European Topic Centre
Inland, coastal, marine waters

MSFD reporting work flow

Technical guidance on provision of spatial data for MSFD reporting

Date: October 2012

Prepared by / compiled by: ETC/ICM
Organisation: IWRS

Notice

This report was produced by the European Environment Agency and its European Topic Centre of Inland, Coastal and Marine Waters, with the support of *Atkins Danmark a/s* for the specific purpose of support for MSFD reporting.

Technical guidance on provision of spatial data for MSFD reporting

Contents

ACRONYMS

CDR	Central Data Repository at EEA ReportNet
E	East
EEA	European Environmental Agency
ETRS89	European Terestrial Reference System 1989
GES	Good Environmental Status
LAEA	Lambert Azimuthal Equal Area projection
MSFD	Marine Strategy Framework Directive
N	North
NUTS	Nomenclature of territorial units for statistics
RS	Reporting Sheets
WG DIKE	Working Group on Data, Information and Knowledge Exchange
WISE	Water Information System for Europe

1 INTRODUCTION

1.1 General

The Marine Strategy Framework Directive (MSFD) requires that by 15 July 2012 Member States complete their Initial Assessment (Article 8), determination of Good Environmental Status (GES) (Article 9) and establishment of a series of environmental targets and associated indicators (Article 10), (Art. 5(2)), and that these are notified (reported) to the European Commission within three months (Art. 9(2) and Art. 10(2)), that is by 15 October 2012 at the latest. The reporting framework and the reporting content and format has been defined via a set of 'reporting sheets' (RS), which were developed through an informal process in the Working Group on Data, Information and Knowledge Exchange (WG DIKE) under the guidance of the Marine Directors. These are to be electronically submitted to the European Environment Agency via ReportNet.

1.2 Purpose of this document

The document "Guidance for 2012 reporting under the Marine Strategy Framework Directive, using the MSFD database tool" provides additional information for Member States to help complete the reporting sheets, giving specific guidance to support completion of reporting using of the MSFD database tool.. When reporting, Member States will need to clearly define the geographic areas to which the different elements of their reports apply. These 'assessment areas' may equate to regions, subregions, subdivisions or other areas as specifically defined by Member States (called 'assessment areas' in the reporting sheets). Section 5 of the referred guidance document gives further information on MS marine waters, on regional, subregional and subdivision boundaries, on defining assessment areas and on providing spatial data for MSFD reporting, namely on the definition of their unique identifier code – the MarineUnitID.

This document is a complementary guidance document, which provides information on how to deliver spatial data files in Reportnet and also further expands on technical background elements related to the provision of spatial data, in both polygon and grid formats. Basic concepts, as well as methods for manipulation of spatial data to a grid system and manipulation of data within a grid system are also described in Appendix 3.

2 REPORTING SPATIAL DATA ON GEOGRAPHIC AREAS

2.1 Reporting Sheet on Geographic Boundaries and associated spatial data

The reporting sheet on Geographic Boundaries (RS 03_4) covers the reporting of the geographic areas by which Member States will organise the reports of their Marine Strategies for the Directive. It is represented in the MSFD database by two tables:

a. DBTable **MSFD4_GeographicalAreasDescription** outlines the approach taken to delimiting relevant geographic boundaries (*e.g.* legal basis or rationale) for:

   i. the Member State's marine waters;

   ii. any boundaries of regions, subregions and subdivisions within MS waters;

   iii. any additional geographic 'assessment areas' used for reporting under Art. 8, 9 and 10 .

b. DBTable **MSFD4_GeographicalAreasID** is used to provide a list of <u>all</u> assessment areas which are in the rest of the reporting sheets. It includes a unique identifier code (MarineUnitID) which is used in the database to link all subsequent reports (for articles 8, 9 and 10) to a specified assessment area (see **"**Guidance for 2012 reporting under the Marine Strategy Framework Directive, using the MSFD database tool" for details on how to define the MarineUnitID).

The reporting sheet also specifies the two sets of spatial data to be provided:

a. For MS marine waters and related boundaries of regions, subregions and subdivisions (*i.e.* relating to a(i) and a(ii) above);

b. For all assessment areas (*i.e.* related to (b) above).

2.1.1 Reporting format

Assessment area reporting format should be in:

- non-grid format (i.e. polygon) or,
- grid format.

If reported as a polygon dataset, data should be delivered in the spheroid coordinates (decimal degrees) in the geodetic datum ETRS-89 (EPGS code 3035, http://www.epsg-registry.org/). For islands not belonging to the European continental landmass the use of ETRS89 may not applicable. For those areas the WGS84 (World Geodetic System 1984) should be used as the geodetic datum. The following format should be applied for reporting polygon datasets:

a. Geography Mark-up Language (GML): is the Extensible Mark-up Language (XML) grammar defined by the OGC (Open Geospatial Consortium) to express geographic features.

b. ESRI shape file: is a geospatial vector data format for GIS; a valid shape file under MSFD reporting is a set of 4 files: geometry (*.shp), index (*.shx), attributes (*.dbf) and projection (*.prj).

If reported as a grid dataset a text file should be provided where the grid cells are linked to a MarineUnitID.

Additional information on reporting using grid formats is found in section 3.

2.1.2 Metadata

It is not required to report metadata if the EEA reference grid data set is used. The metadata information is only required when non grid data is reported. For the non-grid data the metadata should follow the standards developed under the INSPIRE Directive (2007/2/EC) where the minimum requirements can be found in the INPIRE metadata editor: http://inspire-geoportal.ec.europa.eu/editor/. For datasets related to WISE some additional information should be reported as described in the"Updated Guidance on Implementing the Geographical Information System (GIS) Elements of the EU Water policy, Appendix 11: Implementation of the WISE metadata profile DRAFT REVISION" in a Guidance Document No:22 (document can be found under the name "Draft revision to Appendix 11 of the WISE GIS Guidance (Document No 22) Implementation of the WISE metadata profile" [1] .

2.1.3 Attribute template for polygons

For Assessment areas defined in the database table **MSFD4_GeographicalAreasID** and which are reported as polygons the below attribute information should be provided.

Attribute name	Obligation	Type	Description
MarineUnitID	Mandatory	string (42)	Unique EU code for the assessment area following the guidance in section 3.2.2 Eg. ANS-NL-AA-001; MWE-ES-SD-Alboran

For the spatial dataset related to outline of the MS Marine Waters the below attribute information should be provided.

Attribute name	Obligation	Type	Description
CountryCode	Mandatory	Two-character ISO code	Eg. AT, UK

[1] http://icm.eionet.europa.eu/schemas/dir200856ec/resources

Area type	Mandatory	string (2)	from the following options (enter code only): LD - **MS land outside marine waters (mainland) LM -** MS land within marine waters (island/rocks) (to enable calculation of the area of marine waters) WC - MS waters (water column) SS - MS seabed and subsoil MW - MS marine waters (*i.e.* waters and seabed/subsoil) MS - MS marine waters part of a Region or Subregion SD - MS subdivision
Area status	Mandatory	string (2)	from the following options (enter code only): UA - Undisputed area (of MS marine waters, including agreed Continental Shelf areas) DA - Disputed/unresolved area (of MS marine waters with neighbouring state) CS - Continental Shelf area awaiting CLCS outcome JM - Jointly managed area (with neighbouring state)

**Please notice that it is recommended not to include the mainland and islands in the spatial datasets.

2.2 Uploading spatial data files into Reportnet

Spatial data related to geographic assessment areas should be delivered in Reportnet under the sub-folder named "Geographical data and regional cooperation_MSFD4Geo", created within each MS´s envelope for the reporting obligation "MSFD reporting on Initial Assessments (Art. 8), Good Environmental Status (Art.9), Env. targets & associated indicators (Art.10) & related reporting on geographic areas and regional cooperation"

Here MS should upload the spatial dataset representing all assessment areas, defined in the database table MSFD4_GeographicalAreasID. MS might also report a text file with the relation between the ID of the quadratic cells in a chosen grid system and the MarineUnitID, which should also be uploaded to that same folder.

Filenames protocols

When uploading both spatial files for the non-grid system and the text file related to the grid system in Reportnet it is required that the file naming has the following structure for the marine waters:

[Sub_Collection_Code[1]]_[marinewaters]_[YYYYMMDD]
e.g. BALDK_ marinewaters_20121015

for assessment areas

[Sub_Collection_Code[1]]_[assessmentareas]_[YYYYMMDD]
e.g. BALDK_ assessmentareas_20121015

Please notice that the date should be data for when you upload the files in Reportnet.

[1] Sub Collection Codes can be found on Reportnet under the obligation for the relevant Member State

3 REPORTING BASED ON A GRID SYSTEM

3.1 Common European Grid Reference System

A grid system consist of regular and geo-referenced cells, with a specified form, size and associated properties. Projection systems for grid systems used in the Member States are optimized for the territory of the country. Reprojecting national grids into an European projection system may distort national grid cells. It is therefore recommended Member States prepare grid data based on the proposed Common European grid system.

Following a request of the EEA and of the INSPIRE Implementing Strategies Working Group[1] that recommended the adaptation of a Europe-wide reference grid to facilitate the management and analyses of spatial information (and later INSPIRE geographical grid systems for the EEA reference grid), **the ETRS89-LAEA 52N 10E grid reference system** has been proposed and later recommended for MSFD reporting use by EEA. This reference grid is based on ETRS89 Lambert Azimuthal Equal Area projection, also known in the EPSG Geodetic Parameter Dataset (http://www.epsg.org/) under the identifier: EPSG:3035, with parameters: latitude of origin 52° N, longitude of origin 10° E, false northing 3 210 000.0 m, false easting 4 321 000.0 m. The grid origin is at 0 m N, 0 m E (see other properties in Appendix 2). The Geodetic Datum is the European Terrestrial Reference System 1989 (EPSG:6258).

Being based on an equal area projection, the EEA reference grid is suitable for generalising data, statistical mapping and analytical work whenever a true area representation is required. Recommended grid resolutions are **100 m, 1 km, 10 km and 100 km.** Alternatively, 25 m or 250 m resolution can be used for analysis purposes, where the standard 100 m or 1 km grid cell size is not appropriate. Figure 1 illustrates a sample of the 100 km Grid_ETRS89-LAEA. This sample grid covers the area from 900000 meters East, 900000 meters North to 7400000 meters East and 5500000 meters North. The inset map shows some grid cells for Ireland, together with their cell codes.

EEA reference grid is available in EEA data service as vector dataset that can be downloaded at http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-1. For each EEA member country, and for Europe as a whole, three polygon shape files are made available, according to grid resolutions of 1, 10 and 100 km. For full Europe the 1 km vector grid is excluded for size reasons. The grid extent is such that it covers all EU countries; 7500 columns and 5500 rows for 1 km grid cell size.

[1] 1st Workshop on European Reference Grids placed in Ispra October 2003, http://eusoils.jrc.ec.europa.eu/projects/alpsis/Docs/ref_grid_sh_proc_draft.pdf

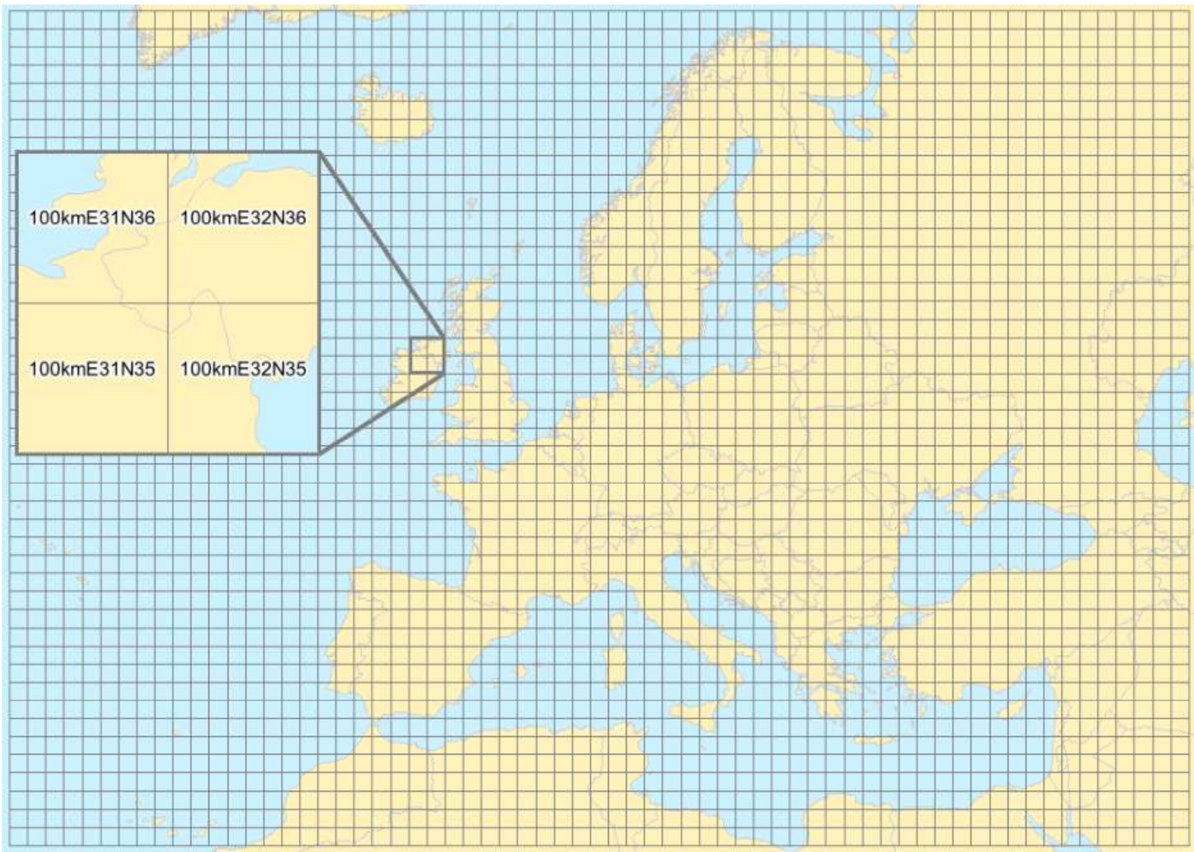


Figure 1: A sample of the 100 km Grid_ETRS89-LAEA

3.2 Coding system for grid cell identifiers

The objective of the cell coding system is to generate unique identifiers for each cell, for any of the recommended resolutions. The cell code is a text string, composed of cell size and cell coordinates. Cell codes start with a cell size prefix. The cell size is denoted in meter (m) for cell sizes below 1000 m and kilometre (km) for cell sizes from 1000 m and above. Examples: a 100 meter cell has an identifier starting with “100m”, the identifier of a 10000 meter cell starts with “10km”.

The coordinate part of the cell code reflects the distance of the lower left grid cell corner from the false origin of CSR (Coordinate Reference System), a coordinate-based local, regional or global system used to locate geographical entities. Easting (E) and Northing (N) values are divided by 10^n (n is the number of zeros in the cell size value). Example for a cell size of 10000 meters: The number of zeros in the cell size value is 4. The resulting divider for Easting and Northing values is 10^4= 10000.

Example #1: Defining a cell code identifier for the 1 kilometre grid
Distance from false origin: 5432000 meters East and 4321000 meters North. Cell size is 1000 meters.
1. Define cell size prefix: 1000 meters => “1km”
2. Identify the number of zeros to remove from Easting and Northing values: The grid size value 1000 has 3 zeros => divide Easting and Northing values by 10^3 = 1000.
3. Define string value based on Easting: 5432000 meters divided by 1000 => “E5432”
4. Define string value based on Northing: 4321000 meters divided by 1000 => “N4321”
5. Concatenate string parts into cell code: “1kmE5432N4321”

Example #2: Defining a cell code identifier for the 250 meter grid
Distance from false origin: 10250 meters East and 220000 meters North. Cell size is 250 meters.
1. Define cell size prefix: 250 meters => "250m"
2. Identify the number of zeros to remove from Easting and Northing values: The grid size value 250 has 1 zero => divide Easting and Northing values by $10^1 = 10$.
3. Define string value based on Easting: 10250 meters divided by 10 => "E1025"
4. Define string value based on Northing: 220000 meters divided by 10 => "N22000"
5. Concatenate strings into cell code: "250mE1025N22000"

The above coding system has been included into the INSPIRE Guidelines on Geographical Grid Systems (http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_GGS_v3.0.1.pdf) as "Requirement 4". The "Requirement 4" says: 'For ensuring compatibility, the grid points of all grids based on ETRS89-LAEA (Lambert Azimuthal Equal Area coordinate reference system is a single projected coordinate reference system for all of the pan-European area. It is based on the ETRS89 geodetic datum and the GRS80 ellipsoid) shall coincide with grid points at Grid_ETRS89-LAEA (Pan-European Grid based on the ETRS89 Lambert Azimuthal Equal Area coordinate reference system with the centre of the projection at the point 52o N, 10o E and false northing: Y0 = 3210000 m, false easting: X0 = 4321000 m). For computations and analysis it might be easier (and faster) to work **with numeric cell identifiers, rather than using the alphanumeric cell codes.** For this purpose, the reference grid shape files include two separate integer attributes named EofOrigin and NofOrigin, representing the Easting and Northing values of the lower left cell corner, in meters.

Properties of the EEA reference grid are given in Appendix 1. In the Appendix 2 an example of codes for 100 km grid size is presented.

3.3 Reporting in a grid environment for the MSFD

In a grid environment there are two possibilities of MSFD reporting regarding the defined assessment area:

1. Assessment area presents one grid cell
2. Assessment area presents a group of grid cells (more grid cells have the same assessment area identification – i.e. same MarineUnitID)

Reporting in grid does not support multi-attribute tables, which means that only one attribute can be related to the dataset. Because of this an attribute table (eg. a text file) must be reported in order to relate the grid id to the MarineUnitID's defined (Figure 2).

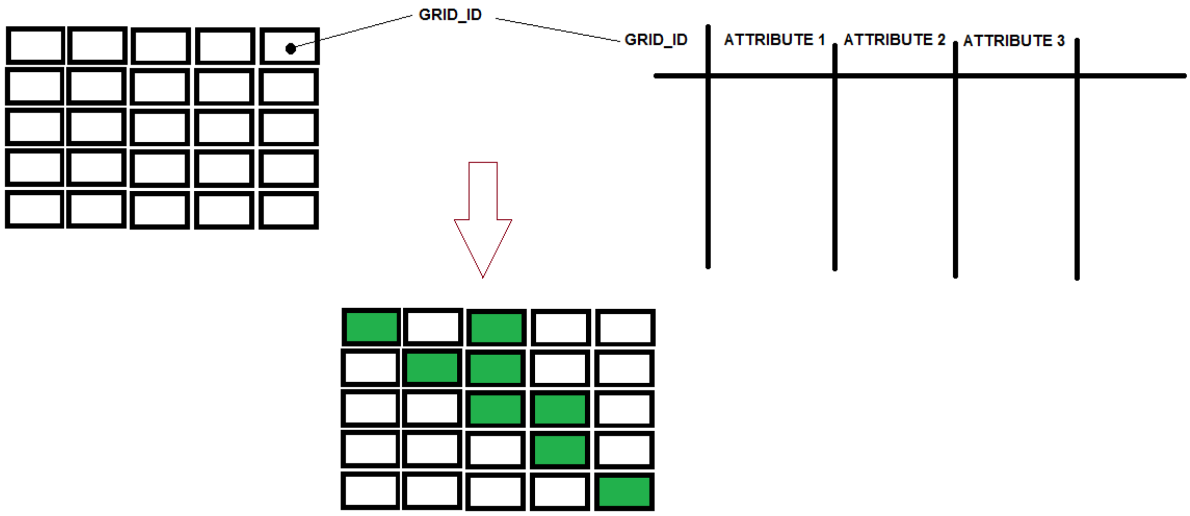


Figure 2: Grid environment and grid attribute reporting

For a boundary grid (grid at the geographical border between two marine regions, subregions, subdivisions or between two Member States) MS makes the decision as to which cells are in its assessment area with indication of method used if appropriate. For a cell or a polygon area and a cell that are between two or more Member States, a common methodology on transboundary assessment and features will be developed at later stage. Possible data handling methodologies for this purpose are explained in Appendix 3.

3.4 Connection between grid data and xml files

An assessment area can be presented with one grid cell or as a group of grid cells. For each assessment area the unique identifier is the MarineUnitID which links the spatial data to the additional information reported through the xml files. Every grid cell has an attribute grid-cell identifier – this is a unique identifier in the attribute table. **If an assessment area includes more than one grid cell – the MarineUnitID should be repeated for all the grid cells defining the assessment area.**

The MS should therefore provide a text file where the grid cell code is linked to a MarineUnitID and should have the following format:

Column name	Obligation	Type	Description
CellCode	Mandatory	string (42)	Grid cell identifiers are accessed in attribute table of the EEA reference Grid. The attribute name is “CellCode” (see figure of attribute table in Appendix 6.3).
MarineUnitID	Mandatory	string (42)	Unique EU code for the assessment area following the guidance in section 3.2.2 Eg. ANS-NL-AA-001; MWE-ES-SD-Alboran Codes MUST be repeated for each grid cell (cellcode) defining the Assessment area.

Each cell in a grid has only one CellCode, but one or many MarineUnitID codes. They all have to be listed in an attribute table of a grid dataset. An example:

CellCode	MarineUnitID
250mE1025N22000	XXX-MS-T1-001
250mE1025N22000	XXX-MS-T1-002
100mE1025N22000	XXX-MS-T2-050
100mE1025N22000	XXX-MS-T2-052
100mE1025N22000	XXX-MS-T3-052
500mE1025N22500	XYY-MS-T1-002
500mE1025N22500	XYY-MS-T2-005
etc	etc

4 LITERATURE

Annoni, A.,Luzet,C., Gubler, E., Ihde, J.(ed.). 2003. Map projections for Europe, Url.: http://mapref.org/LinkedDocuments/MapProjectionsForEurope-EUR-20120.pdf

Disaggregation of socioeconomic data into a regular grid and combination with other types of data. March 2011. Technical report. Espon. EU. Url.: http://database.espon.eu/getProject;jsessionid=42ea4d2382396fb3b3549035cdce?proj=6&file=r&cat=reports

Gallego J. and Peedell S. 2001 Using CORINE land cover to map population density. In Towards agri-environmental indicators. Integrated statistical and administrative data with land cover information. Copenhagen: European Environment Agency, pp. 94-105.

Gallego F.J., 2010 A population density grid of the European Union. Population and Environment 31, 460-473.

http://stats.oecd.org/glossary/detail.asp?ID=4337

http://www.integrated-assessment.eu/guidebook/spatial_disaggregation

http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/documents/ESSnet%20project%20GEOSTAT1A%20final%20report_0.pdf

MSFD 2012 reporting – guidance; MSCG 7/2012/5c. Prepared by DG Environment and MRAG/UNEP-WCMC/URS; 03/05/2012.

Peifer, H. About the EEA reference grid. September 2011. Url.: http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-1EEA

Short Proceedings of the 1st European Workshop on Reference Grids, Ispra, 27-29 October 2003. Url.: http://eusoils.jrc.ec.europa.eu/esdb_archive/etrs_laea_raster_archive/ref_grid_sh_proc_draft.doc

5 APPENDICES

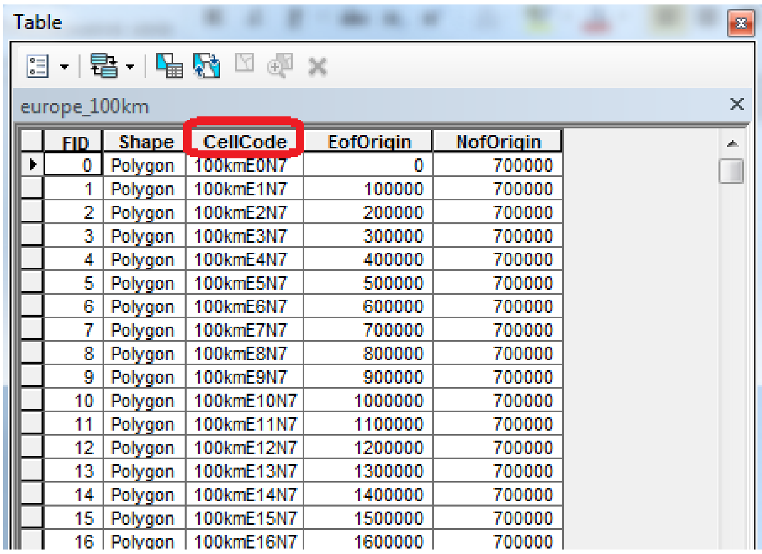
Appendix 1: Properties of the EEA reference Grid

The following properties were identified for the proposed Grid:

1. The Grid should be based on the projection system Lambert azimuthal Equal Area (ETRS-LAEA).
2. Centre of projection: 52N, 10E
3. False Easting: 4321000.0 m, False Northing: 3210000.0 m.
4. Precision: the significant number of digits to be used: Origin for the projection should have an exact location (mm precision), in practical work and operation of the grid and its coordinates number of digits should be at decimeter or meter accuracy.
5. Resolution of grid. Three different options were considered, but 1 km cell size was selected.
6. Metric: 1, 10, 100, 1000, 10.000, 100.000
7. The quad tree approach, 1, 2, 4, 8… gives a higher number of levels.
8. A mixed approach where e.g. the main focus is on the metric, but where one allows the 250 and 500 m cells
9. Grid centre point: LAEA 52N 10E
10. Grid origin: is defined as the south-western corner of the area of validity. This has been stipulated to be 4321000.0 m west of centre point of the projection (52N, 10E), and 3210000.0 m south of projection centre point (52N, 10E). The location of the grid origin far to the west and south would mean that negative values in the coordinates are avoided in all foreseeable uses. The extent is large enough to cover the areas of interest. The false easting and northing values do not have to be moved.
11. Grid shape: rectangular and not hexagonal, more precisely square. The square shape will appear when used in the defined projection, smaller or larger distortions will appear when re-projected to other projections.
12. Grid orientation: south-north, west-east

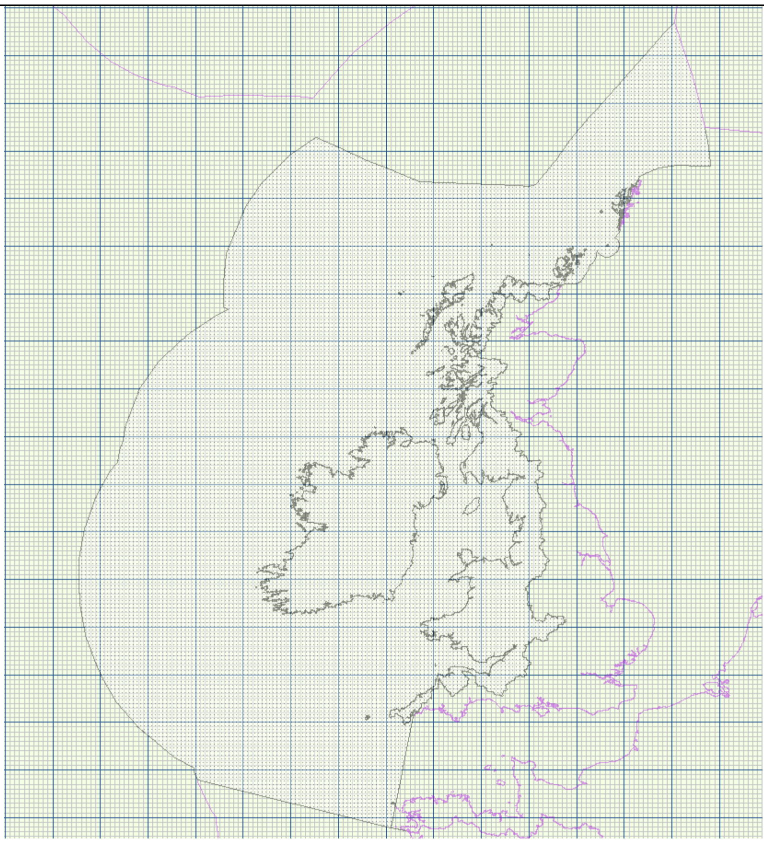
Appendix 2: Grid cell identifiers in EEA reference Grid

Grid cell identifiers are accessed in attribute table of the EEA reference Grid. The attribute name is "CellCode"

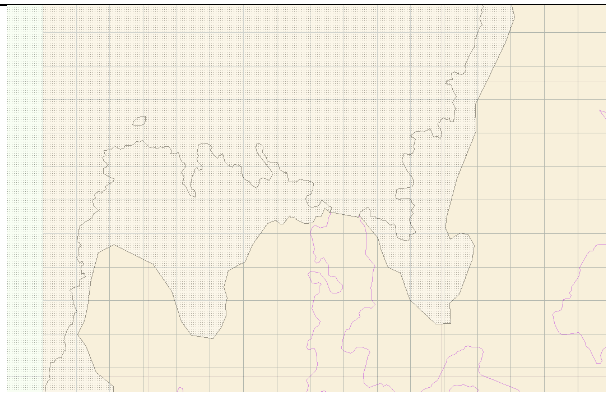


FID	Shape	CellCode	EofOrigin	NofOrigin
0	Polygon	100kmE0N7	0	700000
1	Polygon	100kmE1N7	100000	700000
2	Polygon	100kmE2N7	200000	700000
3	Polygon	100kmE3N7	300000	700000
4	Polygon	100kmE4N7	400000	700000
5	Polygon	100kmE5N7	500000	700000
6	Polygon	100kmE6N7	600000	700000
7	Polygon	100kmE7N7	700000	700000
8	Polygon	100kmE8N7	800000	700000
9	Polygon	100kmE9N7	900000	700000
10	Polygon	100kmE10N7	1000000	700000
11	Polygon	100kmE11N7	1100000	700000
12	Polygon	100kmE12N7	1200000	700000
13	Polygon	100kmE13N7	1300000	700000
14	Polygon	100kmE14N7	1400000	700000
15	Polygon	100kmE15N7	1500000	700000
16	Polygon	100kmE16N7	1600000	700000

Attribute table of EEA reference grid of 100 km size.



Celtic sea sub-region and grids sizes 10km and 100 km



An example of 1 km grid size

Appendix 3: Data handling methodology

AGGREGATION

Aggregation is the process of grouping spatial data at a level of detail or resolution that is coarser than the level at which the data were collected.

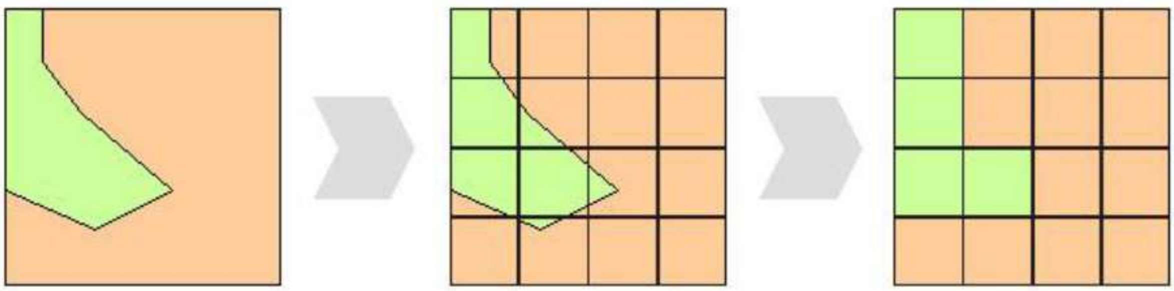
The outcome from aggregation is always the same: There is a loss of spatial and attribute detail through the creation of coarser spatial data consisting of fewer observations. While such data may be a desired outcome for some tasks, this is not always the case. In many instances, it is necessary to work with aggregated spatial data simply because they are the only data available for the task at hand—in other words, there is no choice in the matter. This is especially true when relying on governmental data products such as a national census. Geographic information systems (GIS) not only facilitate aggregation through a variety of techniques, they can also be used to evaluate issues pertaining to the use of aggregate data.

Aggregation methods for polygon data

For transformation from small cells to large cells in a different grid system or to polygons (geographic units), simple methods, such as adding small cells with the centre inside the new cell or polygon, or simple areal weighting (allocation proportional to area) are acceptable. National organizations should be encouraged to collect and stock geo-referenced data and with the maximum possible geographic detail.

MAXIMUM AREA CRITERIA

Maximum area criteria method is a method where the cell takes the value of the unit which covers most of the cell area. It should be a good option for uncountable variables.



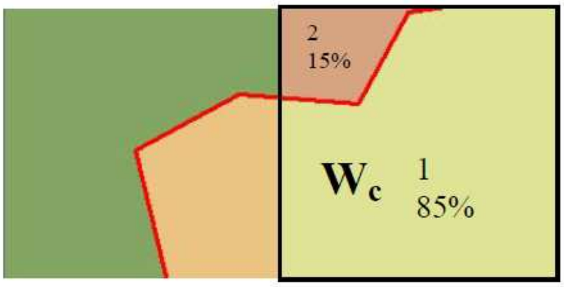
PROPORTIONAL CALCULATION

Proportional calculation method is a method where the cell takes a calculated value depending on the values of the units falling inside and their share within the cell. This method seems very appropriate for countable variables.

Cell value = $\Sigma (V_i * Share_i)$
Where: V_i = Value of unit i
$Share_i$ = Share of unit i within the cell
In the example: $V_1 * 0.85 + V_2 * 0.15$

ROPORTIONAL AND WEIGHTED CALCULATION

Proportional and weighted calculation method is a method where the cell takes also a proportionally calculated value, but this value is weighted for each cell, according to an external variable (e.g. population).



Cell value = $W_c \Sigma (V_i * Share_i)$
Where: V_i = Value of unit i
$Share_i$ = Share of unit i within the cell
W_c = weight assigned to cell c
In the example: $W_c * (V_1 * 0.85 + V_2 * 0.15)$

Whenever it is possible this method should be used to disaggregate the data, as it gives an added value to the source data, providing more interesting results when different data are put together on a cell-by-cell basis.

Aggregation methods for point data

For transformation of point data to data in a certain grid cell the first step is to define which point data affect the value of a defined grid cell. Afterwards they are some methods of aggregating these data for the whole grid cell area. There are generally three types of distribution of point features: random, uniform and clustered. Clustered distribution pattern is expected in the reporting data process. Standard deviation shows how much variation or dispersion exists from the average (mean) value. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values.

AVERAGE OF ORIGINAL VALUE

Average of original value is the average value of all the points selected to be aggregated. In case of grid cells means all values of points on the selected grid cell.

MEDIAN ORIGINAL VALUE

Median original value is the middle occuring value of all points selected to be aggregated. If there is an even number of points, the median value is the average of the two middle values.

DISAGGREGATION

Disaggregation is the breakdown of observations, usually within a common branch of a hierarchy, to a more detailed level to that at which detailed observations are taken (http://stats.oecd.org/glossary/detail.asp?ID=4337).

Spatial disaggregation or downscaling is the process by which information at a coarse spatial scale is translated to finer scales while maintaining consistency with the original dataset (http://www.integrated-assessment.eu/guidebook/spatial_disaggregation).

Disaggregation methods

Although disaggregation can be made from statistical data on administrative units (NUTS) to a finer grid, data computation bottom-up (from detailed data to aggregated) is always preferable.
If data are available only for large geographical units (large cells or polygons), re-allocation of data by methods such as simple area weighting should be avoided. The search of suitable co-variables (proxies) is recommended. For some data (e.g. data linked to biodiversity concerns: species and habitats/ecosystems) only available for large units, defining co-variables is difficult. In this case, analysis should be adapted to these units.

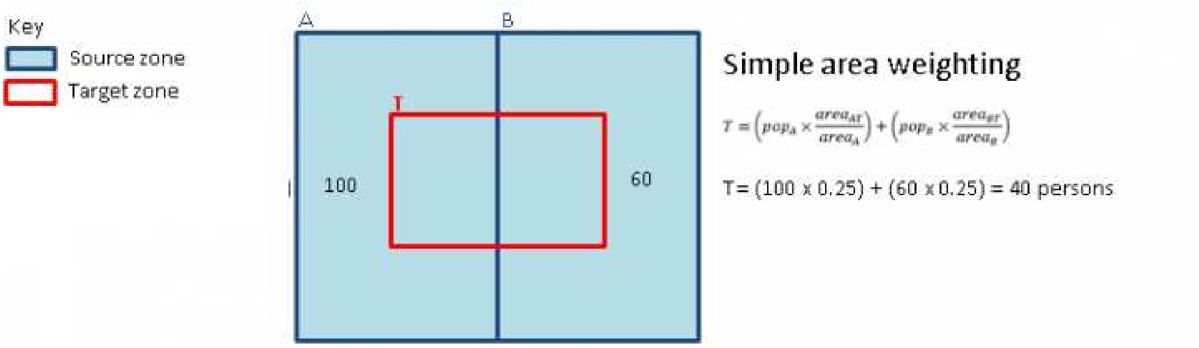
SIMPLE AREA WEIGHTING

Mass-preserving areal interpolation (i.e. area-weighting) redistributes aggregated data based on the proportion of each source zone that overlaps with the target zone according to the following equation:

$$P_t = \sum_{s=1}^{S} \frac{P_s \times A_{ts}}{A_s}$$

where: P_t is the population in target zone t; P_s is the population in source zone s; As is the area of source zone s; and A_{ts} is the area of target zone t overlapping source zone s.

While area-weighting ensures that the total from the source data remains unchanged, it is based on the often incorrect assumption that the phenomena of interest are evenly distributed across the source zones. Population is one such example. Most populations, including that of the EU, are rarely uniform across census tracts, and instead tend to be highly clustered in urban centres surrounded by areas of dispersed rural homesteads.

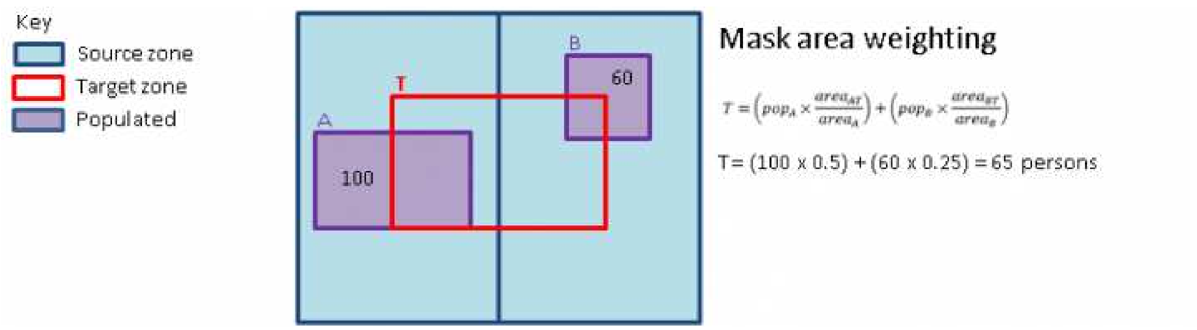


MASK AREA WEIGHTING

Mask area weighting is an improvement on simple area weighting in that it uses a mask to define where, within the target zone, the source data should be allocated. This process is like binary dasymetric mapping, in which each source unit is divided into two sub-regions, e.g. populated and unpopulated – and the information is then allocated only to the populated areas. Land cover can be used to identify populated areas and create the mask. The equation is as follows:

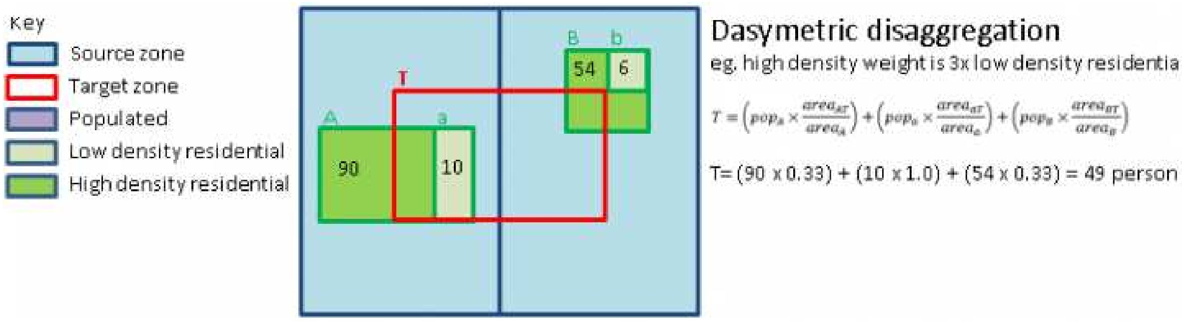
$$P_t = \sum_{s=1}^{S} \frac{P_s . A_{tsp}}{A_{sp}}$$

where: A_{tsp} is the area of populated land that overlaps between the target map unit t and source map unit *s*; and A_{sp} is the area of target map unit *s* that is populated land.



DASYMETRIC DISAGGREGATION

Dasymetric disaggregation is a type of areal interpolation that incorporates ancillary data to facilitate the areal interpolation process. It derives the boundaries from the actual spatial distribution of the variable being mapped (and not on administrative units as it is a case in choropleth mapping). As mentioned in relation to mask area weighting, land cover data, in particular, offer a means by which "residential" areas can be distinguished from "non-residential" areas. Dasymetric disaggregation, however, is an improvement on mask area weighting in that two or more categories can be assigned weights for disaggregation. This is referred to as polycategorical dasymetric disaggregation. The challenge in dasymetric disaggregation thus involves devising an appropriate set of weights that can be applied to the land parcels (or other ancillary data) to reflect population density. Weights may be defined using selective sampling (see Stochastic allocation) or regression analysis.



STOCHASTIC ALLOCATION

Gallego and Peedell (2001) and Gallego (2010) describe a stochastic allocation process by which weights are devised for disaggregating population totals from larger administrative units (NUTS 2 regions) to smaller ones (communes) on the basis of the land cover information. Communes were first stratified, by comparing the commune population density to the average density of the surrounding NUTS 2 region, into one of three levels reflecting population density (i.e. dense, less dense and no urban). The method then involved disaggregating the NUTS 2 totals using an initial set of weights, re-aggregating the population to the commune level and comparing it to the known total, computing a disagreement indicator, and adjusting the weights to reduce the disagreement. The iterative nature of this method involves effort in model tuning to achieve suitable weights. The result is a dasymetric population density grid for the European Union.