European Marine Observation and Data Network

Guidelines for metadata, data and DTM QA/QC

Service Contract for Lot 1 Bathymetry

Modifications

Version	Date	Authors	Comments
1.0	2009/12/03	P. Hunter, B. Loubrieu, E. Moussat and R. Vossen	Initial version
1.1	2010/02/12	E. Moussat	Update
1.2	2010/02/22	Thierry Schmitt, G. Morvan	Update
1.3	2010/03/31	E.Moussat	Update DTM Bounding box and origin DTM production/ Composite DTM Metadata in CDI DTM QC
1.4	2010/04/20	P.Hunter E.Moussat	Text review Inclusion of the GEBCO 30 arc-second ID
1.5	2011/02/04	B. Loubrieu, E.Moussat. D. Schaap	Variables and attributes to be delivered §2.2.1.2 : source dataset identifiers coding
1.6	2015/02/04	D.M.A. Schaap	Updating for new resolution

1. Introduction

The **EMODnet-Bathymetry portal** is being developed in the framework of the European Marine Observation and Data Network (EMODnet) as initiated by the European Commission. It provides services for discovery and requesting access to bathymetric data (survey data sets and composite DTMs) as managed by an increasing number of data providers from government and research. The portal also provides a service for viewing and downloading a harmonised Digital Terrain Model (DTM) for the European sea regions that is generated by the EMODnet Bathymetry partnership on the basis of the gathered data sources.

Bathymetric survey data sets are being catalogued as acquired and managed by European data providers for all regions in the world. A harmonised Digital Terrain Model at a gridsize of 1/8 * 1/8 arc minutes has been produced and is made available for the following European sea regions:

- the Greater North Sea, including the Kattegat and stretches of water such as Fair Isle, Cromarty, Forth, Forties,Dover, Wight, and Portland
- the English Channel and Celtic Seas
- Western Mediterranean, the Ionian Sea and the Central Mediterranean Sea
- Iberian Coast and Bay of Biscay (Atlantic Ocean)
- Adriatic Sea (Mediterranean)
- Aegean - Levantine Sea (Mediterranean).
- Madeira and Azores (Macaronesia)
- Baltic Sea
- Black Sea
- Norwegian and Icelandic Seas
- Canary Islands (Macaronesia)

Users are given access to the following geographical information system layers:

- water depth in gridded form over whole of maritime basin on a grid of 1/8 * 1/8 arc minutes
- water depth in vector form with isobaths at a scale of at least one to one million
- depth profiles along user drawn tracks
- tracklines / coverage of bathymetric surveys, including metadata, underpinning the water depths
- underwater features - wrecks, seabed obstructions etc

These layers are made available in the Bathymetry Viewing and Download service. The portal also provides a CDI Data Discovery and Access service to identify and request access to bathymetric survey data that are acquired and managed by distributed data providers. A selection of these data sets are used for the EMODnet Digital Terrain Model (DTM). Moreover the portal provides a Sextant Catalogue service to describe and identify composite DTMs that were contributed by data providers instead of bathymetric survey data as input for the EMODnet DTM.

This document has a focus on the EMODnet Digital Terrain Model (DTM):

- ➢ the specifications of the DTM
- ➢ the procedure to create it
- ➢ the description of the metadata associated with the datasets and the QA/QC rules to apply to them to produce the DTM
- ➢ the QA and QC to apply to the datasets before merging them in the DTM

- the QC to apply to the resulting DTM

2. DTM specifications

2.1. Data sources to be considered

Source data are preferably **survey datasets** (single and multibeam surveys delivered either as a set of soundings or as a high resolution DTM produced from a single survey as can be in the case of LIDAR or of very shallow MBES).

A number of external providers have opted to deliver only **composite data products** (DTMs with a specific grid resolution), that are derived by the external data providers (e.g. HOs) from multiple surveys. These will be used as alternatives when single survey data are missing.

Elsewhere **GEBCO 30 arc-second** gridded data will be used to complete the coverage.

It is accepted that the accuracy and precision of the gridded data will vary over the basins in question. No new data will be collected specifically for this project. Please note that the gridded product here refers to the GEBCO-2014 grid, released in 2014 and made available to EMODnet by the Geographic Bathymetric Chart of the Oceans (GEBCO) editor, the British Oceanographic Data Centre (BODC).

In the case of several survey datasets being available for a cell, **all** might be used for the calculations **after having been filtered using the QA/QC standards.**

2.2. DTM parameters and attributes to be delivered

The parameters have been chosen to allow the widest use of the DTM together with an evaluation of the accuracy of the interpolated values from observed data.

However, the fact that some data providers have opted for delivering composite data products which are not based on the same specifications (and often on specifications not really well known) limits the possibility to give an estimate of the accuracy over the entire DTM.

For areas covered by soundings, the depth parameters to be delivered for each DTM grid cell have been defined and agreed as follows :

- Mininum cell depth
- Maximum cell depth
- Average cell depth
- Standard deviation of cell depth
- Number of values used for interpolation of cell depth
- Number of elementary surfaces used to compute the average cell depth
- Average depth smoothed by means of a spline function
- An indicator of the offsets between the average and smoothed depth (as a % of the depth).

In addition each grid cell will contain an identifier indicating the source material used for the computation of the depth combined with the ID of the data provider in the European Directory of Marine Organisation EDMO ie :

- the source dataset whose soundings are the most frequently representeded in the cell ie the Local_CDI_ID associated to the description of the survey dataset in the CDI Data Discovery and Access service
- the source composite DTM used to fill the cells without soundings. This ID will allow to find the description of the composite DTM in the Sextant Catalogue service for composite DTMs
- the GEBCO 2014 grid

So there are 3 types of references which have to be distincted. These references must be coded as follows :

- CDI => EDMO-code-provider_Local-CDI-Id
- Composite DTM => EDMO-code-provider_Sextant-CPRD-Id
- GEBCO => GEBCO_version

Note : GEBCO version and composite DTM Ids are recorded in the same reference field while CDI Ids are in another one in the DTM exchange format.

2.3. EMODnet DTM reference framework

The EMODnet Bathymetry partners have agreed on the following DTM reference framework:

2.3.1 Horizontal coordinate reference system

It has been decided, according to the specification to produce the DTM **in angular coordinates at a resolution of 1/8 arc-minute of latitude and longitude.**

2.3.1.1. Datum

Geodetic system WGS84 (EPSG identifier: 4326)

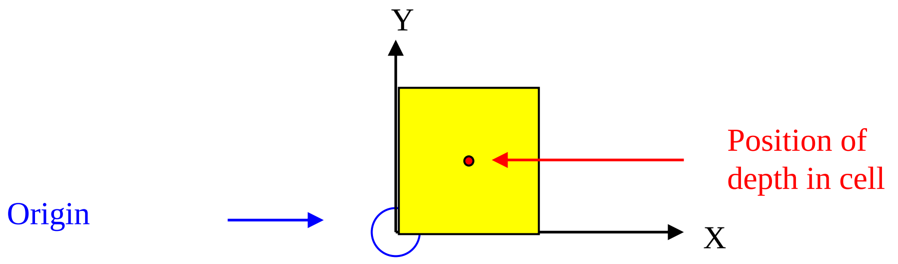
2.3.1.2. Axis conventions

X = longitude in decimal degrees from -180 to +180 with East >0;
Y = latitude in decimal degrees from -90 to +90 with North > 0

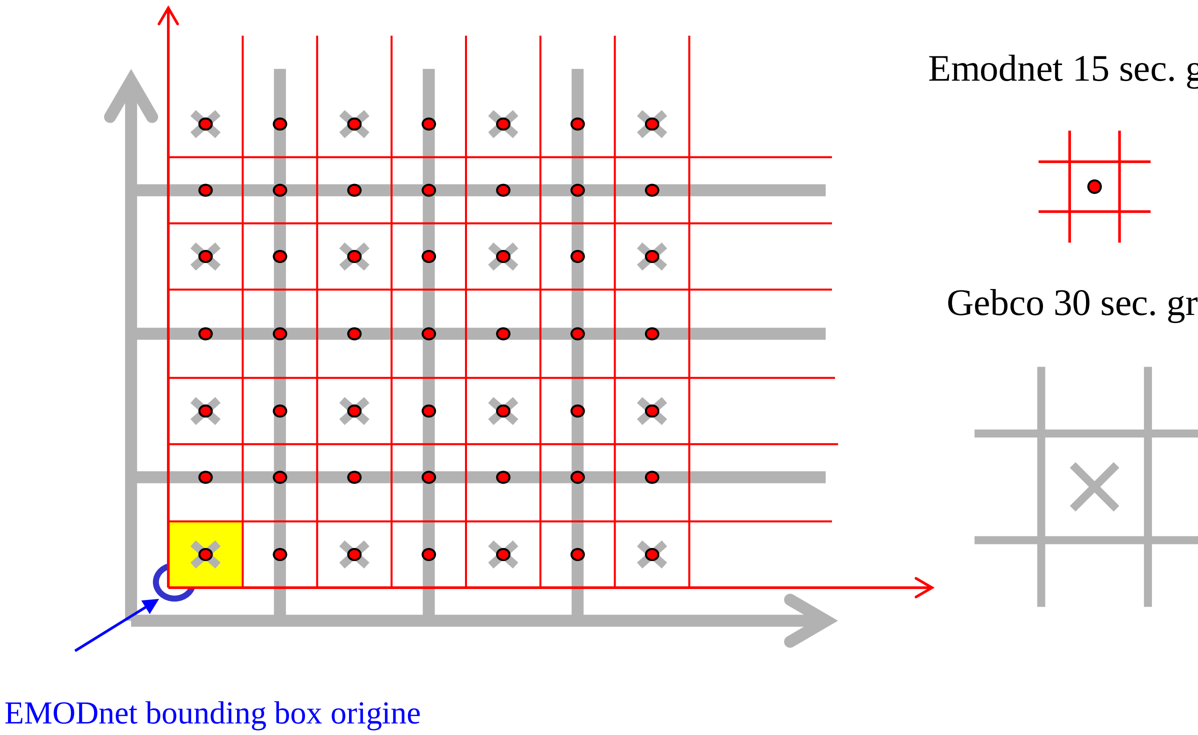
2.3.1.3. Bounding box and origin of DTMs

All DTMs start in the South West corner of South West grid cell and end in the North East corner of North East cell.

.

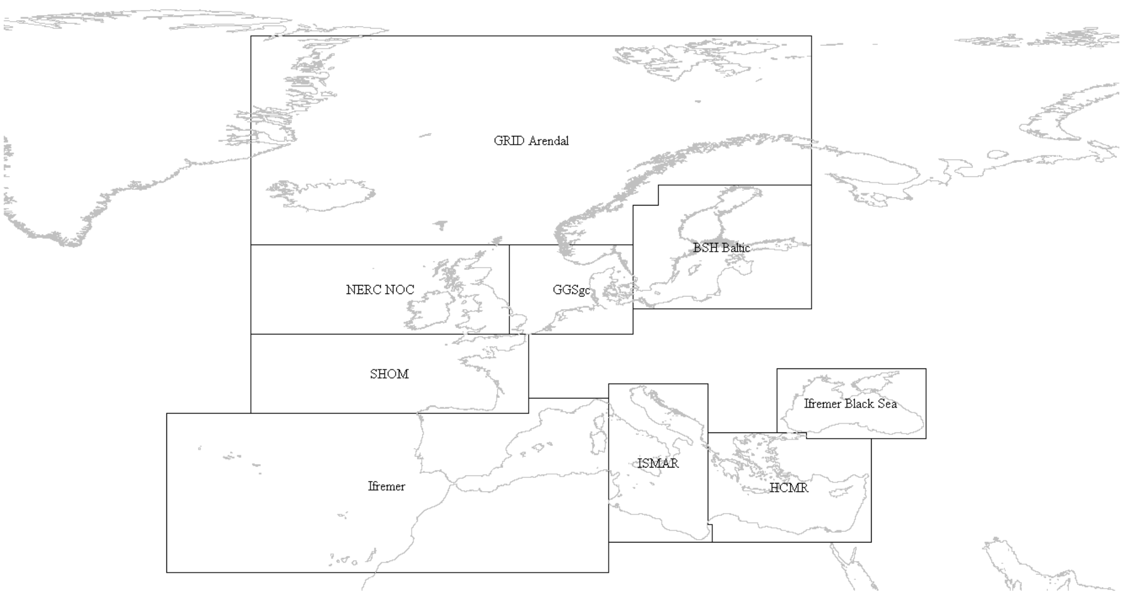


In order to avoid interpolation of the 30 arc-second GEBCO grid values, rows and lines of the EMODnet DTMs must be aligned with rows and lines of the GEBCO grid as illustrated below :

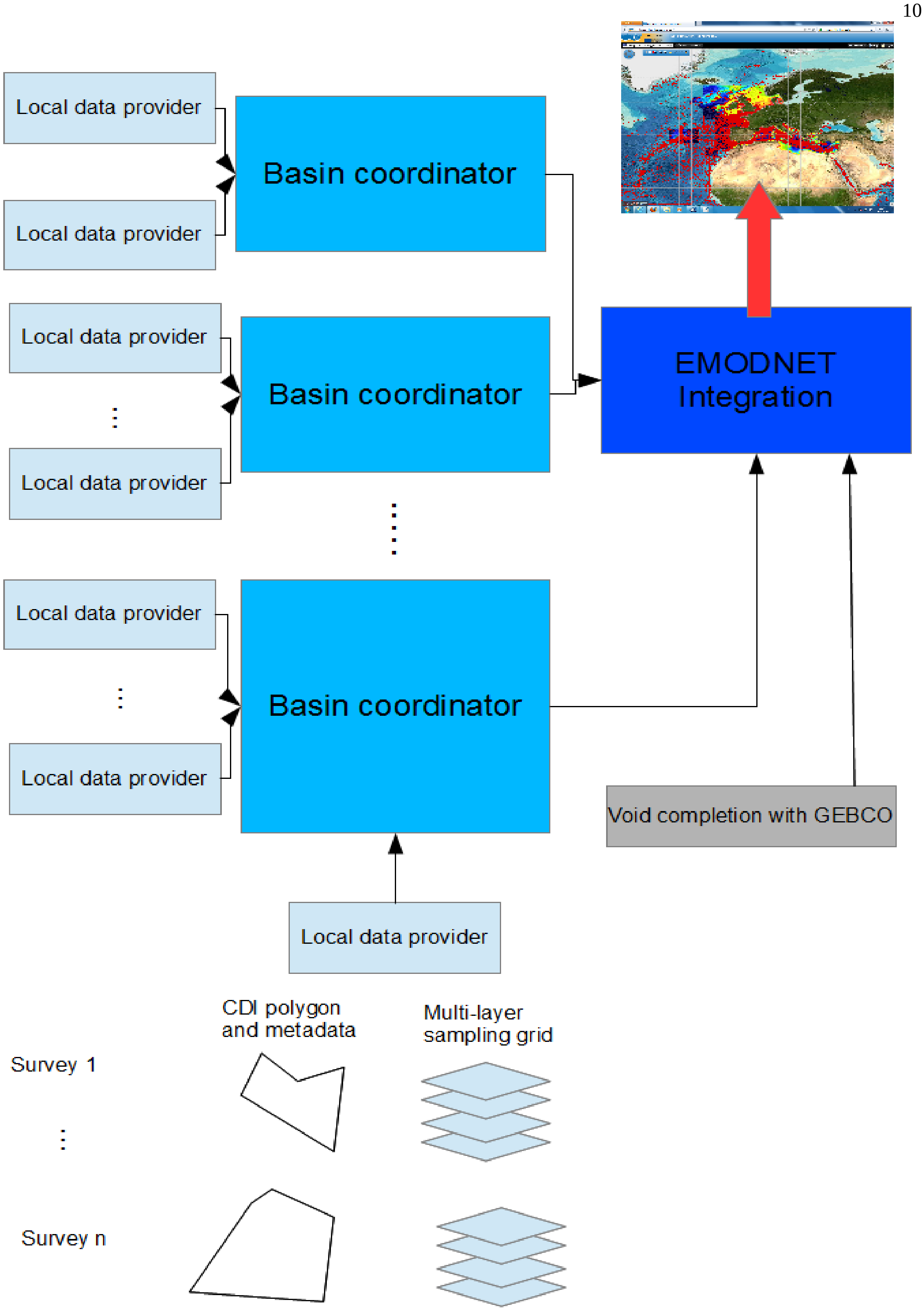


2.3.1.4 EMODnet DTM geographic coverage and division over regions

For producing the EMODnet DTM a division in regional DTMs has been done, each coordinated by a dedicated Regional Coordinator. The EMODnet geographic coverage and the division over regional DTMs are given below.



Coordinator		Upper right		Lower left		Remark
Ifremer + IPMA	*Iberian Coast - Atlantic - Macaronesia - Western Med*	*-34.5*	*43*	*10*	*27*	*In the Mediterranear area's northern lattitu 44.5 degrees*
HCMR	*Eastern Med*	*20*	*41.1*	*36.5*	*30*	*Area includes Sea Marmara and Bosp Strait south of 40.5 deg Area does not include G Sidra*
Ifremer	*Black Sea*	*27*	*47.5*	*42*	*40.5*	*Area* ***does not*** *include S Marmara and Bosp Strait south of 40.5 degr*
CNR-ISMAR	*Central Med*	*10*	*46*	*20*	*30*	*Area includes complete of Sidra*
SHOM	*Channel - part Atlantic - Bay of Biscay*	*-26*	*51*	*2*	*43*	
NERC-NOC	*Celtic Sea - part Atlantic - part North Sea*	*-26*	*60*	*0*	*51*	*Exclusive of the SHOM*
GGSgc	*Greater North Sea*	*0*	*60*	*12.5*	*51*	
BSH	*Baltic*	*12.5*	*66*	*30.5*	*53.5*	*Exclusive of Norwegian*
GRID Arendal	*Norwegian Sea - Icelandic Sea*	*-26*	*81*	*30.5*	*60*	*Exclusive of BSH Baltic*



2.3.2 Depth reference system

2.3.2.1. Reference level

Lowest Astronomical Tide (LAT) is adopted. Mean Sea Level (MSL) is an appropriate surface but LAT has the advantages of being :

- also a potential surface which fits the needs of hydrodynamic applications
- the system recommanded by IHO for data delivery by HOs.

The adoption of MSL would require an additional task which can be difficult to carry out because transformation parameters are rarely supplied.

- In deep waters (more than 200m), this surface converges with Mean Sea level and Sea Surface. Errors will be negligeable in comparison with the uncertainties of the soundings themselves.
- In some areas such as the Mediterranean, this will even be less important due to the small amplitude of tides in these areas. Isoline of the max amplitude of the tide will be derived from tide model in this area to evaluate the limit where not corrected data will be used. (See QA/QC below)

2.3.2.2. Axis convention

Depth in metres, >0 down, (to 2 decimal places)

This convention applies to the DTM produced by the consortium. The EMODnet Bathymetry website offers the possibility to the end user to adopt an opposite convention when retrieving depths.

3. Procedure of production of DTM

It is assumed that all the datasets are:

- in WGS84 (or compatible with WGS84 at the scale of work ie horizontal differences in position are less than 1/16 arc-minute)
- depth values are compatible with the depth reference level and sign convention

3.1. Soundings

3.1.1 Definition

Very shallow multibeam echosounder surveys result in very large datasets requiring filtering of the soundings to be handled. A common solution adopted by data providers is to provide a very high resolution DTM (at a few metres of resolution) of the survey instead of the full set of soundings. This is the most frequent solution adopted for LIDAR surveys.

For the purpose of EMODnet, survey data sets are pre-processed and pre-gridded by data providers, preferably using the GLOBE software as provided by IFREMER to the project partners. This facilitates preparing survey data sets as high resolution DTMs. These will be considered as sets of soundings and as such will be described as observation datasets in the CDI Data Discovery and Access service.

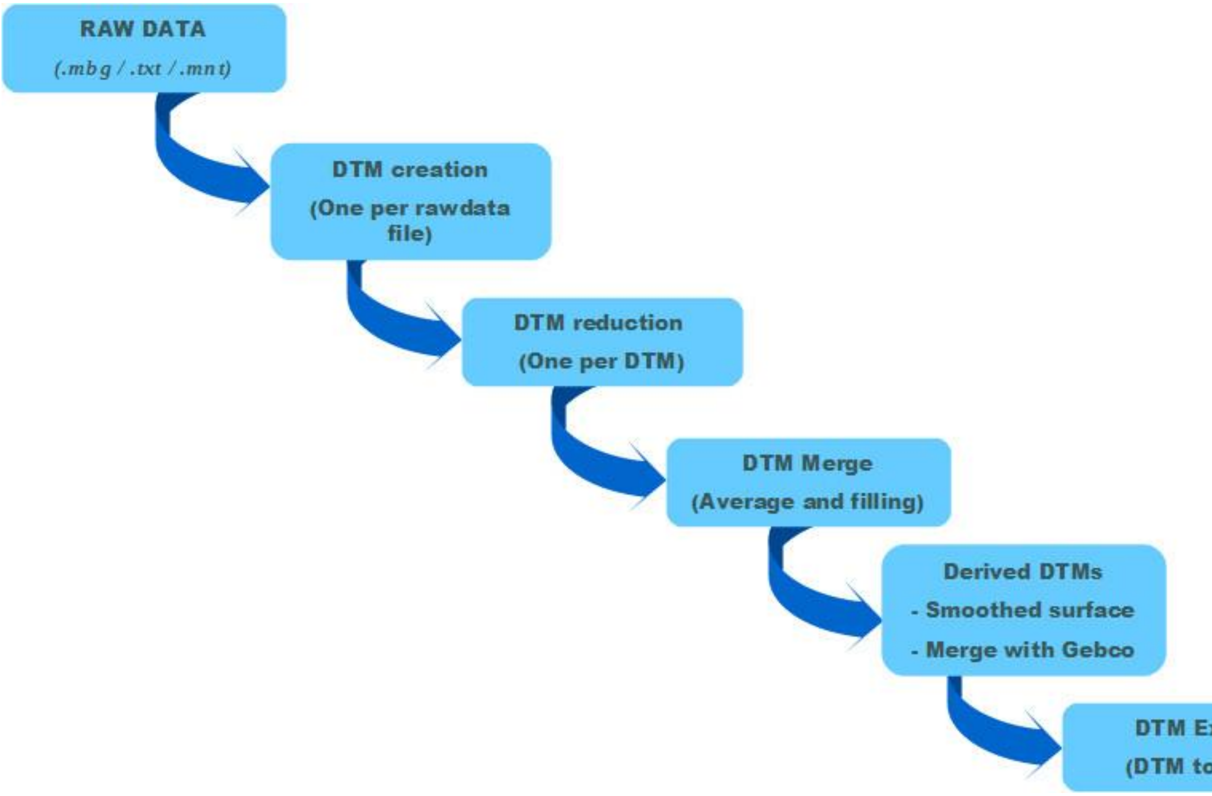
The choice of the grid cell size is dependent on: survey characteristics, data accuracy, depth range, positioning system, internal data policy. This choice is the responsibility of the data provider. The following table gives a guideline.

Level of resolution	**Cell size**	**Approx size in m**	**Single beam**	**Multibeam**
1	1/8'	230	deep sea	deep sea
2	1/16'	115	deep sea and shelf / low density	continental margin

3	1/32'	60	continental shelf	continental shelf
4	1/64'	30	coastal area	coastal area

Table: Pre-sampled grid resolutions accepted vs. context

Note: raw soundings stands for the data that the data provider is ready to provide in his own internal format, while pre-sampled grids stands for these raw soundings gridded at 1/16 (or better) and provided to the regional basin coordinator prior to integration. This is illustrated in the image below.



3.1.2 Sounding dataset selection

The accuracy of the positions of the soundings varies as a function of the source and also as a function of the conditions during acquisition. It is accepted that the accuracy will vary over the DTM. However, in order to keep the resolution of the bathymetric features as accurate as possible (ie 1/8 arc-minute) it is suggested not to take into account data collected with a positioning system accuracy worse than 500m (2 sigma) when better located data are available.

It is not unusual for soundings from different surveys to be present within any given seabed area. This results from the deliberate overlapping of survey edges necessary ot avoid gaps . The combination of soundings from these different surveys to compute a mean seabed surface may result in the creation of artefacts or artificial features (e.g. undulations) Differences in the height of adjacent soundings taken at different times may result from genuine changes in the seabed over time, the survey methodology and/or the method of sounding reduction. The significance of this in the resulting DTM increases with resolution. The selection of soundings from one survey as input to the DTM and the rejection of soundings from other surveys is known as ‘deconfliction’.

Usually the most recent surveys provide the most appropriate data to be used as input to DTM computation and should be used in preference. By definition the most recent surveys provide the latest measure of the seabed and are more likely to be the most accurate(due to improvements in sounding and positioning technologies- improvement in terms of accuracy, coverage and density of soundings). However, this is not always the case, for example, where a different methodology has been used for expediency or comparative purposes.

Although it is possible to derive rules to decide which soundings should be used and which rejected, it is recommended than manual inspection of the data is undertaken by the Basin Coordinator prior to DTM computation and any refinements undertaken as required.

The IHO standard for hydrographic surveys (S-44) provides minimum criteria for hydrographic surveys, so that the highest order of survey (most accurate, greatest feature detection and providing 100% coverage) should be used in preference to a lower order survey. However, the order of survey is not always available or captured within metadata in which case the survey methodology can be used as the primary factor with age being taken as the secondary factor, multi beam data being preferred over single beam and ship track data.

3.1.3 Data processing

There may be instances of sources with a large number of soundings while others may have only one (this is especially the case when providers such as HOs supply subsets of their surveys).
The data sampling process is based on a definition of a set of multiple hierarchical resolution grids for each of the European basins (see image below). Data providers generate pre-sampled grids for one of the levels in this hierarchy. The Regional Coordinators and the overall EMODnet Integrator generate the final bathymetric products by the aggregation of the different resolution grids to produce a grid with a resolution of 1/8 arc-minute.

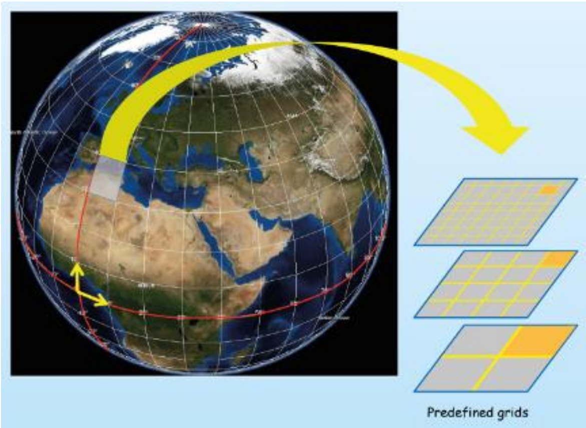


Image: multiple resolutions grid frames

In order for this mechanism to work, the principal characteristics of the pre-sampled grids must use the following structure:

- a common origin : 0° N / 0° E for the 1/8 arc-minute grid, this origin is the low-left corner of the south-western pixel.
- a hierarchy of pixel size: 1/4, 1/8, 1/16, 1/32, 1/64… arc-minutes

- a common definition of the content of the layers
- a common reference of the information within the pixel (i.e. pixel centred values)

The grid origin X0/Y0 is its low-left corner, from which are located all pixels.

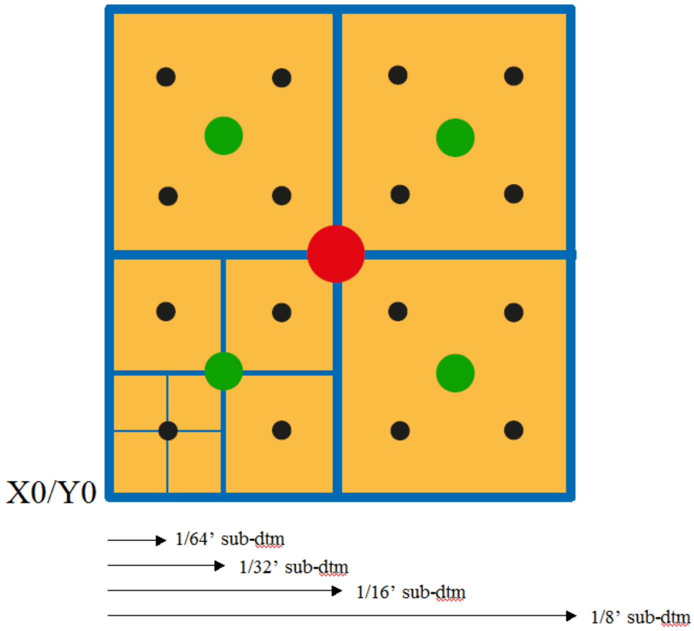


Figure 1 Grid geometry

The origin defines the coverage boundaries of the grid.
XO/YO = 0°+x' / 0°+y'.

The DTM layers are affected to the center of each cell

There may be grid cells without any soundings. In those cases it is agreed to derive the grid cell values from the values of the *bounding* grid cells as far as a value is available in each quarter of the neighbouring space by linear interpolation as a function of the distance.

3.2. Integration of data supplied as composite DTM

3.2.1 Definition

A "composite" DTM is a grid produced from multiple sources of data and surveys.

3.2.2 DTM selection

A number of external providers have opted to deliver only **composite data products** (DTMs with a specific grid resolution), so it will only be possible to fill gaps by deriving depths from compiled DTM values.

These will be integrated as a function of their resolution (the highest resolution first) and as a function of their compatibility with the target reference framework and their consistency with validated observed data.

3.2.3 Composite DTM processing

Each composite DTM will be gridded separately at the resolution of the target DTM. The resulting DTM will be integrated in the regional DTM in the area specified as a polygon by the regional integrator. This will be done in order to erase, if needed, isolated values of the target DTM in order to maintain the consistency of the values from the source composite DTM.

The order to integrate DTMs will depend on their resolution: the higher the resolution, the higher the priority, however the choice of integrating a composite DTM depends on other quality factors the evaluation of which is left to the regional coordinator.

3.3. Inconsistencies between data sources

The use of DTMs of various sources may result in inconsistencies at the limits of their respective coverage as the ways data have been processed vary from one source to another. It may not be possible to correct the resulting offsets between data sources. While it is important to be able to identify these offsets and to preserve in the DTM the results from the sources, they can introduce artificial morphologic features which can preclude the use of the DTM for several applications .

Therefore it is decided to produce also a DTM with :

- depth smoothed by means of a spline function.
- the offset related to the average water depth (%)

3.4 Void filling

Data from composite DTMs and the GEBCO grid are used to fill holes or voids (i.e. areas of no sounding data coverage) in the regional basin DTMs. Only depth estimates sampled from these grids at 1/8 arc-minute intervals by the Basin Coordinator are used for the integration into the EMODnet DTM product. Note that a tight connection is made with GEBCO. The compilation of the individual regional DTM into one average depth grid (containing holes) will be provided to GEBCO. GEBCO will update his model using this grid (using the remove and restore technique). The EMODnet average grid will then be void filled using the updated GEBCO grid.

GEBCO has carried out the void filling procedure using the previous EMODnet regional basin 1/4 arc-minute DTMs. The updated GEBCO grid will then be used for void filling for the new 1/8th grid. This procedure will be repeated in future cycles.

3.5 Generating a smoothed bathymetric surface

As part of the previous EMODnet ¼ arc minutes grid, a smoothed surface was created from the mean depth grid for each regional basin DTMs. Holes in these grids were filed with data from the GEBCO grid. The smoothing algorithm used was a spline function. For the current EMODnet grid development work tests will be carried out to improve the choice and the operation of the spline algorithm and to apply it only at local spots and not overall in order not to lose the higher resolution details.

4. Relevant metadata and associated QA/QC for DTM production

This chapter lists a number of rules to apply when compiling metadata for survey daa sets or composite DTMs.

4.1. General

Rule A: all compiled survey datasets must be described in the CDI Data Discovery and Access service.

The CDI is designed for observation data so composite data products cannot be described in the CDI in spite of the fact that many fields are appropriate.

Rule B: supplied composite DTMs must be documented in the Sextant Catalogue service of Composite DTMs.

Global grids at 30 arc-second of resolution (eg GEBCO) or regional grids at 1km exists. Therefore, the objective of the present 1/8 arc-minute grid is not only to update our knowledge with new bathymetric data sources but to allow a higher resolution and consistency of the representation of bathymetric features.

Therefore, the datasets will be selected for integration only if they are compliant with the criteria below in order to ensure an improvement related to previous works at the scale of the basins.

4.2. Horizontal datum

Rule C: horizontal datum is mandatory. Datasets with unknown horizontal datum will not be taken into account.

Rule D: all datasets must be converted to the target datum (WGS84) unless the offset is negligible related to the accuracy of the positioning before transmission to the data integrator.

4.3. Positioning accuracy

The resolution can be heavily altered by low accuracy of the positions of data. Data position accuracy is frequently not available except for HO survey data. It can be approximated from the positioning systems used during the surveys. Therefore:

Rule E: the identification of the positioning system and its class of precision is mandatory. This must be recorded in the "Instrument" field of CDI.

Rule F: survey data with an accuracy (at 2drms) worse than 500m will have an impact on the quality of the results and will not be integrated when data with better position accuracy exist.

Rule G: when no other data exist, datasets of significant and continuous coverage with position accuracy between 500m and 1km (ie better than the resolution of the GEBCO grid) will be used. These datasets should be compared with the GEBCO grid first to ascertain that they are superior to it.

4.4. Sea level of reference

Rule H: the identification of the sea level of reference (LAT, MSL,sea surface...) is mandatory. This must be recorded in the "Depth datum" field of CDI.

Note: if tide corrections are applied, this implies that a reference level has been adopted (e.g LAT) . When no correction is applied, mention "Sea surface".

Rule I: **soundings and depth of composite data product supplied to data integrator must refer to LAT or a reference level compatible(1) with LAT.**

(1) In deep water (>200m), observed sea surface is compatible with LAT. Areas where sea surface is compatible with LAT will be defined according to a prediction model of the maximum amplitude of tide.

4.5. Sounding systems

This information is relevant for the spatial and depth resolution of the system and the sampling interval during the survey. Distinction between deep, intermediate and shallow water echosounders is recommended as far as this list as been introduced in the instrument list

Rule J: the instrument used for sounding is mandatory and must be recorded in the "instrument field" of CDI.

4.6. Data processing traceability

All relevant information on the way processing and validation has been carried out must be recorded in the abstract field. In particular:

Rule K: sound speed velocity (celerity) applied is mandatory and must be recorded in the "Abstract" field as below:

- "SOUND VP : XXXX m/s"
  for constant velocity (eg. 1500m/s for Seabeam survey of RV Jean Charcot)
- "SOUND VP : observed "
- "SOUND VP : predicted"
- "SOUND VP : unknown"

Rule L: standard applied for data acquisition, processing and validation must be recorded in the "abstract field "
in free text such as : IHO S44 applied

Rule M: recording the horizontal accuracy (2 sigmas) of DATA positions in dataset in ABSTRACT is recommended. It is mandatory when budget error is carried out while positioning system are not available
Example : HO data

- "DATA POS. ACCURACY :" free text. Please give it in metres for 2 sigmas as far as possible

Rule N: **subsetting of survey data must be indicated in the "abstract" field of CDI in indicating the sampling interval of the subset of soundings :**

- “SUBSET SAMPLING INTERVAL : XXX m”

4.7. Age of data in dataset

In order to give an overview of the evolution of the knowledge of the seafloor topography through time and to select the most up-to-date dataset in areas of fast evolution of seafloor topography, the age is very important.

Rule O: start and end date and time of data in data set are mandatory and must be recorded in CDI.

Note: this information is also needed for composite data product

4.8. Spatial Resolution (composite data product)

This information is required to evaluate the usefulness of the data product related to other sources of data. The domain name of the resolution are specified as “time” (temporal resolution) or “track” (spatial resolution). The resolution is defined by a floating point value and a unit from the common vocabulary list P06. As resolution information is not required, the element gridRepInfo is optional, but once included, all information under gridRepInfo is mandatory.

5. Datasets QA/QC

5.1. Spatial CRS

See DTM reference framework for harmonisation.

Note that datasets in projected coordinates as well as those in a different reference datum must be converted in the target reference framework before transmission to the data integrator .

5.2. Depth reference level

See rule in preceding chapter

5.3. Peaks and biases

Most of the datasets are processed datasets. However datasets with poor documentation related to processing traceability will be checked for peaks and bias before transmission to the data integrator.

5.4. Internal consistency

For the same reason as in 5.3, internal consistency will be checked either using automatic comparison at the intersection of track lines or by visual checks (eg. MBES data)

5.5. External consistency

Identification of major anomalies will be carried by comparing soundings of data sets with a spline surface computed with all selected datasets. If offsets greater than 1% of the water depth are observed, source must be analysed to re-evaluate the quality of the source in function of the information available. Doubtful data must be rejected and DTM reprocessed.

6. DTM QC

It is advised that Regional Coordinators perform a quality analysis and quality checking of the Regional DTMs and provide this in a standard report which will help the EMODnet Integator to perform his integration work and it will assist the Regional coordinator for future updating of the DTM. Aspects of QA – QC are as follows.

6.1. Visual expertise

Plotting the DTM:

- with several sun illuminations;
- depth contours;
- In 3 dimensions;

This will help in identifying remaining artefacts. If the sources of the artefacts are easy to track down, the originating data might be simply removed from the selected soundings and the interpolation recomputed.

6.2. Comparison with external soundings

Randomly selected bathymetric data might be deliberately left aside prior to the gridding phase. Variance of the residuals (gridding DTM – soundings) will be an indicator of the overall DTM creation mechanism.

6.3. Comparison with all the soundings

A thorough analysis of the residuals (gridded data – soundings) will allow evaluation specifically of the quality of the sub-sampling method and the gridding algorithm.

6.4. Comparison with local krigged DTM

Although, the support size (grid cell) will be intrinsically different, vertical comparison with the DTM might be carried outfor vertical consistency (distributions of depth).

6.5. Comparison at overlapping area of regional EMODnet DTM:

To ensure a seamless DTM at the overlapping limits of the regional areas of interest a vertical/horizontal comparison of the DTM will be performed. Values of vertical difference and autocorrelation will be defined as a function of the quality of data source in order both to limit the effect of any shift and to ensure the best coverage

7. ANNEX 1 : glossary

CDI :	Common Data Index (see SeaDataNet)
CRS :	Coordinate Reference System
DTM :	Digital Terrain Model
GEBCO :	General Bathymetric Chart of the Oceans
IHO :	International Hydrographic Organisation
HO:	Hydrographic Office
LAT :	Lowest Astronomical Tide level.
MSL :	Mean Sea Level
QA :	Quality Assurance
QC:	Quality Control