D1.4: Data Management Plan

WP1: Management and coordination

Document information

Grant Agreement Number	821926	Acronym	IMMERSE	
Full title	Improving Models for Marine EnviRonment SErvices			
Horizon 2020 call	H2020-SPACE-2018			
Type of action	RIA – Research Innovation Action			
Start date	01/12/2018	**Duration**	48 months	
Project URL	-			
EU Project Officer	Monika KACIK			
Project Coordinator	Julien LE SOMMER			
Deliverable	D1.4 Data Management Plan			
Work package	WP1: Management and coordination			
Date of delivery	**Contractual**	M6	**Actual**	M6
Nature	R - Report	**Dissemination Level**	PU - Public	
Lead Beneficiary	CNRS			
Lead Author	Julien LE SOMMER	**Email**	julien.lesommer@univ-grenoble-alpes.fr	
	CNRS			
Other authors	Victor Van Gucht (CNRS)			
Reviewer(s)				
Keywords	Data management, Horizon 2020, FAIR, data, software			

Version	Date	Description	Authors
1.0	31/05/19	Initial version	J. Le Sommer and V. Van Gucht

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821926.

1. Introduction

This document describes the plans for data management within IMMERSE project. It specifies the data input and output, the data format standards, the means of exploitation and the data curation policy set up within IMMERSE consortium. The target audience of this document are the consortium members of IMMERSE and any third party willing to re-use the results of IMMERSE project. The objective of this document is (i) to help the project partners manage their data and become aware of the FAIR principles, (ii) provide information to third parties as to how IMMERSE consortium approaches data management.

IMMERSE project aims at preparing numerical ocean models for the next generation Copernicus Marine Environment Monitoring Service (CMEMS). In response to the future priorities for CMEMS, IMMERSE develops new capabilities in order to (i) enable the production of ocean forecasts and analyses that exploit upcoming high resolution satellite datasets; (ii) deliver ocean analyses and forecasts with the higher spatial resolution and additional process complexity demanded by users; (iii) exploit the opportunities of new high performance computing (HPC) technology; (iv) allow easy interfacing of CMEMS products with detailed local coastal models. These developments will be delivered in the NEMO ocean model which already forms the basis of the majority of CMEMS analysis and forecast products.

The input for this document is derived from the project proposal, which resulted from many discussions among project partners and from a survey conducted with all the partners at the beginning of the project (see appendix 3). This document is adapted from the “Template Horizon 2020 Data Management Plan (DMP)". It is meant to be a living document to be updated as necessary during the lifespan of the project. The first version of this document has been prepared by the project office and circulated to IMMERSE Scientific Coordination Board for comments. IMMERSE data management plan is available publicly through IMMERSE project website (www.immerse-ocean.eu).

Because a large fraction of the output of IMMERSE project will be delivered in the form of numerical codes, this document not only describes the procedures for FAIR data management within IMMERSE consortium but also covers the question of how IMMERSE consortium follows the principles of FAIR software management with the numerical codes used and developed during the project. With this approach, we expected to reflect more closely how data will be produced and leveraged during the project implementation. In what follows, the term *data* will therefore implicitly refer to both *numerical data* and *numerical codes*.

Section 2 provides a high-level description of the different types of data that IMMERSE project participants deal with. Section 3 describes how IMMERSE consortium manages numerical data. Section 4 describes how IMMERSE consortium manages numerical codes. The allocation of resources is discussed in section 5; data security and ethical aspects are covered in section 6 and 7, respectively. Additional supporting material, including lists of software, is provided in appendix.

2. Typology of data and context of the project

The main outcome of IMMERSE project is delivered in the form of numerical code, a large fraction of which contributes to NEMO ocean circulation model. In a very general sense, these numerical codes are meant to transform numerical data provided as input and to produce numerical data as output. The development and the testing of these numerical codes, which is the main activity of IMMERSE project, uses a large amount of numerical data. This is why we hereafter describe jointly numerical data and numerical codes used and produced during the project.

In what follows, the curation strategy for numerical codes and numerical data is presented depending on the nature of the data considered. These different types of data are defined in order to closely reflect the workflow of ocean model design. This workflow involves the commonly accepted concepts of *model code base* and *model configuration*:

- an *ocean circulation model code base* is the numerical code that implements a discretized version of the set of partial differential equations governing the evolution of ocean circulation; this code is a priori generic enough so that it can be applied in different practical set-up; an ocean circulation model code also includes the numerical codes that specify how numerical data are exchanged during the model integration; IMMERSE projects contributes to developing NEMO ocean circulation model;

- an *ocean circulation model configuration* refers to the practical implementation of an ocean circulation model meant to produce numerical data describing ocean circulation in a particular spatial domain and time-span; a numerical configuration combines (i) an ocean circulation model code, (ii) numerical data describing the specific choice of parameters for the region considered, (iii) numerical data describing the geometry of the domain and (iv) numerical data describing the forcing fields used during the model integration; IMMERSE will implement and use several configurations of NEMO ocean model, some of them are distributed alongside with NEMO ocean model code on NEMO public repository (in such case referred to as NEMO reference configurations).

It is also important to consider that practices in terms of data and code management depend strongly on how the codes are eventually shared and distributed. In the case of IMMERSE project, the following situations may be distinguished:

- *numerical code that is, or is meant to be, distributed as a part of NEMO ocean model through NEMO public repository*; this mostly encompasses NEMO numerical code per se but may also include some other codes closely related to NEMO and supported by NEMO System Team; this type of numerical code will be delivered by IMMERSE WP2, WP3, WP4 and WP5;

- *numerical code that are not meant to be distributed as part of NEMO public repository*; this mostly encompasses (i) pre- and post-processing codes related to NEMO model setup (ii) non-NEMO scientific software used for assessing the impact of IMMERSE activities on downstream applications, and (iii) other scripts used for implementing the project workflow; these non-NEMO numerical codes will be used and developed in IMMERSE WP2, WP6, WP7, WP8.

In terms of data management, IMMERSE project will use and produce different types of data at different stages of the project. The types of data are hereafter defined following the workflow of ocean model design:

- *Type D1 : Data describing numerical model configurations* : this encompasses data describing the geometry (bathymetry, coastlines, ...), data specifying the external forcing (atmosphere, land,...) both in the form of NetCDF files, and numerical parameters in the form of namelists

(ascii files); these data are either produced by the project consortium, or based on pre-existing data;

- *Type D2: Data produced by numerical ocean models* (i.e. model output) : these numerical data mostly consist in NetCDF files. Within IMMERSE, they may either (i) be used internally by a project partner, or (ii) be shared amongst the consortium partners, or (iii) be shared outside the consortium;

- *Type D3 : Data used for assessing the quality of numerical model output* : within IMMERSE, the solutions of numerical models will be compared with external sources of data, mostly in the form of pre-existing numerical data from numerical models or observations; these data will be used in most IMMERSE work-packages and especially in WP6, WP7 and WP8.

In terms of software management, the different types of numerical codes handled during IMMERSE project may be categorized as follows:

- *Type S1 - NEMO ocean model :* which combines NEMO core engines[1] (NEMO-OCE, NEMO-SI3, NEMO-TOP/PISCES), NEMO interfaces (NEMO-OBS, NEMO-ASM, NEMO-TAM) and NEMO Reference Configurations and Test Cases [2];

- *Type S2 - ancillary numerical codes distributed with NEMO ocean model* : which consist in a set of numerical codes provided with NEMO ocean model for setting up configurations and processing data (also referred below as NEMO Tools [3]);

- *Type S3 - third-party numerical codes (not distributed with NEMO)* : which may consist in geoscientifc models (OceanParcels, AeroBulk, FVCOM, SIMBAD,...), components of modelling systems (XIOS, OASIS, AGRIF, …) or pre/post-processing libraries (CDFTOOLS, pyNEMO);

- *Type S4 - other analysis and processing scripts*: day-to-day analysis and processing scripts used and developed by project participants.

3. Data management

As stated in introduction, this document does not strictly follow the official template for H2020 projects data management plans. With this choice, we intend to better reflect the specificities of data management in the ocean modelling community and to better describe the measures that IMMERSE will establish in order to improve further those practices.

3.1 Status of FAIR data practices within the consortium

The status of FAIR data practices within the consortium is established based on the results of the survey (see Appendix 3) and on known workflows in the ocean modelling community (as discussed for instance within DRAKKAR community and within CLIVAR-OMDP). This section is therefore not only reflecting the practices within IMMERSE consortium but also the practices of a wider community.

[1] https://www.nemo-ocean.eu/framework/components/engines/
[2] https://www.nemo-ocean.eu/framework/components/configurations/
[3] https://forge.ipsl.jussieu.fr/nemo/chrome/site/doc/NEMO/guide/html/tools.html

Type D1 data (model configurations)

Type D1 data combine re-used data and new data : re-used data consist mostly in NetCDF files obtained from publicly accessible archives (and possibly slightly modified as for instance model bathymetry); new data mostly consist in model namelist parameters in the form of ascii files.

Are type D1 data findable?
Most pre-existing data, on which type D1 data are based, are associated with digital objects identifiers (DOIs). These DOIs are usually (but not systematically) provided in the scientific papers. Namelists are generally not findable without contacting the data producer.

Are type D1 data openly accessible?
Pre-existing data mostly come from open archives operated by CMEMS, project partners or national facilities. When pre-existing data are modified (ex: bathymetry), the modified files are usually accessible upon request to the project partner.

Are type D1 data interoperable?
Most type D1 data are stored in NetCDF4 format, following CF-compliant metadata standards and naming conventions. Type D1 data are therefore interoperable in NEMO community, and can be used with other models as well after minor changes.

Are type D1 data re-usable?
Type D1 data are generally re-usable in principle but there is generally no specific license associated with it. The main practical barrier to re-use is usually to identify who produced the experiment, which may not always be stated in scientific communications.

Type D2 data (model output)

Type D2 data consist in collections NetCDF files that store the output of specific model experiments produced or used by the project. Each of these collections amount for ~1Gb to ~1Pb, depending on the domain, spatial resolution and output frequency.

Are type D2 data findable?
As for current practices, type D2 data is not systematically available for re-use but a growing fraction of type D2 data used in publications are tagged with DOIs, usually provided through zenodo.

Are type D2 data openly accessible?
Produced data are generally stored over several years in storage facilities attached to supercomputing facilities (eg: GENCI, JASMIN) or in internal archives operated by the partners (MOI, CMCC, UKMO). Some of the smaller datasets may also be duplicated on zenodo. Access policies depend on the centers, and various restriction may apply. Data are usually stored over timescales ~5 years

Are type D2 data interoperable?
Type D2 data follow well-established CF-compliant metadata standards. In the case of NEMO model output, naming conventions follow CMEMS or CMIP naming conventions. Type D2 data can be straightforwardly manipulated with common analysis software (Python, Matlab, Ferret).

Are type D2 data re-usable?
There is generally no restriction nor embargo for re-using type D2 data after publication but only a fraction of the data has a license. For produced data, all participants accept to share the data with the general public. The main barrier to re-use is the size of the datasets and the restriction to access the storage facilities.

Type D3 data (model evaluation data)

Type D3 data are re-used data, mostly obtained from publically available archives and distributed in the form of NetCDF files following well-established metadata standards.

Are type D3 data findable?
The vast majority of type D3 data used in the project consist in published data available through open archives. The datasets are generally identified with DOIs.

Are type D3 data openly accessible?
Type D3 used in the project are openly accessible data available through data portals operated by CMEMS, project partners or national facilities.

Are type D3 data interoperable?
Type D3 data consist in CF-compliant NetCDF files following well-established naming conventions and are therefore fully compatible with existing analysis software.

Are type D3 data re-usable?
Type D3 data can be re-used without restrictions. Specific conditions are described in the associated licenses.

3.2 Critical analysis of data management practices in the consortium

Overall, the ocean modelling community (as gathered within IMMERSE consortium) is well aware of FAIR data principles. Data practices for most re-used data within IMMERSE are compliant with FAIR data principles. Produced data (mostly model data) follow good standards (format, metadata, naming) but there are practical limitations to a full compliance with FAIR data principles.

The main practical barrier to the application of FAIR data principles in ocean modelling is associated with the size of datasets. The total volume of produced data for the project is typically of the order of several hundred of terabytes to several petabytes. It is generally not possible to distribute openly produced model data with publications. Also importantly, model data commonly become obsolete after several years as the code used for generating the data evolves and as other model simulations become available. It is therefore not obvious what fraction of the datasets should be maintained available in the long term.

3.3 Measures to improve data management practices

Given the above analysis, IMMERSE consortium has identified four key objectives regarding data management practices:

- ***(O1)*** *improving the FAIR data practices for type D1 data (model configurations) in order to foster the re-use of model configurations within the consortium and beyond.*
- ***(O2)*** *establishing a robust protocol for the re-use of type D2 (model output) data within the consortium.*
- ***(O3)*** *fostering the use of FAIR practices for sharing type D2 (model) data beyond the consortium.*

- ***(O4)*** *improving FAIR data practices for the use of type D3 (for pre-existing observational or model data) data during the project implementation.*

In practice, the above objectives will be tackled through the following measures:

- Deliverable D2.2 will establish a protocol for describing Type D1 data and improve their reusability. Project participants will be encouraged to use this protocol in all papers and deliverables coming from the project activity.
- Project participants will be encouraged to make sure that the papers and deliverables coming from the project activity include a section describing pre-existing data and produced data with a DOI and a license when possible. This information will also be featured on the project website for each deliverable.
- O2 will be addressed through specific topical discussions during the project General Assemblies.
- Guidelines for sharing Type D2 data in scientific publications and project deliverables will be discussed and communicated during the project General Assemblies.

4. Software management

In this section, we describe the consortium practices in terms of software development, use and overall management. This section is an addition to the standard template for H2020 projects Data Management Plan because a large fraction of the project output will be delivered in the form of numerical codes. This section therefore describes the measures IMMERSE will establish in order to improve FAIR software practices during the project implementation.

4.1 Status of FAIR software practices in the consortium

The status of FAIR software practices within the consortium is established based on the current practices within NEMO consortium (as discussed in NEMO Developers Committee) and on the survey to IMMERSE participants (see appendix 3).

Software management practices in the consortium depend on the type of software being considered. We described below the FAIR practices for each of the type S1-S4 software, listed in section 3.

- *Type S1 software* is findable on NEMO source code repository[4] with a clear versioning. The code is accessible upon registration. An up-to-date documentation (associated with a DOI) is distributed with each release. Type S1 software uses clearly defined IO standard; the API is kept as stable as possible within each release cycle. Type S1 software is distributed under CECILL license[5] and maintained in the long term by NEMO System Team as described in NEMO consortium agreement.
- *Type S2 software* is developed and distributed with similar practices as Type S1 software but the level of documentation may vary from tool to tool. The API of NEMO Tools is not systematically defined.

[4] https://forge.ipsl.jussieu.fr/nemo/browser/NEMO/releases
[5] http://www.cecill.info/licences.fr.html

- *Type S3 software* is developed and maintained following independent threads for each numerical code. All of them are developed and distributed open source with a clear versioning under a well-identified license. A table with the links is provided in Appendix 1. Type S3 software may or may not be findable with DOIs. How long these codes will be maintained in the future depends on each project.

- *Type S4 software* consists in all the day-to-day script and short codes used and developed in the project. Typical languages include Python, Shell, matlab or FORTRAN. Type S4 codes are usually not findable, not accessible nor re-usable outside the group where they have been developed. The level of interoperability varies depending on the author.

4.2 Critical analysis of software management practices in the consortium

Overall, IMMERSE participants are well aware of how FAIR data principles can apply to software management and improve the reproducibility of scientific research. Software management practices in the consortium generally follow FAIR data principles and use quite systematically version control, open source licenses, DOIs, stable API and up-to-date documentations. This is particularly true for NEMO ocean model per se. However the management of day to day scripts and short codes (type S4) does not yet systematically follow FAIR data principles. Moreover, it should be noticed that NEMO ocean model development process is not fully open as the code is only accessible upon registration and internal communications among NEMO developers are not systematically open to the public.

IMMERSE project will deliver numerical code falling into types S1-S4. The main bottleneck to a full compliance with FAIR data principles and to the full reproducibility of IMMERSE results appears to be concentrated on type S4 software (short codes and analysis scripts). Type S4 software is indeed usually not made available with scientific publications in the Ocean Modelling community. It is indeed unclear how an analysis pipeline designed to exploit a large database of model output could be practically shared. Here again, one of the practical barrier to full reproducibility is related to the computational and storage resources required for analyzing large databases. Most other software used and developed during IMMERSE (type S1-S3) follow or will follow FAIR data principles. This is the case for contributions to pre-existing software projects (as instance NEMO) and new software projects (which will be hosted on a specific GitHub organization).

4.3 Measures to improve software management practices

In order to increase FAIR software practices within IMMERSE project, we propose the following objectives:

- ***(O5)*** *following best practices for the development of IMMERSE related non-NEMO software;*

- ***(O6)*** *increasing project participants awareness regarding best practices in software development;*

- ***(O7)*** *encouraging the use of more reproducible/open practices for IMMERSE publications;*

- ***(O8)*** *analyzing the opportunity of adopting fully open development practices for NEMO.*

In practice, we intend to reach the above objectives through the following measures:

- (O5) Establishing a visible GitHub organization for storing new project related software (including type D1 data and type S3-S4 software). The governance of IMMERSE GitHub organization is described in section 3.4 below.

- (O6) Organizing software related events alongside IMMERSE General Assemblies: trainings (from basic training to software version control to more advanced courses on continuous integration) and code-sprints for improving specific pieces of software developed by the project. These events will target specifically project participants outside NEMO System Team.

- (O7): Encouraging the use of D2.2 method for describing configurations in project related publications. Encouraging the open release of S4 type software (analyses scripts) used in scientific publication. These two measures will be implemented through specific communication during general assemblies.

- (O8) organizing discussions on this topic during NEMO Developers Committee and within NEMO System Team.

4.4 Governance and management of IMMERSE GitHub organization

The governance of IMMERSE GitHub organization (https://github.com/immerse-project/) is described below.

- Each project participants can ask to be affiliated to IMMERSE GitHub organization.

- Any software project related to IMMERSE activities is eligible to be hosted under IMMERSE GitHub organization.

- Key software projects associated to IMMERSE activities have been listed by the project office (see in Appendix 2). Core developers of each of them will be encouraged to use GitHub and IMMERSE GitHub organization.

- IMMERSE encourages the use of IMMERSE GitHub organization for storing S4 type code.

- Software developers are encouraged to use continuous integration services and robust testing for their software.

- Developers may choose between two categories of projects: open or restricted ones. By default IMMERSE project office encourages developers to opt for open projects. In both cases the software will be open to all the GitHub accounts under IMMERSE GitHub organization.

- The long-term fate of IMMERSE GitHub organization and associated software projects will be discussed in 2021. GitHub appears stable enough for IMMERSE organization to be maintained beyond the project activity with limited cost.

- The maintenance and evolution of IMMERSE GitHub organization is operated by Julien Le Sommer (incentive and motivation), Victor Van Gucht (access rights) and Nicolas Martin (advice on best practices).

See table in Appendix 2 describing the pieces of software that will be hosted on IMMERSE GitHub organization.

5. Allocation of resources

Data production and preservation is a key aspect of the activity of IMMERSE project partners. The majority of partners already devote significant resources for accessing high performance computing facilities, maintaining local data storage capacity and covering technical support staff. The extra cost associated with IMMERSE activity will mostly be covered in the form of in-kind contributions by the project partners. Therefore, the associated budget requested from the project will be kept at a minimum. The long-term storage of data needs to be discussed within the project in order to define which of the data produced during the project should be preserved in the long-term. Regarding software preservation, no specific resources will be required for using NEMO official software repository nor for using GitHub service (free public plans).

6. Data security

The data produced within the project is not sensitive with respect to the GDPR and personal privacy. As described above, depending on the size of the dataset, the data produced that ought to be shared will be either stored on existing data archives maintained by the project partners or fluxed to zenodo open access archive. The partner institutions which maintain data archives will be responsible for establishing access and usage restrictions if so required (user registration and data licensing) and for setting up regular backups following their internal data policy. Software distribution will be based on the existing NEMO official software repository and GitHub service.

7. Ethical aspects

There are no ethical restrictions with regards to the generated data, possible misuse and ethical implications of misuse are out of scope of this project.

Appendix 1 - List of pre-existing software used in IMMERSE

The table below presents a list of pre-existing (type S2, S3) software that will be used in IMMERSE project, with their associated URL and license.

Name	Link	License
NEMO	https://www.nemo-ocean.eu	CeCILL License
OceanParcels	https://github.com/OceanParcels/parcels	MIT License
AeroBulk	https://github.com/brodeau/aerobulk	GNU General Public License (GPL)
FVCOM	http://fvcom.smast.umassd.edu/fvcom/	GNU General Public License (GPL)
SIMBAD	http://forge.ipsl.jussieu.fr/albatross	CeCILL License
XIOS	https://forge.ipsl.jussieu.fr/ioserver	CeCILL License
OASIS	https://portal.enes.org/oasis	GNU General Public License (GPL)
AGRIF	http://agrif.imag.fr	CeCILL License
CDFTOOLS	https://github.com/meom-group/CDFTOOLS	CeCILL License
pyNEMO	https://github.com/jdha/PyNEMO	GNU General Public License (GPL)
Sesam	https://github.com/brankart/sesam	CeCILL License

Appendix 2 - List of software hosted under IMMERSE GitHub organization

The table below presents a preliminary list of new software that will be developed during IMMERSE project and hosted under IMMERSE GitHub organization. The table also indicates the associated task/work-package in IMMERSE Grant Agreement.

Name	**Description**	**Task**
NEMO test cases	Idealized test cases to be used in NEMO continuous integration suite.	T2.2
NEMO config-tools	Codes that automate the description of NEMO configuration through GitHub and zenodo	T2.3
NEMO configurations	Repositories describing each NEMO configuration implemented during the project	All WPs
Downstream interface	Prototype toolbox for the uptake of CMEMS products in downstream systems	T7.3
(tbd)	Software for the analysis of forecast accuracy in DA systems	T7.2
Validation tools	Series of software and scripts used for assessing high resolution NEMO solutions.	WP6, WP8

Appendix 3 - Data Management Plan Questionnaire

The questionnaire below has been circulated to IMMERSE participants in April 2019 in order to establish a status of FAIR data practices in the consortium and identify the contact points on data management issues for each project partner. The questionnaire is divided into three parts focusing respectively on (i) data usage, (ii) data production and (iii) FAIR software.

The goal of this questionnaire is to gather input for IMMERSE project Data Management Plan. The Data Management Plan is a living document that details, which data is used in the project, where it comes from, and how it is made publicly available for reuse according to the FAIR principles (more on this later).

Please try to give detailed answers especially in the open questions, this allows us to better take into account your needs and wishes.

1. Email address

_____

2. Which institute do you work for?

_____

3. Which work package(s) are you working on? (Tick all that apply).

- ☐WP1
- ☐WP2
- ☐WP3
- ☐WP4
- ☐WP5
- ☐WP6
- ☐WP7
- ☐WP8

Data usage

This section pertains to the data you will use within the IMMERSE project, this is any data that already exists that you will re-use.

4. Please give a short description of the data that you will (re) use.

_____
_____
_____

5. Will you re-use any existing data from others? (within or external to the project) (Mark only one answer).

- ☐Yes, from project partners

☐Yes, from external sources

☐Yes, from both project partners and external sources

☐No

6. What is the origin of this data?

_____

_____

7. Where is this data stored? (Tick all that apply).

☐Public archive

☐ Other

8. Please add any other concerns you have about data re-use here. (Are there any barriers for external data access and re-use, e.g. from other projects)

_____

_____

Data Production

The following section pertains to the data you will produce within the IMMERSE project. Within EU projects the concept of FAIR (Findable, Accessible, Interoperable and Reusable) data is very important. A brief summary of these principles is given here:

- Findable means the data has a global unique identifier (for instance a DOI), are described with metadata and are registered or indexed in a searchable resource.
- Accessible means the identifier can be used to retrieve the data and metadata using standardized communication protocols.
- Interoperable means the data and metadata adhere to formal standards.
- Reusable means the data has a correct license specified in the metadata, have provenance information, and adhere to community standards.

The following questions are both intended to give us an overview of the data that will be produced and how FAIR this data will be.

9. Please give a short description of the data you produce.

_____

_____

10. What is the expected size of your output data in TB?

_____

11. Where will the data be stored?

_____

12. What data formats do you use for your output data? (Tick all that apply).

☐ netCDF3

☐ netCDF4

☐ Other

13. Which Metadata Standards do you use for your output data? (Tick all that apply).

_____

14. What naming conventions do you use for your output data? (Tick all that apply).

_____

15. Do you use any tools to automatically check the adherence to the metadata standards and naming conventions? If so, please provide names or links.

_____

_____

16. Do you use versioning for your data? (Mark only one answer).

☐Yes

☐No

17. Which license do you use for your data? If you don't know about licensing data you can read more about it here: http://www.dcc.ac.uk/resources/how-guides/license-research-data
(Tick all that apply).

☐Creative Commons: Attribution (CC BY)

☐Creative Commons: Attribution Share -Alike (CC BY SA)

☐Open Data Commons Attribution License (ODC-By)

☐Open Government License (OGL)

☐Other

18. Is there any data that you will NOT publicly share? If so, please explain why.

_____

_____

FAIR Software

The following section concerns the sharing of software and other code with the consortium and with the public, under similar principles as FAIR for data. This questionnaire mostly concerns code that is not hosted on NEMO public repository. In practice, this may encompass pre- or post-processing codes for NEMO itself or any other geoscientific model that you may use or develop within IMMERSE.

19. Within IMMERSE project, please describe the sort of non-NEMO code your institution will develop.

_____

_____

20. For this non-NEMO software, do you use software versioning? If so, which one? (Tick all that apply).

☐Git
☐SVN
☐Mercurial
☐None
☐Other

21. For this non-NEMO software, which programming languages do you use for your code?
(Tick all that apply).

☐Fortran
☐Python
☐C++
☐Matlab
☐R
☐Java
☐Julia
☐Other

22. For this non-NEMO software, which licenses do you use for your code? (Tick all that apply).

☐MIT
☐Apache-2.0
☐GPL
☐Mozilla
☐Other

23. Are you willing to share the following software with the consortium? (Tick all that apply).

☐Geoscientific model code
☐Pro-processing code
☐Post-processing code
☐Other

24. Which of the aforementioned non-NEMO software that you developed are you willing to share with the public?

_____

25. If you do NOT wish to share your code please explain why.

_____

_____