UNESCO: UNESCO Ocean Project or Programme Template

Section 3. Data Creation, Collection and/or Acquisition

1. What data will be collected, created and/or acquired?

Guidance:

What parameters or attributes are being collected or created in these data, e.g. temperature, salinity, abundance of Mytilus Edulis, modelled temperature or salinity. Are these new data, or are the data re-used from existing resources?

Will any third party data be acquired? If so, from where and under what terms and conditions?

2. What methods will be used in data collection or creation?

Guidance:

What methods or best practices are used in the initial data collection or creation.
Where appropriate, provide references to journal papers or links to documentation, such as from the Ocean Best Practices repository.

Also describe any platforms (such as research vessels or models) or instruments used in the data collection and how the data will be transferred from the platforms to the storage location.

3. What file formats will the data be collected, created or acquired in?

Guidance:

Please list the file formats used to store the data (e.g. Microsoft Excel, netCDF, OceanDataView, SQL Database, TIFF, MP4).

If possible, please indicate the expected volume of data to be produced (for example 10Mb; 1Tb etc…).

4. What conventions or standards, if any, are used at the time of data collection, creation or acquisition?

Guidance:

The preparation of data files often follows conventions or standards such as the Climate and Forecast (CF) conventions for NetCDF file formats, SeaDataNet for physical and biogeochemical data, including litter, or Darwin Core for biodiversity data.

Please list the conventions or standards used in preparation of the data files here.

5. What ontology/vocabulary, if any, is used to name the parameters in the data files from the time of data creation, collection and/or acquistion?

Guidance:

What ontology or vocabulary will be used for the data parameters?

6. What is the backup strategy used to ensure the data files from this stage of the data lifecycle can be recovered in the event of an IT disaster?

Guidance:

Please describe how continued access to the data will be ensured if there is an IT disaster, such as a security breach or corrupted local drive?

7. (To be completed after data collection) What differences or discrepancies occurred between planning and defining the objective of the data creation or collection and the completion of data creation or collection?

Guidance:

Differences or discrepancies may occur for operational reasons (such as a change in survey plans) or for other valid reasons. Giving details here may help later assessments of the appropriateness or completeness of the data for further analysis.

8. Following data collection, creation or acquisition but before processing and storage, who may the data be shared with, under what conditions and when?

Guidance:

At this stage of the data lifecycle, with whom may the data be shared (for example a project team or across an institution)? What restrictions are placed on any users at this stage of the data lifecycle

(e.g. not for publication)? How soon after data creation, collection or acquisition may the potential users listed gain access to the data?

9. How is the data creation, collection and/or acquisition to be funded?

Guidance:

How is this stage of the data lifecycle funded?

10. People, Institutions, Contact Details, Roles involved in Data Creation, Collection and/or Acquisition

Guidance:

List the names, organisational affiliations, contact details and roles of all associated with the data.

Roles are outlined in Annexe Data Roles on "Plan overview".

Section 4. Data Storage and Processing

1. What data will be collected, created and/or acquired?

Guidance:

What parameters or attributes are being collected or created in these data, e.g. temperature, salinity, abundance of Mytilus Edulis, modelled temperature or salinity. Are these new data, or are the data re-used from existing resources?

Will any third party data be acquired? If so, from where and under what terms and conditions?

2. What file formats will the data be in (or available in) once processing is complete?

Guidance:

Please list the file formats used to store the data (e.g. plain text formats, Microsoft Excel, netCDF, JSON, Application Programming Interface, Open Geospatial Consortium Web Services, Raster or Vector Geospatial Formats, OceanDataView, SQL Database, TIFF, MP4).

If possible, please indicate the expected volume of data to be produced (for example 10Mb; 1Tb etc…).

How often will the data be updated? For example, is the data static, or does it receive real-time or near-real time updates and on what frequency?

3. What conventions or standards, if any, are used in the preparation of the data file(s) at this stage of the data lifecycle?

Guidance:

The preparation of data files often follows conventions or standards such as the Climate and Forecast (CF) conventions for NetCDF file formats, SeaDataNet for physical and biogeochemical data, including litter, or Darwin Core for biodiversity data.

Please list the conventions or standards used in preparation of the data files here.

4. What ontology/vocabulary, if any, is used to name the parameters in the data files at this stage of the data lifecycle?

Guidance:

What ontology or vocabulary will be used for the data parameters?

5. How will the quality of the data be assessed and reported?

Guidance:

What methods or best practices are used in assessing or quantifying the quality of the data. How will any errors or outliers in the data be identified and reported? How will any missing data values be identified and reported? Do these methods for data quality assessment and reporting meet the ends of the expected end users of the data.

Where appropriate, provide references to journal papers or links to documentation, such as from the Ocean Best Practices repository. Please describe any data flagging schemes in use, such as the Argo data quality flags.

Please detail how you will record both in the data, accompanying metadata (data about the data) and documentation what quality control steps have been taken and who (person, role and/or organisation) will undertake the quality control.

6. What is the backup strategy used to ensure the data files from this stage of the data lifecycle can be recovered in the event of an IT disaster?

Guidance:

Please describe how continued access to the data will be ensured if there is an IT disaster, such as a security breach or corrupted local drive?

7. Following data processing and storage but before data analysis, who may the data be shared with, under what conditions and when?

Guidance:

At this stage of the data lifecycle, with whom may the data be shared (for example a project team or across an institution)? What restrictions are placed on any users at this stage of the data lifecycle (e.g. not for publication)? How soon after data processing and storage may the potential users listed gain access to the data?

8. How will the data storage and processing be funded?

Guidance:

How is this stage of the data lifecycle to be paid for?

9. People, Institutions, Contact Details, Roles involved in Data Storage and Processing

Guidance:

List the names, organisational affiliations, contact details and roles of all associated with the data.

Roles are outlined in Annexe Data Roles on "Plan overview".

Section 5. Data Analysis

1. What methods will be used in data analysis?

Guidance:

What methods or best practices are used in the analysis of the data. Where appropriate, provide references to journal papers or links to documentation, such as from the Ocean Best Practices repository.

2. What file formats will the data be in once analysis is complete?

Guidance:

Please list the file formats used to store the data (e.g. plain text formats, Microsoft Excel, netCDF, JSON, Application Programming Interface, Open Geospatial Consortium Web Services, Raster or Vector Geospatial Formats, OceanDataView, SQL Database, TIFF, MP4).

If possible, please indicate the expected volume of data to be produced (for example 10Mb; 1Tb etc…).

3. What conventions or standards, if any, are used in the preparation of the data file(s) at this stage of the data lifecycle?

Guidance:

The preparation of data files often follows conventions or standards such as the Climate and Forecast (CF) conventions for NetCDF file formats, SeaDataNet for physical and biogeochemical data, including litter, or Darwin Core for biodiversity data.

Please list the conventions or standards used in preparation of the data files here.

4. What ontology/vocabulary, if any, is used to name the parameters in the data files at this stage of the data lifecycle?

Guidance:

What ontology or vocabulary will be used for the data parameters?

5. How will any further assessments of the quality of the data be undertaken and reported?

Guidance:

What methods or best practices are used in assessing or quantifying the quality of the data. How will any errors or outliers in the data be identified and reported? How will any missing data values

be identified and reported? Do these methods for data quality assessment and reporting meet the ends of the expected end users of the data.

Where appropriate, provide references to journal papers or links to documentation, such as from the Ocean Best Practices repository. Please describe any data flagging schemes in use, such as the Argo data quality flags.

Please detail how you will record both in the data, accompanying metadata (data about the data) and documentation what quality control steps have been taken and who (person, role and/or organisation) will undertake the quality control.

6. What is the backup strategy used to ensure the data files from this stage of the data lifecycle can be recovered in the event of an IT disaster?

Guidance:

Please describe how continued access to the data will be ensured if there is an IT disaster, such as a security breach or corrupted local drive?

7. Following data analysis but before data publication, who may the data be shared with and under what conditions?

Guidance:

At this stage of the data lifecycle, with whom may the data be shared (for example a project team or across an institution)? What restrictions are placed on any users at this stage of the data lifecycle (e.g. not for publication)? How soon after data analysis may the potential users listed gain access to the data?

8. How is the Data Analysis to be funded?

Guidance:

How is this stage of the data lifecycle to be paid for?

9. People, Institutions, Contact Details, Roles involved in Data Analysis

Guidance:

List the names, organisational affiliations, contact details and roles of all associated with the data.

Roles are outlined in Annexe Data Roles on "Plan overview".

Section 6. Data Publication

1. What metadata (data about the data) standards to describe the data are used by either the research community or the funding agencies?

Guidance:

Metadata standards are often set at a national or international level. Often the metadata standards are based on a profile of a broader standard (such as the EU's INSPIRE Spatial Data Infrastructure and the US' Earth Observing System Data and Information System based on ISO19115; or OBIS' use of the Ecological Metadata Language).

These metadata standards will often describe the way in which the data are geo-referenced, so please describe how your data will be located in space and time.

2. What keywords will you use to describe the data, and which vocabulary will you take them from?

Guidance:

Keywords support the findability or discovery of a data. Metadata standards often describe which vocabulary (set of keywords) to use in describing a data (such as from the NERC Vocabulary Server, the CF Standard Names list or the Environment Ontology (EnvO).

Please list here the vocabularies and keywords you will use in describing the data.

3. What other documentation can be provided to support the data?

Guidance:

Are there any other documents which will provide additional contextual information supporting the data? These may include but are not limited to research vessel survey reports or data description papers.

How will the processing of the data (who did what to the data, when and why) from collection to

final data product be described?

4. How will the metadata be published to make it generally available?

Guidance:

How will the metadata created for the data be published and made generally, publicly available? Which catalogue or system will be used to allow others to discover the existence of the data?

5. What services are used to publish the data?

Guidance:

Please list any web services which are used to make the underlying data available (not the metadata), such as Web Map Services, Esri ArcGIS Online, Erddap data servers, cloud storage (such as Amazon S3 or Microsoft Azure object storage.

6. Does the data require any further transformation (file formats, attributes, conventions) in order to be published? If yes, please describe.

Guidance:

Please list any further changes in file formats, the naming of attributes or alterations to the conventions used in the preparation of the data in order to allow the Data Publication lifecycle phase to be achieved.

7. What licence will the data be published under?

Guidance:

An open license (for example a Creative Commons or Open Government Licence) is recommended for data sharing, especially for alignment with the FAIR principles. However, this is not always possible.

Please provide a link to the license under which the data will be shared.

8. Who may the data be shared with or published to?

Guidance:

Open sharing of the data under an open licence is recommended, but if this is not possible (for example the data contains sensitive or personal data) please list the individuals or organisations with whom the data may be shared and under what arrangements or conditions (e.g. formal data sharing agreements, contractual agreements, etc...).

9. Are there any embargo periods to be placed on the data sharing before external publication?

Guidance:

Is there a time during which the data will have been collected and processed but which should not be shared wider? If so, please provide the details and an explanation of why this embargo period is in place here.

10. When the data are shared or published, are there any uses for which the data are not suitable?

Guidance:

For example, some bathymetry data may not be suitable for navigation.

11. Are there any further terms of use to be placed on the data?

Guidance:

Please list any further terms of use that are to be placed on the data, for example for warnings or flood forecasts it may be appropriate to add a caveat such as “Users may not redistribute the material if has been modified” or “material content of the warnings may not be altered, and the warnings be kept up to date”

12. How is Data Publication to be funded?

Guidance:

How is this stage of the data lifecycle to be paid for?

13. People, Institutions, Contact Details, Roles involved in Data Publication

Guidance:

List the names, organisational affiliations, contact details and roles of all associated with the data.

Roles are outlined in Annexe Data Roles on "Plan overview".

Section 7. Data Archival

1. Which parts of the data will be selected for long-term archival / preservation?

Guidance:

If there are parts of the data which will not be selected for long-term preservation or archival, such as part-processed data, please provide a description and a reason here.

How will you make the version history of the data (from raw data to harmonised data to fully processed data) transparent to end users and make data from any stage of this process available to end users?

2. What is the long-term importance or value of the data?

Guidance:

Please give a description of the anticipated long-term value of the data, for example will the data be used in sustainable growth or sustainable development, health or food security?

3. What is the long-term preservation plan for the data?

Guidance:

Please describe which archive, data facility or data assembly centre the data will be submitted to for long-term preservation and indicate if the submission of the data has been discussed with them.

The archive, data facility or data assembly centre should be encouraged to publish the data's metadata to the Ocean Data and Information System (ODIS) and to openly publish the data (where applicable) using web standards.

Any data aggregators that the archive, data facility or data assembly centre is connected to and which may validate the data and metadata should also be documented.

Are there any other intended archival or preservation approaches, such as formal data publication with a digital object identifier (doi)?

4. Does the data require any further transformation (file formats, attributes, conventions) or quality control in order to be archived? If yes, please describe.

Guidance:

Please list any further changes in file formats, the naming of attributes or alterations to the conventions used in the preparation of the data in order to allow the Data Archival lifecycle phase to be achieved. For example, is there anything further required to make the data meet the FAIR principles, or are geometry validations required?

5. Will the data license, terms and conditions or any other restrictions on use change between Data Publication and Data Archival? If so, please indicate how.

Guidance:

Questions 6.7- 6.11 detail various licensing and restrictions on the data. Please indicate any changes to the answers to those questions resulting from the Data Archival phase of the lifecycle.

6. How is Data Archival to be funded?

Guidance:

How is this stage of the data lifecycle to be funded?

7. People, Institutions, Contact Details, Roles involved in Data Archival

Guidance:

List the names, organisational affiliations, contact details and roles of all associated with the data.

Roles are outlined in Annexe Data Roles on "Plan overview".

Section 8. Data Reuse

1. What other documentation can be provided, or linked to, to support the data?

Guidance:

Are there any other documents which will provide additional contextual information supporting the data? These may include but are not limited to research vessel survey reports or data description papers.

How will the processing of the data (who did what to the data, when and why) from collection to final data product be described?

2. What metrics are in place to monitor the reuse of the data and its impact?

Guidance:

How will the reuse of the data be tracked? For example through citations of a digital object identifier for the dataset.

Who will be responsible for maintaining and reporting any chosen metrics and how will they be funded?

Section 9. Data Erasure

1. What foreseeable circumstances, excluding hardware failure or cybersecurity issues, may trigger the intended complete erasure of the data?

Guidance:

Are there intended circumstances in which the data may be erased. This could be a superseding of model data with a new analysis, or for data containing personally identifiable information a request for a data subject.

Please do not include unintended circumstances here, as they should be covered elsewhere by data backup questions.

2. What would be the consequences of erasing the data?

Guidance:

What would be the consequences to any users of the data if the intended erasure of the data was undertaken.

3. How would users of the data be informed of an intended erasure of the data?

Guidance:

What steps would be taken to give users warning of the erasure of the data? How long a period of notice or warning of the intended erasure would be given? What alternative sources of similar data exist which users could switch to?

4. People, Institutions, Contact Details, Roles involved in Data Erasure

Guidance:

List the names, organisational affiliations, contact details and roles of all associated with the data.

Roles are outlined in Annexe Data Roles on "Plan overview".

Section A. Ethical and Legal Compliance

1. Is there an ethics policy which was required to be followed during the creation or processing and analysis of the data?

Guidance:

If yes, please provide a link to the policy and explain how the data meets the requirements of that policy.

2. Is there a diversity policy which was required to be followed during the creation or processing and analysis of the data?

Guidance:

If yes, please provide a link to the policy and explain how the data meets the requirements of that policy.

3. Are there any considerations related to Collective Benefit for indigenous peoples associated with the data?

Guidance:

The CARE Principles consider the following topics under Collective Benefit. Please address these here if applicable to the data:

- Inclusive development and innovation
- Improved governance and citizen engagement
- Equitable outcomes

4. Are there any considerations related to the Authority of indigenous peoples to Control the data?

Guidance:

The CARE Principles consider the following topics under Authority to Control. Please address these here if applicable to the data:

- Recognising rights and interests
- Data for governance
- Governance of data

5. Are there any considerations related to Responsibility towards supporting indigenous peoples' self-determination and collective benefit associated with the data?

Guidance:

The CARE Principles consider the following topics under Responsibility. Please address these here if applicable to the data:

- Positive relationships
- Expanding capability and capacity
- Indigenous languages and worldviews

6. Are there any considerations related to the Ethics of indigenous peoples' rights and wellbeing associated with the data?

Guidance:

The CARE Principles consider the following topics under Responsibility. Please address these here if applicable to the data:

- For minimising harm and maximising benefit
- For justice
- For future use

7. Are there any further ethical considerations associated with the data?

Guidance:

Please list any further ethical considerations or issues associated with the data.

8. Is there a data policy associated with the data describing access and intellectual property rights of the data?

Guidance:

If yes, please provide a link to the data policy and explain how the data meets the requirements of that policy and what are the implications for Intellectual Property Rights associated with the data.

9. Which individual or organisation owns the copyright of the data?

Guidance:

List the name and contact details of the individuals or organisations who own the copyright and/or intellectual property rights associated with the data.