Process: Updating the Seamap Australia National Benthic Habitat Layer

Contents

Introduction

Seamap Australia is an ongoing program, hosted by the Institute for Marine and Antarctic Studies (IMAS), University of Tasmania, with a core focus on delivering access to marine seafloor habitat data. The program provides three primary resources:

- The Seamap Australia National Benthic Habitat Classification Scheme;
  …applied to
- The Seamap Australia National Benthic Habitat layer, a nationally synthesised and uniformly classified seafloor habitat data product
  …hosted on
- The Seamap Australia data portal which brings together seafloor habitat and ancillary data in an interactive web mapping platform.

The Seamap Australia National Benthic Habitat Layer (hereafter NBHL) is the key data output from the Seamap Australia program, and represents a single national-scale seafloor habitat dataset spanning the Australian continental shelf. The dataset has been constructed from of a range of local- and regional-scale data collections collated from various state, government, and research agencies. Each source dataset has been classified according to a common scheme (the Seamap Australia National Benthic Habitat Classification Scheme, hereafter HCL) and synthesised into a single uniformly classified, national habitat mapping product.

This aim of this document is to describe the technical aspects and workflows associated with maintaining and updating the NBHL. It is intended to be used as a practical guide for maintaining the data asset, and thus contains levels of technical detail that will be unnecessary for most audiences. If you wish to contribute data to the NBHL or the Seamap Australia data portal, the ***Seamap Australia Data Acceptance Guidelines*** have been written for this purpose.

Frequency of updates

The Seamap Australia NBHL is a 'live' dataset that may undergo minor updates or amendments at any time. Historically, updates to the NBHL have been made on an ad-hoc basis. We recognise the need to have a formal and transparent process around updates and versioning of the national habitat asset to facilitate decision-making, reporting and planning. This also helps to establish clear expectations for data providers and stakeholders.

Moving forward, any major updates (including changes to the data structure, or addition of new datasets) will be associated with discrete *versioning events*, which will scheduled and occur **once per year**. Unscheduled updates to the NBHL will be at the sole discretion of the Steering Committee.

Superseded versions of the NBHL will be retained in the public domain for transparency but will not be promoted through the Seamap Australia data portal. The most recent version of the NBHL will always represent the most current and accurate state of knowledge of Australian seafloor habitat.

Assessing candidate habitat datasets for suitability

The Seamap Australia NBHL is a national authority for seafloor habitat. Data contributions must meet minimum quality standards and acceptance criteria to be considered for inclusion in the NBHL to maintain the integrity of the national asset (see ***Seamap Australia Data Acceptance Guidelines*** for specific **acceptance criteria**).

The Steering Committee is responsible for vetting candidate habitat datasets and assessing them against the Seamap Australia acceptance criteria. The Steering Committee must reach consensus regarding data to be accepted for assimilation into the NBHL prior to a *versioning event*.

Publishing source datasets to the Seamap Australia data portal

All habitat datasets contributed to Seamap Australia are published as stand-alone products available for viewing and download through the Seamap Australia website, irrespective of whether they meet the criteria for inclusion in the NBHL. The steps below detail the process that should be followed to publish source data:

1) Import vector data to Seamap Australia database

- Data must have primary key (name: ID) and geometry column (name: geom)
- Data must use EPSG:3112 CRS (for subsequent computation of metric area statistics)
- Table name in database should match the convention "SeamapAus_<STATE>_<short description with underscores>_<end date (year), if available>"
  e.g. SeamapAus_NSW_HABMAP_subtidal_marine_habitat_2016
- ogr2ogr (GDAL) is the most efficient mechanism for inserting data and handles geometry creation and reprojection

2) Prepare database table for publication

- Ensure all geometries are valid (.MakeValid())
- Add new column called SM_HAB_CLS for styling purposes, and populate with values from the most appropriate source data classification (note: this will usually be a single column, but in some cases it may be appropriate to concatenate multiple columns)
- Expose table to IMAS Geoserver (grant 'select' privileges)
- NOTE: do not make any modifications to the table other than the specified above. All source habitat data published by Seamap Aus should be 'unaltered' (with the exception of the SM_HAB_CLS attribute addition for WMS styling purposes).

3) Publish table as layer in IMAS Geoserver

- Add table as new layer to store **seamap:IMASSeamapAU**
- Layer **name** should match database table name, avoiding spaces and special characters
- Layer **title** should follow the convention "Seamap Australia – STATE – short title, date range (responsible org)"
  e.g. Seamap Australia - NSW - HABMAP subtidal marine habitats 2005-2016 (NSW DPIE)
- Change *Declared SRS* to EPSG:4326 and select *Reproject native to declared* in bounding box computation options. EPSG:4326 is currently required for datasets published to the AODN portal in order for spatial subsetting to work correctly (this may change in the future).
- Create and assign polygon style (.sld file) using SM_HAB_CLS attribute. Where possible, re-use colours for common habitat classifications (mangroves, seagrass, coral etc) from existing Seamap Australia-hosted source datasets.
- Create content.ftl file to style popup.

4) Create metadata record and publish to IMAS Geonetwork

All metadata records for source habitat datasets, regardless of whether they already exist online in an external metadata repository, are required to be re-hosted in the IMAS Catalogue. This is necessary to standardise the XML structure (using a single ISO standard, applied in a single way)

which enables reliable performance and consistent representation content (e.g. licensing and citation information). The benefit of this is that metadata records will be specifically configured to interact with the AODN Portal (including filters, subsetting, multi-WMS records etc) and source data will additionally be discoverable for viewing and download via the AODN portal, as well as harvested to the AODN Catalogue, Research Data Australia, and other aggregation services.

- For duplicated records, the XML can be directly exported from the source catalogue and inserted into the IMAS Geonetwork catalogue, *including maintaining the source record UUID*. This is important to maintain provenance of the source record. When both versions of the record are harvested into an external aggregation service, they will be grouped together.
- Follow the template of existing records for Seamap Australia-hosted datasets, including acknowledgement, use statements, and link to original record (if it exists). If no metadata record exists, refer to supporting documentation in order to generate an adequately descriptive record.
- Add WMS/WFS connection parameters [for Geoserver] to record – so that data can be accessed/discovered via the metadata record.
- Add Point-of-Truth link [to Geonetwork] metadata record in the Geoserver layer definition – so that metadata can be accessed via the Geoserver layer (exposed via GET requests and desktop GIS software).

Assimilating source data into the National Benthic Habitat Layer

There are five main steps to assimilating new data into the NBHL. These should be applied in the order specified below.

1) Apply Seamap Australia biota and substrata classifications

The habitat component of the Seamap Australia National Benthic Habitat Classification Scheme is typically determined from a single habitat attribute in the source data, but in some cases may be derived from two or more attributes. Data should be inspected (may need to reference supporting documentation) to ascertain what attribute(s) are relevant in determining the Seamap Australia classification.

If the mapping from source classification to Seamap Australia HCS is unclear, extract all distinct original classification terms, and contact the data provider for clarification on translating their source scheme to the Seamap HCS.

2) Apply contextual attributes

The Seamap Australia HCS contains a number of contextual attributes (see pages 1 and 2 of the HCS visual guide) that help to classify the data based on its spatial relationship with established physical or biogeographical boundaries. Their description, and rulesets for their application, are well described in Lucieer et al. (2019) (see especially Table 2).

These attributes can either be applied via desktop GIS software using inbuilt tools (ArcGIS: IDENTIFY or UNION), or in SQL Server via spatial transactions (.STIntersects, UnionAggregate). Both approaches will require the following polygon datasets, either as Shapefile/Geodatabase (for desktop GIS calculations), or as database tables (for SQL Server calculations):

- Oz Estuaries 100k (Geoscience Australia, 2012 – see file 'OzEstuariesReclassified100K'),

- [30m depth contour](#) (derived from the [Australian Bathymetry and Topography Grid](#) (Geoscience Australia, 2009), processed as polygon by the Institute for Marine and Antarctic Studies, 2017)
- [Australian Lowest Astronomical Tide (LAT)](#) (derived from the AHO S57 dataset, processed by the Institute for Marine and Antarctic Studies, 2017)
- [Marine Ecoregions of the World](#) (MEOW)
- Integrated Marine and Coastal Regionalisation of Australia (IMCRA) v4.0 [Provincial](#) and [Meso-Scale](#) Bioregions

Aquatic Settings
- **AS_System** = Aquatic Setting – System. *Describes the aquatic setting in terms of freshwater influence.*
  - Possible values = 2 (Coastal Waterway, Marine)
  - *Unknown* permitted: NO
  - Determined by: spatial relationship with the [*Oz Estuaries 100k*](#) dataset (Geoscience Australia, 2012). Where the Oz Estuaries source dataset identifies an estuarine or delta habitat, AS_SYSTEM = *Coastal Waterway*. All other areas = *Marine*.
- **AS_SubSys** = Aquatic Setting – Subsystem. *Describes the aquatic setting in terms of geomorphology or depth.*
  - Possible values = 2 (Nearshore, Offshore)
  - *Unknown* permitted: NO
  - Determined by: spatial relationship with [30m depth contour](#) (derived from the Australian Bathymetry and Topography Grid; Geoscience Australia, 2009) (for <30m AS_SYSTEM = *Nearshore*, for >30m AS_SYSTEM = Offshore)
  - QC checks: where AS_SYSTEM = Coastal Waterway, AS_SUBSYS = Nearshore *unless unique coastal circumstances, e.g. very deep river mouth*
- **AS_TidalZ** = Aquatic Setting - Tidal Zone. *Describes the tidal zone.*
  - **Possible values = 2** (Subtidal, Intertidal)
  - *Unknown* permitted: NO
  - Determined by: spatial relationship with the [Lowest Astronomical Tide (LAT)](#) and Highest Astronomical Tide (HAT) (as defined by the Permanent Committee on Tides & Mean Sea Level (PCTMSL), 2014). For shallower than LAT, AS_TIDALZ = *Intertidal*, for deeper than LAT, AS_TIDALZ = *Subtidal*.
  - QC checks: where AS_SUBSYS = Offshore, AS_TIDALZ = Subtidal *with a handful of exceptions for localised exposed reefs identified in PCTMSL 2014 but not GA 30m depth contours*, *which and miss some localised features emerging from deep water.*
- **AS_BDepth** = Aquatic Setting – Biotic Depth Zone. *Describes the photic zone of the benthos with regard to the biota that are typically able to grow and survive in those conditions.*
  - Possible values = 3 (Littoral, Infralittoral, Circalittoral), also *Unknown (Sublittoral)*
  - *Unknown* permitted: YES
  - Determined by: data are assigned to Benthic Depth Zones (AS_BDepth) according to tidal influence and photic zone. Rules can be applied to classify shallow zines (see below), but any further information must come from the source habitat data (or else be assigned *Unknown*).
  - Possible rules to apply;
    - WHERE AS_TidalZ = Intertidal, AS_BDepth = Littoral

- WHERE BC_Level2 = Vegetation AND AS_TidalZ = Subtidal AND AS_BDepth !=
Littoral, AS_BDepth = Infralittoral
i.e. an algal dominated area of the sublittoral zone
- WHERE BC_Level2 = Invertebrates AND AS_TidalZ = Subtidal AND AS_BDepth !=
Littoral, AS_BDepth = Circalittoral
*i.e. an invertebrate (non-algal) dominated area of the sublittoral zone; below
euphotic)*
- WHERE AS_BDepth IS NULL, AS_BDepth = Unknown (Sublittoral)
*i.e. no definitive classification is possible, but we know that this falls outside the
littoral zone, therefore it is sublittoral*

Biogeographic Settings

- **BS_IMCRA_P** = IMCRA Provincial Bioregion
  - o Possible values = 41 (see source data)
  - o *Unknown* permitted: NO – use '-' for data outside bioregion bounds
  - o Determined by: spatial relationship with IMCRA v4.0 Provincial Bioregions
- **BS_IMCRA_B** = IMCRA Meso-scale Bioregion
  - o Possible values = 61 (see source data)
  - o *Unknown* permitted: NO – use '-' for data outside bioregion bounds
  - o Determined by: spatial relationship with IMCRA v4.0 Meso-scale Bioregions
- **BS_MEOW_R** = MEOW Realm
  - o Possible values = 12 (see source data)
  - o *Unknown* permitted: NO (all data will fall inside bioregion bounds)
  - o Determined by: spatial relationship with MEOW
- **BS_MEOW_P** = MEOW Province
  - o Possible values = 62 (see source data)
  - o *Unknown* permitted: NO (all data will fall inside bioregion bounds)
  - o Determined by: spatial relationship with MEOW
- **BS_MEOW_E** = MEOW Ecoregion
  - o Possible values = 232 (see source data)
  - o *Unknown* permitted: NO (all data will fall inside bioregion bounds)
  - o Determined by: spatial relationship with MEOW

3) Appending data provenance information

Each individual polygon in the NBHL retains provenance information on the data source, and the
original habitat classification (from the native scheme). Note that biotic and substratum
classifications may be derived from a single dataset, or two different datasets (i.e. Data_BC and
Data_SC may be the same, or different).

The following attributes should be created and populated:

- Data_BC / Data_SC: Title of source dataset used for biotic/substratum classification
- Date_BC / Date_SC: Date of collection of data used for biotic/substratum classification
- Source_BC / Source_SC: Source of data used for biotic/substratum classification
- Info_BC / Info_SC: URL for metadata record containing information on source data used for
biotic/substratum classification

4) Ranking/prioritising datasets

Habitat datasets will frequently spatially overlap. A transparent and repeatable set of prioritisation criteria is required to rank datasets and determine which will be included / assigned priority in the NBHL.

Four main criteria are used to rank data when constructing the NBHL:

1. The accuracy of the methodology employed to collect the data (for example singlebeam acoustic data (point coverage) will be superseded by multibeam acoustic data (full coverage) it will be replaced.
2. The temporal resolution of the data (in general, newer = higher priority). However, if the most recent data is collected by a less accurate sampling methodology, the most accurate and highest resolution dataset will be retained.
3. The degree of field validation/ground-truthing (e.g. by AUV, BRUV, drop cameras, manta tows). A dataset that is validated by more observations is deemed more accurate and will take priority over a sparsely validated dataset.
4. The taxonomic resolution of the data. E.g. dataset with a source classification that enables a classification to level 3 or 4 of the Seamap Australia HCL, or classifies biota where only substrata classification was possible previously, will receive priority.

While all attempts are made to apply prioritisation criteria in a fully transparent and repeatable manner, discretion is often required when assessing new habitat data contributions against the existing NBHL. It is unlikely that a fully automated approach to data ranking and updates will ever be possible due to the unavoidable need for expert assessment at this step.

The figure below (extracted from from Scientific Data publication) illustrates the basic data processing workflow, and point at which decisions must be made regarding data priority.
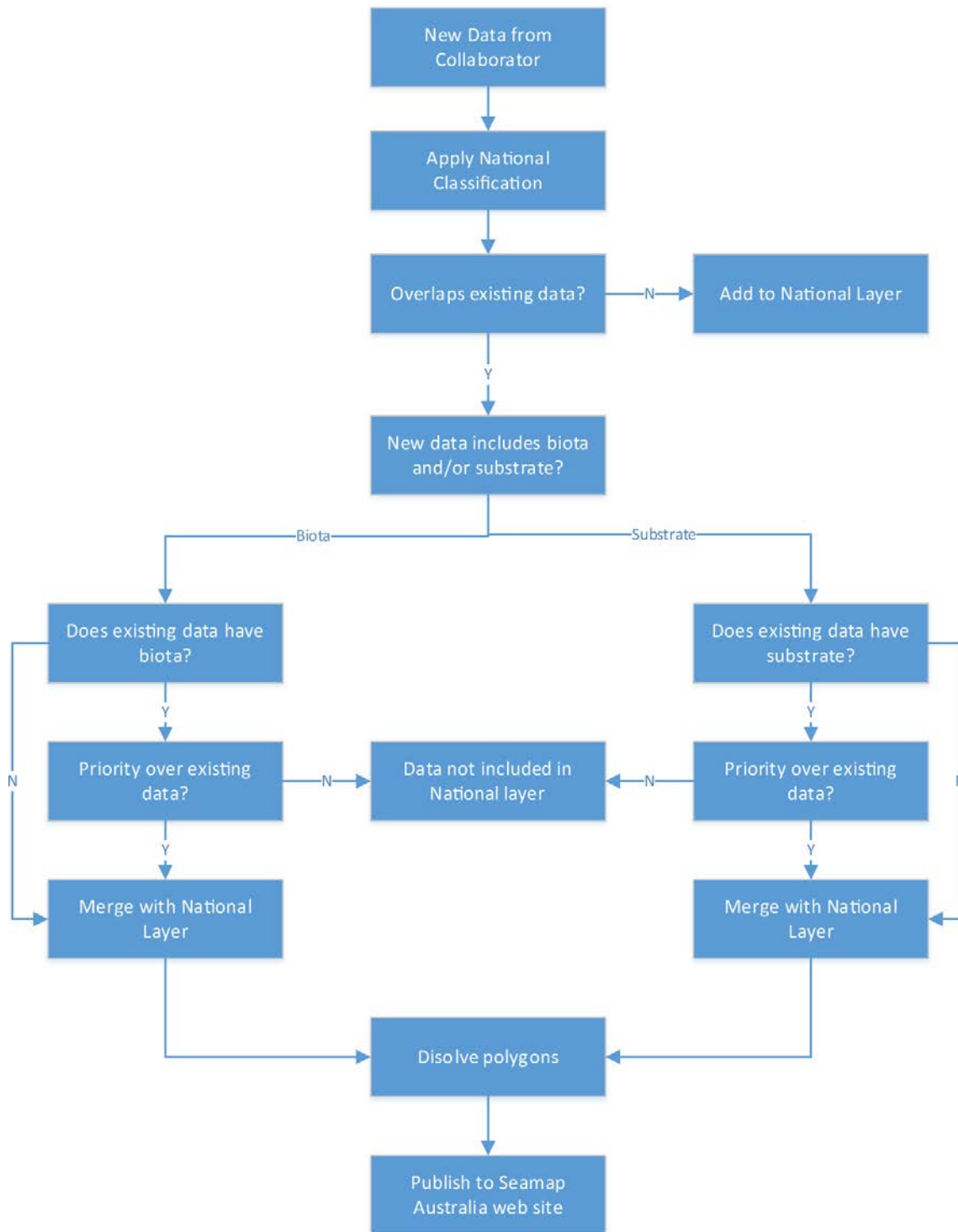
Figure 1. Seamap Australia data processing workflow (taken from [Lucieer et al. 2019](#)).

5) Spatial operations for adding new data to the NBHL

Once the order of data priority has been determined, new data must be assimilated into the NBHL using the prioritisation framework specified in the previous section. Spatial operations for this final assimilation step are currently carried out in desktop GIS software before pushing the data back into the database server. There are only four possible scenarios (based on both layer priority and proximity of new to existing data) to consider when determining the operation workflow for assimilating new data into the NBHL. The instructions below list the steps for ArcGIS (ESRI) software, but other alternatives (QGIS, gdal etc) may also be used.

Case 1: new habitat data does not overlap existing NBHL.

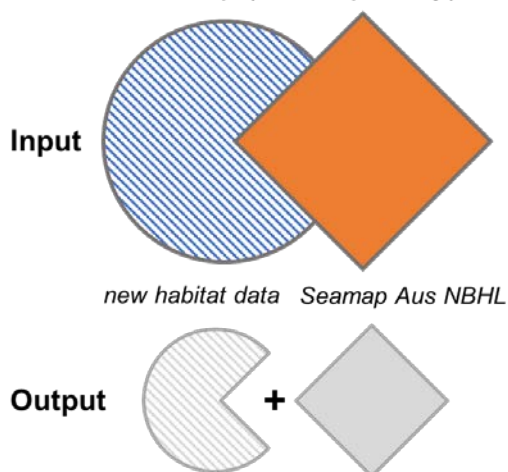- MERGE new habitat data with existing data (*input:* new data, NBHL)
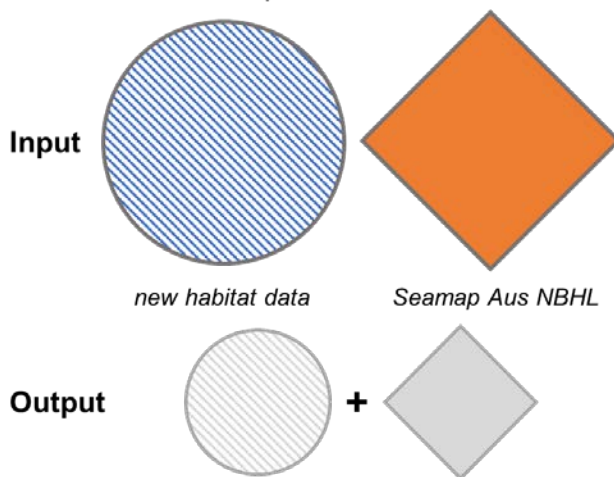
Case 2: new habitat data is *lesser* priority than existing NBHL.

- ERASE new habitat data with existing data (*input:* new data; *erase feature:* NBHL)
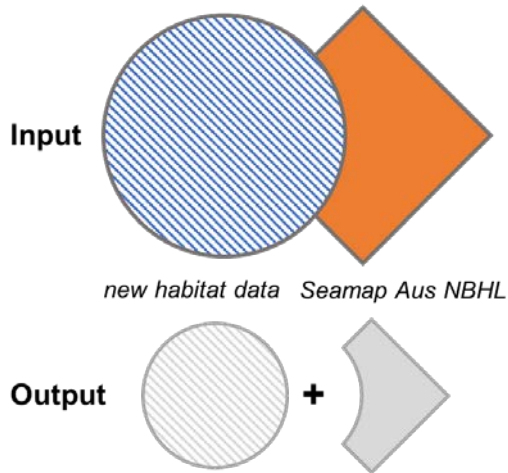- MERGE output with NBHL

Case 3: new habitat data is *greater* priority than existing NBHL, and classifies both biota and substrata

- ERASE existing data with new habitat data (*input:* NBHL; *erase feature:* new data)
- MERGE output with new habitat data

Case 3. overlap (greater priority)

new habitat data   Seamap Aus NBHL

Case 4: mixed model. *Either* new habitat data is greater priority than existing biota but not substrata (or vice versa); *or* new habitat data is greater priority than existing data but only classifies biota or not substrata (not both).

- UNION existing data with new habitat data (*inputs*: NBHL, new data)
- SELECT BY ATTRIBUTE where new data exists (i.e. FID <> -1)
- Use Field Calculator to assign values as required (may be replacing existing biota/substrata, or populating previously-unmapped areas)

Case 4. mixed (partial) model

new habitat data   Seamap Aus NBHL
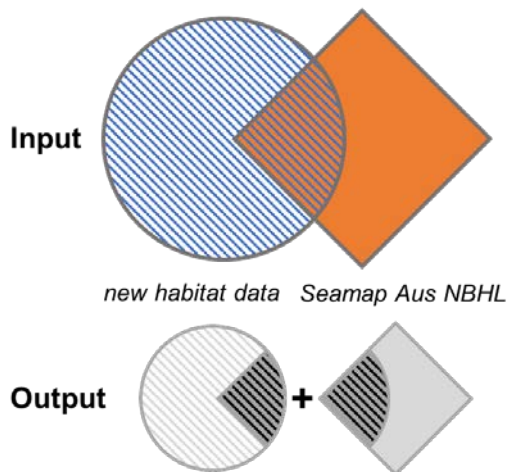
General data cleaning

Quality control of datasets contributed to Seamap Australia for inclusion in the NBHL is the responsibility of the data providers. The Seamap Australia team will undertake basic automated data cleaning (REPAIR GEOMETRY, remove overlapping polygons) before assimilating source datasets into the NBHL. Non-geometry attributes will not be modified.

The process of assimilating new data into the NBHL can introduce spatial artefacts in the form of sliver polygons, overlapping polygons, or invalid geometries (empty or null geometries, self-intersecting segments etc). The NBHL is a large dataset and maintaining clean and valid geometry is important for both accuracy and performance. Each time significant changes are made to the NBHL, cleaning steps should be applied to prevent accumulation of unwanted spatial artefacts. The instructions below list the steps for ArcGIS (ESRI) software, but analogous tools are available in other software.

1. REPAIR GEOMETRY
2. ELIMINATE (by longest border) all polygons with area < 3.0m (the accuracy of EPSG:3112). Note this a computationally intensive process, it may help to split polygons before attempting to eliminate them (convert multipart to singlepart).
3. Inspect remaining polygons not removed by eliminate (i.e. those disconnected from other features) – these may be real data (e.g. small 'islands' from MBES) or residual artefacts. Delete if/as required.
4. Remove overlapping sections of polygons using the following technique:
   a. INTERSECT (with NBHL as single input) to select overlapping polygons
   b. DELETE IDENTICAL. Alternatively, overlapping polygons (output of Step (a)) can be manually inspected and removed.
   c. ERASE NBHL with output from Step (b)
   d. MERGE outputs from Step (b) and Step (c)
5. SIMPLIFY POLYGON using 'preserve bend' algorithm, setting the tolerance appropriately low (e.g. 2.0m curve diameter). This will not significantly affect polygon shape but will help to manage filesize.
6. INTEGRATE may be used to remove sliver polygons, but care must be taken in setting tolerance so as not to collapse geometries – threshold should be set very low (<0.1m)
7. DISSOLVE and create MULTIPART POLYGONS (this is important for performance reasons)

Supporting Documents and Resources

- ***Seamap Australia Data Acceptance Guidelines***
- Seamap Australia National Benthic Habitat Classification (visual guide here, registered vocabulary API here)
- *Scientific Data* publication on the technical aspects of the NBHL
- Report fully detailing the derivation and development of the classification scheme, and its relevance to other schemes (includes term descriptions)

References

Lucieer, V., Barrett, N., Butler, C., Flukes, E., Ierodiaconou, D., Ingleton, T., Jordan, A., Monk, J., Meeuwig, J., Porter-Smith, R., Smit, N., Walsh, P., Wright, A. and Johnson, C. A seafloor habitat map for the Australian continental shelf. *Sci Data* **6**, 120 (2019). https://doi.org/10.1038/s41597-019-0126-2

Butler, C., Lucieer, V., Walsh, P., Flukes., E and Johnson, C (2017) Seamap Australia [Version 1.0] the development of a national benthic marine classification scheme for the Australian continental shelf. *Final Report to the Australian National Data Service (ANDS) High Values Collection #19*. Institute for Marine and Antarctic Studies, University of Tasmania. 52 pgs.

Version	Revision Date	Author	Action
Draft	01/07/2022	Emma Flukes	Developed
V1.0	04/01/2023	Emma Flukes	Finalised