Eurofleets+: Eurofleets+ cruise DMP Template - Phase 1: Preliminary DMP

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

What types and formats of data will the project generate/collect?

How is the original data gathered on board and how do you transfer it to shore?

What processing on the raw data do you plan? Please differentiate between data quality assurance (handling of outliers, missing and suspect values, null observations) and data harmonisation (code, label and QC flag explanations, consistent use of headers, data formatting, conforming to standards).

When do you plan to perform these processing steps?

2.1 Making data findable, including provisions for metadata

Can you list some search keywords? The purpose of keywords is to optimize the findability of the datasets.

The Eurofleets+ data repository will allow you to create the metadata when uploading the data. For a list of metadata elements, please refer to the data management guidelines. How will you document all this information before submission, especially lineage information (i.e. processing and QC steps)?

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions or embargo), explain why, clearly separating legal and contractual reasons from voluntary restrictions. Do this for each type of dataset you will create.

Do you plan to make the data and metadata available on another repository than the EuroFleets/SeaDataNet data repository, for instance an institutional, national or general data repository?

2.3. Making data interoperable

The data management guidelines document the specific meta-information surrounding the data that is needed. Please specify how you plan to capture and store the specified individual meta-information elements.

Guidance:

Reference data centres will make your data interoperable with the standards used in European marine research. A large part of this work is to connect the information surrounding the data in the original datasets to standardized definitions (stored in vocabularies). Therefore it is important that the operations during a scientific cruise are noted down in a detailed way, so that they can be interpreted correctly. For a list of elements and assistance on how to do this, please refer to the data management guidelines.

2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest (by any party) re-use possible?

Guidance:

Horizon 2020 wants to stimulate openly accessible data, with maximal reusability. In order to facilitate this, open, up-front and machine readable licenses are preferred, such as those on https://creativecommons.org/licenses.

Furthermore, we want to highlight that EurofleetsPlus data must be as open as possible, which means that the CC0 (public domain) license is the most appropriate.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

3. Allocation of resources

4. Data security

Before the data is transferred to the Eurofleets+ data repository, what provisions are in place for data security (including backups, secure storage and transfer)?

5. Ethical aspects

6. Other issues

On top of the infrastructure and procedures that Eurofleets+ provides, which national/sectorial/departmental procedures for data management are you following?

Eurofleets+: Eurofleets+ cruise DMP Template - Phase 2: Full DMP

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

What types and formats of data will the project generate/collect

How is the original data gathered and stored on board and how do you transfer it to shore?

What processing on the raw data do you plan? Please differentiate between data quality assurance (handling of outliers, missing and suspect values, null observations) and data harmonisation (code, label and QC flag explanations, consistent use of headers, data formatting, conforming to standards).

When do you plan to perform these processing steps?

What is the expected size (Megabyte to Terabyte range) of the data?

Who will be the principal users of the data? Users are both active (those that clean up or analyse the data) and passive (those that read or assess the data).

2. 1. Making data findable, including provisions for metadata

What naming conventions for your data files will you follow?

Can you list some search keywords? The purpose of keywords is to optimize the findability of the datasets.

Do you foresee a need for different versions of the data? Both for your own internal use and when publishing the data? E.g. for some analyses the data might need reorganisation from a common ancestor. Which versioning scheme do you have in mind?

The Eurofleets+ data repository will allow you to create the metadata when uploading the data. For a list of metadata elements, please refer to the data management guidelines. How will you document all this information before submission, especially lineage information (i.e. processing and QC steps)?

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions or embargo), explain why, clearly separating legal and contractual reasons from voluntary restrictions. Do this for each type of dataset you will create.

Do you plan to make the data and metadata available on another repository than the EuroFleets/SeaDataNet data repository, for instance an institutional, national or general data repository?

What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

2.3. Making data interoperable

The data management guidelines document the specific meta-information surrounding the data that is needed. Please specify how you plan to capture and store the specified individual meta-information elements.

Guidance:

Reference data centres will make your data interoperable with the standards used in European marine research. A large part of this work is to connect the information surrounding the data in the original datasets to standardized definitions (stored in vocabularies). Therefore it is important that the operations during a scientific cruise are noted down in a detailed way, so that they can be interpreted correctly. For a list of elements and assistance on how to do this, please refer to the data management guidelines.

Notwithstanding the work the reference data centres will perform, do you plan to already make use of standardized definitions (stored in vocabularies) to store the above meta-information?

Guidance:

For a list of formatting and terminology-level standards used in marine data management, please refer to the data management guidelines.

Do you estimate that you will use uncommon terminology or will generate novel (e.g. new sampling techniques and devices) or project specific scientific terminology? If yes, how will you communicate this information in the above meta-information and make sure they are seen as novel?

2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest (by any party) re-use possible?

Guidance:

Horizon 2020 wants to stimulate openly accessible data, with maximal reusability. In order to facilitate this, open, up-front and machine readable licenses are preferred, such as those on https://creativecommons.org/licenses.

Furthermore, we want to highlight that EurofleetsPlus data must be as open as possible, which means that the CC0 (public domain) license is the most appropriate.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

3. Allocation of resources

Who will be responsible for data management in your project?

4. Data security

Before the data is transferred to the Eurofleets+ data repository, what provisions are in place for data security (including backups, secure storage and transfer)?

5. Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing?

6. Other issues

On top of the infrastructure and procedures that Eurofleets+ provides, which national/sectorial/ departmental procedures for data management are you following?