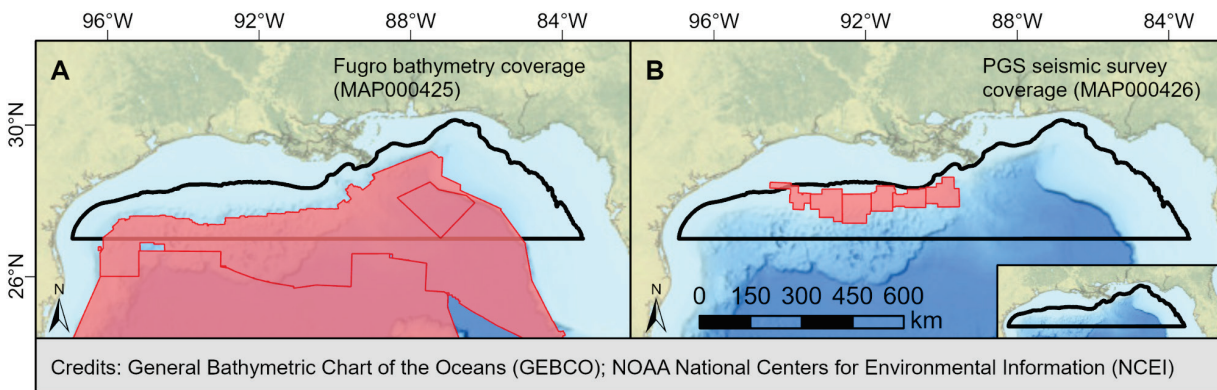


Figure 35. Polygons corresponding to existing A) Fugro bathymetry data and B) PGS seismic survey data.

### 8.1.8. Imagery

In addition to footprints and derived products, the seafloor mapping inventory also included 737 raster images. These raster images were the images used to generate spatial footprints and to calculate associated footprint attributes, such as resolution, depth, and survey area. Since the raster images were the basis for footprint generation, the images likewise span multiple sources, decades, and depths. Imagery ranged from fine resolution to coarse resolution and was mainly bathymetric or backscatter surfaces collected using multibeam echosounders (Figure 36) and intensity mosaics collected using side-scan sonar (Figure 37).

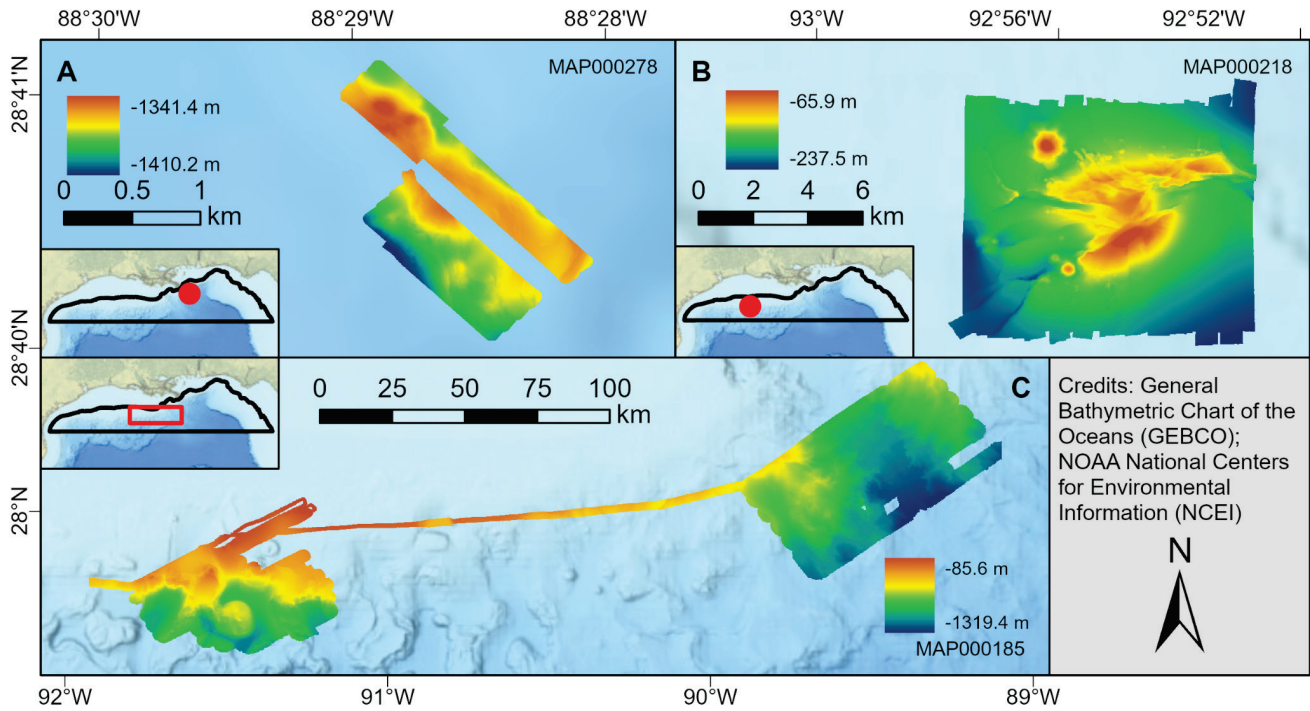


Figure 36. Representative seafloor mapping bathymetric imagery collected with multibeam echosounders from fine to coarse resolutions: A) multibeam collected from an autonomous underwater vehicle (Marine Geoscience Digital Library); B) shipboard multibeam of Elvers Bank in the Flower Gardens National Marine Sanctuary (NOAA); and C) shipboard multibeam of Green and Mississippi Canyons (NOAA). Black polygons represent the inventory geographic scope.

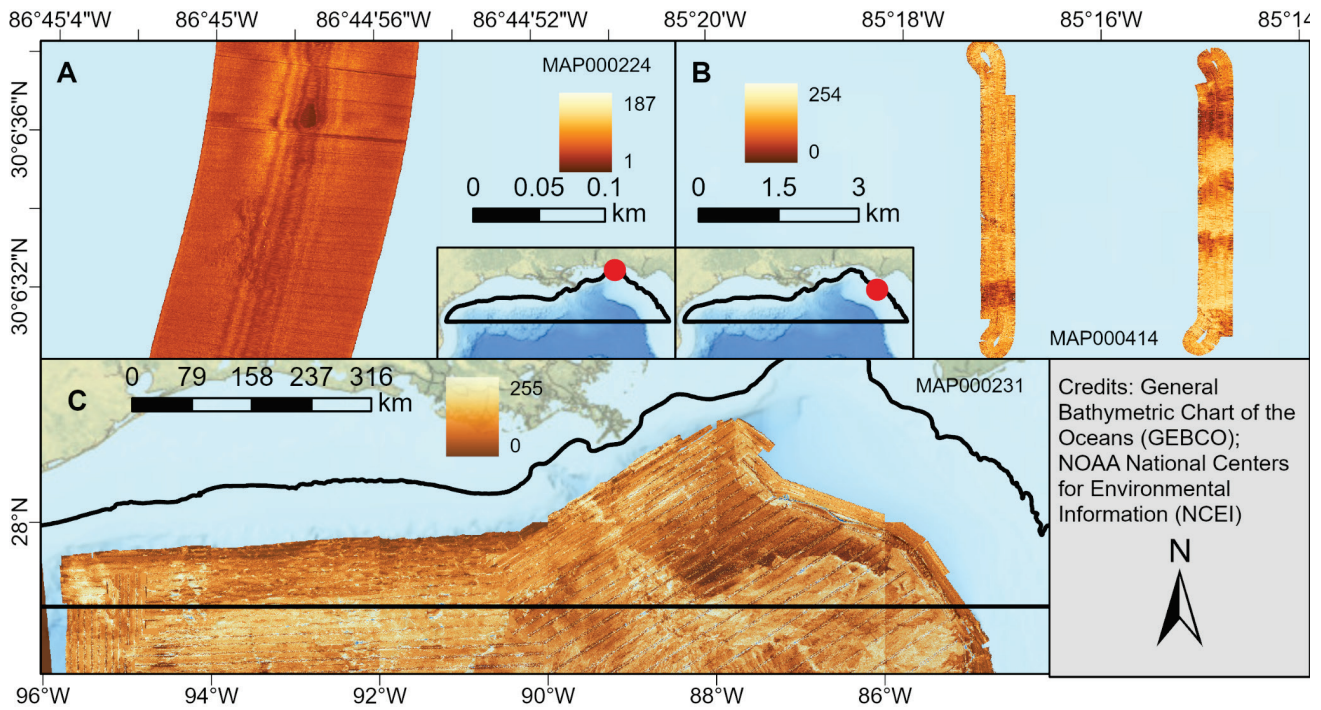


Figure 37. Representative seafloor mapping mosaic imagery collected with side-scan sonar from fine to coarse resolutions collected by: A) FWC/FWRI, B) NOAA NMFS Panama City Lab, and C) USGS GLORIA survey. Black polygons represent the inventory geographic scope. Imagery color represents intensity.

### 8.1.9. Mapping Datasets in Inventory Queue

The inventory process is ongoing, and datasets are added to the data catalog as they are discovered and processed for inventory compilation as they are acquired. Future updated versions of inventory products will encompass these additional seafloor mapping datasets, as we receive them. For example, we are currently working on adding the following datasets to the mapping inventory:

- Datasets from literature review for which data acquisition is still underway via external engagement with dataset points of contact
- Datasets from BOEM ESPIS for which data acquisition is still underway via external engagement with BOEM partners
- Underlying data corresponding to BOEM high-resolution industry survey polygons
- Underlying data corresponding to NSF and Columbia University Lamont Doherty Earth Observatory (LDEO) survey polygons
- NAVY datasets from the Naval Oceanographic Office
- Industry datasets, including those from Fugro and PGS
- Seismic and sub-bottom surveys acquired but in a proprietary format (.seggy) or held in other repositories besides those searched
- Data collected but not yet archived, published on, or discovered through engagement

Additionally, multibeam echosounder surveys that included bathymetry were included in the inventory, and if backscatter was available, those surfaces were also included. In cases when only bathymetry data were available, it is possible that backscatter data were also collected but not processed and thus available for inventory following additional post-processing. In the future, we plan to revisit the bathymetry surveys and determine whether backscatter was also collected and whether backscatter products may be generated and added to the inventory.

## 8.2. Ground-truthing Results

### 8.2.1. Overall

During the ground-truthing inventory process, 75 datasets were considered for inclusion. Of these, 67 datasets passed coarse-scale inclusion criteria and were further screened based on fine-scale inclusion criteria (Table 19; Figure 38). In total, 57 datasets passed fine-scale inclusion criteria and were either incorporated into the inventory or are in the process of being compiled into the inventory, as data acquisition or processing are underway (Table 19; Figure 38).

Table 19. Number of ground-truthing datasets screened according to coarse and fine scale inclusion criteria. Numbers of datasets excluded and included according to the inclusion criteria are also provided.

Screening Level	Datasets Screened	Datasets Excluded	Datasets Included
coarse scale screening	75	8	67
fine scale screening	67	10	57
datasets for accretion*	14	0	14

\*These are datasets that are in the process of being incorporated into the ground-truthing inventory

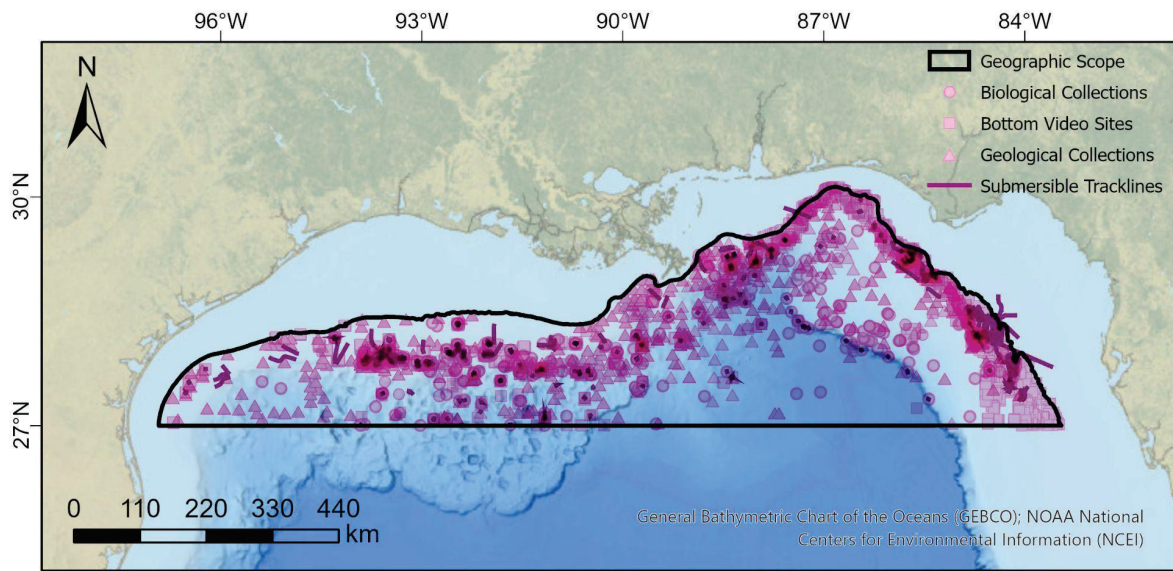


Figure 38. Ground-truthing surveys represented by biological or geological collection surveys, bottom video sites, or submersible tracklines.

### 8.2.2. Data Collection Type

Ground-truthing datasets were represented by the collection method(s) used, which include biological collections, geologic collections, bottom video surveys, and submersible surveys. There were 10 datasets representing biological collection surveys (Figure 39), 10 datasets representing bottom video surveys (Figure 40), eight datasets representing geological collection surveys (Figure 41), and 17 datasets representing submersible surveys (Figure 42).

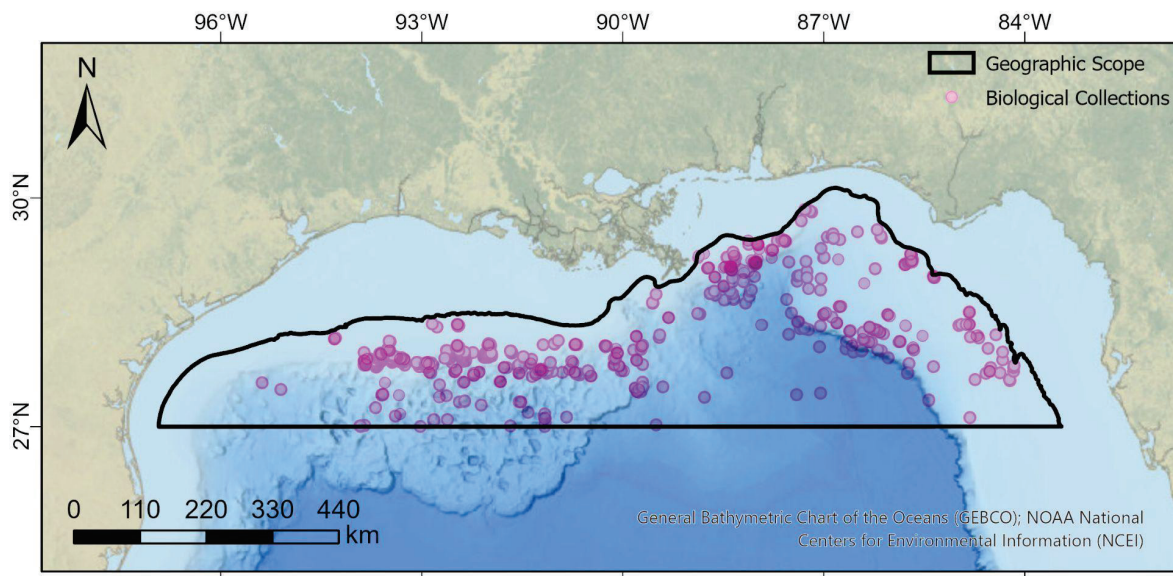


Figure 39. Biological collections conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.

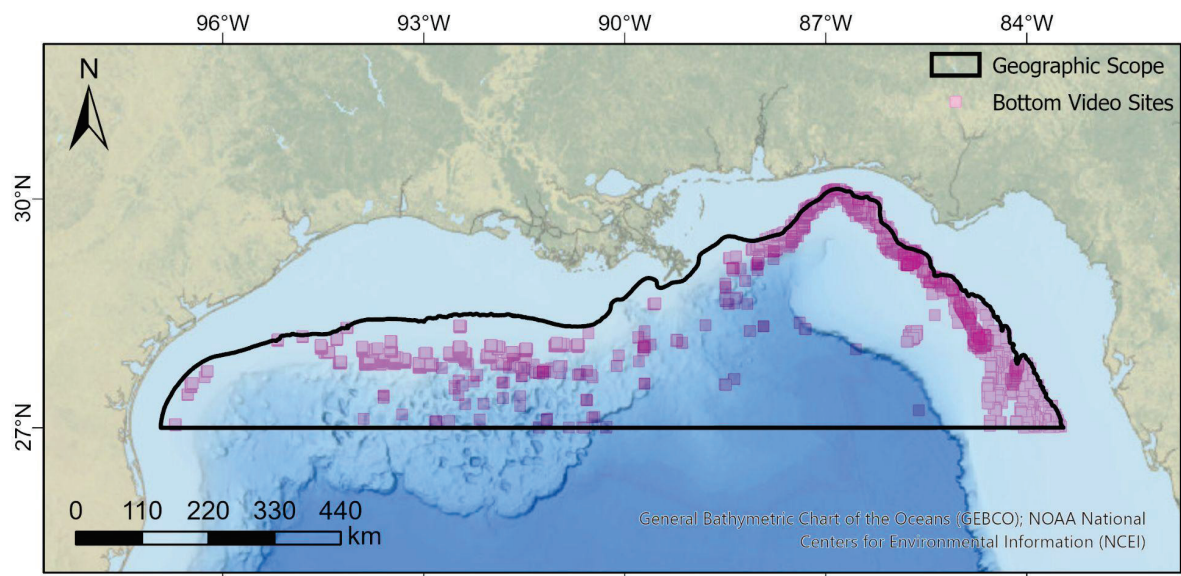


Figure 40. Bottom video sites conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.

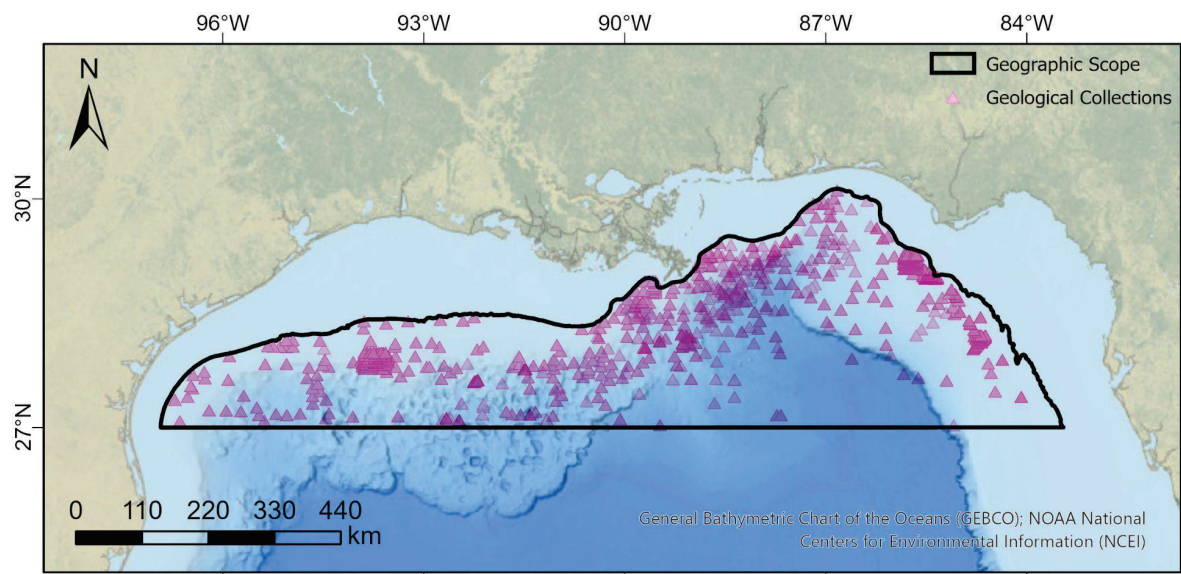


Figure 41. Geological collections conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.

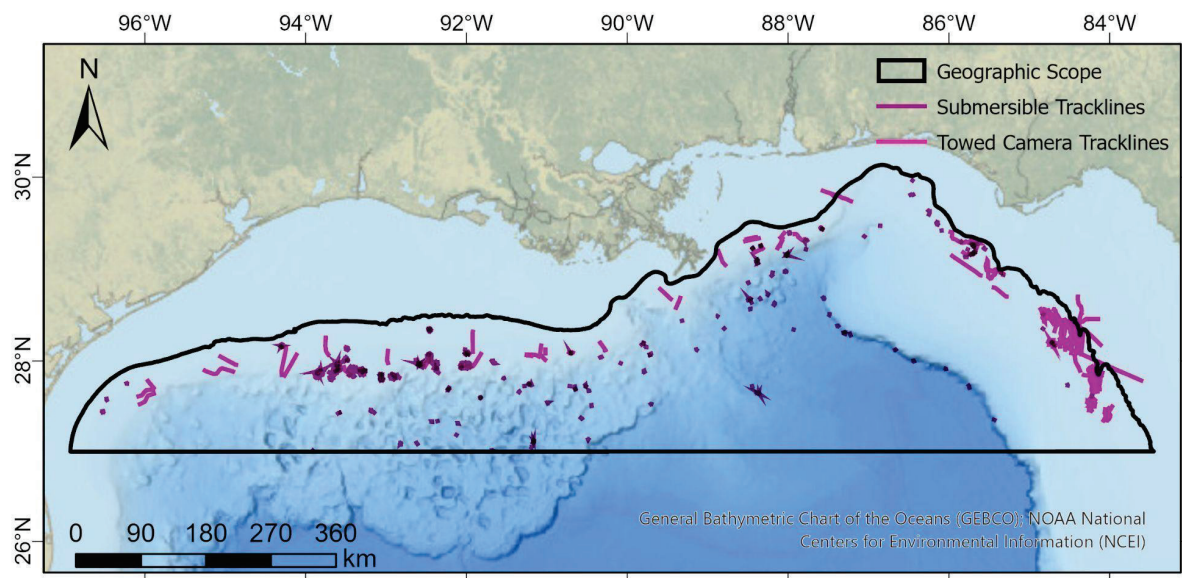


Figure 42. Submersible vehicle and towed camera system tracklines conducted during various ground-truthing surveys in the northern Gulf of Mexico from 1980 to present.

### 8.2.3. Data Source

There were a variety of sources utilized for capturing ground-truthing surveys conducted in the northern GOM from 1980 to present (Figure 43; Table 20). Datasets that were sourced from NOAA repositories were representative of three biological collection surveys, five bottom video surveys and nine submersible surveys, for a total of 17 datasets (31.5%, respectively; Figures 43 and 44; Table 20). Datasets that were sourced from USGS repositories were representative of one biological collection survey, four geological collection surveys and three submersible surveys, for a total of eight datasets (14.8%, respectively; Figures 43 and 45; Table 20). Datasets that were acquired from academic sources were representative of three biological collection surveys, one bottom video survey, and 12 submersible surveys, for a total of 16 datasets (29.6%, respectively; Figures 43 and 46; Table 20). The Smithsonian and BOEM repositories provided a biological collection survey and bottom video survey, contributing 1.9% of the total count, respectively (Figures 43, 47, and 48; Table 20). Datasets that were acquired from other sources were representative of two biological collection surveys, three bottom video surveys, four geological collection surveys and two submersible surveys, for a total of eleven datasets (20.4%, respectively; Figures 43 and 49; Table 20).

Table 20. The total count as well as percentages were calculated per each source of surveys inventoried for fish ecology datasets. The other category indicates sources that were outside of those already listed, such as state agencies.

Source	Number of datasets	Percentage of datasets
NOAA	17	31.5 %
USGS	8	14.8 %
Academia	16	29.6 %
Smithsonian	1	1.9 %
BOEM	1	1.9 %
Other	11	20.4 %

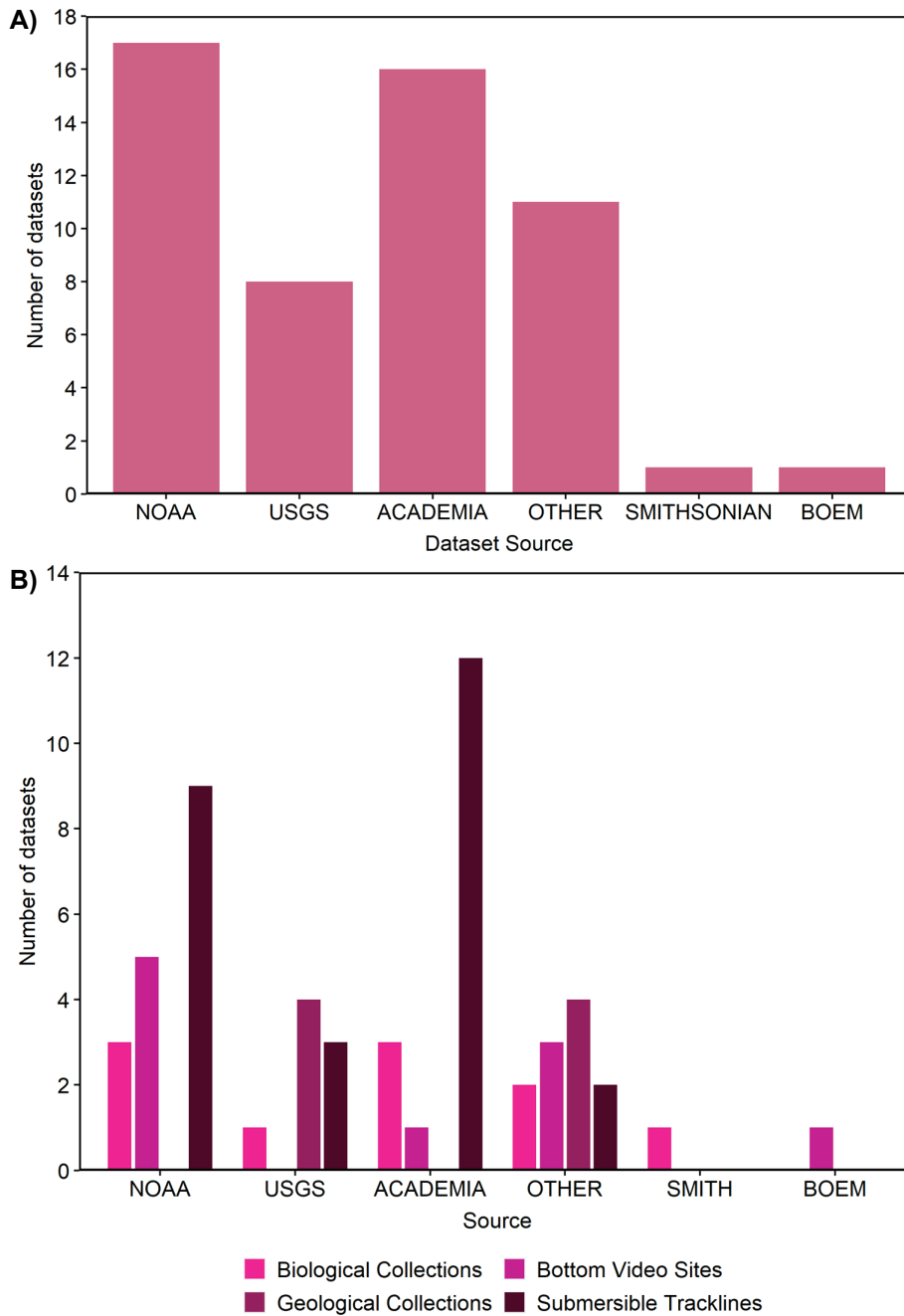


Figure 43. Bar plots identifying the total number of ground-truthing datasets that were acquired from various sources used for the purpose of this inventory (A), and the number of datasets acquired per source and the ground-truthing data type (B).



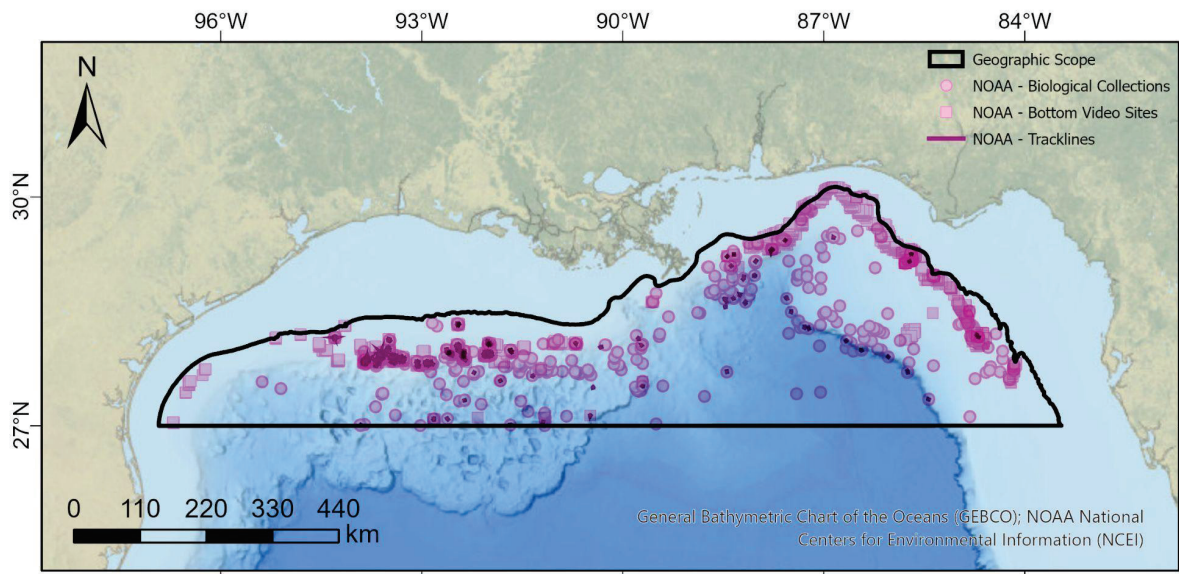


Figure 44. Ground-truthing surveys that were acquired from NOAA sources as biological collection surveys, bottom video surveys, and submersible surveys conducted in the northern Gulf of Mexico from 1980 to present.

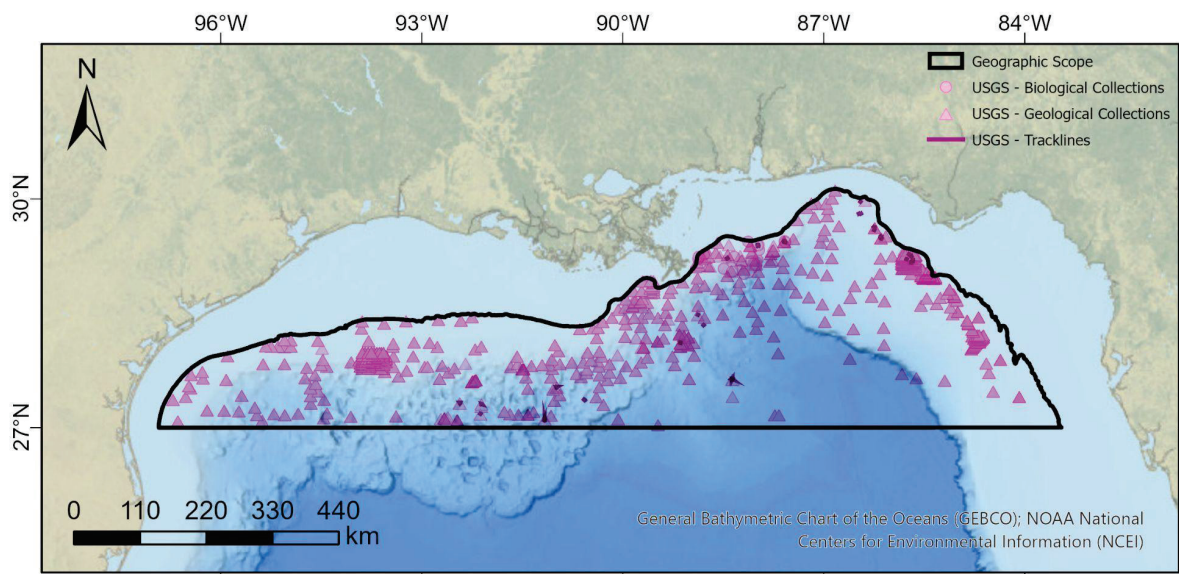


Figure 45. Ground-truthing surveys that were acquired from USGS sources as biological and geological collection surveys and submersible surveys conducted in the northern Gulf of Mexico from 1980 to present.

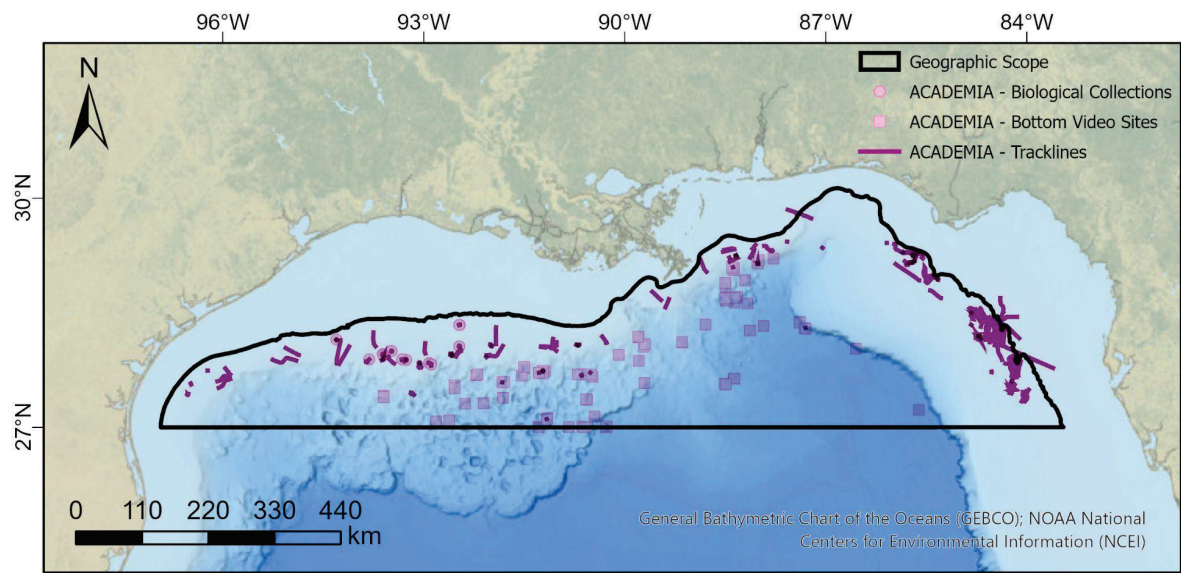


Figure 46. Ground-truthing surveys that were acquired from academic sources as biological collection surveys, bottom video surveys, and submersible surveys conducted in the northern Gulf of Mexico from 1980 to present.

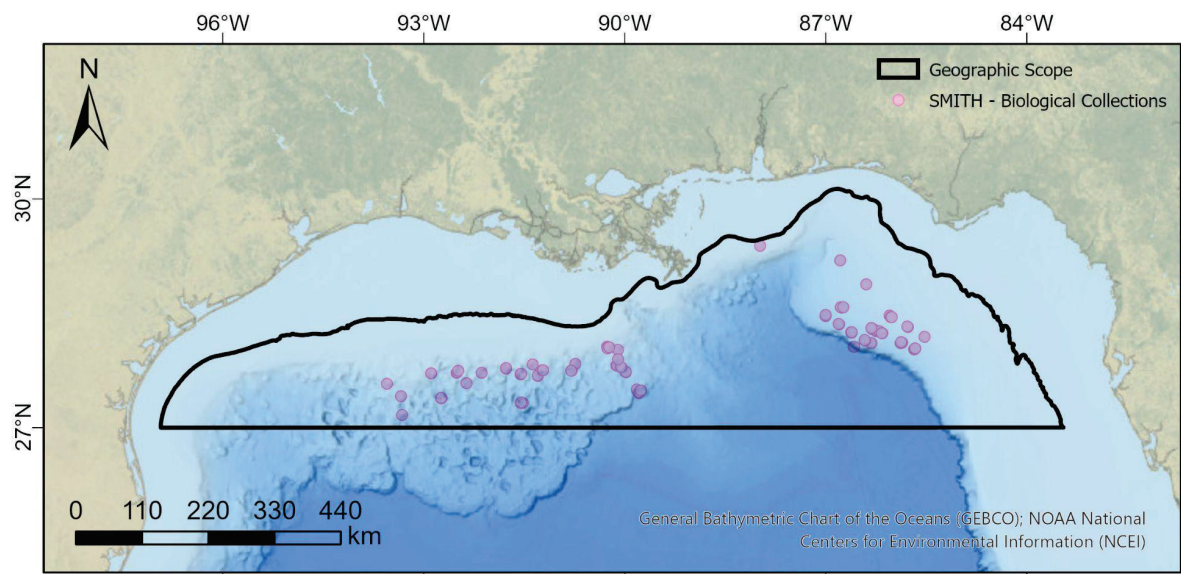


Figure 47. Ground-truthing surveys that were acquired from the Smithsonian (SMITH), as biological collection surveys conducted in the northern Gulf of Mexico from 1980 to present.

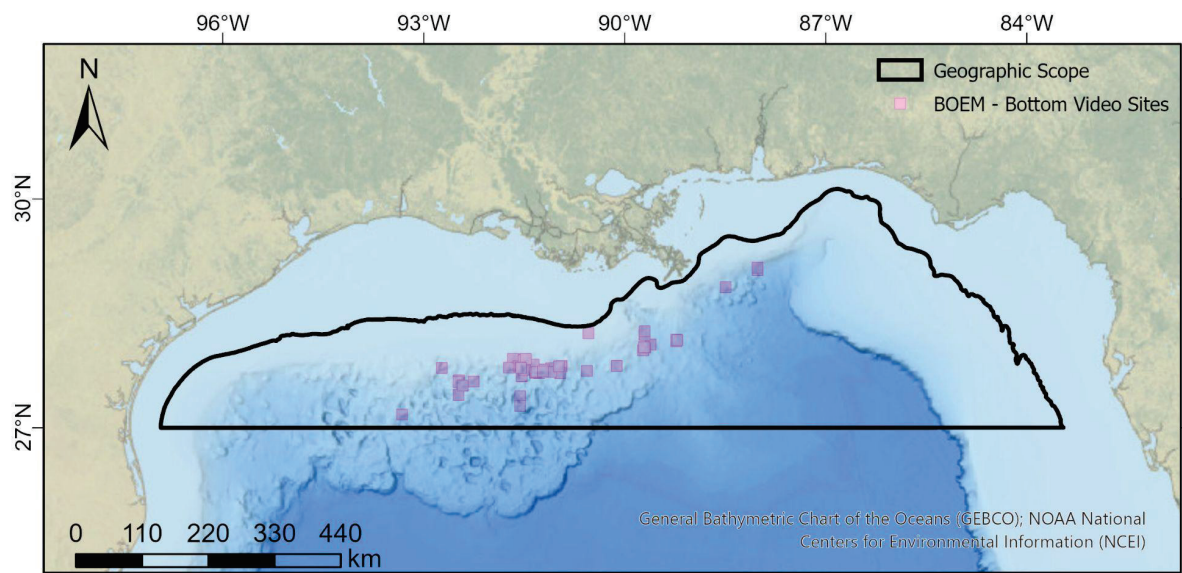


Figure 48. Ground-truthing surveys that were acquired from BOEM, as bottom video surveys conducted in the northern Gulf of Mexico from 1980 to present.

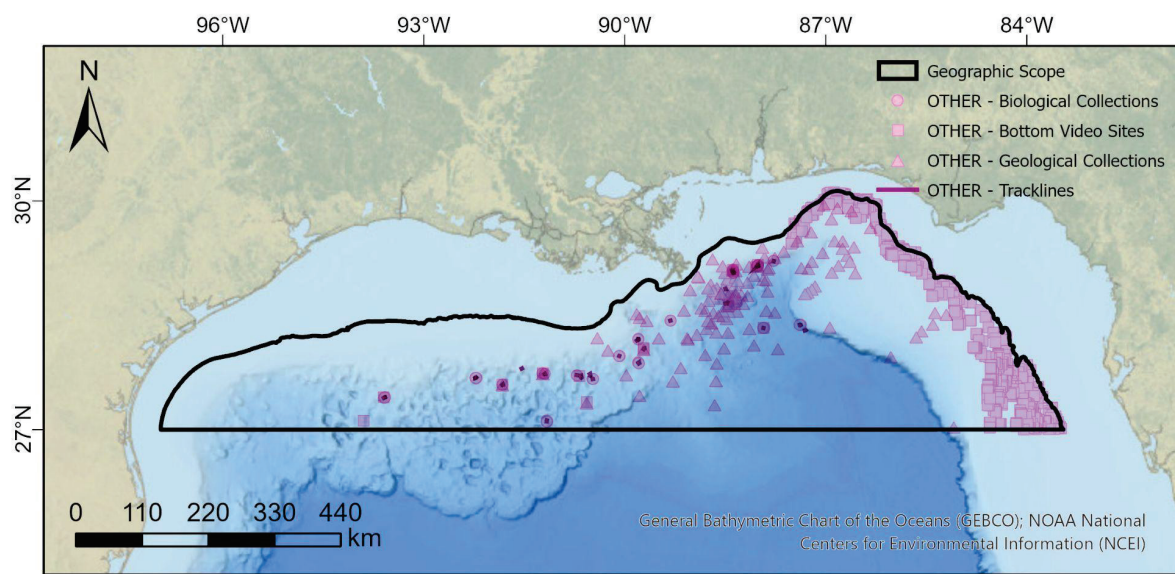


Figure 49. Ground-truthing surveys that were acquired from other sources outside of external MDBC partners and proprietary efforts, as biological and geological collection surveys, bottom video surveys, and submersible surveys in the northern Gulf of Mexico from 1980 to present.

#### 8.2.4. Data Collection per Decade

The majority (20) of the ground-truthing surveys collected for this inventory spanned multiple decades (Figures 50 and 51). These represent four biological collection surveys, four bottom video surveys, three geological collection surveys, and nine submersible surveys (Figures 50 and 51). Ground-truthing surveys conducted from 2010 to 2019 account for a total of 18 datasets, with two biological collection surveys, three bottom video surveys, three geological collection surveys, and 10 submersible surveys (Figures 50 and 52). Ground-truthing surveys conducted from 2000 to 2000 account for a total of 16

datasets, with four biological collection surveys, two bottom video surveys, three geological collection surveys, and seven submersible surveys (Figures 50 and 53). There were no ground-truthing surveys collected for this inventory that represent the decades 1980 to 1989 and 1990 to 1999.

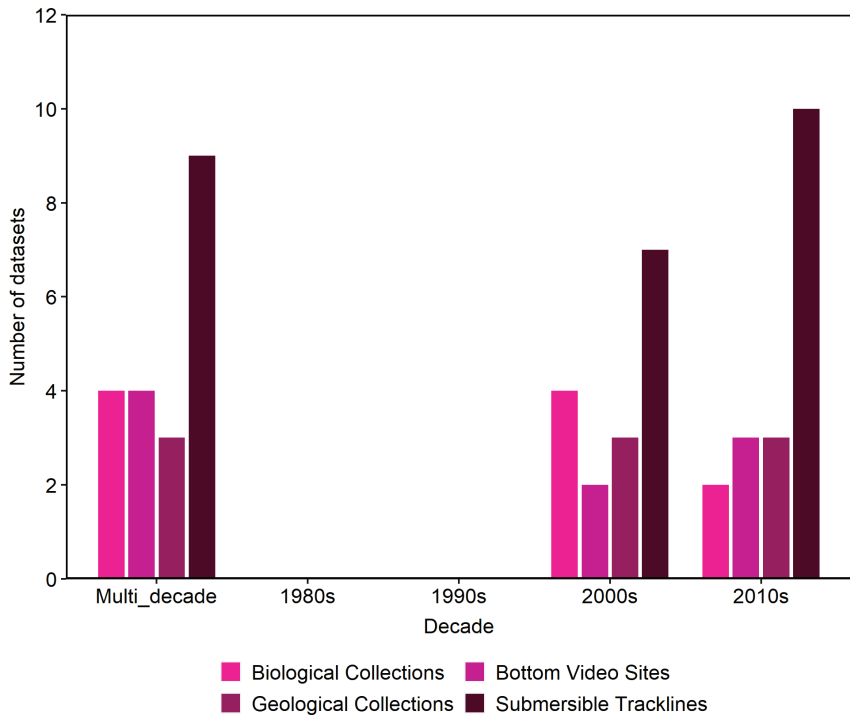


Figure 50. A bar plot showing the types of ground-truthing data collected per decade between 1980 to present. The Multi\_decade datasets are those surveys that spanned over multiple decades.

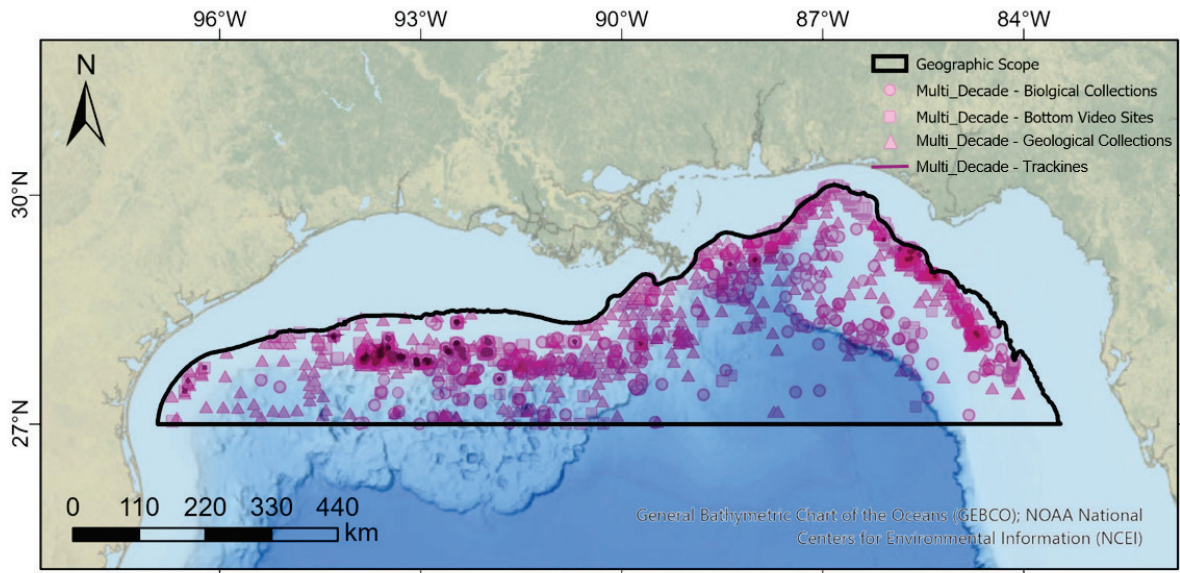


Figure 51. Ground-truthing surveys that were conducted over multiple decades (Multi\_Decade) as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.

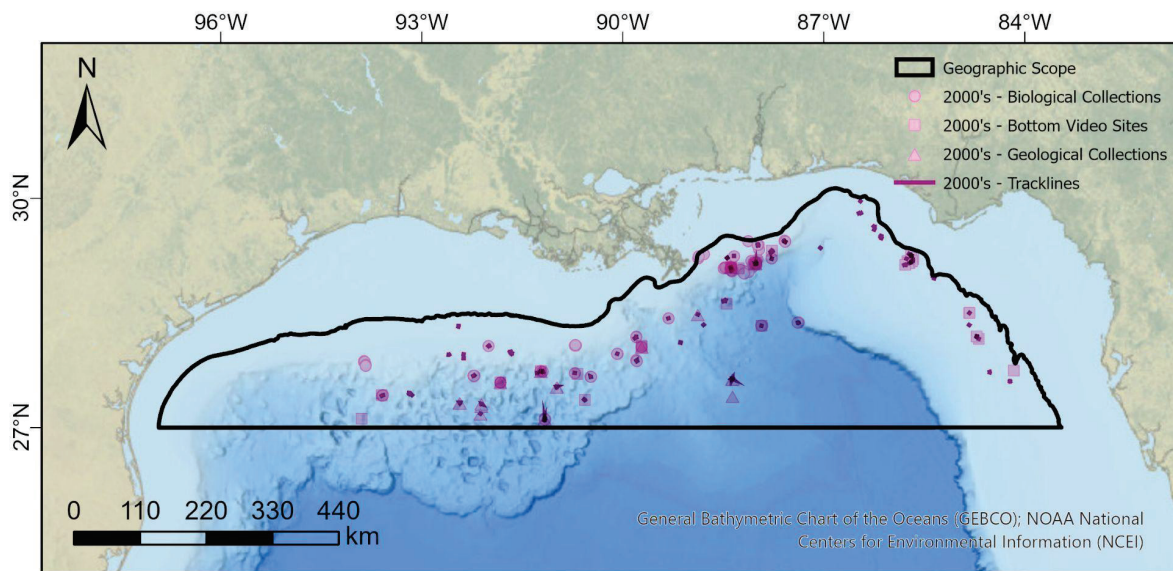


Figure 52. Ground-truthing surveys that were conducted from the year 2000 to 2009 as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.

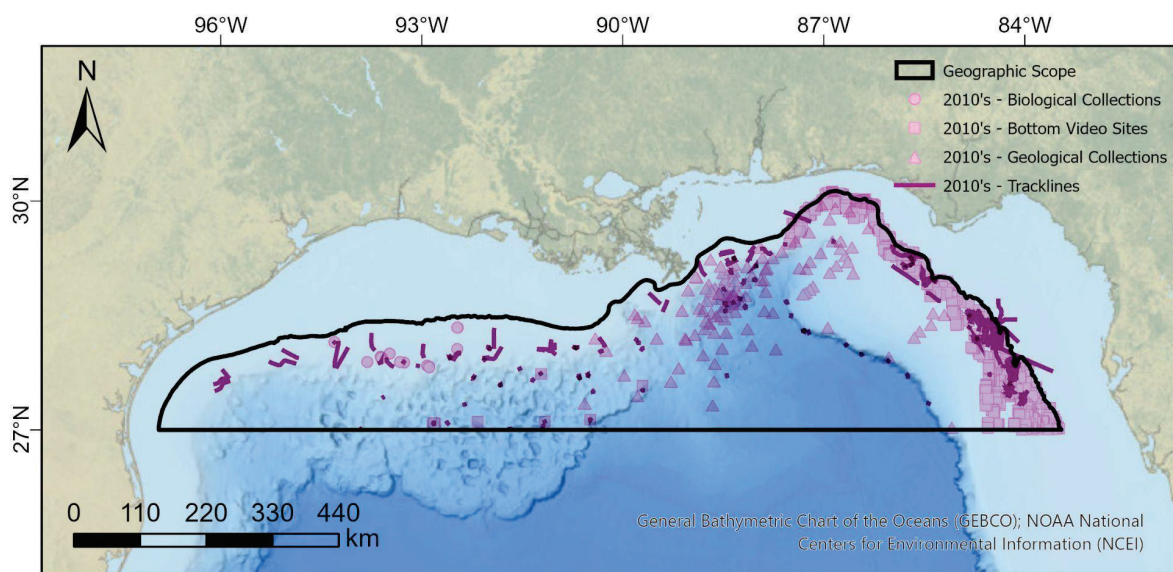


Figure 53. Ground-truthing surveys that were conducted from the year 2010 to 2019 as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.

### 8.2.5. Data Collection by Depth

The majority of ground-truthing surveys collected for this inventory were conducted in mesophotic depths and represent two biological collection surveys, four bottom video surveys, and 12 submersible surveys (Figures 54 and 55). Ground-truthing surveys conducted in waters deeper than 300 m account for a total of 17 datasets, representing two biological collection surveys, four bottom video surveys, four geological collection surveys and seven submersible surveys (Figures 54 and 56). There were 15 ground-truthing datasets that represent various surveys conducted in both mesophotic and deep depths in the northern GOM (Figures 54 and 57). This included seven biological collection surveys, two

bottom video surveys, four geological collection surveys, and two submersible surveys (Figures 54 and 55). All ground-truthing surveys collected for this inventory were also mapped with smoothed depth contours of 30 m, 300 m, 800 m, and 2500 m, shown in Figures 54 and 58.

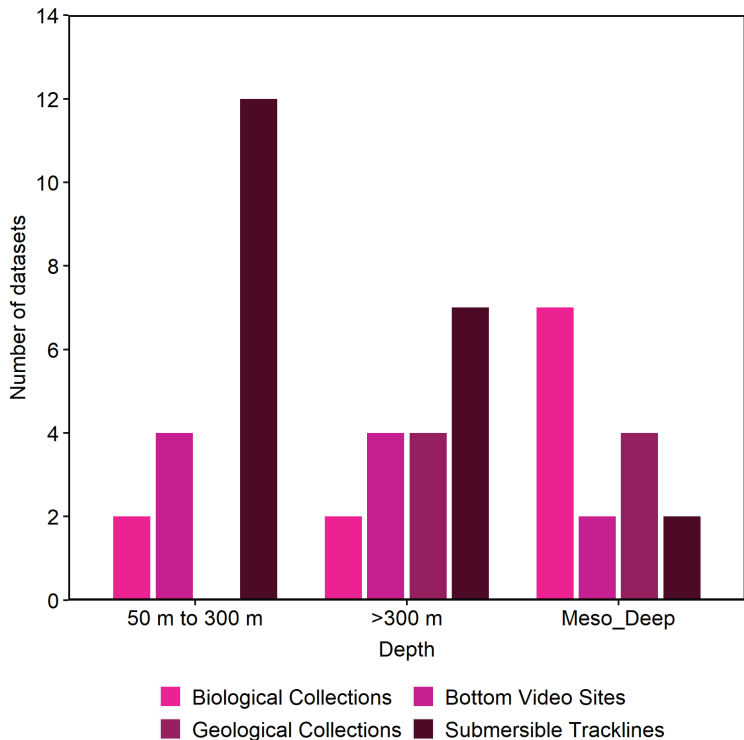


Figure 54. A bar plot showing different ground-truthing dataset surveys conducted over mesophotic (50m to 300m), deep (300m and greater), and both mesophotic and deep (Meso\_Deep) depth ranges.

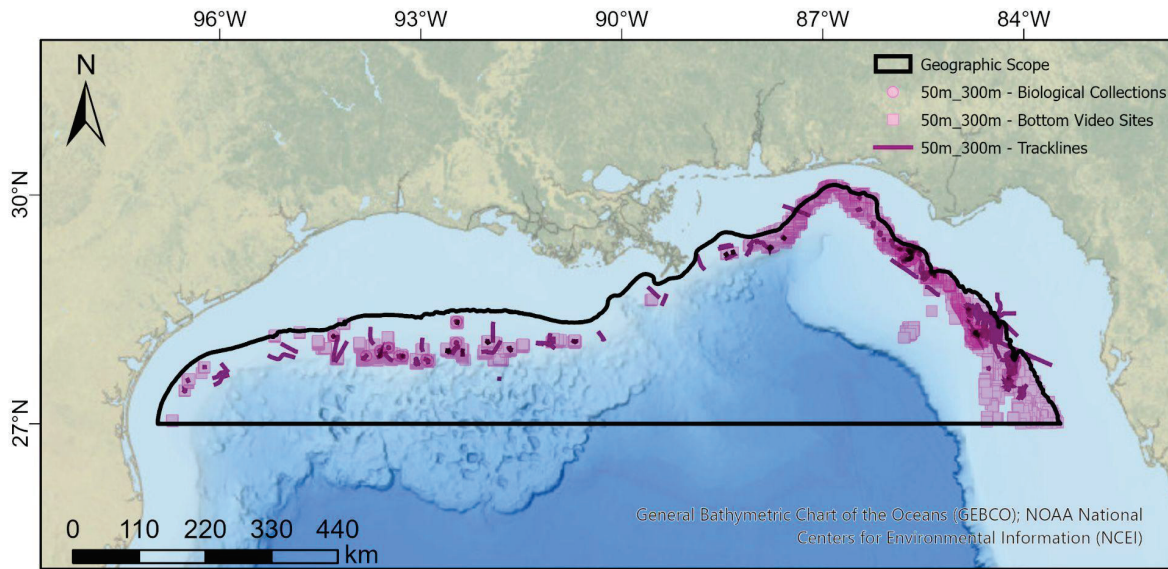


Figure 55. Ground-truthing surveys that were conducted in mesophotic depth range between 50m and 300m as biological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.

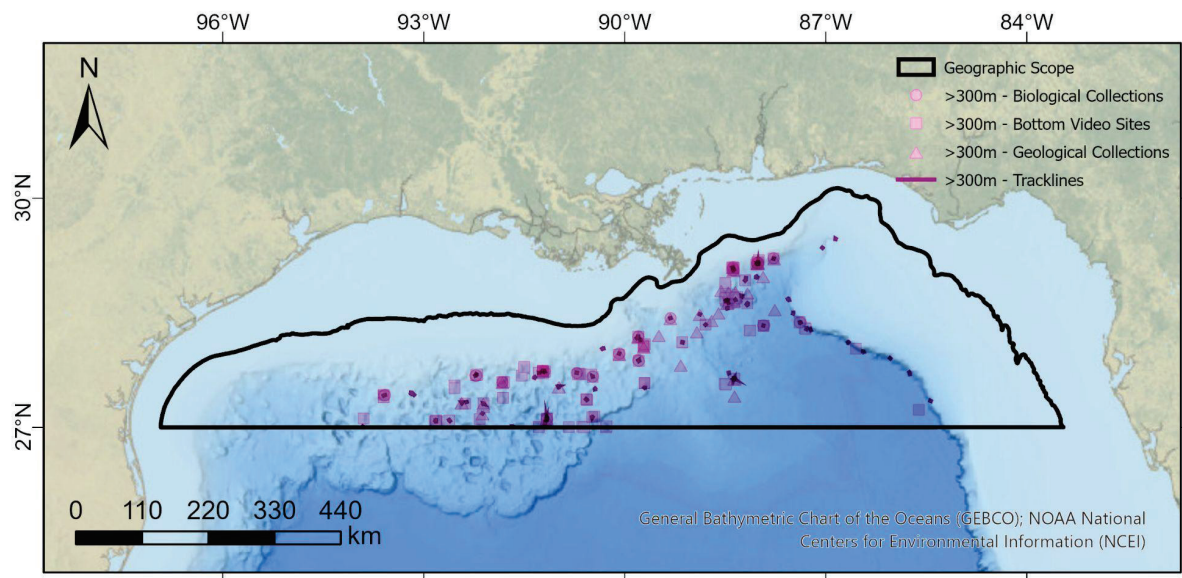


Figure 56. Ground-truthing surveys that were conducted in a depth range greater than 300m as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.

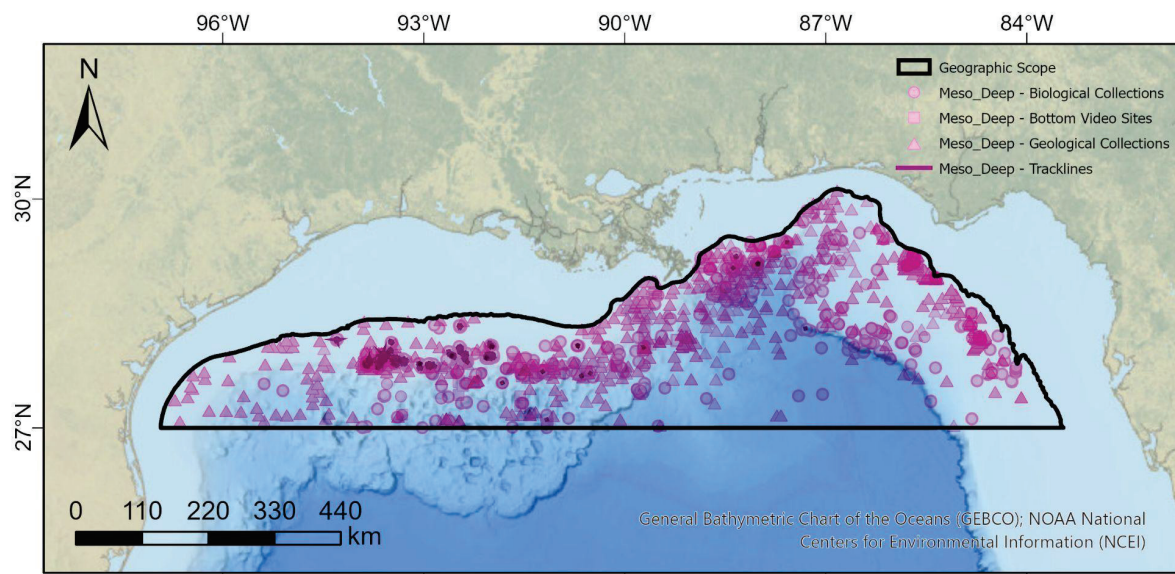


Figure 57. Ground-truthing surveys that were conducted in both mesophotic and greater depth range, 50m and greater as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico.

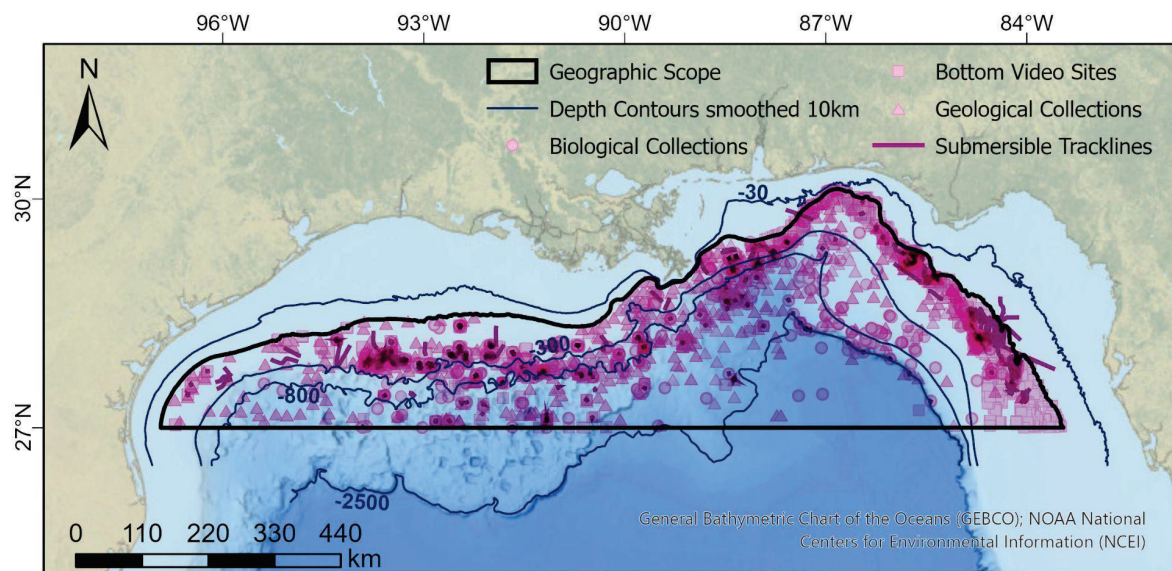


Figure 58. Ground-truthing surveys that were conducted in depths greater than 50m as biological and geological collection, bottom video, and submersible surveys in the northern Gulf of Mexico. The depth contours are represented as 30m, 300m, 800m, and 2500m.

### 8.2.6. Ground-truthing Datasets in Inventory Queue

As stated in seafloor mapping's results Section 8.1.9., this inventory process is ongoing, and datasets are added to the data catalog as they are discovered and processed. There will be updated versions of inventory products that encompass additional datasets, as we receive them. For example, we are currently working on adding the following ground-truthing datasets to the inventory:

- Datasets from literature review for which data acquisition is still underway via external engagement with dataset points of contact
- Datasets from other MDBC project inventories (i.e., Habitat Assessment Project)
- Data collected but not yet archived, published on, or discovered through engagement

## 8.3. Predictive Habitat Modeling Results

During the predictive habitat modeling inventory process, numerous datasets were considered for inclusion. For example, the literature search identified 39 potential datasets (Table 9). Of the datasets acquired and screened thus far, 12 passed the inclusion criteria and were incorporated into the inventory. As of January 2022, the predictive habitat modeling inventory contained 12 datasets with their associated footprint layers and a total of 198 model prediction rasters. These datasets originated from a range of sources including government agencies (NOAA, USGS, and U.S. Naval Research Laboratory), non-governmental organizations (Institute of Zoology London and Marine Conservation Institute), and academic institutions (Florida State University, Texas A&M University, and University of South Florida). All of these modeling datasets were published during the past decade, the earliest being from 2012, although the original data modeled span a longer time period, in some cases dating back to the late 1800s.



The spatial extent of predictive habitat modeling datasets in the inventory varied widely from global (including the GOM) to relatively small focal areas within the Gulf (Figure 59). There were two global datasets, three regional datasets covering the entire study area and ranging from 488,199 to 666,555 km<sup>2</sup>, and six local datasets covering only parts of the study area and ranging from 20 to 6,129 km<sup>2</sup> (Figure 60).

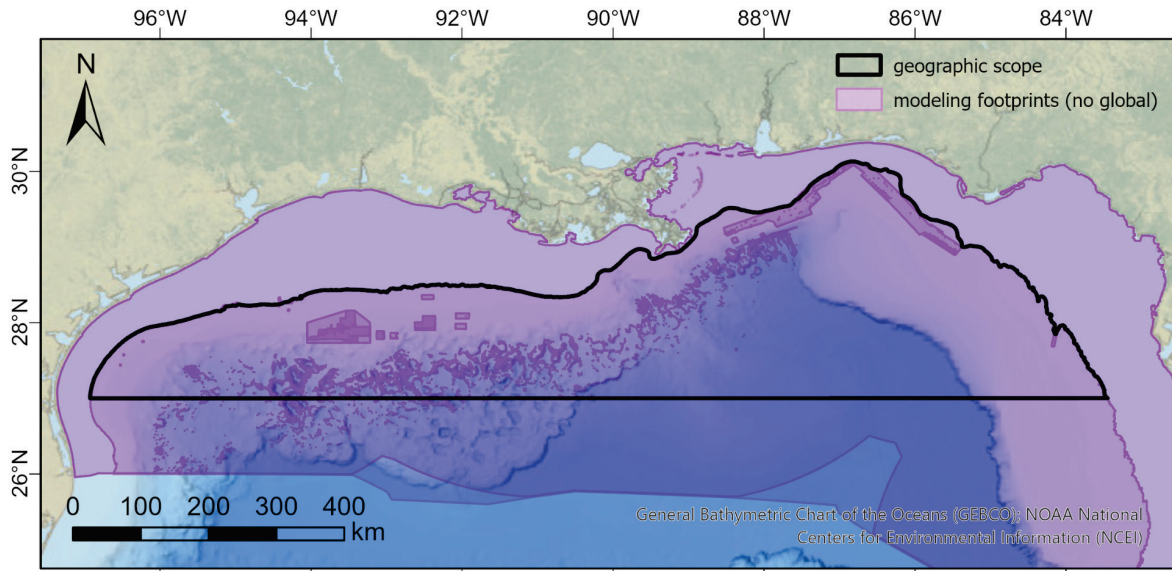


Figure 59. Footprints for predictive habitat modeling data included in the inventory, excluding datasets with global coverage.

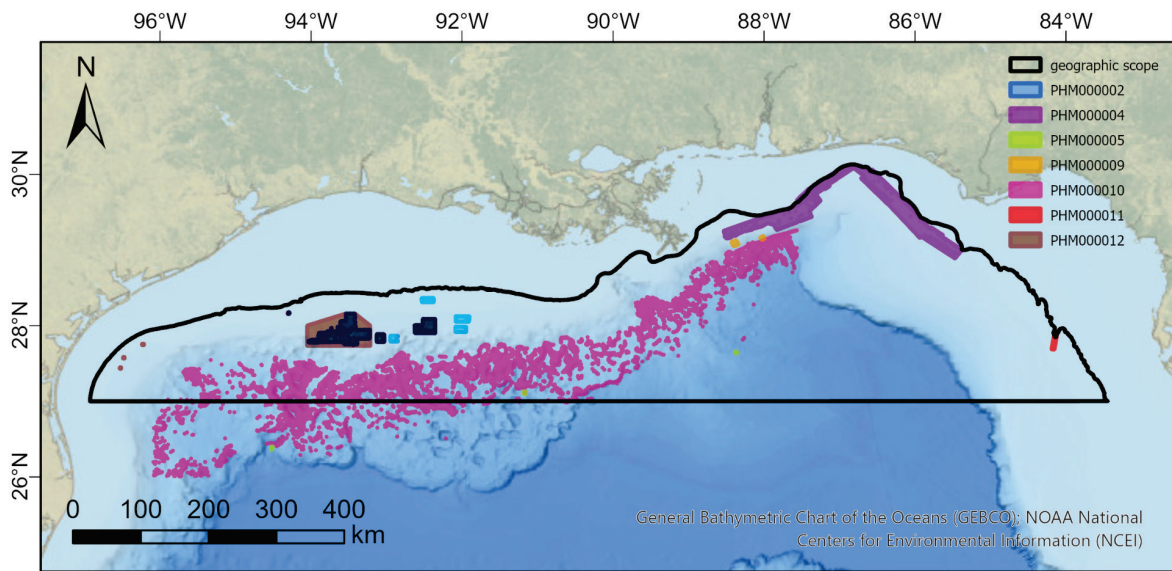


Figure 60. Footprints for relatively local predictive habitat modeling data included in the inventory.

The predictive habitat modeling datasets in the inventory represented three types of models: 1) models of the spatial distributions of deep-sea corals, 2) models of the spatial distributions of physical and biological habitat types, and 3) models of the spatial distributions of seep habitats. Models of deep-sea corals were the majority (Figure 61A) and had varying taxonomic resolution from individual species to

higher levels like families or even orders. There were four habitat models, two of which were focused on seep habitats (Figure 61A). Of the non-seep habitat models, one modeled hard versus soft substrate, and the other modeled multiple physical and biological habitat categories. The quantity predicted by most of the models in the inventory was habitat suitability, a relative measure of where the modeled entity was more or less likely to be found (Figure 61B). For the two non-seep habitat models, the predicted quantity was the most likely habitat type, and three other models predicted probability of occurrence, an absolute measure of the likelihood of finding the modeled entity at a given location. The modeling datasets in the inventory were derived using a range of modeling methodologies, but by far the most common technique was maximum entropy modeling implemented in the software Maxent (Phillips et al. 2022) along with a single instance of each of boosted regression tree, nearest neighbor learning, occupancy, and random forest techniques (Figure 61C, Table 21). The spatial resolution of model predictions ranged from meters to tens of kilometers and tended to correlate with the spatial extent of the model predictions whereby global models had the coarsest resolution and local models had the finest resolution (Figure 61D).

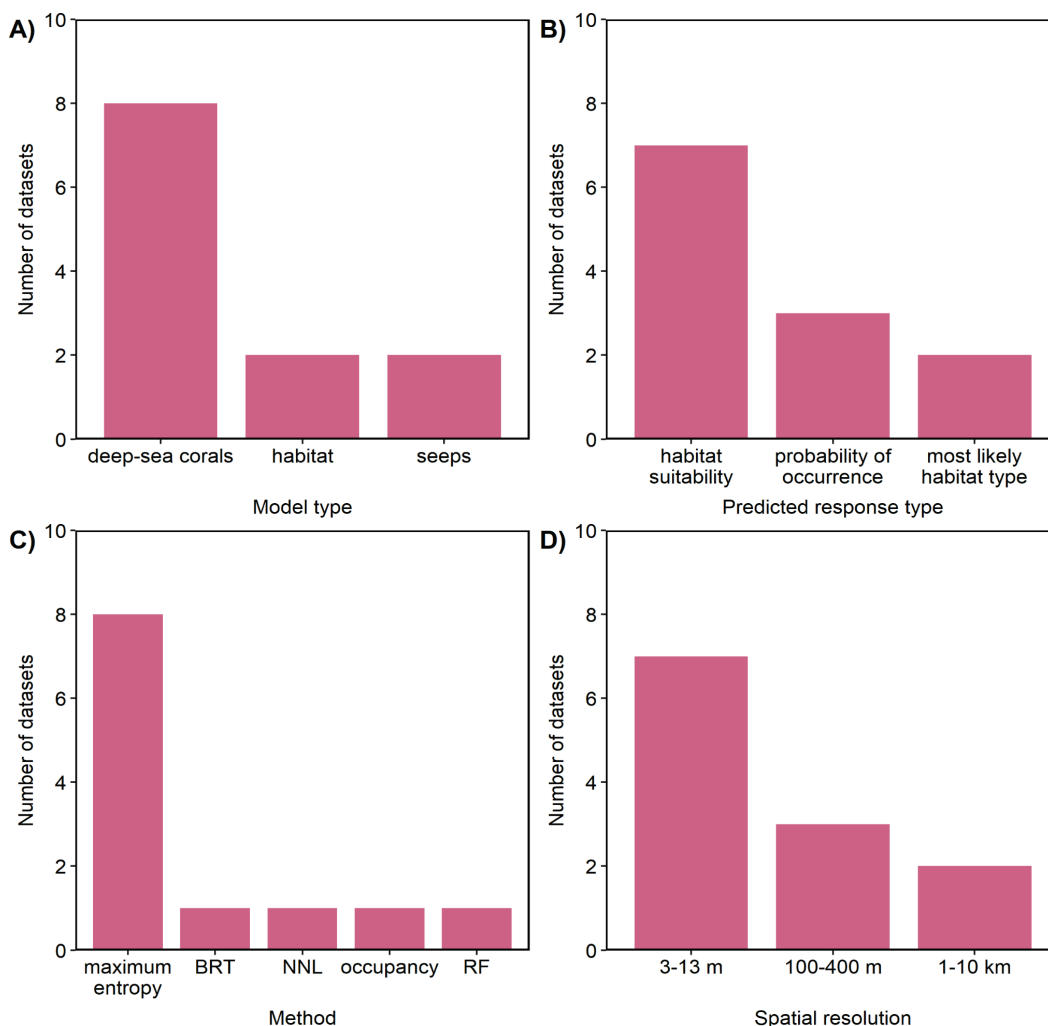


Figure 61. Number of predictive habitat modeling datasets by A) model type, B) predicted response type, C) method, and D) spatial resolution. BRT refers to boosted regression tree modeling, NNL refers to a nearest neighbor learning method, and RF refers to random forest modeling. The dataset derived from BRT modeling also included maximum entropy/habitat suitability predictions (PHM000012), but only the BRT models are represented in this Figure.

Table 21. Methods employed by predictive habitat modeling datasets in the inventory.

Method	Description	Software	Datasets
maximum entropy	machine learning technique for modeling species niches and distributions; typically uses presence-background data and predicts habitat suitability	Maxent <sup>1</sup>	PHM000002, PHM000003, PHM000004, PHM000005, PHM000007, PHM000008, PHM000009, PHM000010, PHM000012
boosted regression tree	decision tree analysis combined with boosting	R <sup>2</sup> (gbm package <sup>3</sup> )	PHM000012
nearest neighbor learner (with ball tree algorithm)	machine learning technique	Python <sup>4</sup> (scikit-learn <sup>5</sup> )	PHM000006
occupancy	model of species occurrence accounting for imperfect detection	JAGS <sup>6</sup> , R <sup>2</sup> (rjags package <sup>7</sup> )	PHM000001
random forest	decision tree analysis combined with bagging	R <sup>2</sup> (RStoolbox <sup>8</sup> and ranger <sup>9</sup> packages)	PHM000011

<sup>1</sup> Phillips et al. (2022)

<sup>2</sup> R Core Team (2020)

<sup>3</sup> Greenwell et al. (2020)

<sup>4</sup> Van Rossum and Drake (2020)

<sup>5</sup> Pedregosa et al. (2020)

<sup>6</sup> Plummer (2017)

<sup>7</sup> Plummer (2019)

<sup>8</sup> Leutner et al. (2019)

<sup>9</sup> Wright and Ziegler (2017)

All of the predictive habitat modeling datasets in the inventory had associated estimates of model performance. For all datasets, predictive performance was assessed by fitting the model to ‘training’ data and then using the fitted model to compare its predictions to ‘test’ data that were not used in model fitting. In some cases, the training/test data split was repeated multiple times; i.e., cross-validation. Different metrics were used to quantify performance across datasets, but nearly all reported the area under the receiver operating characteristic curve (AUC). Threshold-dependent performance metrics (e.g., confusion matrix statistics) were reported for several models as was Akaike’s Information Criterion (AIC). These performance metrics represent a form of validation of model predictions; however, most, if not all, of the modeling datasets are lacking validation in the form of new data collected for the express purpose of model validation.

While all of the predictive habitat modeling datasets had associated measures of model performance, only a third of them reported measures of precision or uncertainty in the model predictions. Reported measures of uncertainty included bootstrapped confidence intervals and Bayesian posterior credible intervals.

A wide range of environmental predictor variables were employed across the predictive habitat modeling datasets in the inventory. The number of predictors used also varied across datasets. All models used depth and seafloor topography predictor variables. Seafloor substrate and oceanographic predictors were also common. The native spatial resolutions of seafloor topography predictors tended to be similar to the resolutions of model predictions. In contrast, oceanographic predictors were typically coarser in native spatial resolution than model predictions, especially for regional and local modeling datasets. All datasets reported associated measures of predictor variable importance (e.g., percent contribution), but only a little over half reported the estimated relationships between model predictions and predictor variables.

## 9. Discussion and Conclusion

### 9.1. Overall Patterns

The data inventory described here represents a comprehensive compilation of existing seafloor mapping, ground-truthing, and predictive habitat modeling data for a large part of the northern GOM dating back to 1980. The resulting inventory is one of the most comprehensive for this geographic and temporal scope and purpose because it included data from a variety of sources. Thorough and reproducible data discovery protocols were applied to search a range of data sources including project partners, external contacts, scientific literature, and online data portals. Reproducible protocols were also applied to screen, acquire, standardize, store, and document inventoried datasets. In many cases, custom computer scripts and tools were developed to facilitate data processing and storage. Summary products, for example spatial footprints, were derived from the original datasets to provide simplified, standardized views of the inventoried data across datasets and data types.

Generally speaking, the seafloor mapping, ground-truthing, and predictive habitat modeling datasets in the inventory cover a large proportion of the study area. Nevertheless, there are large parts of the study area with few or no seafloor mapping or ground-truthing data, particularly when datasets are subset by technology, time period, etc. Much of the study area is also lacking habitat model predictions with a spatial resolution  $\leq 10$  m. The intersection of all three data types provides the most information about MDBC, and in many places one or more data types is lacking, for example areas of model predictions without corresponding high-resolution seafloor mapping data or ground-truthing data. The inventory provides a framework for identifying these data gaps and directing future data collection and modeling accordingly.

### 9.2. Lessons Learned and Best Practices

Each of the three data types inventoried here—seafloor mapping, ground-truthing, and predictive habitat modeling—entailed unique considerations. The development of standardized, reproducible protocols across data types facilitated their combination in the inventory. The inventory also captured the spatial aspects of these data types in a standard format (e.g., spatial footprint layers in a common coordinate reference system). The protocols developed have established a workflow whereby new datasets will be incorporated into the inventory throughout the duration of the project.

Despite the standardized protocols, each data type had particular challenges with respect to data acquisition and processing. For example, most of the seafloor mapping and ground-truthing data were discovered and acquired from large public sources like NOAA NCEI, while the process of discovering and acquiring predictive habitat modeling datasets was primarily searching the literature and engaging external contacts or online sources on a case-by-case basis. The literature search and external engagement for seafloor mapping and ground-truthing data involved a large time investment and facilitated discovery of some datasets but most had already been discovered from broad sources, like NOAA NCEI archives. All three data types faced challenges in data acquisition, but the reasons differed. Seafloor mapping data from private industry was proprietary and when available was costly. In contrast, some predictive habitat modeling and ground-truthing datasets were difficult to acquire given the sometimes-ephemeral nature of those studies and difficulties contacting the original investigators. Finally, the level of processing required often differed among the data types. As evidenced by the protocols and tools developed here, seafloor mapping datasets often required processing and reformatting, and some mapping datasets still require further processing for full interpretability (e.g.,

seismic data). For ground-truthing, it was difficult to acquire processed annotations, as most datasets included spatial data or live annotations. In comparison, while predictive habitat modeling datasets were often in different file formats, they were all rasters, which eased standardization.

Given the challenges in acquiring data, there are a multitude of datasets that we have not been able to acquire and add to the inventory. These datasets are described in more detail in Section 8.1.9 for seafloor mapping and Section 8.2.6 for ground-truthing. Several of the predictive habitat modeling datasets were not served publicly or the publicly available data were incomplete, thus direct external engagement with the principal investigators was necessary. We successfully established contact and received full datasets from approximately half of those investigators, but the remainder either did not receive or did not respond to our inquiries and data requests. The publicly archived modeling datasets greatly facilitated this inventory, and we would encourage other predictive habitat modelers to archive and serve their model products publicly when possible, to maximize the use and value of those studies.

### **9.3. Updating the Inventory**

The inventory will be updated and expanded for at least the duration of the MDBC restoration project (i.e., through 2028). New datasets are continually being acquired, processed, and integrated into the inventory, and updates to the inventory will be made periodically.

Steps taken to assemble mapping, ground-truthing, and predictive habitat modeling data will be repeated every six months to one year at a minimum to incorporate new data being acquired during the life of the project. These steps include: coordinate data search and POC engagement, discover and acquire data, process data as needed, integrate source data and any derived products into geospatial database, assemble draft inventory and standardized summaries, and inventory review, which could be in the form of a workshop to better engage subject experts.

### **9.4. Planned Uses of Data Inventory**

The data inventory serves as a foundation for MDBC project teams and associated partners to prioritize and support DWH NRDA restoration activities for MDBC, especially in relation to seafloor mapping, ground-truthing, and predictive habitat modeling. In particular, the inventory has and will facilitate the following project components:

- Document abundance and distribution of MDBC from existing data;
- Inform future MDBC mapping by comparing survey specifications of data within the inventory to evaluate whether they meet mapping requirements for reuse;
- Develop a programmatic data management framework that accommodates access to existing data and data that will be collected during upcoming phases of the project;
- Define MDBC mapping, ground-truthing, and predictive habitat model requirements needed to inform MDBC management and enhance MDBC resiliency; and
- Conduct a mapping, ground-truthing, and predictive habitat modeling prioritization exercise (Kendall et al. 2022) with key stakeholders and seafloor mapping experts to refine data gaps and prioritize future seafloor surveys, to identify initial priority sites for other MDBC projects, and to iteratively refine or expand the initial priority sites throughout project implementation.
- Complete a gap analysis quantifying gaps in existing mapping, ground-truthing, and predictive habitat modeling data to help inform data collection and modeling needs

# References

- Bureau of Ocean Energy Management (BOEM). 2021. Biological environmental background report for the Gulf of Mexico OCS Region. Gulf of Mexico Regional Office. OCS Report BOEM 2021-15.
- Continental Shelf Associates, Inc. (CSA) and Texas A&M University (TAMU), Geochemical and Environmental Research Group. (2001). Mississippi/Alabama Pinnacle Trend Ecosystem Monitoring, Final Synthesis Report. U.S. Department of the Interior, Geological Survey, Biological Resources Division, USGS BSR 2001-0007 and Minerals Management Service, Gulf of Mexico OCS Region, New Orleans, LA, OCS Study MMS 2001-080.
- Deepwater Horizon Natural Resource Damage Assessment Trustees (DWH NRDA Trustees). (2016). Deepwater Horizon oil spill: Final Programmatic Damage Assessment and Restoration Plan and Final Programmatic Environmental Impact Statement.
- Environmental Systems Research Institute (ESRI). (2020). ArcGIS Pro (Version 2.8). Redlands, CA.
- Etnoyer, P. J., Wickes L. N., Silva M., Dubick J. D., Balthis L., Salgado E., MacDonald I. R. (2016). Decline in condition of gorgonian octocorals on mesophotic reefs in the northern Gulf of Mexico: before and after the Deepwater Horizon oil spill. *Coral Reefs*, 35, 77–90.
- Gardner, J. V., Sulak, K. J., Dartnell, P., Hellequin, L., Calder, B., & Mayer, L. A. (2000). Cruise Report RV Ocean Surveyor Cruise O-1-00-GM The Bathymetry and Acoustic Backscatter of the Pinnacles Area, Northern Gulf of Mexico May 23, through June 10, 2000 Venice, LA to Venice, LA (00-350). U.S. Geological Survey Open-File Report 00-350. Retrieved from <https://pubs.usgs.gov/of/2000/0350/>
- Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers. (2020). gbm: Generalized Boosted Regression Models. R package. Retrieved from <https://CRAN.R-project.org/package=gbm>
- Hijmans, R. J. (2020). raster: Geographic data analysis and modeling. R package. Retrieved from <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J. (2021). terra: Spatial data analysis. R package. Retrieved from <https://CRAN.R-project.org/package=terra>
- Hourigan, T. F., Etnoyer, P. J., McGuinn, R. P., Whitmire, C., Dorfman, D. S., Dornback, M., . . . Sallis, D. (2015). An Introduction to NOAA's National Database for Deep Sea Corals and Sponges. NOAA Technical Memorandum NOS NCCOS 191. Silver Spring, MD.
- Howard, B. E., Phillips, J., Tandon, A., Maharana, A., Elmore, R., Mav, D., . . . Shah, R. R. (2020). SWIFT-Active Screener: accelerated document screening through Active Learning and integrated recall estimation. *Environment International*, 138.
- Kahng, S. E., Garcia-Sais, J. R., Spalding, H. L., Brokovich, E., Wagner, D., Weil, E., . . . Toonen, R. J. (2010). Community ecology of mesophotic coral reef systems. *Coral Reefs*, 29, 255-275.
- Kendall, M.S., B.L. Williams, K. Buja, and D. Dorfman. 2022. Priority Areas Suggested by Stakeholders for Restoration and Protection of Mesophotic and Deep Benthic Communities Injured by the *Deepwater Horizon* Oil Spill. NOAA Technical Memorandum DWH MBDC 001. Silver Spring, MD. 28 pp. doi: 10.25923/61r3-gk12
- Kinlan, B. P., Poti, M., Etnoyer, P., Siceloff, L., Jenkins, C., Dorfman, D., & Caldow, C. 2013. Digital data: Predictive models of deep-sea coral habitat suitability in the U.S. Gulf of Mexico. Downloadable digital data package. Department of Commerce (DOC), National Oceanic and Atmospheric Administration (NOAA), National Ocean Service (NOS), National Centers for Coastal Ocean Science (NCCOS), Center for Coastal Monitoring and Assessment (CCMA), Biogeography Branch. Released August 2013. Available at: <https://coastalscience.noaa.gov/project/deep-coral-habitat-modeling-atlantic-gulf-mexico/>

- Leutner, B., Horning, N., & Schwalb-Willmann, J. (2019). RStoolbox: Tools for Remote Sensing Data Analysis. R package. Retrieved from <https://CRAN.R-project.org/package=RStoolbox>
- Marshak, A. R., et al. (2018). Report from the Using Acoustic Multibeam Echosounder (ME70) Technologies for Habitat Mapping Workshop. NOAA Technical Memorandum NMFS-F/SPO-191.
- Open Ocean Trustee Implementation Group (OO TIG). (2019). Deepwater Horizon Oil Spill Natural Resource Damage Assessment, Open Ocean Trustee Implementation Group, Draft Restoration Plan 2/ Environmental Assessment: Fish, Sea Turtles, Marine Mammals, and Mesophotic and Deep Benthic Communities.
- Pebesma, E. (2018). Simple Features for R: Standardized support for spatial vector data. *The R Journal*, 10(1), 439-446. Retrieved from <https://doi.org/10.32614/RJ-2018-009>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. G., O., Blondel, M., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Phillips, S. J., Dudík, M., & Schapire, R. E. (2022). Maxent software for modeling species niches and distributions. Retrieved from [https://biodiversityinformatics.amnh.org/open\\_source/maxent/](https://biodiversityinformatics.amnh.org/open_source/maxent/)
- Pierce, D. (2019). ncd4: Interface to Unidata netCDF (version 4 or earlier) format data files. R package. Retrieved from <https://CRAN.R-project.org/package=ncdf4>
- Plummer, M. (2017). JAGS. Retrieved from <https://mcmc-jags.sourceforge.io/>
- Plummer, M. (2019). rjags: Bayesian Graphical Models using MCMC. R package. Retrieved from <https://CRAN.R-project.org/package=rjags>
- Puglise, K. A., Hinderstein, L. M., Marr, J. C. A., Dowgiallo, M. J., & Martinez, F. A. (2009). Mesophotic Coral Ecosystems Research Strategy: International Workshop to Prioritize Research and Management Needs for Mesophotic Coral Ecosystems, Jupiter, Florida, 12-15 July 2008. NOAA Technical Memorandum NOS NCCOS 98 and OAR OER 2. Silver Spring, MD.
- Pyle, R. I. (1996). The twilight zone. *Natural History*, 105, 59-62.
- R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Silva M., Etnoyer P. J., MacDonald I. R. (2016). Coral injuries observed at Mesophotic Reefs after the Deepwater Horizon oil discharge. *Deep-Sea Research II*, 129, 96-107.
- Sulak, K. J., & Dixon, P. M. (2015). Change in Gulf of Mexico mesophotic reef fish community structure across the time threshold of the Deepwater Horizon oil spill event in 2010: Mesophotic reef fish community impacts of the DWH oil spill. DWH Deepwater Benthic NRDA Technical Working Group Report.
- The EndNote Team. (2020). EndNote (Version EndNote 20) [64 bit]. Philadelphia, PA: Clarivate Analytics.
- Van Rossum, G., & Drake, F. L., Jr. (2020). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.
- Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17.

# Appendix

A list of data sources and processes used and generated during the MDBC MGM inventory process as standard operation procedures (SOPs). For accessing these documents, please use the search tool and search by the SOP title served on the following website:

<https://www.diver.orr.noaa.gov/web/guest/diver-admin-record/14408>

These documents are also available under the Standard Operation Procedures section (categorized by mapping, ground-truthing, and predictive habitat modeling) at the bottom of the page.

Text in bold is also the name of the file within the DIVER data package.

**Portal Catalog** – Information about portals relevant to each data type documented in the MDBC Data Portals Catalog

**Data Catalog** – List of attributes, sources, and corresponding definitions associated with each dataset in the MDBC Data Catalog

**Data Organization and Processing** – Steps detailing the data download and organization, directory structure required for successful execution of the Python scripts/tools, and setting up of the ArcGIS Pro template

**Qimera Processing** – Overview of QPS Qimera raster processing

**ME70 Multibeam Processing** – Overview of ME70 multibeam processing



