

Data Management Plan (v.03)

Deliverable 5.6

GES4SEAS WP5: Learning sites: Testing and synthesis
Task 5.1 Stocktaking of pressures and impacts



Funded by
the European Union



UK Research
and Innovation

Document Summary

Topic	HORIZON-CL6-2921-BIODIV-01-04
Grant Agreement Number	101059877
Start date of the project	01.09.2022
Project Coordinator	Angel Borja
Work Package	WP5
WP Co-Leaders(s)	Aarhus University and Thünen Institute
Deliverable Title	Data Management Plan (v.03)
Deliverable Number	5.6
Deliverable Beneficiary	Aarhus University
Authors	Andreas Holbach, Deevesh Ashley Hemraj, Jesper Harbo Andersen, Torsten Berg, Silvana Birchenough, Angel Borja, Giulio Farella, Tomaso Fortibuoni, Annalisa Minelli, Samuli Korpinen, Luminita Lazar, Miguel Leal, Ciarán Joseph Murray, Miquel Ortega, Nadia Papadopoulou, Debbi Pedreschi, Riikka Puntila-Dodd, Sofia Reizopoulou, Yolanda Sagarminaga, Alina Spinu, Jacqueline Tamis, Ibrahim Tan, Heliana Teixeira, Fábio L. Matos, Christina O'Toole, Jacob Carstensen
Dissemination level	Public
Submission date	31 st August 2025

How to cite | Holbach A., Hemraj D.A., Andersen J.H., Berg T., Birchenough S., Borja A., Farella G., Fortibuoni T., Minelli A., Korpinen S., Lazar L., Leal M., Murray C.J., Ortega M., Papadopoulou N., Pedreschi D., Puntila-Dodd R., Reizopoulou S., Sagarminaga Y., Spinu A., Tamis J., Tan I., Teixeira H., Matos F.L., O'Toole C., Carstensen J., 2025. *Data Management Plan (v.03)*. *GES4SEAS Deliverable 5.6*. 25 pp.

GES4SEAS (Achieving Good Environmental Status for maintaining ecosystem services, by assessing integrated impacts of cumulative pressures) project is funded by the European Union under the Horizon Europe program (grant agreement no. 101059877). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or UK Research and Innovation. Neither the European Union nor the granting authority can be held responsible for them.

Contents

Document Summary	2
List of Figures	4
List of Tables.....	4
1. GES4SEAS project summary.....	5
2. Data summary	7
2.1. Data levels and data flow	8
2.2. Data sources.....	10
2.3. Data size	12
2.4. Data utility.....	13
2.5. Relevant legal framework and policies	13
2.6. Data governance and responsibilities	13
2.7. Informed consent template	14
3. FAIR data	14
3.1. Making data findable, including provisions for metadata	14
3.2. Making data openly accessible	15
3.3. Making data interoperable	15
3.4. Increase data reuse (through clarifying licenses).....	17
3.5. Data Publication Strategy and Repository Transition	17
4. Allocation of resources.....	18
5. Data security.....	19
6. Ethical aspects	20
7. Other issues	22
8. References.....	23
Appendix 1: informed consent template	24

List of Figures

Figure 1. Schematic illustration of data-flow (black arrows) within the GES4SEAS project. Dashed arrows symbolize partial data-flows. HABs: Harmful Algal Blooms..... 8

List of Tables

Table 1. Summary of estimated data layers, spatial resolution and file size for the GES4SEAS Learning Sites. Entries marked with ‘*’ are initial guesses and are to be defined during the project. For LS11 estimates cannot be provided yet at this early stage of the project. 12

1. GES4SEAS project summary

Human activities at sea (e.g., maritime transport, extraction of living and non-living resources, energy infrastructures, etc.) and coastal areas (e.g. urbanization, agriculture, leisure and recreation, etc.) have expanded considerably, leading to an increased level of pressures and subsequent degradation of ocean health and, ultimately, human health. Single and cumulative impacts of these activities are likely to increase, driven by human demands and enhanced by climate change.

Human activities evolve following socio-economic drivers leading to pressures, which often are studied in isolation from each other even though their impacts on marine ecosystems can interact, making the effects cumulative (e.g. synergistic, antagonistic or a combination). Knowledge on these interactions and their cumulative effects in the marine environment has increased in recent years, but huge challenges still remain to be solved. Thus, there is little predictability with which to inform decision-making processes, especially on ecological tipping points, which, if exceeded, could lead to a point of no-return for the system. In this context, an ecosystem-based management (EBM) approach to the management of human activities at sea and on land should ensure that the combined pressure of such activities is kept within levels compatible with the requirements of Good Environmental Status (GES) (within the Marine Strategy Framework Directive – MSFD), against a background of climate change. This means that the capacity of marine and coastal ecosystems to respond to human-induced changes is not compromised, enabling the sustainable use of marine goods and services by present and future generations.

Thus, the main objective of GES4SEAS is to inform and guide marine governance in minimizing human pressures and manage their impacts on marine biodiversity and ecosystem functioning, while maintaining the sustainable delivery of ecosystem services. This will be achieved through developing an innovative and flexible toolbox, tested, validated, demonstrated and upscaled, in the context of an adaptive EBM approach. The toolbox will allow competent authorities to assess and predict the effect of multiple stressors (including climate change) and pressures from human activities, at the national, sub-regional, regional and European level. Ultimately, this will ensure they achieve GES, and support different policies at national, European and global levels (e.g. Birds and Habitats Directives (BHD), Biodiversity Strategy 2030, Maritime Spatial Planning Directive (MSPD), United Nations Sustainable Development Goals (SDG)).

Stakeholders and the key competent authorities (including national, regional and European levels) are integrated in a Practitioner Advisory Board (PAB) to co-create and validate the toolbox and the EBM approach. This will result in a real problem-solving approach with iterative and incremental development steps.

GES4SEAS will also rely on existing multi-actor networks to involve and engage with stakeholders. This multi-actor approach will ensure that the research and deliverables are relevant to marine managers all around the world. Lastly, it is important to highlight that the toolbox will be tested and demonstrated at 11 Learning Sites (LSs) covering all European regional seas (and also overseas), and environments. Thus, it is expected that GES4SEAS will achieve Technological and

Societal Readiness Levels 6. This will be achieved by the participation of 20 partners, covering the four European regional seas and Canada.

It is expected that GES4SEAS will:

- Operationalize integrative and holistic solutions for EBM, based upon a software toolbox for analyzing, assessing and mapping cumulative pressures, GES and ecosystem services.
- Provide evidence (and training) to key stakeholders of the benefits of using the toolbox that will be developed in GES4SEAS for assessing the environmental status of marine waters and the ecosystem services considering the effects of multiple pressures so opt for using it.
- Ensure the EBM approach and guidelines for the management of Invasive Alien Species (IAS), Harmful Algal Blooms (HABs) and jellyfish, the approach for monitoring top predators are used by end-users.
- Investigate, using models, the best ways to obtain thresholds of GES/non-GES status and tipping points (system breaking points).
- Reach and engage a wider society, and specifically young generations and educators, on key messages steaming from this project, so GES4SEAS contributes to societal ocean literacy and responsible behaviours.

2. Data summary

This Data Management Plan will be the guide to be followed by all GES4SEAS partners during the life of the project, as well as for its legacy, regarding the collection, processing, storage, use and publication of data, ensuring proper management of any existing, obtained or developed data in GES4SEAS. This will be a 'blue print document', which will be updated regularly during the life of the project, especially for each 18-months reporting period.

To achieve the objectives of the project described in the previous section, GES4SEAS will make use of existing datasets on human activities, related pressures (including climate change), ecosystem components, ecosystem impacts and state changes, and corresponding ecosystem services. These datasets will have to be collected, quality controlled, possibly processed, and harmonized in order to build and test the toolbox. The toolbox will relate all these components across the complex underlying network of relationships between the different levels. This will be done for a representative set of eleven so-called Learning Sites (LS) distributed all across European seas and one site in French Polynesia. In particular, GES4SEAS aims at describing and understanding complex and cumulative multi-variate impacts of various pressures on marine ecosystems, ecosystem services, and resulting societal benefits.

GES4SEAS research and development will mostly build upon pre-existing datasets that will be used for:

- empirical analysis and modelling to parametrize the complex network of relationships and interactions between pressures, impacts and state changes, ecosystem services and societal benefits supply, and management measures (WP3),
- development of the individual tools and the integrated toolbox,
- simulating scenarios with the toolbox developed,
- defining new and testing existing thresholds set by policy (e.g. spatial extent thresholds by MSFD),
- identifying and testing tipping points between major relationships (e.g. impacts on ecosystem service supply),
- documenting the current 'state of play', gaps and improve management of invasive species, Harmful Algal Blooms (HABs), jellyfish outbreaks and top predators' distribution.

Each task and LS will require specific data to cover information on: (i) human activities and pressures, (ii) climate change and extreme events, (iii) ecosystem components (state, vulnerability, resilience to different pressures, ecosystem services provided), and (iv) measures

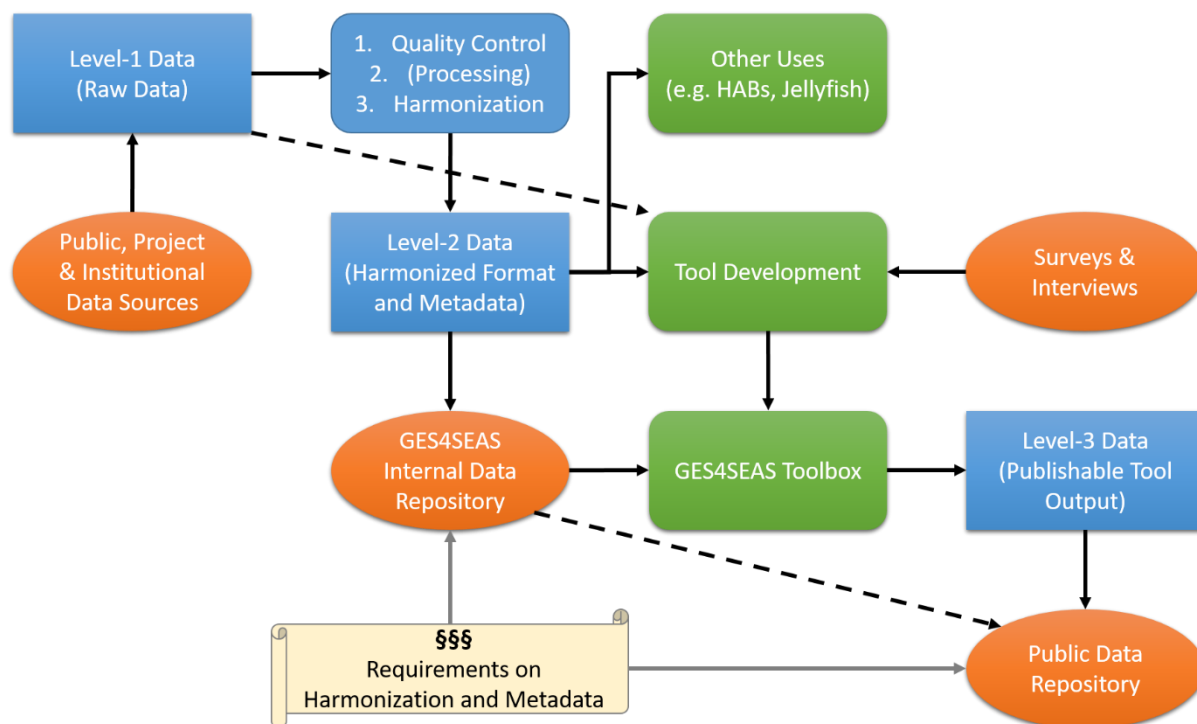


Figure 1. Schematic illustration of data-flow (black arrows) within the GES4SEAS project. Dashed arrows symbolize partial data-flows. HABs: Harmful Algal Blooms.

(source and impact on pressures or states), with the most suitable spatio-temporal extent, resolution, and quality.

2.1. Data levels and data flow

GES4SEAS is dealing with different levels of data throughout the project. Figure 1 schematically illustrates the data flow within GES4SEAS as we see it at the initial phase of the project.

Level-1 data

Raw spatial data from public, other project, and institutional data sources available to the different LSs/institutions involved will be classified as “level-1”. Level-1 data requires **quality control**, will potentially require data **processing** steps, and needs **harmonization** before being “promoted” to “level-2”. The individual LSs/institutions, in accordance with their requirements for data handling and management, shall manage this level-1 data.

GES4SEAS will primarily collect and generate spatial distribution data. The raw input data will consist of different formats of spatial data describing distributions of variables as spatial points, polygons, meshes, or pixels. Therefore, we will have to handle a number of common spatial data formats, such as rasters/grids (e.g. in Geopackage, NetCDF and GeoTIFF formats) describing variables across regularly distributed pixels and vector data (e.g. Geopackage or ESRI shape files).

Quality Control: Some level-1 datasets will be well-documented, having been subject to relatively rigorous checks and analysis before submission to GES4SEAS. In such cases, a simple note describing the lineage of the data will suffice as quality control. This will be the case for certain

standardized/harmonized datasets from certain public or other project-related repositories. Other level-1 datasets will be more unrefined and require more extensive quality checks. Suitable methods for quality control will have to be defined and documented case-by-case by the LSs with respect to the specific type of content and origin of the dataset.

Processing: Some level-1 datasets will require data processing. Processing could include, i.e. conversion of units, conversion between parameters, spatial and/or temporal aggregation, spatial interpolation, etc. Suitable processing steps and methods will have to be defined and documented case-by-case by the LSs with respect to the specific requirements and research questions for each LS.

Harmonization: Level-1 data will be of very different formats and manifold structures, thus only of limited use to other parties. Regardless of other pre-processing steps required, spatial data will be converted to a common data format. In GES4SEAS, we aim for the Geopackage format as a wrapper for the different types of spatial datasets into a platform-independent, portable, self-describing, and compact format. Each dataset will represent a certain point or range in time, so several spatial layers for the same variable might also be available and used/generated.

For all European LSs, harmonization will also include the re-projection of the spatial data onto a common coordinate reference system (CRS). All European LSs will use the projection of the EEA reference grid (ETRS89-LAEA Europe, EPSG:3035). For LS11 in French Polynesia, a suitable common CRS for the re-projection of level-1 datasets will have to be defined during the project.

Level-2 data

The quality-controlled, pre-processed, and harmonized datasets, derived from all level-1 datasets, will, for each LS, lead to a specific set of spatial data on pressures, impacts, ecosystem components, and ecosystem services. These different sets of level-2 data gathered from various sources (see below) will be used to develop the individual tools that are going to form elements of the final toolbox. Likewise, level-2 data will also be used for exemplary applications of the final toolbox with respect to LS-specific research questions. Level-2 data will be stored as a respective zip-archive on the GES4SEAS Zenodo repository. This zip-archive will contain folders for each LS, as well as a central register of all the available datasets and metadata. The detailed structure of data storage for the LSs will be developed in close coordination with WP4, as the toolbox is meant to be able to directly access and import data from the repository.

Linkage matrices and frameworks

In addition to the spatial datasets described above, linkage matrices and/or frameworks connecting the different categories (activities, pressures, ecosystem components, ecosystem services) of spatial data will be generated in GES4SEAS through, e.g. sensitivity scores, non-linear response functions, impact chains, etc. The values used within these linkage datasets will be based on either science-based expert judgement or data-driven analyses. The exact format of this type of data and how it will be implemented in the toolbox is still subject to the development under the different activities and will be defined as the project progresses.

GES4SEAS toolbox and level-3 data

GES4SEAS aims to develop the toolbox using a representative set of eleven LSs. The formerly described quality control, processing, and harmonization will facilitate an efficient development of the individual tools and the toolbox, as well as cross-LS comparative studies.

The GES4SEAS toolbox will be applied by accessing the harmonized level-2 datasets from the project-internal data repository to generate output datasets, classified as “level-3”. The specific contents and formats of the generated output data of the GES4SEAS toolbox will need to be defined in close coordination between the involved WPs (especially WP4) and partners within the course of the project and cannot be specified in detail yet.

The final version of the GES4SEAS toolbox is meant to include a tool able to handle un-harmonized level-1 data, as well. So, the toolbox should finally adopt data harmonization procedures developed for data management in WP5 and could then also be applied to the generation of harmonized level-2 data by the end-users. Hence, un-harmonized level-1 data will also contribute to the development of a data management tool inside the GES4SEAS toolbox.

In addition, qualitative assessments (i.e. opinions and expert judgements) and quantitative data will be obtained in WP1 through surveys and interviews from the Practitioners Advisory Board (PAB), stakeholders and participants in GES4SEAS workshops and meetings. This information will substantially contribute to the development of the layout, capabilities, and individual features of the GES4SEAS toolbox. Level-3 data will be stored in a respective folder structure on the GES4SEAS Zenodo repository.

2.2. Data sources

As far as possible, GES4SEAS research will build upon pre-existing datasets. Only in cases where it appears necessary in the course of the project will new input datasets be created.

Many datasets are already available from several open data platforms maintained by international and EU thematic platforms and data centres, national or institutional data centres, and portals that are largely enumerated and described in the European catalogue of marine data and information portal. GES4SEAS will also look for datasets created within previous research projects, datasets created and published as outputs from scientific research, and web portals created under civil society initiatives as and when needed and suitable for the task’s purposes. The main data sources to be used are listed below:

- EMODnet (biology, seabed habitats, chemistry, physics, human activities, geology, bathymetry);
- Copernicus (marine, climate change, land, atmosphere, emergency and security/maritime surveillance), and Joint EMODnet and Copernicus Marine Portfolio for the MSFD descriptors;
- Eionet’s Reportnet infrastructure: data reported by EU MS under different directives, i.e. BHD, Urban Waste, Nitrates Directives, etc., including measures set up;
- EEA data centre (= WISE Marine);
- the four RSCs;

- ICES Data Centre: with commercial fish catches and environmental data;
- JRC Data Collection Framework, including fisheries data and reports by STECF - Scientific, Technical and Economic Committee for Fisheries;
- Biodiversity Information System for Europe (BISE);
- Ocean Biodiversity Information System Spatial Ecological Analysis of Megavertebrate Populations (OBIS-SEAMAP);
- Protected Planet website, includes the World Database on Protected Areas (WDPA), Global Database on Protected Area Management Effectiveness (GD-PAME);
- Information system on aquatic non-indigenous and cryptogenic species (AquaNis);
- Harmful Algal Event Database (HAEDAT);
- HABMAP-OBIS, geographic range of HAB species;
- Jellyfish Database Initiative (JeDI);
- Jellywatch, sightings of jellyfish and other organisms;
- EUROSTAT Data portal: EU statistics of socio-economic indicators;
- Geographic Information System of the Commission (GISCO), EU statistics of socio-economic indicators;
- National data infrastructures, e.g. Spanish Marine Information System (INfomar), the Italian National Oceanographic Data Centre (NODC);
- Partner data infrastructures (e.g. CEFAS Data Hub; Sistema Informativo Centralizzato Dati di Monitoraggio MSFD);
- ClimateFish: Abundance of selected coastal fish species as candidate indicators of climate change in the Mediterranean Sea (<https://www.seanoe.org/data/00756/86784/>);
- GBIF | Global Biodiversity Information Facility;
- FAO fisheries and aquaculture statistics (<https://www.fao.org/fishery/statistics-query/en/home>);
- Global Fishing Watch;
- GEBCO (General Bathymetric Chart of the Oceans), providing the most authoritative, publicly available bathymetry data sets for the world's oceans;
- T-MEDNet, an initiative devoted to developing an international observation network on climate change effects in marine coastal ecosystems by spreading the acquisition of standard monitoring protocols on different indicators over a large scale and long-term;
- SeaDataNet, a distributed Marine Data Infrastructure for the management of large and diverse sets of data deriving from *in situ* of the seas and oceans;
- Marine Geoscience Data System (MGDS), a trusted data repository that provides free public access to a curated collection of marine geophysical data products and complementary data related to understanding the formation and evolution of the seafloor and sub-seafloor;
- FAO Vulnerable Marine Ecosystems Database: A global inventory of fisheries measures to protect vulnerable marine ecosystems in areas beyond national jurisdiction;
- MAPAMED, the database of MARine Protected Areas in the MEDiterranean;
- The OSPAR Data and Information System (ODIMS), an online tool providing a single point of access to all the data and information gathered through OSPAR's Joint Assessment and Monitoring Programme across the different thematic work areas of the Convention of the North-East Atlantic;
- The ACCOBAMS Survey Initiative (ASI) project repository, implemented by the ACCOBAMS Permanent Secretariat, to collect data on marine wildlife (cetaceans and other marine

megafauna species) and anthropic pressures thanks to regional surveys conducted throughout the Mediterranean and the Black Sea.

- Any other source (e.g. MBON, GEOSS, GFCM) if necessary;
- The nautical maps of the Italian Hydrographic Marine Institute.

Each LS will address specific research and application questions to be answered with the toolbox and represent different types and sizes of marine ecosystems. The availability of level-1 data will differ greatly in the number of variables and the corresponding degree of detail between the LSs. Furthermore, several general datasets available with great spatial coverage will be reused across the LSs.

Currently, we estimate to generate ~405 level-2 datasets (Table 1), which would equal at least the same number of level-1 datasets. For the individual LSs, however, it is not yet clear at this stage of the project which level-1 datasets will be applied and promoted to level-2.

2.3. Data size

The expected total size of the input level-2 data is hard to estimate beforehand. As a first attempt, we generated a conservative estimate based on the assumption that all datasets will consist of two layers (data and uncertainty) in uncompressed GeoTIFF format with 16-bit depth, and we will generate two versions of each dataset in the course of the project. From each LS, we included the approximate number of layers to be included, as well as the size and intended spatial working resolution for the study area (Table 1). Based on this rough first assessment, we estimate a total file size of level-2 input data for LS1 to LS10 in the order of magnitude of 350 MB of data. Estimates for LS11 and level-3 data have not yet been included in this estimate. Therefore, we estimate a total file size of GES4SEAS level-2 and level-3 datasets of ~500 MB.

Table 1. Summary of estimated data layers, spatial resolution and file size for the GES4SEAS Learning Sites (LS). Entries marked with '' are initial estimates and are to be defined during the project. For LS11 estimates cannot be provided, yet, at this early stage of the project.*

Learning Site		No. Activities	No. Pressures	No. of Ecosystem components	No. Ecosystem Services	LS Area [km ²]	Spatial Resolution [kmxkm]	Estimated file size [MB]
LS1	Baltic Sea (Gulf of Bothnia)	9	15	30	2	104 000	1x1	47
LS2	Baltic-Atlantic (Kattegat)	6	7	9	6	30000	0.5x0.5	27
LS3	Atlantic (North Sea - Netherlands)	12	12	4	11	750000	28x28	0.30
LS4	Atlantic (Celtic Sea)	5	6	26	3	200000	1x1*	62
LS5	Deep Atlantic - Mediterranean transition	5	8	7	12	12000	1x1	3.2
LS6	Western Mediterranean	10	9	9	7	846000	5x5	9.5
LS7	Adriatic Sea	15	12	21	2	60500	1x1	17

LS8	Aegean Sea	15	7	8	4	214000	1x1*	58
LS9a	Black Sea (Romania)	10	7	7	3	29600	0.5x0.5	26
LS9b	Black Sea (Türkiye)	16	7	8	3	685	0.5x0.5	0.74
LS10	Whole Europe's Seas	0	15	30	0	13689994	10x10	96
LS11	French Polynesia	?	?	?	?	?	?	?
Sum								~350

2.4. Data utility

Both the harmonized input dataset collection (level-2) as well as the output datasets (level-3) will be useful for numerous practitioners involved with ecosystem-based marine management, marine cumulative pressures and impacts assessments, and Maritime Spatial Planning (MSP), including national and international institutions and organizations, Knowledge Centre for Biodiversity (KCB), Joint Research Center (JRC), European Environment Agency (EEA), International Council for the Exploration of the Sea (ICES), Non-Governmental Organizations (NGOs), as well as research institutions.

2.5. Relevant legal framework and policies

Ordinary personal data from surveys and interviews from the PAB, stakeholders and participants in GES4SEAS workshops and meetings may be acquired and processed in GES4SEAS. Personal data acquisition and processing will only take place when necessary for the research and in accordance with Article 6(1)(e) of the General Data Protection Regulation. In these cases, informed consent will be sought from the participants.

Compliance with the legal and ethics requirements of level-1 datasets remains the responsibility of each data handler and respective LS.

2.6. Data governance and responsibilities

WP5 in GES4SEAS is responsible for the data management. During the course of the project, data access will only be granted internally for exchange between participants in GES4SEAS, including PAB members and the commission. Different partners and LSs, according to their roles in Task 5.1, under the coordination and assistance of the data management team (WP5 leaders), will do the quality control, processing, and harmonization of level-1 datasets. Preserving the working dataset versions will be made by the data management team under the GES4SEAS Zenodo repository. LSs are responsible for the creation and validity check of metadata before submitting. Data managers (WP5) will be checking the validity of the metadata once more and potentially rectifying issues together with the LSs. After the project period, parts of level-2 (depending on data redistribution rights) and all level-3 data will be made publicly available after an embargo period of 6 months. Data will be securely stored and later published on GES4SEAS Zenodo repository.

Long-term preservation of data has not been conclusively discussed yet, but the GES4SEAS intention is that the KCBD can receive those data for the benefit of future research projects and needs for management.

Currently, we do not expect to store personal data beyond the institutional and contact information of data authors (as part of the metadata) on the data repository. Instead, we expect all other personal information obtained for research purposes in GES4SEAS to be anonymized beforehand by the respective data controllers.

2.7. Informed consent template

An informed consent template for application within GES4SEAS can be found in Appendix 1.

3. FAIR data

3.1. Making data findable, including provisions for metadata

All the data produced and/or used in GES4SEAS will be discoverable with metadata, identifiable and locatable by means of Digital Object Identifiers (DOI). All spatial level-2 and level-3 datasets generated and used within GES4SEAS will be accompanied by metadata compliant with the 'Dublin Core' standard (ISO 15836-1:2017), and supplemented with geospatial metadata elements as defined in ISO 19115:2014. This combined approach ensures both general discoverability and domain-specific interoperability for marine spatial data. Metadata for non-spatial data will likewise follow an implementation of the 'Dublin Core' standard (ISO 15836-1:2017). To optimize possibilities for reuse, the metadata for each dataset will be equipped with corresponding search keywords describing the content of the data and making it findable through metadata-based search engines. The metadata will be generated in human-readable YAML format. In principle, it is possible to bind both the metadata and datasets within the Geopackage with the advantage that the metadata will always accompany datasets. With respect to findability, however, it is an advantage to have separately accessible metadata files alongside, too. Individual metadata files can be searched and inspected without the need to install the toolbox and download the data. The feasibility of binding datasets and metadata in a Geopackage within GES4SEAS and, thus, the final data structure will be decided in the course of the project.

Within GES4SEAS, we have developed a preliminary naming convention for level-2 and level-3 datasets generated and used within the project. The convention is based on standardized vocabulary, mainly based on MSFD Terminology synthesized in WP2 (Smith et al., 2022).

Metadata and dataset names will consist of 4 mandatory parts separated by an underscore:

- The LS number.
- A root composed of MSFD compliant classification.
- A prefix indicating if it is a dataset (ds) or metadata (md).

- A suffix indicating the date of upload into the project's internal repository in YYYYMMDD format, thus also indicating the sequence of versions.

Example: "LS01_Pressure-Eutrophication-Nutrients-D5C1_ds_20230312"

Likely, there will be several versions of level-2 and level-3 data for individual datasets generated and/or used in GES4SEAS and the different LSs. The version of each dataset can be derived from the generation date in its name. This preliminary naming convention will be extended/adapted whenever this seems appropriate in the course of the project.

3.2. Making data openly accessible

All research output datasets at level-3 will be made publicly available on a GES4SEAS Zenodo repository, where all publishable outputs of GES4SEAS (papers, datasets, software, etc.) will be available in open access, either in green or gold open access. Personal registration will be needed to access data on the repository after an embargo period of 6 months after finalizing the project. During the course of the project, all level-2 and level-3 data are intended for project-internal usage only. Which of the level-2 datasets after the project will be made publicly available requires case-by-case evaluation, as GES4SEAS, in many circumstances, might have the right to use the data but no license for their redistribution. However, numerous level-1 datasets are already publicly available from data repositories or available on request from respective owners. As the final GES4SEAS toolbox is meant to include a data-harmonization tool for the generation of harmonized level-2 data, end-users will be able to access these datasets directly from the original data owner and apply them in the toolbox.

In specific cases, e.g. for several datasets derived from national authorities, these might be considered sensitive with respect to national security. Consequently, access to these datasets by third parties might be restricted. Which specific datasets will be an element of such restrictions needs to be determined in the course of the data collection procedure during the project. In any case, the lineage of respective datasets, incl. original data sources, as well as quality control, processing, and harmonization steps, will be adequately documented within the metadata.

Level-2 datasets in Geopackage format will be compatible with standard and broadly used GIS handling software (i.e. ArcGIS, QGIS, Python, R, etc.). Detailed documentation for the developed toolbox will be provided along with the toolbox software, which will be fully and publicly accessible through the GES4SEAS web page. We have decided to keep the source code closed while GES4SEAS is running, and no papers are out on the specific implementation. At a later stage, however, it might be decided to publish the toolbox source code in open source. Codes (e.g. R and Python) for the harmonization of level-1 data for the GES4SEAS toolbox will be published in open-source code, and at a later stage, adopted as a data management tool within the GES4SEAS toolbox.

3.3. Making data interoperable

The data produced in GES4SEAS will be interoperable to support data exchange and reuse across researchers, institutions, organizations, and countries, in line with the principles of the INSPIRE

Directive (2007/2/EC). For each LS, level-2 (toolbox input) and level-3 (toolbox output) datasets will be harmonized to enable consistent geospatial processing. Across the European LSs (LS01–10), a common projection and coordinate reference system—ETRS89-LAEA Europe, EPSG:3035—will be used to facilitate simultaneous GIS handling and cross-LS comparisons. Data will be provided in the GeoPackage format, ensuring broad compatibility and ease of use.

While full INSPIRE compliance—including alignment with specific data models—is not yet finalized, this will be addressed in the final version of the DMP, where we will specify:

- The relevant INSPIRE data themes for each dataset type
- The corresponding data models to be used
- The procedures for compliance verification
- This phased approach ensures that GES4SEAS datasets are developed with interoperability in mind, while allowing flexibility during the project's active phase.

GES4SEAS will follow data and metadata vocabularies and standards to make data interoperable. As already mentioned above, the spatial data will be in the common and standardized Geopackage data format, directly compatible with broadly used software for spatial data analysis. Within GES4SEAS, we will follow an implementation of the standardized Dublin Core metadata format (which includes a vocabulary for its metadata items) to harmonize the metadata used between all datasets and LSs. The Vocabulary used will mainly be based on MSFD classifications and definitions (Smith et al., 2022).

GES4SEAS addresses issues related to the MSFD (but is not restricted to the MSFD). To support a common understanding and coherent use of the MSFD terminology and official reporting lists within the project, WP2 (Conceptual Models) in GES4SEAS has developed a standardized vocabulary summary (Smith et al., 2022), which will be implemented in the datasets and metadata and extended/adapted whenever this seems appropriate in the course of the project.

In case it is unavoidable to use vocabulary other than the harmonized terms in datasets, we will develop respective mapping ontologies within GES4SEAS, and both vocabulary terms will appear in the metadata description for transparency.

Intended vocabulary references:

- Countries: ISO codes as described in the ISO 3166 international standard Alpha-2I
- Marine Regions: Regional seas around Europe
- MSFD: Descriptors, Criteria, Activities, Pressures, Ecosystem Components, Species Groups (Smith et al., 2022)
- Parameters: to be identified
- Units: as far as possible, based on SI units
- Species: World Register of Marine Species (WORMS)
- Habitats: EUNIS habitat classification (2021-22 and further updates) and crosswalks with other relevant classification lists (e.g. Red List, MSFD Habitat Types, Helcom Hub regional classification system habitat types, Barcelona Convention classification system habitat types)

- Others: as far as identified as necessary

3.4. Increase data reuse (through clarifying licenses)

The licensing of the openly accessible level-2 and -3 datasets has not been conclusively decided yet and will be subject to future GES4SEAS internal discussions. However, to promote transparency and facilitate reuse, we anticipate adopting open licenses such as Creative Commons (e.g., CC BY or CC0) for most datasets, in alignment with previously established data licenses and specific dataset requirements. Licensing decisions will be documented per dataset to reflect any constraints or conditions.

To enable publication of results prior to open data release, we anticipate an embargo period of 6 months following the conclusion of the GES4SEAS project. After this period, datasets generated and used in GES4SEAS will be made available for reuse by third parties (e.g., KCBD, EEA, JRC, ICES, RSCs). The data collections might also form a guideline for third parties to generate corresponding data for their own assessment areas of interest, as GES4SEAS works on exemplary LSs only. National and international assessments are out of the scope of GES4SEAS.

How long the data remains re-usable will depend on the temporal variability of the underlying variable. However, for future comparisons with, e.g. historical conditions, the re-usability will persist.

We anticipate that most of the datasets will be generated along with a documented uncertainty/quality assessment. Where appropriate, we will publish individual datasets alongside scientific papers or pure data papers, providing respective quality assurance documentation. Certain level-2 datasets will be restricted from reuse (see section 2.2.). Metrics of data reuse, such as dataset download counts and citation statistics, will be tracked and reported as part of the project's impact and dissemination activities.

3.5. Data Publication Strategy and Repository Transition

During the active phase of the GES4SEAS project, all datasets and toolbox components will be published on the **GES4SEAS Zenodo community repository**. Zenodo provides a flexible and accessible platform for versioning, embargo management, and internal collaboration while data and tools are still under development. This approach supports iterative refinement and ensures that project partners and stakeholders can access evolving resources in a controlled and traceable manner.

Following the completion of the project and the expiration of the embargo period (6 months post-project), **final harmonized datasets**—particularly those relevant to marine pressures, impacts, ecosystem components, and services—will be transitioned to **EMODnet thematic portals**. EMODnet offers a domain-specific, policy-aligned infrastructure for long-term data stewardship and discoverability, especially for marine managers, EU agencies, and regional sea conventions.

This dual-repository strategy ensures:

- Flexibility during development (Zenodo)
- Maximum visibility and reuse post-project (EMODnet)
- Alignment with EU marine data governance frameworks

Toolbox documentation, software, and supplementary materials will remain available on Zenodo, while EMODnet will host the curated datasets intended for broader policy and scientific use.

4. Allocation of resources

The total costs for making data FAIR within GES4SEAS will consist of

- Fees for publishing data papers in Open Access journals.
- Fees for project internal data repository.
- Fees for long-term open data repository.
- Working time allocated for data generation and curation.
- Working time allocated/used for data management.

These costs will primarily be covered by GES4SEAS project funds, and, in some cases, some partners can allocate their own funds whenever necessary or appropriate to maximize impact or co-fund with other projects for collaborative outputs. The total cost to make the GES4SEAS data FAIR cannot be outlined yet. In GES4SEAS, data management is part of WP5, and Aarhus University (Ashley Hemraj and Andreas Holbach) is responsible for data management as part of task 5.1.

Long-term preservation of data has not been conclusively discussed yet. However, following the transition of curated datasets to EMODnet thematic portals (as outlined in section 3.5), GES4SEAS also intends to make relevant datasets available to the Knowledge Centre for Biodiversity (KCBD), ensuring accessibility for future research and marine management needs. This dual approach supports both policy-aligned stewardship and broader scientific reuse.

5. Data security

To a limited extent, some LSs within GES4SEAS will handle datasets considered sensitive for the national security of the corresponding countries. The safe handling and storage of these level-1 data will only happen on internal institutional repositories of the partner institutions which were granted access and usage rights for these datasets. LSs are responsible for the adequate and secure handling of sensitive data, which will not be made publicly available. For level-2 datasets derived from sensitive level-1 datasets, access restrictions might be implemented on the GES4SEAS Zenodo repository. This decision will require case-by-case evaluation with respective data owners, who will have to decide on the sensitivity of the contents of the generated level-2 data. After a certain level of abstraction during data processing, derived level-2 data might not be considered sensitive anymore.

Acquired ordinary personal data through interviews, questionnaires and the like will only be handled and stored in repositories intended and approved for personal data classified as sensitive in the EU General Data Protection Regulation (GDPR). All personal information will be anonymized before being submitted to the GES4SEAS Zenodo repository.

The GES4SEAS Zenodo repository is equipped with an automatic backup archive, ensuring possible data recovery and secure storage. Hence, Zenodo is a safe and certified data repository for long-term preservation and curation.

6. Ethical aspects

GES4SEAS has been assigned one independent and external ethics advisor (ETHICAL & LEGAL PLUS SL (PlusEthics)).

The following tasks in GES4SEAS have been identified in the 'Ethics Report' (Deliverable 8.2) to be related to collecting and handling personal data, which requires particular attention with respect to the use of data, personal data protection, as well as equity and gender behavioural guidelines:

- T1.1. Coordination of the Practitioners' Advisory Board (PAB) with target stakeholders
- T1.2. Co-creation and design thinking process
- T1.3. Iterative development and validation of the beta toolbox version.
- T3.4. Building analytical capacity
- T5.3. Testing and identification of information gaps
- T6.5. Blue Schools engagement
- T7.3. Establish effective internal communication through annual meetings, workshops, and remote meetings.
- T7.4. Coordinate participation in conferences and meetings
- T7.5. Contact other relevant scientific projects and networks

Data protection regulations in GES4SEAS are in line with EU data protection policies following the GDPR, and all GES4SEAS consortium partners commit themselves that they will not:

- collect sensitive or conviction-related data on a large scale,
- use covert observation or deception to obtain sensitive data,
- use personal data without consent or legitimate justification,
- engage in activities that threaten the privacy or security of data subjects,
- engage in profiling or intrusive systematic monitoring, or
- transfer sensitive data to third countries.

Even though the most sensitive issue about personal data in GES4SEAS is, according to the 'Ethics Report', identified to be the potential transfer to third countries or data collection outside the European Union, to address this, GES4SEAS' approach to GDPR includes the following elements:

- In the unexpected case that a GES4SEAS project partner identifies the need to transfer personal data to non-EU countries or to collect personal data from research subjects outside the EU, informed consent from the data subjects will always be sought. Data subjects have to be fully aware of how their data will be used and their rights.
- Recipients of personal data in non-EU countries have to maintain data protection standards equivalent to EU law. Data-transfer agreements, adherence to EC standard contractual

clauses, or the establishment of binding corporate rules. Ensure that the same level of data protection required by EU law is upheld.

- Thorough records and documentation of all data transfers and data collection processes are kept. Compliance with both EU regulations and the laws of the host country where the research is conducted has to be documented.
- Ethical standards have to be maintained irrespective of the research location by ensuring that research participants are comfortable, well-informed about data handling, and free from any undue pressure to participate.
- The legal and data protection requirements of the host country where the research is conducted (i.e. United Kingdom, Norway, Turkey, Canada, and French Polynesia) have to be followed. Authorities have to be notified if required, and necessary permissions for the research have to be sought. Any potential data sovereignty provisions that may restrict data transfers, especially for sensitive information, have to be considered.
- Advice from GES4SEAS partners host institution's Data Protection Officer, or a qualified expert has to be sought in any case of uncertainty about how to handle international data transfer issues. Credibility and compliance with data protection regulations will be ensured this way.

Currently, we are expecting to acquire only small amounts of ordinary personal data through participatory events, interviews, questionnaires and the like. Personal data acquisition and processing will only take place when necessary for the research and in accordance with Article 6(1)(e) of the General Data Protection Regulation. In all these cases, informed consent will be sought from the participants by applying an informed consent template (Appendix 1). No transfer of personal data outside the EU and EEA areas is expected in GES4SEAS. As described in section 5 on 'Data security', ordinary personal data will not be stored centrally on the GES4SEAS Zenodo repository but only on repositories intended and approved for personal data classified as sensitive in the EU General Data Protection Regulation (GDPR) under the responsibility of the respective data handlers.

Some countries of participating institutions in GES4SEAS have restrictions for the use and sharing of spatially detailed marine data because of their sensitivity to national security. GES4SEAS will obey these sensitivity issues, and all members commit themselves to not sharing such data outside of the LS/institution that was granted respective access and usage rights.

In any case, as this DMP is a live document, the next version of the DMP will potentially update other relevant ethical issues detected, as well as further measures imposed by the external ethics advisor in relation to the processing of data, especially personal data.

7. Other issues

If it turns out that other national/funder/sectorial/departmental procedures for data management are needed for specific databases in the course of the project, it would be provided by the respective national organizations.

8. References

Smith C., Papadopoulou N., Elliott M., Franco A., Barnard S., Borja Á., Uyarra M.C., 2022. Marine Strategy Framework Directive Terminology Definitions and Lists. GES4SEAS Project. 29pp. https://www.ges4seas.eu/wp-content/uploads/2022/11/GES4SEAS__Report-Definitions-and-Lists-17112022final.pdf

Appendix 1: informed consent template

How [insert name of institution] within task [X.X] of GES4SEAS process your personal data

In connection with your participation in task [X.X] of GES4SEAS, pursuant to the General Data Protection Regulation, we are required to inform you about how your personal data will be processed.

The data controller	[insert name, address and CVR no. of the data controller] are the data controllers responsible for the processing of personal data in task [X.X] of GES4SEAS. Task [X.X] of GES4SEAS is headed by [insert information on the researcher responsible for the project] who can be contacted at [insert: address, phone and email].
Data protection officer at [insert name of institution]	[insert name of DPO] Data protection officer/DPO [insert email of DPO]
Title of the research project	GES4SEAS
The purpose of the project and of processing your personal data	[Describe your purpose for processing the participants' personal data. Describe the purpose in a way that clearly indicates the scientific field of the research, and describe the purpose in such a way that it will "accommodate" all your intended processing operations. It is important that you think in terms of broad purposes, and that you consider all the situations in which it may be relevant to process personal data in your research.]
Which personal data will be processed in the project?	The project will process the following information about you as a participant: ☐ Name ☐ Age ☐ Gender ☐ Affiliation ☐ [insert more types if relevant]
Use of automated processing (profiling)	Profiling is the automated processing of your personal data. For example processing determined by an algorithm. See below whether processing of your personal data will involve automated processing. ☐ Your personal data will be subject to automated processing. ☐ Your personal data will <u>not</u> be subject to automated processing. [If you intend to use automated processing, indicate the purpose and any risks/consequences for the participants.]
For how long do we store your other personal data?	At present, we cannot say for how long we will be processing your personal data. Your personal data will be processed by [insert name of institution] in a personally identifiable form for as long as required by the research purpose and the rules on storage according to responsible conduct of research. When we no longer need your personal data for processing, the data will be anonymized or erased. or Your personal data will be stored at [insert place] in a personally identifiable form until [insert date].

	After this date, your personal data will be anonymized or erased.
Will personal data be made available or disclosed to others, e.g. researchers at other universities?	☐ Your personal data collected for the project will not be disclosed to others. ☐ Your personal data collected for the project will be processed by one or more external data processors [insert information (name, address, CVR no., etc.) on the data processor(s)] pursuant to the rules in Article 28 of the General Data Protection Regulation. ☐ Your personal data collected for the project will be included in a research collaboration with researchers in task [X.X] of GES4SEAS and will therefore be shared with data controllers [insert information (name, address, CVR no., etc.) on the recipient(s)].
The personal data has been obtained:	☐ From you ☐ From you and others* ☐ From others* [*If persons other than the participant have provided information about the participant, then insert the name, address, etc. of the person who provided the data.]
We are entitled to process your personal data pursuant to the rules of the General Data Protection. We are obligated to inform you about the rules that apply to our work with your personal data.	☐ Article 6(1)(e) entitles [insert name of institution] to process non-sensitive personal data about you without your consent, because the research project is a task carried out in the public interest and because processing of personal data is necessary for the performance of the research project.
Your rights under the General Data Protection Regulation	You have the following rights if [insert name of institution] processes your personal data as part of a research project carried out in the public interest: • Right to erasure or the “right to be forgotten”. • Right to data portability - in some cases, you have the right to receive your personal data and to request that the personal data be transferred from one data controller to another. • Right not to be subject to an automatic decision based solely on automated processing, including profiling. Note that your rights may be limited by other legislation or be subject to exemptions, e.g. in relation to research and the exercising of public authority.
Complaints	If you wish to complain about the processing of your personal data, you can do so by contacting the supervisory authority: [insert name and address of the national data protection agency]



2025