D9.1
Marine subdomain FAIRness roadmap

Work Package	WP9
Lead partner	MARIS
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30-06-2019
Submission date	31-08-2019

Deliverable abstract

This report, deliverable **D9.1 - Marine subdomain FAIRness roadmap**, is the result of Task 9.2 "Analysis and priorities to enhance RI data FAIRness" activities which were led by MARIS with involvement of all RIs within the Marine subdomain, and tuning with WP5-WP7. Task 9.2 has analysed the FAIRness of each of the Marine RIs and made a list of priorities for enhancing their FAIRness by means of a roadmap. This deliverable D9.1 is to be considered as an analysis and assessment providing a major basis for formulating the next deliverable **D9.2 - Marine subdomain implementation plan**. This will be prepared in the coming months interacting closely with WP5 (Community standards and catalogue of services) and WP7 (Common implementation and support).

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Peter Thijsse and Dick Schaap	MARIS	22-07-2019 08-08-2019 17-08-2019
Contributing Authors	Barbara Magagna; Markus Stocker; Sylvie Pouliquen; Katrina Exter; Ivan Rodero; Alex Vermeulen; Thierry Carval; Benjamin Pfeil; Justin Buck	Umweltbundesamt; TIB; Ifremer; Lifewatch ERIC; EMSO ERIC; ICOS ERIC; Euro-Argo ERIC; ICOS ERIC; Euro-Argo ERIC	16-07-2019 16-07-2019 26-07-2019 29-07-2019 30-07-2019 11-08-2019 16-08-2019 16-08-2019 16-08-2019
Reviewer(s)	Daniela Franz	FZJ	22-08-2019
Approver	Andreas Petzold	FZJ	31-08-2019

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	22-07-2019	Draft – to be checked by Marine RIs	Peter Thijsse and Dick Schaap
V 0.2	08-08-2019	Suggestions and comments processed from WP9 leader, LifeWatch (Marine) RI and EMSO RI representatives, and ENFRI-FAIR coordinating team. Tables updated by Umweltbundesamt in section 4.4 and appendix 2 for LifeWatch (Marine) RI and EMSO RI.	Peter Thijsse and Dick Schaap
V 1.0	17-08-2019	Finalised with processing of suggestions and comments as received from ICOS (Marine) and Euro-Argo RI representatives.	Peter Thijsse and Dick Schaap
V2.0	22-08-2019	Finalised after receiving edits and suggestions from final reviewer.	Peter Thijsse and Dick Schaap

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu.

TERMINOLOGY

A relevant project terminology is included as an Appendix. The latest version of the master list of the terminology is here:
https://confluence.egi.eu/pages/viewpage.action?pageId=14452608

PROJECT SUMMARY

ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity.

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.

This goal is reached by: (1) well defined community policies and standards on all steps of the data life cycle, aligned with the wider European policies, as well as with international developments; (2) each participating RI will have sustainable, transparent and auditable data services, for each step of data life cycle, compliant to the FAIR principles. (3) the focus of the proposed work is put on the implementation of prototypes for testing pre-production services at each RI; the catalogue of prepared services is defined for each RI independently, depending on the maturity of the involved RIs; (4) the complete set of thematic data services and tools provided by the ENVRI cluster is exposed under the EOSC catalogue of services.

TABLE OF CONTENTS

1. Introduction 6
2. RIs in the Marine subdomain 6
3. Approaches applied for gathering information and roadmap analyses 8
   3.1 Increasing our knowledge about FAIR 10
   3.2 Generic questionnaire about state of FAIRness in each RI 13
   3.3 Initial analyses of the responses to the generic questionnaire 13
   3.4 Second questionnaire aimed at collecting machine readable information 14
   3.5 Organise subdomain workshops 14
   3.6 Next steps as undertaken towards the roadmap 15
4. Quantified analysis 16
   4.1 Explanation of analysis approach 16
   4.2 Questionnaire responses in XLS files 18
   4.3 YAML representation files 18
   4.4 Output of the SparQL queries (reporting tables) 20
      4.4.1 Findability 21
      4.4.2 Accessibility 24
      4.4.3 Interoperability 27
      4.4.4 Reusability 32
5. Analysis of gaps and solutions 35
   5.1 ICOS (Marine) 35
      5.1.1 Gaps and planned activities 35
      5.1.2 Strengths 36
   5.2 Euro-Argo 36
      5.2.1 Gaps and planned activities 36
      5.2.2 Strengths 37
   5.3 SeaDataNet 37
      5.3.1 Gaps and planned activities 37
      5.3.2 Strengths 37
   5.4 LifeWatch (Marine) 38
      5.4.1 Gaps and planned activities 38
      5.4.2 Strengths 38
   5.5 EMSO 38
      5.5.1 Gaps and planned activities 39
      5.5.2 Strengths 40
6. Selection of priority EOVs 40
   6.1 EOV framework 40
   6.2 Selection of EOVs for ENVRI-FAIR Marine subdomain 42
   6.3 Benefits of enhancing FAIRness for selected EOVs 44
7. Conclusions and next steps 46

Appendix 1: Glossary 47

Appendix 2: XLS spreadsheets from RIs 49
ICOS (Marine) 49
SeaDataNet CDI 53
SeaDataNet Data products catalogue 57
Euro-Argo 61
LifeWatch Marine Data Archive 73
LifeWatch - EUROBIS 77
EMSO 81
Appendix 3: YAML content 90
ICOS (Marine) 90
SeaDataNet 93
Euro-Argo 97
LifeWatch (Marine) 100
EMSO 105

1. Introduction

The ENVRI-FAIR project is engaging Research Infrastructures (RIs) in the environmental domain covering the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems. The overarching goal of ENVRI-FAIR is that all participating research infrastructures (RIs) will improve their FAIRness and become ready for connection to the European Open Science Cloud (EOSC).

WP9 has a focus on the RIs in the Marine subdomain, which are represented in ENVRI-FAIR by Euro-Argo, EMSO, and the marine component of ICOS and LifeWatch, as RIs listed on the ESFRI roadmap, as well as SeaDataNet as European marine data management infrastructure. The overall aim of WP9 is to analyse the FAIRness of each of these RIs and to implement within each RI the necessary actions to improve its FAIRness. The latter is critical for the Marine subdomain as it will provide a coherent base for developing the integrated services systems required by a broad variety of research, regulatory and operational communities. Climate change and anthropogenic impacts are among the key issues that already affect and concern European and worldwide societies, but there are others that also have an enormous socio-economic impact (natural hazards, water quality, acidification). Therefore, "Essential Ocean Variables" (EOV) as identified by the Global Ocean Observing System (GOOS) will provide guidance to the WP9 activities for setting priorities at data type level.

Considering the ENVRI-FAIR challenge of multiple RIs and multiple subdomains, the agreed way forward is that the FAIR principles will be implemented within each RI to improve RI FAIRness at three levels: 1) to better serve its own users; 2) to facilitate the development of cross RI services at Marine sub-domain level; and 3) to facilitate the development of cross subdomain services within ENVRI-FAIR. This implicates that in ENVRI-FAIR a bottom-up approach is followed, respecting the autonomy of RIs concerning requirements and solutions, however in close and regular interaction with WP5-WP7 which consider common standards, training, common implementation options at environmental domain level, as well as with WP8-WP11 which concern analysis and implementation activities in each of the subdomains.

This report, deliverable **D9.1 - Marine subdomain FAIRness roadmap**, is the result of Task 9.2 "Analysis and priorities to enhance RI data FAIRness" activities which is led by MARIS with involvement of all RIs within the Marine subdomain, and tuned with WP5-WP7. Task 9.2 has analysed the FAIRness of each of the Marine RIs and made a list of priorities for enhancing their FAIRness by means of a roadmap. This deliverable D9.1 is to be considered as an analysis and assessment providing a major basis for formulating the next deliverable **D9.2 - Marine subdomain implementation plan**. This will be prepared in the coming months interacting closely with WP5 (Community standards and catalogue of services) and WP7 (Common implementation and support).

2. RIs in the Marine subdomain

The Marine subdomain in ENVRI-FAIR is represented by Euro-ARGO, ICOS (Marine), EMSO, and LifeWatch (Marine) as RIs as listed on the ESFRI roadmap, and SeaDataNet as European marine data management infrastructure.

Euro-Argo - European Contribution to the international Argo Program
The objectives of the Euro-Argo ERIC are to optimize, sustain and improve the European contributions to Argo and to provide a world-class service to the research (ocean and climate) and operational oceanography (COPERNICUS Marine Service) communities. Euro-Argo also aims at preparing the next phase of Argo with an extension to deeper depths, biogeochemical parameters and observations of the polar regions. The Euro-Argo RI comprises a central facility and distributed national facilities. On May 2014, the EC awarded European legal status (ERIC) to the central facility. Euro-Argo aims at developing a capacity to procure, deploy and monitor one quarter of the global network (350 floats per year including the new phase of

Argo to abyssal ocean, high latitudes, European marginal seas and biogeochemical parameters) and ensure that all the data can be processed and delivered to users (both in real-time and delayed-mode).
Coordinating institution: EURO-Argo ERIC, France;
Representing organisations in ENVRI-FAIR: Euro-Argo ERIC, IFREMER, UKRI-BODC

EMSO - European Multidisciplinary Seafloor and Water Column Observatory
EMSO is a large-scale European ESFRI RI for strategically placed, deep sea observatories with the essential scientific objective of real-time, long-term monitoring of environmental processes related to the interaction between the geosphere, biosphere, and hydrosphere. It is a geographically distributed infrastructure composed of several deep-seafloor and water-column observatories, which is deployed at key sites in European waters, spanning from the Arctic, through the Atlantic and Mediterranean, to the Black Sea. Observatories are platforms equipped with multiple sensors, placed along the water column and on the seafloor. They constantly measure different biogeochemical and physical parameters, that address natural hazards, climate change and marine ecosystems. EMSO offers data and services to a large and diverse group of users, from scientists and industries to institutions and policy makers. It provides relevant information for defining environmental policies based on scientific data. EMSO is a consortium of partners sharing a common strategic framework of scientific facilities (data, instruments, computing and storage capacity). EMSO ERIC has been established in January 2017.
Coordinating institution: EMSO ERIC, Italy;
Representing organisation in ENVRI-FAIR: EMSO ERIC

ICOS (Marine) – Integrated Carbon Observation System
ICOS provides consistent, long-term and high-quality observations required to understand the present state and predict future behavior of the global carbon cycle and greenhouse gas emissions. The objectives of ICOS RI are to provide effective access to a single and coherent data set to facilitate research into multi-scale analysis of greenhouse gas emissions, sinks and the processes that determine them. ICOS provides essential information to research in order to understand regional budgets of greenhouse gas sources and sinks, their human and natural drivers, and the controlling mechanisms. All this to inform policy makers and the general public on the state of the greenhouse gas budget of Earth's atmosphere as main driver of climate change. ICOS RI tracks carbon fluxes in Europe and adjacent regions by observing the ecosystems, the atmosphere and the oceans through integrated national station networks, European central facilities and distributes the GHG data to the users via the ICOS Carbon Portal. ICOS ERIC was established in November 2015.
Coordinating institution: ICOS ERIC, Finland;
Representing organisations in ENVRI-FAIR: UHEL, FMI, ULUND, UVSQ, CMCC, UiB

LIFEWATCH - European e-Science Infrastructure for Biodiversity and Ecosystem Research
LifeWatch is operating an e-Infrastructure for basic research on biodiversity and ecosystems, but also supports research for the protection, management and sustainable use of biodiversity. The infrastructure includes facilities for data integration and interoperability; capabilities to create workflows of analytical and modelling tools; and a Service Centre providing special services for scientific and policy users, including training and research opportunities for young scientists. Its architecture supports access to and the integration of external resources such as data from associated infrastructures and distributed computational capacity from high performance clusters. User groups may create their own e-laboratories or e-services within the common architecture of the infrastructure. LifeWatch enables distributed large scale and collaborative systems research on biodiversity complexity. LifeWatch ERIC was established in May 2017.
Coordinating institution: LifeWatch ERIC, Spain;
Representing organisation in ENVRI-FAIR: LifeWatch ERIC

SEADATANET - Pan-European Infrastructure for Ocean and Marine Data Management
SeaDataNet (SEADATANET) is a major operational infrastructure for managing, indexing and providing access to ocean and marine data sets and data products, acquired by European organisations from research cruises and other observational activities worldwide. It also promotes common standards and tools for the marine domain. Since the mid-1990s, SEADATANET has expanded and matured; at present, it provides federated discovery and access to more than 110 data centres for physics, chemistry, geology, bathymetry, and biology. It works closely with EuroGOOS, CMEMS, Euro-Argo, ICES, and EurOBIS a.o., is a major driver of EMODnet, and a principal initiator of the international Ocean Data Interoperability Platform (ODIP). SEADATANET is further developing its discovery, access, ingestion, publishing and visualisation services in the current SeaDataCloud project, working together with EUDAT and also considering the EOSC challenge.
Coordinating institution: IFREMER, France;
Representing organisations in ENVRI-FAIR: MARIS, IFREMER, RBINS, OGS, CSIC, UKRI-BODC

3. Approaches applied for gathering information and roadmap analyses

Task 9.2 - Analysis and priorities to enhance RI data FAIRness - has the overall objective to analyse the level of FAIRness of the Marine RIs. This analysis and assessment must identify priorities for improving the FAIRness at each of the RIs (resulting in deliverable D9.1), then to be followed by activities for formulating an implementation plan (deliverable D9.2) for common development and implementation. The actual implementations will be carried out by the Marine RIs as part of other Tasks within WP9, while finally a demonstration of the achievements by means of a more coherent Marine subdomain will be given around selected Essential Ocean Variables (EOVs).

ENVRI-FAIR is aiming at improving the overall FAIRness of the RIs in the environmental domain, in particular in the four subdomains. However, the RIs are quite different from each other with respect to aims, technical structure, services, organization, funding, and other relevant aspects. Therefore, an overall common solution with each RI adopting the same technical and organizational principles is unrealistic and not considered an option. As a feasible approach, ENVRI-FAIR has agreed on an overall bottom-up approach, starting with analysing the RIs in each of the subdomains in WP8-WP11, while interacting with the more horizontal activities in WP5-WP7 for common standards, training, and implementation options.

The agreed way forward is that the FAIR principles will be implemented within each RI to improve RI FAIRness at three levels:
1) to better serve its own users;
2) to facilitate the development of cross RI services at Marine sub-domain level;
3) to facilitate the development of cross subdomain services within ENVRI-FAIR.

For Task 9.2 the overall approach has been translated into a more detailed approach which will be explained below. This translation has been done by MARIS as leader of Task 9.2 in communication with IFREMER as WP9 coordinator, with representatives of the Marine RIs, and with members of WP5 in order to establish a common FAIRness analysis methodology which should be adopted by all four subdomains. The formulation of an effective analysis methodology started in January 2019 at the ENVRI-FAIR kick-off meeting in Prague – Czech Republic, which included a WP9 subdomain meeting in which MARIS presented its initial ideas for implementing Task 9.2. At the ENVRI-FAIR kick-off meeting also the planned interaction with WP5-WP7 activities was discussed and started. As follow-up, WP5 and WP7 have hosted and continue to host short regular web conferences to discuss and monitor progress of their activities and to tune these with activities in WP8-WP11. For that reason, each subdomain from WP8-WP11 is represented in WP5 and WP7 by persons that function as linking pins.

This has already proven to be instrumental for achieving cohesion between the activities at subdomain level and the activities at overarching environmental domain as all four subdomains have adopted a common FAIRness analysis methodology.

Also, within WP9 regular communication was performed since the ENVRI-FAIR kick-off meeting by means of IFREMER circulating and inviting marine RIs to review the minutes and actions of the WP9 kick-off meeting, regular emails of MARIS as part of Task 9.2, and the organization of a dedicated Task 9.2 meeting in May 2019 in Amsterdam – The Netherlands by MARIS with participation of marine RIs, and leaders of WP5 – WP7.

In this process Task 9.2 activities have first focused on increasing our general understanding and knowledge of FAIR principles for data and services, together with WP5, followed by creating within WP9 an overview of the state-of-the-art in FAIRness of marine RIs, identifying FAIRness strengths and weaknesses of each of the marine RIs, and finally, proposing a priority list of possible improvements per marine RI, to overcome current weaknesses. In addition, Task 9.2 has worked out an analysis for establishing which EOVs are of common priority interest to all five marine RIs.

In the ENVRI-FAIR context, the term ''weakness" should not be explained as being negative about the RI and its FAIRness, considering the development stage of most RIs and the field of FAIRness. However, it should be seen as a way to indicate items for potential improvement or even in some cases as ambitions for improvements.

The interaction of Task 9.2 with WP5 has so far mostly focused on formulating the common FAIRness analysis methodology, while the tuning also with WP7 will become more relevant in the next steps of Task 9.2 for formulating the marine implementation plan (deliverable D9.2). WP7 will consider the roadmaps of each of the four subdomains and look for possible common issues between RIs and associated solution options. For that purpose, WP7 has set-up and is now populating a Knowledge Base (KB) of technical solutions and best practices for implementing these. RIs will be able to query the KB for possible solutions for identified FAIRness weaknesses, which might be adopted by RIs in the subdomain implementation plans. The KB completed with all options and full descriptions of how to apply these technical solutions is planned to be delivered by M20 (D7.3); however already a basic list of possible technical solutions is being gathered, so that the KB can already support formulating the implementation plans with conceptual solutions at RI and subdomain levels.

In this part of the process, WP5 will look for common standards and assist RIs with guidance where possible and desired. While in the implementation process, WP7 can offer technical support to RIs for their implementation.

For the common FAIRness analysis methodology, it was agreed in principle at the ENVRI-FAIR Kick-Off meeting in January 2019 that MARIS would seek cooperation with the GO FAIR initiative as they were developing analytical methods for assessing FAIRness of data and services. This was followed up by MARIS in January 2019 contacting the GO FAIR office in the Netherlands to discuss the basics and meaning of FAIR and the best approach for collecting information from the RIs in the ENVRI-FAIR project. The discussion soon resulted in a clear commitment from the NL GO FAIR office to cooperate with the Marine subdomain as a pivot for ENVRI-FAIR as it became clear that the motivation behind the GO FAIR approach for checking the FAIRness of their "Implementation networks" is similar to that of ENVRI-FAIR. Therefore, it was decided in dialogue of Task 9.2 with WP5 and the linking pins of WP8-WP11 that ENVRI-FAIR would adopt the GO FAIR analysis tools and benefit from the experience already made by GO FAIR, while GO FAIR on its turn would benefit and learn from the additional analysis activities that ENVRI-FAIR plans to undertake. The GO FAIR methodology includes completing survey questionnaires that GO FAIR has developed for gathering information on the FAIRness of an infrastructure.

In the following paragraphs more details will be given of the steps that were undertaken within Task 9.2 to gather information from the marine RIs about their current FAIRness.

3.1 Increasing our knowledge about FAIR

The FAIR concept relates to "Data and services that should be Findable, Accessible, Interoperable, and Re-usable, both for machines and for people." The emphasis is on machine FAIRness. Technological advances provide innovative opportunities for new forms of science, which is one of the drivers behind the European Open Science Cloud (EOSC). However, this demands well-described, accessible data that conforms to community standards. The FAIR principles articulate the attributes data need to have to enable and enhance reuse, by humans and machines. There were a few predecessors before coming to the formulation of the FAIR Guiding Principles as now in use. An influential document was the OECD's 2007 "Principles and Guidelines for Access to Research Data from Public Funding". The seminal Royal Society report of 2012, "Science as an Open Enterprise" argued that research data being open was not sufficient as data need to be accessible, assessable, interoperable and usable too. The 2013 G8 Science Ministers' statement in 2013 says: "Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards." These criteria were adopted in the initial data guidelines for the EU Horizon 2020 framework programme later the same year. Echoing these criteria, the FAIR principles were conceived at the Lorentz conference in 2014 and published following consultation via the FORCE11 Group. The most influential document is the article of Wilkinson et al. (2016) in Nature Scientific Data, "The FAIR Guiding Principles for scientific data management and stewardship" (see Figure 1):

Figure 1. Wilkinson et al. (2016)[1] about FAIR Guiding Principles.

This article introduced a table of FAIR Guiding Principles which can be used by RIs for checking their FAIRness and formulating measures for improving their FAIRness. The table published in the article is shown in Figure 2. Most of the principles are related to metadata. Moreover, the emphasis is machine FAIRness for data and their associated services. Another recent (2018) and prominent publication about FAIR is "Turning FAIR into reality" from an EU committee of experts. The following has been extracted and summarised from their report.

[1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018, doi:10.1038/sdata.2016.18.

Data are **Findable** when they are described by sufficiently rich metadata and registered or indexed in a searchable resource that is known and accessible to potential users. Additionally, a unique and persistent identifier should be assigned such that the data can be unequivocally referenced and cited in research communications. The identifier enables persistent linkages to be established between the data, metadata and other related materials in order to assist data discovery and reuse. Related materials may include the code or models necessary to use the data, research literature that provides further insights into the creation and interpretation of the data and other related information.

Accessible data objects can be obtained by humans and machines upon appropriate authorisation and through a well-defined and universally implementable protocol. Anyone should be able to access at least the metadata. It is important to emphasise that Accessible in FAIR does not mean Open without constraint. Accessibility means that the human or machine is provided - through metadata - with the precise conditions by which the data are accessible and that the mechanisms and technical protocols for data access are implemented such that the data and/or metadata can be accessed and used at scale, by machines, across the web.

Interoperable data and metadata are described community and/or domain standards for technical interoperability and vocabularies for semantic interoperability, and they include qualified references to other data or metadata. It is this that allows the data to be 'machine-actionable'. Interoperability is an essential feature in the value and usability of data. Legal interoperability of data has to be considered as well. In FAIR, legal interoperability falls under the principle that data should be 'Reusable ´.

For data to be **Reusable**, the FAIR principles reassert the need for rich metadata and documentation that meet relevant community standards and provide information about provenance, reporting how data was created and information about consecutive data reduction or transformation processes to make data more usable, understandable or 'science-ready'. The ability of humans and machines to assess and select data on the basis of criteria relating to provenance information is essential to data reuse, especially at scale. Reusability also requires that the data be released with a 'clear and accessible data usage license': in other words, the conditions under which the data can be used should be transparent to both humans and machines.

Box 2: The FAIR Guiding Principles
To be Findable: F1. (meta)data are assigned a globally unique and persistent identifier F2. data are described with rich metadata (defined by R1 below) F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource To be Accessible: A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available To be Interoperable: I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data To be Reusable: R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

Figure 2. FAIR Guiding Principles as defined in Wilkinson et al. (2016).

As the FAIR concept and terminology are quite complex, as part of the analysis in Task 9.2 marine RIs were advised to read more on the backgrounds of FAIR and get themselves acquainted with the FAIR terminology and the ways to measure/describe FAIRness. In communication with the GO FAIR office the following documentation was provided to the RIs:

- Background of the FAIR metrics: www.fairmetrics.org
- Framework template: http://fairmetrics.org/fairmetricform.html

Moreover, background information and examples on the two GO FAIR questionnaires were provided which were going to be used for assessing the FAIRness of each of the RIs as follows:

- background for first questionnaire: GO FAIR reference document with some example content. https://drive.google.com/file/d/1UqchXgR-WFeOy0dwUgP_vOz3Zvl5toQi/view
- background for the second questionnaire consisting of a well-documented questionnaire for the first generation metrics and FAIRness levels of nine resources.

Recently, end July 2019 and early August 2019, the following interesting scientific papers have been composed, partly with input of ENVRI-FAIR WP9 partners, and published by Frontiers publishing in the context of the coming OceanOb 2019 Conference which will take place in September 2019 and which will give quite some attention to FAIRness of RIs. These publications can be retrieved at:

- https://www.frontiersin.org/articles/10.3389/fmars.2019.00442/full
- https://www.frontiersin.org/articles/10.3389/fmars.2019.00428/full
- https://www.frontiersin.org/articles/10.3389/fmars.2019.00440/full

3.2 Generic questionnaire about state of FAIRness in each RI

In agreement with WP5, early March 2019 the first questionnaire was sent out by MARIS to the marine RIs as part of Task 9.2. This survey aims to collect mostly qualitative information focused on implementation and solutions in place at the RIs.

Survey 1: the survey form can be found at:
https://docs.google.com/forms/d/e/1FAIpQLSdDHAcmsIfzGOpTNkl0vacWsc418KTFv0BDqDD1zMIfWwJdkQ/viewform
The marine RIs were asked to fill out one full form for an example dataset of each representative "data entity", where an entity can be considered as having its own metadata schema and or its own service. As example: SeaDataNet offers the CDI metadata catalogue for discovery and access of observation data, and it offers the Sextant catalogue with metadata of data products / data aggregations. These are two different entities, with a different metadata model and different data discovery and access service. Therefore, SeaDataNet completed survey 1 two times, one for each catalogue service.

All responses from the marine RIs were collected in a Google spreadsheet.

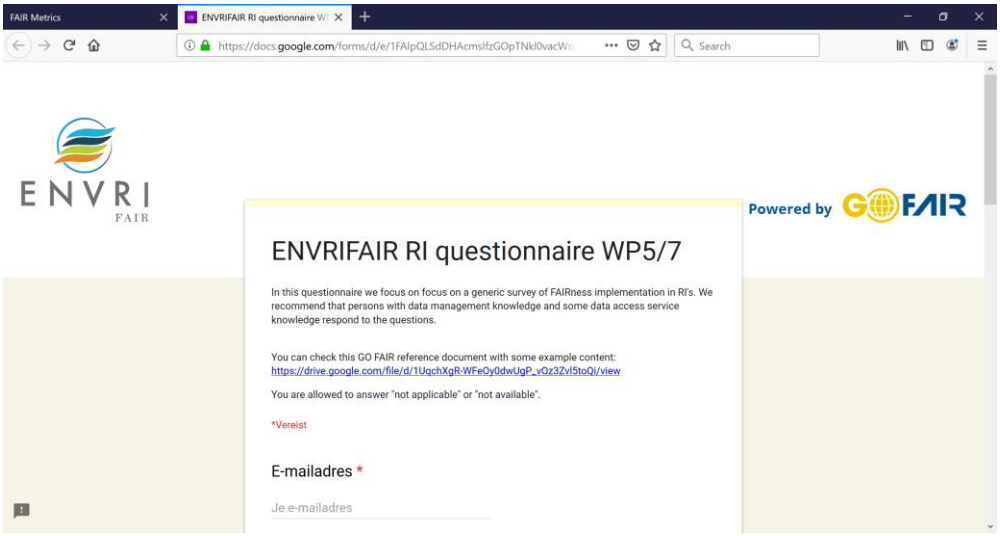

Figure 3. Homepage of questionnaire 1.

While responding to this first questionnaire, the partners gained more knowledge of the FAIR jargon that would help them in next steps. As preparation for discussion on possible implementations, also a second survey has been sent out with a more technical character. More details will follow below.

A comparable action was organized as part of WP5 through the linking pins for WP8, WP10, and WP11 for setting out the survey 1 at the other subdomains.

3.3 Initial analyses of the responses to the generic questionnaire

Per subdomain, end March/Early April 2019, completed surveys have been compiled by the WP8-WP11 linking pins in WP5, each for their RIs and subdomain. The subdomain coordinators, MARIS for the Marine subdomain (WP9), have made an initial analysis of the received responses together with the WP5 and WP7 teams, paying special attention to missing knowledge, misinterpreted questions, and clarity of responses. The initial completion

of survey 1 provided some first insights in the FAIRness of each of the RIs. By reviewing the responses, the WP5-WP7 partners became better informed about the leading data services and solutions as in use in the RIs. However, it also became clear that survey 1 had not been completed in a coherent way by all RIs and that questions had been interpreted in different ways. This underpinned the need for organizing subdomain meetings in which the RIs together would review the questions and given answers, possibly amending earlier answers, in order to achieve a more stable and balanced set of answers.

3.4 Second questionnaire aimed at collecting machine readable information

Early April 2019, as a second step a more detailed and more technical survey of the FAIRness was made available to the RIs via an online questionnaire based of the FAIR metrics. More information on the FAIR metrics was given by referring to the following document: https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen1/ALL.pdf

The RIs were urged to complete survey 2 before end April 2019, so that draft results would be available before the subdomain meetings.

The form of survey 2 can be found at:
https://docs.google.com/forms/d/e/1FAIpQLSeqvH7sQRYOfXpmyo_IGZ_4Ko_OGELKgIZrlTfDE1ovTO-oeg/viewform
The form contains instructions per question and reference to documentation of the concept.

As with survey 1, all responses have again been collected in a spreadsheet. In practice, many of the RIs have completed the survey for their RI as part of WP8-WP11 with guidance and support from the WP5-WP7 team. However, in a later stage during the review of the survey 2 form together with the GO FAIR office at the Marine subdomain meeting (see paragraph 3.5), it appeared that the second questionnaire was more a checklist on the common technical internet basics of the RI configurations and hardly on the technical aspects of the RI services. Therefore, in dialogue with WP5 it was decided to skip this questionnaire as it would not provide any more useful information, as the RIs are already mastering the internet basics and can be considered as advanced.

Only a small selection of questions of questionnaire 2 was kept and added to questionnaire 1, which resulted in May 2019 in one consistent questionnaire 1 (note: the URL provided in paragraph 3.2 leads to the final extended version).

3.5 Organise subdomain workshops

The next step in the analysis approach was to organize subdomain meetings to review together the results of the questionnaire(s). For the Marine subdomain, this meeting was organized by MARIS as part of Task 9.2 in Amsterdam – The Netherlands, 13 – 14 May 2019. It was the first subdomain meeting and functioned as a forerunner for the other planned subdomain meetings as the meeting was not only joined by marine RIs, but also by the leaders of WP5 – WP7 and by the NL GO FAIR office. In total 22 persons participated, most in person, and a few persons attended remotely. The Amsterdam meeting was used to learn more about FAIR and the GO FAIR activities, to review the analysis methodology of using the questionnaires, and to plan the follow-up actions for the actual analysis and drafting of deliverable D9.1.

The presentations of the meeting have been made available at the ENVRI-FAIR REDMINE system for internal use and also concise minutes were made. The agenda, list of participants and all presentations can be found at:

https://iagos-comm.iek.fz-juelich.de/projects/wp-09-marine-subdomain-implementation/dmsf?folder_id=240
while the minutes can be found at:
https://docs.google.com/document/d/11Fvu2ZVcVxxfqnVTy72uP2yWtlt4LmVCyWr8XiCRypc/edit?usp=sharing

A major part of the Marine subdomain meeting was spent on reviewing the two FAIRness survey forms in order to get a better understanding of the meaning of the questions and what kind of answers were expected for a consistent response. Having a representative of the NL GO FAIR office at the meeting was very helpful as they are the authors of the questionnaires and have undertaken multiple sessions with other RIs for completing the survey forms and making an initial analysis of the answers.

The completed questionnaire 1 of Euro-Argo – ERIC was taken as example and while doing, several answers were amended as certain questions had other meaning than earlier expected. As a result, the newly completed survey form 1 for Euro-Argo has become a reference model that the other RIs in the Marine subdomain and later also in the other subdomains should follow for reviewing and amending their earlier answers. At the meeting the status of the survey forms 1 for the other RIs were also discussed and reviewed to a certain level, thereby already getting some better understanding between the RIs about possible weaknesses / gaps and options for improvement. But, as an overall action, it was agreed that all RIs should review and amend their answers to questionnaire 1 completely in the following two weeks in order to provide a consistent and coherent set of answers for the further planned analysis of FAIRness per RI and their strengths and weaknesses. Moreover, it was decided to include in the survey form 1 not only information about the current situation, but also extra information about ideas, existing plans or implementation steps that RIs have underway or planned for improving their FAIRness as this would be very useful information for the roadmap (deliverable D9.1).

As already indicated in paragraph 3.4, in the Amsterdam meeting also time was spent time on reviewing questionnaire 2 together with the experts of the NL GO FAIR office, who are developing an online application for validating specific answers. However, since it became apparent that the second questionnaire is a checklist on common technical internet basics of the RI configuration for which all ENVRI-FAIR RIs will qualify with a high score, it was decided together with WP5 to withdraw this questionnaire 2. Only a small selection of questions of questionnaire 2 was kept and added to questionnaire 1, which thereafter was set-out again under the RIs of WP8-WP11 for reviewing and amending earlier answers, considering the Euro-Argo reference.

The final part of the Amsterdam meeting was dedicated to brainstorming about the next steps in the process towards deliverable D9.1 and D9.2, interacting between WP9 and WP5-WP7. As part of the analysis WP5 colleagues from Umweltbundesamt (Austria) proposed to develop and apply a FAIRness analysis tool which works with YAML files. More details on this are given in paragraph 3.6 and chapter 4.
Overall the Amsterdam meeting for Task 9.2 was an intensive workshop which provided much better insights for the marine RIs and representatives of WP5 – WP7 of what is required for FAIRness. Moreover, it resulted in a better understanding of the current status of the marine RIs and their ideas and/or plans for improving their FAIRness. Furthermore, it resulted in an amended questionnaire 1 which should be completed by all RIs, thereby considering the Euro-Argo reference form.

3.6 Next steps as undertaken towards the roadmap

In each subdomain, the completed forms of the extended survey 1 provide the basis input for the further analyses towards formulating their roadmap.

For the Marine subdomain and Task 9.2 the following analysis activities were undertaken for completing the Marine roadmap (deliverable D9.1):

1. the WP5 colleagues from Umweltbundesamt have developed a method to summarise direct answers with a FAIRness analysis tool, modelled on the extended survey form 1, by means of YAML files. In dialogue with WP5 the Task 9.2 lead agreed to develop and apply this tool. Using the completed forms from the marine RIs, this tool has generated resulting tables which more or less reflect the FAIRness strengths and weaknesses of each marine RI. More about this tool and the results is given in Chapter 4;
2. the YAML tool only uses the direct answers using a fixed schema while for a complete picture also the textual information in the forms of survey 1 is required, more focusing on the ideas and planned improvement activities as reported by the RIs. Therefore, in chapter 5 both will be combined in order to provide insight on non-covered weaknesses and to give an opinion about proposed ways forward, considering solutions as in use or planned by other RIs.

It is planned with WP5 and its linking pins for WP8-WP11 that each of the subdomains will execute these analysis activities in order to achieve roadmaps for each of the subdomains.

4. Quantified analysis

4.1 Explanation of analysis approach

A key task for ENVRI-FAIR in general is to assess and to monitor over the project's lifetime the FAIR maturity of participant RIs. This assessment enables the identification of current gaps and thus informs the customized consultation and support to RIs by the Information and Communications Technology (ICT) experts of the project. The assessment must be as efficient as possible, given the number of involved RIs, the fact that most RIs are distributed, and thus require different assessments for their centres (which may be in different domains, e.g. ICOS with its Atmospheric, Marine and Ecosystem Thematic Centres), and that such assessment should be performed repeatedly during the course of the project in order to monitor the development. Not only should it be efficient in collecting the required information, but also in analysing the collected information. Making the assessment efficient has been and continues to be a challenge led by WP5.

In this chapter the current implementation of the assessment is described for the Marine subdomain, explaining first the required information collection and then the information analysis (see Figure 4 for the schematic overview).

Information collection
In order to assess the FAIR maturity of RIs, it is important to collect current information about RI services and their data. Specifically, information is required about their data and metadata repositories and how these systems deal with the FAIR Data Principles. For instance, of interest is whether data repositories of an RI use persistent identifiers.

To collect such information, the basis has been a questionnaire originally developed by GO FAIR (see Chapter 3). This questionnaire had several purposes, including:

- Assess the state of RI data and services in terms of the FAIR data principles
- Detect information and implementation gaps
- Increase understanding of the FAIR principles and their advantages for the RIs
- Discover strengths
- Compare different implementations by RIs
- Evaluate possible technology take-ups for improvements
- Prioritize FAIR improvements
- Include chosen FAIR improvements in RI plans

All responses from the marine RIs were collected in a google spreadsheet. The spreadsheet tables can be found in Appendix 2. From a first analysis during preparation for the Marine subdomain Amsterdam workshop, it became quickly obvious that the responses to the questionnaires by RI representatives were not directly usable for downstream analysis without substantial post-processing and harmonization (e.g. of the vocabulary used). It thus became necessary to review the answers and to post-process them to extract the key information needed or to mark answers as insufficient. The result of such post-processing informed subsequent interaction with RI representatives aimed at addressing the insufficiently addressed questions so that the questionnaire could be completed and is quality controlled and improved. The interaction was done face-to-face during the subdomain meetings, and afterwards in videoconferences.

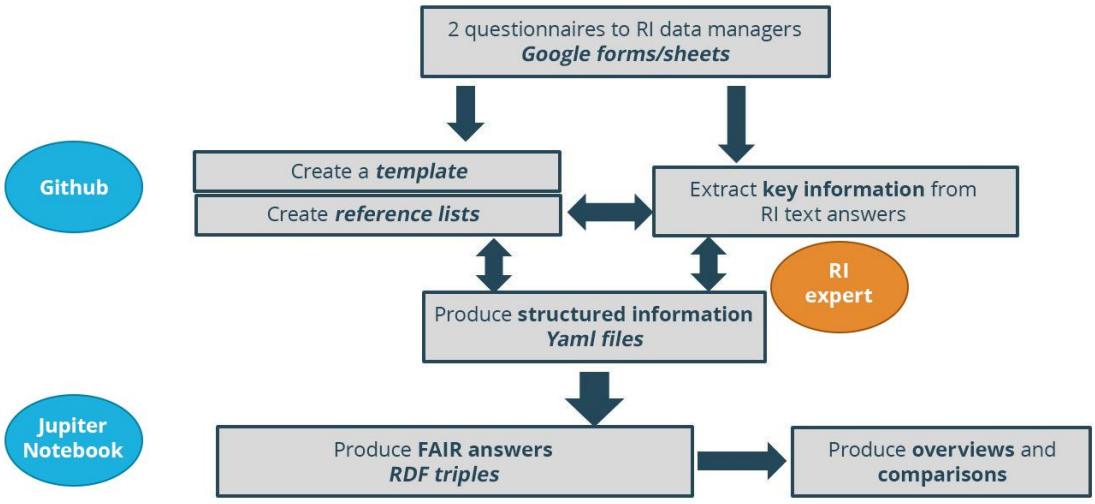

Figure 4. Approach from questionnaires to FAIRness overviews.

In order to support the face-to-face interviews with RI representatives aimed at resolving the outstanding issues in questionnaire responses, in the post-processing the answers from the questionnaires in XLS were converted and the extracted key information was transformed into a structured form in YAML format following a template, also written in YAML. This format was chosen for its conciseness and readability as well as for the fact that it requires minimal extra information to encode answers. And while making this conversion the answers were translated as much as possible from free text to reference lists (same name for same concept/responses), and if that was not possible at least to condensed answers. The original responses in XLS were kept (and used again in later stage), because they contain additional information about planned implementations, gaps a.o., but for the automated repeatable FAIRness check these are not used. This work thus supported the efficient editing of answers during face-to-face interviews.

The structured YAML format of the information collected through questionnaire responses also served a second purpose. Compared to questionnaire answers in natural language text, the structured information better suits as input to further processing of the information. Given the requirement to analyse the collected information efficiently, it was decided to build a database for this information. Concretely, a knowledge base was implemented in the form of a triple store using RDF as the data format. Hence, as an additional step the information in YAML was converted into RDF. Each YAML document was converted into an RDF document using a fully automated script implemented in Python as a Jupyter notebook that can be executed on EGI Notebooks service. The resulting YAML documents are provided in paragraph 4.3.

To give some more explanation of the terms used: a triple store or RDF store is a purpose-built database for the storage and retrieval of triples through semantic queries. A triple is a

data entity composed of subject-predicate object, like "Bob is 35" or "Bob knows Fred". Much like a relational database, one stores information in a triple store and retrieves it via a query language. Unlike a relational database, a triple store is optimized for the storage and retrieval of triples. In addition to queries, triples can usually be imported/exported using Resource Description Framework (RDF) and other formats. RDF is a family of World Wide Web Consortium (W3C) specifications and is in use as a general method for conceptual description or modeling of information that is implemented in web resources. It is also used in knowledge management applications. SparQL is a semantic query language for triple or RDF stores and facilitates retrieving and manipulating data stored in RDF format. For analysis use can be made of Jupyter notebooks, which is an open-source web application that allows users to create and share documents that contain live code, equations, visualisations and narrative text.

Information Analysis
The resulting RDF documents have been loaded into a triple store. An in-memory triple store implementation provided by the Python rdflib package has been used. By doing this, it gained the possibility to formulate declarative queries in SPARQL that implement user requests for information, not only about individual RIs but also across RIs. SPARQL queries have also been formulated in Jupyter. The results can be stored as Excel sheets and downloaded for future processing. The results of the first SparQL queries are included in paragraph 4.4.

Current and Future Developments
There is close contact with GO FAIR to share experiences with the questionnaire created to collect information about FAIRness and translate this into "metrics". ENVRI-FAIR members are now part of a working group to develop the next generation FAIR metrics questionnaire led by GO FAIR, and to bring in the knowledge of translating questionnaire answers into YAML and creating the SparQL queries. Later in the ENVRI-FAIR project the FAIRness of RI's will be measured again. At that time the new questionnaire will be used, while the current responses will be converted into the new data model for comparison.

4.2 Questionnaire responses in XLS files

As already described in Chapter 3 it was necessary to review and amend both the survey form 1 and the earlier received answers from the marine RIs as follow up to the Amsterdam Marine meeting. The final results of this effort by marine RIs are included in **Appendix 2**. Thereafter, these results have been used to convert to YAMLs. The textual content has been used in the analysis of gaps, strengths, and planned activities for increasing FAIRness, as will be detailed in Chapter 5.

4.3 YAML representation files

As explained above, the RI responses as present in the XLS files, have been converted into the structured YAML format. The basis for the YAML is a template and a related list of references that are allowed to be used in certain elements. The YAML template can be found on GITHUB: https://github.com/envri-fair/fairness-assessment/blob/master/template.yaml

The template provides guidance on:

- The field type (Boolean, free text, number, date, reference list value [vocabulary], etc.)
- The linked FAIR principle (F1, R1.1 etc). This is used later in sparQL queries for grouping answers.

Furthermore, the template shows clearly the structure of the information with a generic part describing the RI details, followed by the responses for each repository.

---

survey:
  date: yyyy-mm-dd
  version: number
  creator:
    name: free
    email:
infrastructure:
  acronym: ref
  name: free
  website:
  domain: list ref
  URL/IRI of dataset:
  URL of discovery portal:
  repositories: list
    URL: F4
    name: free
    kind: list ref
    allocation: ref F4
    software: list ref F4
    identifier: list
      kind: ref F1
      system: ref F1
      landing page: bool F1
      assigned: ref F1
      provider: ref F1
      includes metadata schema: list ref F2
    certification methods: list ref
    policies: list ref R1.1
    registries: list ref F4
    persistency-guaranty: ref A1
    access mechanisms:
      authentication method: ref A1.2
      access protocol URL: A1.1
      access without costs: bool A1.1
      own user database maintained: bool A1.2
      person identification system: ref A1.2
      major access technology supported: ref A1.1
      authorization technique: ref A1.2
      authorization for accessing content needed: bool A1.2
      data licenses in use: list ref R1.1
      data license IRI: R1.1
      metadata openly available: bool R1.1
    data:
      type name: list ref I1
      preferred formats: list
        format name: ref I1
        metadata types in data headers: list ref I1
      registered data schema: ref I1
      search on data: bool F4
    metadata:
      schema: list
        URL:
        name: ref I1
        provenance fields included: list ref F2
      machine readable provenance: bool R1.2
      categories defined in registries: bool I2
      PIDs included: bool F3

```
    primary storage format: ref I1
    export formats supported: list ref I1
    search engine indexing: bool F4
    exchange/harvesting methods: list ref I1
    local search engine URL: F4
    external search engine types supported: list ref F4
    access policy statements included: bool R1.1
    metadata longevity plan URL:
    machine actionable: bool I1
    IRI of machine-readable metadata of dataset: F2A
  vocabularies: list
    IRI: I2
    name: ref
    type: ref I1
    topic: ref I1
    specification language: ref I1
  data management plans:
    specific DMP tools used: ref
    data publishing steps applied: list ref
    compliance validation service: bool R1.3
  data processing:
    special data processing steps applied: list ref
    workflow frameworks applied: list ref
    distributed workflows tools used: list ref
    other analysis services offered: list ref A1
    data products offered: list ref
  fairness:
    data findability:
      data findable: bool
      gaps: list ref
    data accessibility:
      data accessible: bool
      gaps: list ref
    data interoperability:
      data interoperable: bool
      gaps: list ref
    data re-usability:
      data reusable: bool
      gaps: list ref
```

The YAML representations for all marine RI's can be found in Appendix 3, as well as on GitHub:
https://github.com/envri-fair/fairness-assessment/tree/marinecomplete/descriptions

4.4 Output of the SparQL queries (reporting tables)

The YAML files for each of the marine RIs have been converted to RDF files ("RDF store") that can be queried via SparQL. Keeping in mind that the goal of the exercise is to evaluate the FAIRness in a repeatable method, and finding gaps and strengths of each RI, useful queries have been created by WP5 to extract the FAIRness information. In next subchapters the results can be found, that are grouped based on the FAIR principles: https://www.go-fair.org/fair-principles/ (F1, F2, etc), as far as covered by the questions and answers for the marine RIs.

4.4.1 Findability

Table 1. F1. (Meta)data are assigned a globally unique and persistent identifier.

ID	RI	Predicate	Value
0	LIFEWATCH (Marine) Marine Data Archive	usesProvider	LocalService
1	LIFEWATCH (Marine) EUROBIS	usesProvider	DataCite
2	EMSO data	usesProvider	Planned
3	ICOS (Marine) Carbon Portal	usesProvider	DataCite
4	LIFEWATCH (Marine) Marine Data Archive	usesProvider	DataCite
5	ICOS (Marine) Carbon Portal	usesProvider	Handle
6	Euro-Argo Data	usesProvider	SEANOE
7	SeaDataNet Common Data Index (CDI)	usesProvider	EUDAT
8	SeaDataNet Central Data Products	usesProvider	DataCite
9	LIFEWATCH (Marine) EUROBIS	usesProvider	LocalService
10	LIFEWATCH (Marine) Marine Data Archive	isAssigned	automatically
11	LIFEWATCH (Marine) EUROBIS	isAssigned	manually
12	EMSO data	isAssigned	manually
13	ICOS (Marine) Carbon Portal	isAssigned	manually
14	LIFEWATCH (Marine) Marine Data Archive	isAssigned	manually
15	ICOS (Marine) Carbon Portal	isAssigned	automatically
16	Euro-Argo Data	isAssigned	automatically
17	SeaDataNet Common Data Index (CDI)	isAssigned	automatically
18	SeaDataNet Central Data Products	isAssigned	manually
19	LIFEWATCH (Marine) EUROBIS	isAssigned	automatically
20	LIFEWATCH (Marine) Marine Data Archive	usesIdentifierSystem	None
21	LIFEWATCH (Marine) EUROBIS	usesIdentifierSystem	DigitalObjectIdentifier
22	EMSO data	usesIdentifierSystem	Planned
23	ICOS (Marine) Carbon Portal	usesIdentifierSystem	DigitalObjectIdentifier
24	LIFEWATCH (Marine) Marine Data Archive	usesIdentifierSystem	DigitalObjectIdentifier
25	ICOS (Marine) Carbon Portal	usesIdentifierSystem	Handle
26	Euro-Argo Data	usesIdentifierSystem	DigitalObjectIdentifier
27	SeaDataNet Common Data Index (CDI)	usesIdentifierSystem	B2HANDLE
28	SeaDataNet Central Data Products	usesIdentifierSystem	DigitalObjectIdentifier
29	LIFEWATCH (Marine) EUROBIS	usesIdentifierSystem	None

Table 2. F2. Data are described with rich metadata.

ID	RI	Predicate	Value
1	LIFEWATCH (Marine) Marine Data Archive	includesMetadataSchema	None
2	ICOS (Marine) Carbon Portal	includesMetadataSchema	https://meta.icos-cp.eu/ontologies/cpmeta/instanceschema.rdf
3	LIFEWATCH (Marine) EUROBIS	includesMetadataSchema	None
4	SeaDataNet Common Data Index (CDI) metadata	includesMetadataSchema	http://schemas.seadatanet.org/Standards-Software/Metadata-formats/SDN2_CDI_ISO19139_10.0.1.xsd
5	EMSO data metadata	includesMetadataSchema	None
6	SeaDataNet Common Data Index (CDI) metadata	hasMachineReadableDatasetMetadata	https://cdi.seadatanet.org/report/15222/xml
7	EMSO data metadata	hasMachineReadableDatasetMetadata	None
8	LIFEWATCH (Marine) EUROBIS metadata	hasMachineReadableDatasetMetadata	http://www.lifewatch.be/en/imis?module=dataset&dasid=1841&show=eml
9	LIFEWATCH (Marine) Marine Data Archive metadata	hasMachineReadableDatasetMetadata	http://www.lifewatch.be/en/imis?module=dataset&dasid=1841&show=eml
10	SeaDataNet Central Data Products metadata	hasMachineReadableDatasetMetadata	https://sextant.ifremer.fr/geonetwork/srv/eng/xml.metadata.get?uuid=90ae7a06-8b08-4afe-83dd-ca92bc99f5c0
11	Euro-Argo Euro-Argo Data metadata	hasMachineReadableDatasetMetadata	https://doi.org/10.17882/42182
12	ICOS (Marine) Carbon Portal metadata	hasMachineReadableDatasetMetadata	https://hdl.handle.net/11676/-ffoiHjX5NDN0Vq_fKuVmas0

Table 3. F3. Metadata clearly and explicitly include the identifier of the data they describe.

ID	RI	Predicate	Value
1	SeaDataNet Common Data Index (CDI) metadata	persistentIdentifiersAreIncluded	true
2	EMSO data metadata	persistentIdentifiersAreIncluded	true
3	LIFEWATCH (Marine) EUROBIS metadata	persistentIdentifiersAreIncluded	true
4	LIFEWATCH (Marine) Marine Data Archive metadata	persistentIdentifiersAreIncluded	true
5	SeaDataNet Central Data Products metadata	persistentIdentifiersAreIncluded	true
6	Euro-Argo Data metadata	persistentIdentifiersAreIncluded	true
7	ICOS (Marine) Carbon Portal metadata	persistentIdentifiersAreIncluded	true

Table 4. F4. (Meta)data are registered or indexed in a searchable resource.

ID	RI	Predicate	Value
0	SeaDataNet Common Data Index (CDI) metadata	supportsExternalSearchEngineTypes	GeoPortal
1	SeaDataNet Common Data Index (CDI) metadata	supportsExternalSearchEngineTypes	IODEOceanDataPortal
2	EMSO data metadata	supportsExternalSearchEngineTypes	GeoPortal
3	EMSO data metadata	supportsExternalSearchEngineTypes	IODEOceanDataPortal
4	LIFEWATCH (Marine) EUROBIS metadata	supportsExternalSearchEngineTypes	GBIF
5	LIFEWATCH (Marine) EUROBIS metadata	supportsExternalSearchEngineTypes	EDMED
6	LIFEWATCH (Marine) EUROBIS metadata	supportsExternalSearchEngineTypes	DataCite
7	LIFEWATCH (Marine) EUROBIS metadata	supportsExternalSearchEngineTypes	GCMD
8	LIFEWATCH (Marine) Marine Data Archive metadata	supportsExternalSearchEngineTypes	Planned
9	SEADATANET Central Data Products metadata	supportsExternalSearchEngineTypes	None
10	Euro-Argo Data metadata	supportsExternalSearchEngineTypes	CswEndpoint
11	Euro-Argo Data metadata	supportsExternalSearchEngineTypes	GoogleDatasetSearch
12	ICOS (Marine) Carbon Portal metadata	supportsExternalSearchEngineTypes	SPARQLEndpoint
13	ICOS (Marine) Carbon Portal metadata	supportsExternalSearchEngineTypes	DataCite
14	SEADATANET data products repository	usesSoftware	Sextant
15	SEADATANET data repository	usesSoftware	B2SAFE
16	Euro-Argo repository	usesSoftware	NetAPP
17	EMSO repository	usesSoftware	ApacheHTTP
18	LIFEWATCH (Marine) repository	usesSoftware	RelationalDatabase
19	LIFEWATCH (Marine) repository	usesSoftware	iRODS
20	ICOS (Marine) repository	usesSoftware	iRODS
21	ICOS (Marine) repository	usesSoftware	B2SAFE
22	SEADATANET data products repository	hasAllocation	CentralDataRepository
23	SEADATANET data repository	hasAllocation	CentralDataRepository
24	Euro-Argo repository	hasAllocation	CentralDataRepository
25	EMSO repository	hasAllocation	CentralDataRepository
26	LIFEWATCH (Marine) repository	hasAllocation	CentralDataRepository
27	LIFEWATCH (Marine) repository	hasAllocation	CentralDataRepository
28	ICOS (Marine) repository	hasAllocation	CentralDataRepository
29	SEADATANET data products repository	searchOnData	Planned
30	SEADATANET data repository	searchOnData	Planned
31	Euro-Argo repository	searchOnData	true
32	EMSO repository	searchOnData	Planned
33	LIFEWATCH (Marine) repository	searchOnData	true
34	LIFEWATCH (Marine) EUROBIS repository	searchOnData	true
35	ICOS (Marine) repository	searchOnData	true
36	SEADATANET Common Data Index (CDI) metadata	hasLocalSearchEngineUrl	https://cdi.seadatanet.org/search

ID	RI	Predicate	Value
37	EMSO data metadata	hasLocalSearchEngineUrl	Planned
38	LIFEWATCH (Marine) EUROBIS metadata	hasLocalSearchEngineUrl	http://www.vliz.be/en/integrated-marine-information-system
39	LIFEWATCH (Marine) Marine Data Archive metadata	hasLocalSearchEngineUrl	http://www.vliz.be/en/integrated-marine-information-system
40	SEADATANET Central Data Products metadata	hasLocalSearchEngineUrl	https://www.seadatanet.org/Products#/search
41	ICOS (Marine) Carbon Portal metadata	hasLocalSearchEngineUrl	https://meta.icos-cp.eu/sparqlclient/
42	ICOS (Marine) Carbon Portal metadata	hasLocalSearchEngineUrl	https://data.icos-cp.eu/portal
43	SEADATANET data products repository	inRegistries	None
44	SEADATANET data repository	inRegistries	GEOSS
45	SEADATANET data repository	inRegistries	IODEOceanDataPortal
46	Euro-Argo repository	inRegistries	GEOSS
47	EMSO repository	inRegistries	None
48	LIFEWATCH (Marine) repository	inRegistries	FairsharingOrg
49	LIFEWATCH (Marine) repository	inRegistries	None
50	ICOS (Marine) repository	inRegistries	re3data
51	SEADATANET data products repository	hasRepositoryUrl	https://www.seadatanet.org/Products#/search
52	SEADATANET data repository	hasRepositoryUrl	https://cdi.seadatanet.org/search
53	Euro-Argo repository	hasRepositoryUrl	http://doi.org/10.17882/42182
54	EMSO repository	hasRepositoryUrl	http://data.emso.eu/files/
55	LIFEWATCH (Marine) repository	hasRepositoryUrl	http://www.eurobis.org/
56	LIFEWATCH (Marine) repository	hasRepositoryUrl	http://www.marinedataarchive.org/
57	ICOS (Marine) repository	hasRepositoryUrl	https://data.icos-cp.eu/portal

4.4.2 Accessibility

Table 5. A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.

ID	RI	Predicate	Value
0	LIFEWATCH (Marine) EUROBIS data processing	otherAnalysisServicesOffered	RPackages
1	LIFEWATCH (Marine) Marine Data Archive data processing	otherAnalysisServicesOffered	None
2	SEADATANET Common Data Index (CDI) data processing	otherAnalysisServicesOffered	Planned
3	EMSO data processing	otherAnalysisServicesOffered	Planned
4	Euro-Argo Data data processing	otherAnalysisServicesOffered	Spark
5	Euro-Argo Data data processing	otherAnalysisServicesOffered	Cassandra
6	Euro-Argo Data data processing	otherAnalysisServicesOffered	Pangeo
7	Euro-Argo Data data processing	otherAnalysisServicesOffered	Elasticsearch
8	ICOS (Marine) Carbon Portal	otherAnalysisServicesOffered	JupiterNotebooks

ID	RI	Predicate	Value
	data processing		
9	SEADATANET Central Data Products data processing	otherAnalysisServicesOffered	Planned
10	SEADATANET data products repository	hasPersistencyGuaranty	stewardship
11	SEADATANET data repository	hasPersistencyGuaranty	stewardship
12	Euro-Argo repository	hasPersistencyGuaranty	50 years
13	EMSO repository	hasPersistencyGuaranty	None
14	LIFEWATCH (Marine) repository	hasPersistencyGuaranty	stewardship
15	LIFEWATCH (Marine) repository	hasPersistencyGuaranty	stewardship
16	ICOS (Marine) repository	hasPersistencyGuaranty	20 years

Table 6. A1.1 The protocol is open, free, and universally implementable.

ID	RI	Predicate	Value
0	LIFEWATCH (Marine) EUROBIS access mechanism	accessWithoutCosts	true
1	LIFEWATCH (Marine) Marine Data Archive access mechanism	accessWithoutCosts	true
2	Euro-Argo Data access mechanism	accessWithoutCosts	true
3	SEADATANET Common Data Index (CDI) access mechanism	accessWithoutCosts	true
4	ICOS (Marine) Carbon Portal access mechanism	accessWithoutCosts	true
5	SEADATANET Central Data Products access mechanism	accessWithoutCosts	true
6	EMSO data access mechanism	accessWithoutCosts	true
7	LIFEWATCH (Marine) EUROBIS access mechanism	supportsAccessTechnology	HTTPPOST
8	LIFEWATCH (Marine) Marine Data Archive access mechanism	supportsAccessTechnology	HTTPPOST
9	Euro-Argo Data access mechanism	supportsAccessTechnology	FTP
10	SEADATANET Common Data Index (CDI) access mechanism	supportsAccessTechnology	CAS
11	ICOS (Marine) Carbon Portal access mechanism	supportsAccessTechnology	HTTPGET
12	SEADATANET Central Data Products access mechanism	supportsAccessTechnology	CAS
13	EMSO data access mechanism	supportsAccessTechnology	HTTP
14	LIFEWATCH (Marine) EUROBIS access mechanism	hasAccessProtocolUrl	https://doi.org/10.17487%2FRFC2616
15	LIFEWATCH (Marine) Marine Data Archive access mechanism	hasAccessProtocolUrl	https://doi.org/10.17487%2FRFC2616
16	Euro-Argo Data access mechanism	hasAccessProtocolUrl	https://doi.org/10.17882/42182
17	SEADATANET Common Data Index (CDI) access mechanism	hasAccessProtocolUrl	https://www.seadatanet.org/Data-Access/Data-policy
18	ICOS (Marine) Carbon Portal access mechanism	hasAccessProtocolUrl	https://data.icos-cp.eu/licence
19	SEADATANET Central Data Products access mechanism	hasAccessProtocolUrl	https://www.seadatanet.org/Data-Access/Data-

ID	RI	Predicate	Value
			policy
20	EMSO data access mechanism	hasAccessProtocolUrl	VOID

Table 7. A1.2 The protocol allows for an authentication and authorisation procedure, where necessary.

ID	RI	Predicate	Value
0	LIFEWATCH (Marine) EUROBIS access mechanism	hasAuthenticationMethod	account manually
1	LIFEWATCH (Marine) Marine Data Archive access mechanism	hasAuthenticationMethod	account manually
2	Euro-Argo Data access mechanism	hasAuthenticationMethod	VOID
3	SEADATANET Common Data Index (CDI) access mechanism	hasAuthenticationMethod	Marine-ID
4	ICOS (Marine) Carbon Portal access mechanism	hasAuthenticationMethod	OAuth paired with eduGain
5	ICOS (Marine) Carbon Portal access mechanism	hasAuthenticationMethod	Email/password
6	ICOS (Marine) Carbon Portal access mechanism	hasAuthenticationMethod	ORCID
7	SEADATANET Central Data Products access mechanism	hasAuthenticationMethod	Marine-ID
8	EMSO data access mechanism	hasAuthenticationMethod	VOID
9	LIFEWATCH (Marine) EUROBIS access mechanism	contentAccessAuthorizationRequired	false
10	LIFEWATCH (Marine) Marine Data Archive access mechanism	contentAccessAuthorizationRequired	false
11	Euro-Argo Data access mechanism	contentAccessAuthorizationRequired	false
12	SEADATANET Common DAta Index (CDI) access mechanism	contentAccessAuthorizationRequired	false
13	ICOS (Marine) Carbon Portal access mechanism	contentAccessAuthorizationRequired	false
14	SEADATANET Central Data Products access mechanism	contentAccessAuthorizationRequired	false
15	EMSO data access mechanism	contentAccessAuthorizationRequired	false
16	LIFEWATCH (Marine) EUROBIS access mechanism	maintainsOwnUserDatabase	true
17	LIFEWATCH (Marine) Marine Data Archive access mechanism	maintainsOwnUserDatabase	true
18	Euro-Argo Data access mechanism	maintainsOwnUserDatabase	false
19	SEADATANET Common Data Index (CDI) access mechanism	maintainsOwnUserDatabase	true
20	ICOS (Marine) Carbon Portal access mechanism	maintainsOwnUserDatabase	true
21	SEADATANET Central Data Products access mechanism	maintainsOwnUserDatabase	true
22	EMSO data access mechanism	maintainsOwnUserDatabase	true
23	LIFEWATCH (Marine) EUROBIS access mechanism	usesAuthorisationTechnique	SSQLService
24	LIFEWATCH (Marine) Marine Data Archive access mechanism	usesAuthorisationTechnique	SSQLService
25	Euro-Argo Data access mechanism	usesAuthorisationTechnique	VOID

ID	RI	Predicate	Value
26	SEADATANET Common Data Index (CDI) access mechanism	usesAuthorisationTechnique	OAUTH
27	ICOS (Marine) Carbon Portal access mechanism	usesAuthorisationTechnique	OAUTH
28	SEADATANET Central Data Products access mechanism	usesAuthorisationTechnique	OAUTH
29	EMSO data access mechanism	usesAuthorisationTechnique	VOID

A2. Metadata are accessible, even when the data are no longer available
This item is not covered in questionnaire. However, one should be aware that keeping all metadata and landing pages after termination of the data or even termination of the data repository, is a requirement for FAIRness. It is very relevant if data has been used in publications or data products. Then the metadata should still be available as part of provenance information, for instance for evaluating the use.

4.4.3 Interoperability

Table 8. I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

ID	RI	Predicate	Value
0	Euro-Argo Data vocabularies	hasTopic	DomainSpecific
1	SEADATANET Central Data Products vocabularies	hasTopic	ProjectSpecific
2	ICOS (Marine) Carbon Portal vocabularies	hasTopic	General
3	LIFEWATCH (Marine) EUROBIS vocabularies	hasTopic	DomainSpecific
4	LIFEWATCH (Marine) Marine Data Archive vocabularies	hasTopic	DomainSpecific
5	SEADATANET Common Data Index (CDI) vocabularies	hasTopic	ProjectSpecific
6	ICOS (Marine) Carbon Portal vocabularies	hasTopic	General
7	Euro-Argo Data vocabularies	hasTopic	ProjectSpecific
8	LIFEWATCH (Marine) EUROBIS vocabularies	hasTopic	DomainSpecific
9	EMSO data vocabularies	hasTopic	DomainSpecific
10	LIFEWATCH (Marine) EUROBIS vocabularies	hasTopic	DomainSpecific
11	SEADATANET Central Data Products vocabularies	hasTopic	INSPIRE
12	LIFEWATCH (Marine) Marine Data Archive vocabularies	hasTopic	DomainSpecific
13	SEADATANET Common Data Index (CDI) vocabularies	hasTopic	INSPIRE
14	SEADATANET Central Data Products vocabularies	hasTopic	DomainSpecific
15	ICOS (Marine) Carbon Portal vocabularies	hasTopic	DomainSpecific
16	SEADATANET Common Data Index (CDI) vocabularies	hasTopic	DomainSpecific

ID	RI	Predicate	Value
17	LIFEWATCH (Marine) Marine Data Archive vocabularies	hasTopic	DomainSpecific
18	Euro-Argo Data vocabularies	hasSpecificationLanguage	http://cfconventions.org/Data/cf-standard-names/64/src/cf-standard-name-table.xml
19	SEADATANET Central Data Products vocabularies	hasSpecificationLanguage	RDF
20	ICOS (Marine) Carbon Portal vocabularies	hasSpecificationLanguage	Owl
21	SEADATANET Common Data Index (CDI) vocabularies	hasSpecificationLanguage	RDF
22	ICOS (Marine) Carbon Portal vocabularies	hasSpecificationLanguage	RDF
23	Euro-Argo Data vocabularies	hasSpecificationLanguage	http://cfconventions.org/Data/cf-standard-names/64/src/cf-standard-name-table.xml
24	EMSO data vocabularies	hasSpecificationLanguage	https://www.w3.org/RDF/
25	SEADATANET Central Data Products vocabularies	hasSpecificationLanguage	https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/
26	SEADATANET Common Data Index (CDI) vocabularies	hasSpecificationLanguage	https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/
27	SEADATANET Common Data Index (CDI) metadata	hasPrimaryStorageFormat	ExtensibleMarkupLanguage
28	EMSO data metadata	hasPrimaryStorageFormat	ExtensibleMarkupLanguage
29	LIFEWATCH (Marine) EUROBIS metadata	hasPrimaryStorageFormat	RelationalDatabase
30	LIFEWATCH (Marine) Marine Data Archive metadata	hasPrimaryStorageFormat	RelationalDatabase
31	SEADATANET Central Data Products metadata	hasPrimaryStorageFormat	ExtensibleMarkupLanguage
32	Euro-Argo Data metadata	hasPrimaryStorageFormat	NetcdfCfArgo
33	ICOS (Marine) Carbon Portal metadata	hasPrimaryStorageFormat	RDF
34	SEADATANET Common Data Index (CDI) metadata	isMachineActionable	true
35	EMSO data metadata	isMachineActionable	true
36	LIFEWATCH (Marine) EUROBIS metadata	isMachineActionable	partially
37	LIFEWATCH (Marine) Marine Data Archive metadata	isMachineActionable	partially
38	SEADATANET Central Data Products metadata	isMachineActionable	partially
39	Euro-Argo Data metadata	isMachineActionable	true
40	ICOS (Marine) Carbon Portal metadata	isMachineActionable	true
41	EMSO data metadata schema	hasSchemaName	ISO_19115_19139
42	SEADATANET Central Data Products metadata	hasSchemaName	SeaDataNetCommunityProfile

ID	RI	Predicate	Value
	schema		
43	LIFEWATCH (Marine) EUROBIS metadata schema	hasSchemaName	EML
44	SEADATANET Common Data Index (CDI) metadata schema	hasSchemaName	ISO_19115_19139
45	LIFEWATCH (Marine) Marine Data Archive metadata schema	hasSchemaName	EML
46	LIFEWATCH (Marine) Marine Data Archive metadata schema	hasSchemaName	ISO_19115
47	SEADATANET Central Data Products metadata schema	hasSchemaName	ISO_19115_19139
48	SEADATANET Common Data Index (CDI) metadata schema	hasSchemaName	SeaDataNetCommunityProfile
49	ICOS (Marine) Carbon Portal metadata schema	hasSchemaName	ISO_19115
50	Euro-Argo Data metadata schema	hasSchemaName	NetCDFCFChecker
51	Euro-Argo Data metadata schema	hasSchemaName	ArgoUserManual
52	LIFEWATCH (Marine) EUROBIS metadata schema	hasSchemaName	ISO_19115
53	ICOS (Marine) Carbon Portal metadata schema	hasSchemaName	GeoDCAT
54	LIFEWATCH (Marine) EUROBIS data	dataSchemaIsRegistered	DarwinCoreArchive
55	SEADATANET Central Data Products data	dataSchemaIsRegistered	SEADATANET NetCDF CF
56	ICOS (Marine) Carbon Portal data	dataSchemaIsRegistered	None
57	EMSO data data	dataSchemaIsRegistered	None
58	LIFEWATCH (Marine) Marine Data Archive data	dataSchemaIsRegistered	None
59	ICOS (Marine) Carbon Portal data	dataSchemaIsRegistered	None
60	LIFEWATCH (Marine) EUROBIS data	dataSchemaIsRegistered	None
61	SEADATANET Common Data Index (CDI) data	dataSchemaIsRegistered	SEADATANET ODV
62	EMSO data data	dataSchemaIsRegistered	Planned
63	LIFEWATCH (Marine) EUROBIS data	dataSchemaIsRegistered	OBISEventScheme
64	SEADATANET Common Data Index (CDI) data	dataSchemaIsRegistered	SEADATANET NetCDF CF
65	SEADATANET Common Data Index (CDI) data	dataSchemaIsRegistered	MedAtlas
66	SEADATANET Common Data Index (CDI) metadata	hasHarvestingMethods	OGC_CSW
67	EMSO data metadata	hasHarvestingMethods	OAI-PMH
68	EMSO data metadata	hasHarvestingMethods	OGC_CSW
69	LIFEWATCH (Marine)	hasHarvestingMethods	Planned

ID	RI	Predicate	Value
	EUROBIS metadata		
70	LIFEWATCH (Marine) Marine Data Archive metadata	hasHarvestingMethods	Planned
71	SEADATANET Common Data Index (CDI) metadata	hasHarvestingMethods	OAI-PMH
72	SEADATANET Central Data Products metadata	hasHarvestingMethods	OGC_CSW
73	Euro-Argo Data metadata	hasHarvestingMethods	OpenDap
74	Euro-Argo Data metadata	hasHarvestingMethods	OAI-PMH
75	Euro-Argo Data metadata	hasHarvestingMethods	OGC_CSW
76	ICOS (Marine) Carbon Portal metadata	hasHarvestingMethods	SPARQLEndpoint
77	SEADATANET Central Data Products metadata	hasHarvestingMethods	OpenDap
78	SEADATANET Common Data Index (CDI) metadata	supportedExportFormats	ExtensibleMarkupLanguage
79	SEADATANET Common Data Index (CDI) metadata	supportedExportFormats	HyperTextMarkupLanguage
80	EMSO data metadata	supportedExportFormats	ExtensibleMarkupLanguage
81	EMSO data metadata	supportedExportFormats	HyperTextMarkupLanguage
82	LIFEWATCH (Marine) EUROBIS metadata	supportedExportFormats	HyperTextMarkupLanguage
83	LIFEWATCH (Marine) EUROBIS metadata	supportedExportFormats	Json
84	LIFEWATCH (Marine) EUROBIS metadata	supportedExportFormats	RSS
85	LIFEWATCH (Marine) EUROBIS metadata	supportedExportFormats	ExtensibleMarkupLanguage
86	LIFEWATCH (Marine) Marine Data Archive metadata	supportedExportFormats	ExtensibleMarkupLanguage
87	LIFEWATCH (Marine) Marine Data Archive metadata	supportedExportFormats	RSS
88	LIFEWATCH (Marine) Marine Data Archive metadata	supportedExportFormats	Json
89	LIFEWATCH (Marine) Marine Data Archive metadata	supportedExportFormats	HyperTextMarkupLanguage
90	SEADATANET Central Data Products metadata	supportedExportFormats	ExtensibleMarkupLanguage
91	SEADATANET Central Data Products metadata	supportedExportFormats	HyperTextMarkupLanguage
92	Euro-Argo Data metadata	supportedExportFormats	NetCDF
93	Euro-Argo Data metadata	supportedExportFormats	CommaSeparatedValues
94	ICOS (Marine) Carbon Portal metadata	supportedExportFormats	Txt
95	ICOS (Marine) Carbon Portal metadata	supportedExportFormats	ExtensibleMarkupLanguage
96	ICOS (Marine) Carbon Portal metadata	supportedExportFormats	Json
97	ICOS (Marine) Carbon Portal metadata	supportedExportFormats	HyperTextMarkupLanguage
98	ICOS (Marine) Carbon	supportedExportFormats	Turtle

ID	RI	Predicate	Value
	Portal metadata		
99	SEADATANET Common Data Index (CDI) metadata	supportedExportFormats	Json

Table 9. I2. (Meta)data use vocabularies that follow FAIR principles.

ID	RI	Predicate	Value
0	SEADATANET Common Data Index (CDI) metadata	categoriesAreDefinedInRegistries	true
1	EMSO data metadata	categoriesAreDefinedInRegistries	true
2	LIFEWATCH (Marine) EUROBIS metadata	categoriesAreDefinedInRegistries	true
3	LIFEWATCH (Marine) Marine Data Archive metadata	categoriesAreDefinedInRegistries	true
4	SEADATANET Central Data Products metadata	categoriesAreDefinedInRegistries	true
5	Euro-Argo Data metadata	categoriesAreDefinedInRegistries	Planned
6	ICOS (Marine) Carbon Portal metadata	categoriesAreDefinedInRegistries	true
7	Euro-Argo Data vocabularies	hasVocabularyIri	http://vocab.nerc.ac.uk/collection/P01/current/PSALST01/
8	SEADATANET Central Data Products vocabularies	hasVocabularyIri	https://edmo.seadatanet.org
9	ICOS (Marine) Carbon Portal vocabularies	hasVocabularyIri	http://www.w3.org/ns/prov
10	LIFEWATCH (Marine) EUROBIS vocabularies	hasVocabularyIri	http://www.marinespecies.org/
11	LIFEWATCH (Marine) Marine Data Archive vocabularies	hasVocabularyIri	http://www.marinespecies.org/
12	SEADATANET Common Data Index (CDI) vocabularies	hasVocabularyIri	https://edmerp.seadatanet.org
13	SEADATANET Central Data Products vocabularies	hasVocabularyIri	https://cdi.seadatanet.org
14	ICOS (Marine) Carbon Portal vocabularies	hasVocabularyIri	http://purl.org/dc/elements/1.1/
15	Euro-Argo Data vocabularies	hasVocabularyIri	http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P06
16	LIFEWATCH (Marine) EUROBIS vocabularies	hasVocabularyIri	http://www.marinespecies.org/
17	EMSO data vocabularies	hasVocabularyIri	http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery_1-3
18	LIFEWATCH (Marine) EUROBIS vocabularies	hasVocabularyIri	http://www.marineregions.org/
19	SEADATANET Central Data Products vocabularies	hasVocabularyIri	ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml
20	LIFEWATCH (Marine) Marine Data Archive vocabularies	hasVocabularyIri	http://www.marinespecies.org/
21	SEADATANET Common Data Index (CDI) vocabularies	hasVocabularyIri	ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml

ID	RI	Predicate	Value
22	SEADATANET Central Data Products vocabularies	hasVocabularyIri	https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/
23	ICOS (Marine) Carbon Portal vocabularies	hasVocabularyIri	http://meta.icos-cp.eu/ontologies/cpmeta/
24	SEADATANET Common DAta Index (CDI) vocabularies	hasVocabularyIri	https://edmo.seadatanet.org
25	SEADATANET Common Data Index (CDI) vocabularies	hasVocabularyIri	http://vocab.nerc.ac.uk/collection/V22/current/
26	LIFEWATCH (Marine) Marine Data Archive vocabularies	hasVocabularyIri	http://www.marineregions.org/
27	SEADATANET Common Data Index (CDI) vocabularies	hasVocabularyIri	https://csr.seadatanet.org
28	SEADATANET Common Data Index (CDI) vocabularies	hasVocabularyIri	https://edmed.seadatanet.org

I3. (Meta)data include qualified references to other (meta)data
<not covered by questions/mapping>

4.4.4 Reusability

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

Table 10. R1.1. (Meta)data are released with a clear and accessible data usage license.

ID	RI	Predicate	Value
0	SEADATANET Common Data Index (CDI) metadata	includesAccessPolicyStatements	true
1	EMSO data metadata	includesAccessPolicyStatements	true
2	LIFEWATCH (Marine) EUROBIS metadata	includesAccessPolicyStatements	partially
3	LIFEWATCH (Marine) Marine Data Archive metadata	includesAccessPolicyStatements	partially
4	SEADATANET Central Data Products metadata	includesAccessPolicyStatements	true
5	Euro-Argo Data metadata	includesAccessPolicyStatements	true
6	ICOS (Marine) Carbon Portal metadata	includesAccessPolicyStatements	true
7	SEADATANET data repository	hasPolicies	DataUsage
8	SEADATANET data products repository	hasPolicies	DataUsage
9	Euro-Argo repository	hasPolicies	DataAccess
10	EMSO repository	hasPolicies	None
11	LIFEWATCH (Marine) repository	hasPolicies	TermsOfUse
12	LIFEWATCH (Marine) repository	hasPolicies	FAQ
13	LIFEWATCH (Marine) repository	hasPolicies	TermsOfUse
14	LIFEWATCH (Marine)	hasPolicies	FAQ

ID	RI	Predicate	Value
	repository		
15	ICOS (Marine) repository	hasPolicies	Usage
16	LIFEWATCH (Marine) EUROBIS access mechanism	usesDataLicenses	CCBY_CC0
17	LIFEWATCH (Marine) Marine Data Archive access mechanism	usesDataLicenses	CCBY_CC0
18	Euro-Argo Data access mechanism	usesDataLicenses	CC-BY4.0
19	SEADATANET Common Data Index (CDI) access mechanism	usesDataLicenses	LocalLicense
20	ICOS (Marine) Carbon Portal access mechanism	usesDataLicenses	CC-BY4.0
21	SEADATANET Central Data Products access mechanism	usesDataLicenses	LocalLicense
22	EMSO data access mechanism	usesDataLicenses	CC-BY4.0
23	LIFEWATCH (Marine) EUROBIS access mechanism	openAccessMetadata	true
24	LIFEWATCH (Marine) Marine Data Archive access mechanism	openAccessMetadata	true
25	Euro-Argo Data access mechanism	openAccessMetadata	true
26	SEADATANET Common Data Index (CDI) access mechanism	openAccessMetadata	true
27	ICOS (Marine) Carbon Portal access mechanism	openAccessMetadata	true
28	SEADATANET Central Data Products access mechanism	openAccessMetadata	true
29	EMSO data access mechanism	openAccessMetadata	true
30	Euro-Argo Data access mechanism	dataLicenseIri	https://creativecommons.org/licenses/by/4.0/
31	SEADATANET Common Data Index (CDI) access mechanism	dataLicenseIri	https://www.seadatanet.org/Data-Access/License/1.0
32	ICOS (Marine) Carbon Portal access mechanism	dataLicenseIri	https://creativecommons.org/licenses/by/4.0/
33	SEADATANET Central Data Products access mechanism	dataLicenseIri	https://www.seadatanet.org/Data-Access/License/1.0

Table 11. R1.2. (Meta)data are associated with detailed provenance.

ID	RI	Predicate	Value
0	EMSO data metadata schema	includesProvenanceFields	TextOnly
1	SEADATANET Central Data Products metadata schema	includesProvenanceFields	Planned
2	LIFEWATCH (Marine) EUROBIS metadata schema	includesProvenanceFields	TextOnly
3	SEADATANET Common Data Index (CDI) metadata schema	includesProvenanceFields	TextOnly
4	LIFEWATCH (Marine) Marine Data Archive metadata schema	includesProvenanceFields	TextOnly
5	LIFEWATCH (Marine) Marine Data Archive metadata schema	includesProvenanceFields	TextOnly
6	SEADATANET Central Data Products metadata schema	includesProvenanceFields	TextOnly
7	SEADATANET Common Data Index (CDI) metadata schema	includesProvenanceFields	Planned
8	ICOS (Marine) Carbon Portal metadata schema	includesProvenanceFields	TextOnly
9	Euro-Argo Data metadata schema	includesProvenanceFields	codes
10	Euro-Argo Data metadata schema	includesProvenanceFields	TextOnly
11	LIFEWATCH (Marine) EUROBIS metadata schema	includesProvenanceFields	TextOnly
12	EMSO data metadata schema	includesProvenanceFields	Planned
13	ICOS (Marine) Carbon Portal metadata schema	includesProvenanceFields	SimplifiedProvO
14	SEADATANET Common Data Index (CDI) metadata	http://envri.eu/ns/hasMachineReadableProvenance	FALSE
15	EMSO data metadata	http://envri.eu/ns/hasMachineReadableProvenance	FALSE
16	LIFEWATCH (Marine) EUROBIS metadata	http://envri.eu/ns/hasMachineReadableProvenance	FALSE
17	LIFEWATCH (Marine) Marine Data Archive metadata	http://envri.eu/ns/hasMachineReadableProvenance	FALSE
18	SEADATANET Central Data Products metadata	http://envri.eu/ns/hasMachineReadableProvenance	FALSE
19	ICOS (Marine) Carbon Portal metadata	http://envri.eu/ns/hasMachineReadableProvenance	FALSE

Table 12. R1.3. (Meta)data meet domain-relevant community standards.

ID	RI	Predicate	Value
0	Euro-Argo Data data management plans	hasComplianceValidationService	TRUE
1	LIFEWATCH Marine Data Archive data management plans	hasComplianceValidationService	FALSE
2	EUROBIS data management plans	hasComplianceValidationService	FALSE
3	EMSO data data management plans	hasComplianceValidationService	FALSE
4	SEADATANET Central Data Products data management plans	hasComplianceValidationService	FALSE
5	SEADATANET Common Data Index (CDI) data management plans	hasComplianceValidationService	FALSE
6	Carbon (Marine) Portal data management plans	hasComplianceValidationService	FALSE

5. Analysis of gaps and solutions

Combining the automatic SparQL results, the more specific feedback in the XLS's per RI, and the RIs foreseen activities from the description of work, this chapter provides an overview of the identified strengths and weaknesses. Moreover, as part of the roadmap it summarises planned and/or suggested directions for solutions to overcome weaknesses. These plans and suggestions will be evaluated and if positive, worked out into detailed specifications in the following months as part of formulating Deliverable 9.2 - Marine subdomain implementation plan.

5.1 ICOS (Marine)

5.1.1 Gaps and planned activities

Findability

- Search at data level, such as finding all observations with a specific parameter value higher than X; (developing this function is already planned by ICOS [Marine])

Accessibility:

- Access policy statements are implicit part of the portal metadata, but is included in the datacite DOI metadata. Access of any data object through the portal enforces the acceptance of the data license.

Interoperability:

- Schemas could be registered in common registry
- Needs to work more on the metadata mode to support better the attribution and provenance information, applying relevant standards (especially wider use of vocabularies) and making sure the definitions and attribute names used are registered.

Reusability:

- Need to include more machine interpretable provenance information in the "to be enriched" metadata (including more links to ICOS observation and data processing protocols).
- Missing compliance validation service

5.1.2 Strengths

- Strong data identification: all data objects from raw data up to elaborated products are minted Handles and DOis: vital for transparency, reproducibility and reusability
- Focus on attribution of data to data providers coupled to data usage statistics as part of the metadata
- Data and metadata are well findable and accessible
- Linked data implementation, RDF and SparQL endpoints
- Use of identifiers and persistent storage
- Open data access and user registration
- Provenance using Simplified PROV-O format
- Portal is multi-domain integrating atmosphere, marine and ecosystem data in one system using a common ontology

5.2 Euro-Argo

5.2.1 Gaps and planned activities

Findability

- The global Argo dataset is well findable. However, a rich and efficient search service at data level is missing; a local search engine should be developed.

Accessibility:

- Need for a rich and efficient local search engine (planned using Elastic search).
- On top of the collection of NetCDF files, API cloud services should be developed/enhanced and implemented for data discovery, visualization, download, subscription with rich sub-setting for:
    - Individual scientist end users, with direct access to the original Argo NetCDF files, through a web GUI (Graphic User Interface)
    - Virtual Research Environment with cloud data and metadata (API queries instead of NetCDF files transfers)
- Improvements in the field of subsetting access with OGC WMS and SOS services.
- Introduce semantic web capabilities through Linked data/SparQL endpoint development.

Interoperability:

- Schemas to be registered in common registry (e.g. on github). Apart from that implementing a data processing HUB: public and shared codes for tools such as data processing chains (Argo floats decoders), file format checkers, decimation or standardization services.
- Argo vocabulary tables are not yet implemented in the SeaDataNet vocabulary service, and not yet harmonised with other vocabularies. Euro-Argo will implement a vocabulary server for Argo metadata to solve this. The provenance of data in the Argo Data System is underpinned by rich metadata which is standardised across the data system using vocabularies currently held in manuals and associated spreadsheets. The accuracy, controlled evolution and semantic value of this metadata will be further enhanced by migrating to a controlled vocabulary management environment and server compliant with W3C standards.
- The Argo vocabulary server will also lead to having all categories in the schema registered.

Reusability:

- Provenance info is now text only, should be coded and registered in the SeaDataNet Vocabulary service (see above).

5.2.2 Strengths

- Metadata and global dataset well findable and accessible
- Very open data access and no user registration
- Supporting many protocols to access data and metadata
- Analytics of use via Spark, Cassandra, Elastic

5.3 SeaDataNet

For SeaDataNet two core data repositories have been analysed: CDI metadata catalogue to discover and access data and the Sextant dataproduct catalogue to discover and access data products (including data collections).

5.3.1 Gaps and planned activities

Findability

- none

Accessibility

- Sextant and CDI: search on data via ERDDAP (planned)
- CDI: export formats – Json and RDF (planned)
- CDI: metadata exchange/access - Extension with SPARQL endpoint (planned)
- CDI: API development for improved machine access
- Sextant: Registration of repository in official registry
- Sextant and CDI: Explicit persistency policy in metadata
- Sextant: Missing access to dataplots via WPS

Interoperability

- CDI: Storage format of metadata: Extension with RDF - SPARQL (planned)
- Sextant: Not all categories in metadata marked up with vocabularies, some still in text.

Reusability

- CDI: Categories mostly supported by controlled vocabularies, but a few with free text (especially concerning provenance and quality)
- Sextant and CDI: Provenance info - Metadata enrichment planned for structured provenance metadata, using linked data principles and SeaDataNet directories and vocabularies. Machine readable and interpretable information needed e.g. for quality info, processing info. This info is known but not captured and provided in structured metadata.

5.3.2 Strengths

For both repositories:

- Data and metadata well findable
- Linked data representation in development. SEADATANET already maintains pan-European metadata directories with user interfaces for cruises (CSR), data collections (EDMED), projects (EDMERP), organisations (EDMO), observing systems (EDIOS), and data sets (CDI). These directories are linked where appropriate and make use of a rich set of controlled vocabularies for semantic interoperability. In SeaDataCloud developments are underway for adopting 'Linked Data' principles, whereby each of the Directories has been modelled in Linked Data terms using existing patterns, and RDF - SPARQL endpoints are being set up for publishing the Directories as machine readable Linked Data resources. This will already improve the machine findability of each Directory considerably.
- Extensive management and use of vocabularies, also available for expansion and use by other RI's (e.g. for provenance metadata)
- Workflows in place for metadata and data control before publishing

- Use of PID's (CDI) and DOI (Sextant)
- User authentication system for monitoring use

5.4 LifeWatch (Marine)

For LifeWatch (Marine) two important repositories were analysed: Marine Data Archive (MDA) and EUROBIS.

5.4.1 Gaps and planned activities

Findability
- MDA and EUROBIS: Use of identifier system, PID's instead of/on top of of local IMIS ID's

Accessibility
- MDA and EUROBIS: Overview of tools, services and data resources: Discovery metadata for the data resources will be standardized according to generally accepted metadata standards (ISO 19115, GCMD, EML, etc.) and made accessible through an OAI-PMH compliant interface (Integrated Marine Information System).
- MDA: Registration of repository in official registry
- MDA and EUROBIS: Explicit persistency policy in metadata

Interoperability
- MDA and EUROBIS: Registration of metadata schema in common registry, no online explanation of metadata.
- MDA: Most data are in standard file formats, but if not, like for specific data publication creation, no checks are run on the format of the actual data therein. Increased interoperability is planned by continued web service development adhering OGC web services where applicable and implementing recent standard developments in the framework of OBIS-EVENT IODE projects standardisation while applying internationally accepted ontologies and vocabularies (World Register of Marine Species, Marine Regions, EUNIS habitat typology, etc.). Next steps will cover the entire data life-cycle by putting protocols and platforms in place to support data storage, harmonization and long-term preservation of raw and processed data.

Reusability
- MDA and EUROBIS: Provenance info now text only, not part of schema
- MDA: DMP is planned but not available yet
- MDA: Metadata to aid re-useability are provided but not for specific data publication creation; no check is added on what has been done

5.4.2 Strengths

- Use of vocabularies like WORMS and Marine Regions
- Harmonised data formats
- Open License
- DarwinCore format

5.5 EMSO

The selected method of GO FAIR questionnaire analysis, although effective for other RIs, was not fully appropriate for EMSO ERIC due to a different level of integration (i.e., EMSO ERIC currently does not provide a single data repository) and, as a result, the assessment in an automatic manner does not deliver responses in full depth (Will be better fit when the integrated data management system is transitioned from development to pre-production in

Fall 2019). The responses using a single EMSO repository do not represent the entire ERIC; however, responses for an interim integrated solution were provided and used in the analysis. Additionally, the FAIR Evaluation tool developed by GO FAIR (https://w3id.org/AmIFAIR) has already been used and scored 5 succeeded tests out of 22 as of May 14, 2019.

For the analysis below both the YAML for the single repository, and well as the questionnaire responses and feedback during the workshop were used.

5.5.1 Gaps and planned activities

Findability

- Data access only via HTTP protocol, not via a metadata catalogue (which has been developed as part of H2020 EMSODEV project and is expected to be made publicly available in Fall 2019).
- No use of common vocabularies, i.e., different vocabularies are currently used by EMSO ERIC facilities.
- Central search on metadata and data (planned). EMSO metadata catalogue is expected to use an OAI-PMH interface. The EMSO dataset catalogue will be assessed against FAIR principles for improving findability for harvest or indexing with emphasis on metadata enrichment and metadata will be made publicly available (including to other networks).

Accessibility

- Lack of a single point of access (e.g., API). An interim solution based on files and a data portal is available; however, the developed API as part of H2020 EMSODEV project and is expected to be made publicly available in Fall 2019.
- Data access only via HTTP protocol, not metadata catalogue. EMSO data web services already in place but still heterogeneous. An abstraction layer will homogenize and standardize data access across EMSO facilities, which will make it interoperable with the subdomain.
- Data and metadata are not homogenous across regional facilities. The major gap is the integration of the different regional facilities in a way that data findable in a more uniform manner. Data format will be harmonized to bring accessibility and, when consistent and relevant, it will be improved through enriched metadata.
- PID registration (planned, not centralised, depending on source). While DOI labelling of EMSO published datasets has been initiated, granularity and versioning of dynamic datasets shall be analysed and investigated in order to adhere to a PID approach based on best practices.
- Central EMSO repository not certified.
- Central EMSO repository not centrally registered.
- No central metadata harvesting (metadata only in regional facility).
- EMSO ERIC does not have a centralized user database at this point.

Interoperability

- EMSO ERIC metadata is machine actionable; however, this is not uniform across facilities.
- Integration efforts are required to improve interoperability, including more mature standardization, better semantics and standard metadata.
- Harmonization needs completion.

Reusability

- Provenance info is partially included.
- EMSO ERIC does not have an integrated data curation process at this time. A curation process has been developed as part of H2020 EMSODEV project and it is expected to transition from development to pre-production in Fall 2019.

5.5.2 Strengths

- Access from decentral repositories.
- Very open data access and no user registration

A data management system that addresses several FAIR principles have been developed as part of H2020 EMSODEV project and is expected to be made publicly available in Fall 2019.

6. Selection of priority EOVs

One additional element in Task 9.2 was to select Essential Ocean Variables (EOVs) that will be given priority in the Marine subdomain for improving FAIRness and developing demonstration products as part of Task 9.8 to underpin the final WP9 FAIRness achievements.

6.1 EOV framework

The costs involved in ocean observation are high and there is a need to avoid duplication of efforts, across observing platforms and networks, and to adopt common standards for data collection, processing, and distribution. For that reason, the EOV framework has been formulated to identify and set Essential Ocean Variables (EOVs). These are identified by **Global Ocean Observing System (GOOS) Expert Panels**, based on the following criteria:

- **Relevance:** variable is effective for overall GOOS Themes – Climate, Operational Ocean Services, and Ocean Health.
- **Feasibility:** Observing or deriving variable on a global scale is technically and scientifically feasible.
- **Cost effectiveness:** Generating and archiving data on the variable is affordable

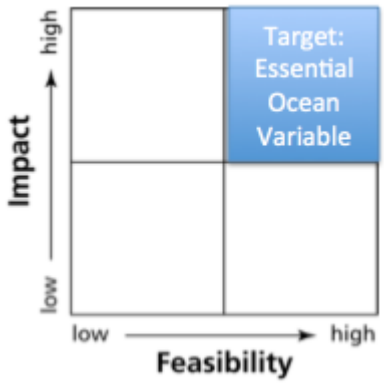

Figure 5. Principle of determining EOVs[2].

Currently there are the following **GOOS Expert Panels**:

- **Physics and Climate Expert Panel**: provided by the Ocean Observations Panel for Climate (OOPC)
- **Biogeochemistry Expert Panel**: provided by the International Ocean Carbon Coordination Project (IOCCP)
- **Biology and Ecosystems Expert Panel**

[2] Source: http://www.goosocean.org/index.php?option=com_content&view=article&id=14&Itemid=114)

They have formulated and produced EOV specification sheets for most of the EOVs as can be derived from www.goosocean.org/eov.

Table 13. Overview of EOVs and URLS to the specification sheets.

Physics	Biogeochemistry	Biology and ecosystems
Sea state	Oxygen	Phytoplankton biomass and diversity
Ocean surface stress	Nutrients	Zooplankton biomass and diversity
Sea ice	Inorganic carbon	Fish abundance and distribution
Sea surface height	Transient tracers	Marine turtles, birds, mammals abundance and distribution
Sea surface temperature	Particulate matter	Hard coral cover and composition
Subsurface temperature	Nitrous oxide	Seagrass cover and composition
Surface currents	Stable carbon isotopes	Macroalgal canopy cover and composition
Subsurface currents	Dissolved organic carbon	Mangrove cover and composition
Sea surface salinity	Ocean colour	Ocean Sound
Subsurface salinity		Microbe biomass and diversity (*emerging)
Ocean surface heat flux		Benthic invertebrate abundance and distribution (*emerging)

Remark: Ocean colour and Ocean sound are cross disciplinary and not just associated with one of the fields.

The inclusion of observing components in the Global Ocean Observing System is based on their readiness level that is their level of maturity in terms of three aspects of observing elements:

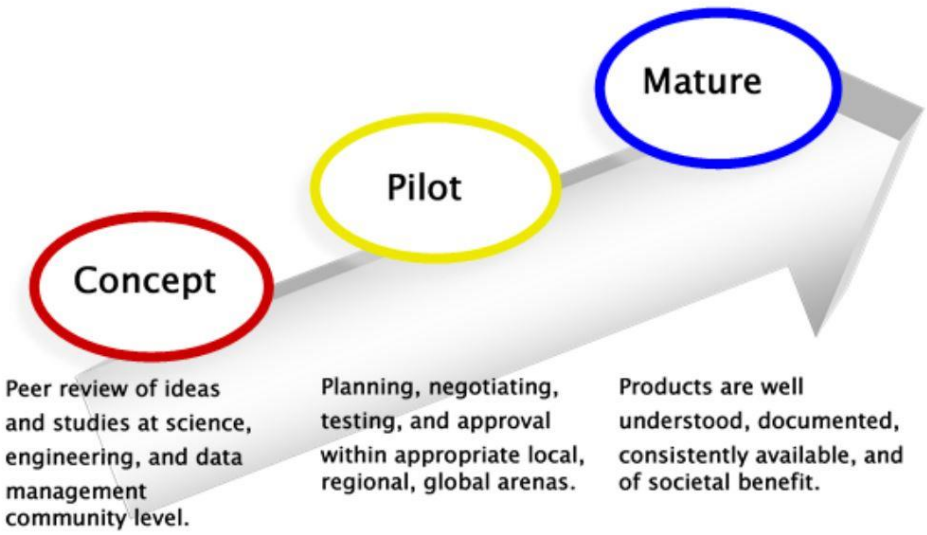

Figure 6. Flow diagram of EOV definition process[3].

6.2 Selection of EOVs for ENVRI-FAIR Marine subdomain

For ENVRI-FAIR WP9 it is essential that the focus will be on EOVs which are mature and managed within the marine RIs. This will allow to use the selected EOVs in the FAIRness analysis and implementation plans, in particular as part of Task 9.8 which aims at developing a demonstrator for underpinning the final WP9 FAIRness achievements. It was agreed to build upon the landscape analysis performed in AtlantOS H2020 project to identify what are the EOV as measured by each RI and what are the services available. Therefore, MARIS has adapted the AtlantOS survey method for ENVRI-FAIR WP9 and has circulated the questionnaire in March 2019 to the marine RIs. The received responses have been summarised by MARIS in the following table.

From this table it has been agreed that WP9 for Task 9.8 will focus on the following EOVs: **Temperature, Salinity, and Chlorophyll-A** (or alternatively **Oxygen**) as they have the strongerst overlap in all the marine RIs in ENVRI-FAIR. These EOVs are also integrated in operational services within the **Copernicus Marine Environmental Monitoring Service (CMEMS)** and the **European Marine Observation and Data Network (EMODnet)**, which are overarching marine data infrastructures driven by the EU DG GROW respectively EU DG MARE.

In addition to these selected physical EOVs, it might be worthwhile to expand the selection with an **EOV for biology** that probably will only involve EMSO and LifeWatch (Marine). Furthermore, consideration might be given to adding the EOV **Inorganic Carbon**, which will involve ICOS (Marine), EMSO and SeaDataNet. This EOV includes subvariables: Dissolved Inorganic Carbon (DIC), Total Alkalinity (TA), Partial pressure of carbon dioxide (pCO2) and pH. Both suggestions will be evaluated in the framework of the planned activities for formulating D9.2 - Marine subdomain implementation plan.

[3] source: http://www.goosocean.org/index.php?option=com_content&view=article&id=20&Itemid=119

Table 14. EOVs covered by marine RIs.

AtlantOS Essential Ocean Variables		ENVRI-FAIR Marine RI´s/ e-infrastructures				
Identifier	Label	Euro-Argo	EMSO	ICOS (Marine)	LifeWatch (Marine)	SeaDataNet
EV_AIRHUM	Air humidity		X	(X)		X
EV_AIRTEMP	Air temperature		X			X
EV_AIRPRESS	Atmospheric pressure		X	(X)		X
EV_BATHY	Bottom depth		X			X
EV_13C	Carbon isotope 13C		X			X
EV_CO2	Inorganic carbon		X	X		X
EV_CHLA	Chlorophyll-a and fluorescence	X	X		X	X
EV_CURR	Currents	X	X			X
EV_DOM	Dissolved organic matter	X				X
EV_N2O	Nitrous oxide					X
EV_NUTS	Nutrients	X	X		X	X
EV_OXY	Oxygen	X	X	(X)		X
EV_RADFLX	Radiative fluxes					
EV_RAIN	Rainfall		X			
EV_SALIN	Salinity	X	X	X	X	X
EV_SEALVL	Sea Level		X			X
EV_POM	Suspended particulates	X	X			X
EV_SEATEMP	Seawater Temperature	X	X	X	X	X
EV_TTRACE	Transient tracers					X
EV_WAVES	Waves		X			X
EV_WDIR	Wind direction		X	(X)		X
EV_WSPD	Wind speed		X	(X)		X
	Phytoplankton biomass and diversity	X		X		
	Zooplankton biomass and diversity		X		X	
	Fish abundance and distribution		X		X	
	Conductivity		X			
	Seismic Data (X, Y, Z)		X			
	Live Video		X			
	Acoustic Data		X			
	Sound pressure Level data (MSFD)		X			
	Sound Speed		X			
	Sea water pressure		X			
	Acoustic pressure		X			
	Vertical acceleration		X			
	Ground motion		X			
	Magnetic field		X			

Remark: The table is built upon the AtlantOS EOVs as was indicated in the introduction of this paragraph 6.2. One should be aware that the AtlantOS EOVs are not completely aligned with the GOOS EOVs. However for the analysis in ENVRI-FAIR this slight difference is not relevant.

6.3 Benefits of enhancing FAIRness for selected EOVs

The objective of Task 9.8 will be to demonstrate the improved FAIRness of the Marine subdomain. This is illustrated in Figure 7, giving the overall workplan of WP9. Task 9.2 is included as the upper purple bar with "Analysis and priorities to enhance marine RI data and services FAIRness". Following the resulting roadmap (D9.1) and implementation plan (D9.2), the FAIRness improvements will be developed and deployed for each marine RI in the five vertical green coloured pillars which refer to Tasks 9.3 – 9.7. Thereafter, the benefits of the improved FAIRness will be demonstrated by developing a global demonstration product for the selected EOVs, using input from the five marine RIs and in synergy with CMEMS and selected EMODnet thematic lots, in particular EMODnet Chemistry, EMODnet Physics, and EMODnet Biology. The demonstration activity is included in the workplan figure as the lower purple bar which refers to Task 9.8. Finally, the outcome of the demonstration will be analysed in Task 9.9 for successes, shortcomings, new challenges and ideas, which is illustrated in the schematic workplan with the orange vertical pillar with label "Progress synthesis and strategy for future development of Marine subdomain".

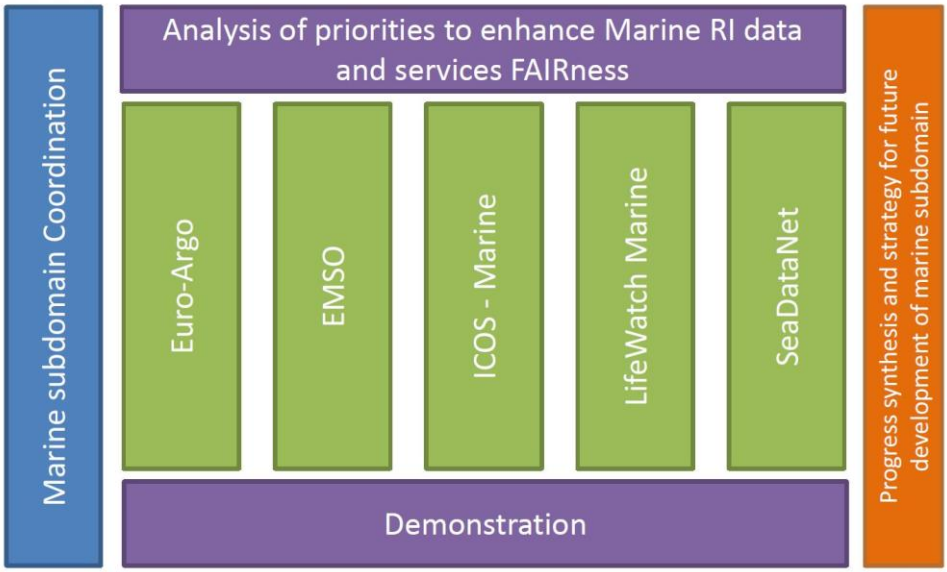

Figure 7. Schematic overview of WP9 workplan.

CMEMS and **EMODnet** are very important and major users of marine data and data products in the Marine subdomain. **CMEMS** is the marine component of **Copernicus**, the long-term European Union's Earth Observation Programme, which looks at our planet and its environment for the ultimate benefit of all European citizens. It develops and offers information services based on satellite Earth Observation and in situ (non-space) data. Copernicus is coordinated by EU DG GROW. CMEMS deploys pan-European capacity for Ocean Monitoring and Forecasting. For that challenge it is operating a number of ocean forecasting service components for sea basins, including a global modelling component. The CMEMS-INSTAC service is the component which gathers and distributes near real-time data from the EuroGOOS network, additional national and European oceanography operators such as Euro-Argo, and international sources, such as ARGO, NOAA, WMO, and others. The focus is on physical met-ocean parameters with recent expansion into biogeochemical parameters, in order to support ecosystem modelling. Next to near real-time data sets, CMEMS has large interest in archived data collections, which are aggregated, harmonised, and validated, as additional input and as calibration data for the CMEMS forecasting model components.

The **European Marine and Observation Data Network (EMODnet)** is a long-term marine data initiative of EU DG MARE since 2008. It is a top down initiative for establishing an overarching network in support of implementation of the Marine Strategy Framework Directive (**MSFD**) and Marine Knowledge 2020 communication. Through a step-wise development EMODnet has become operational with > 150 organisations working together for collating marine data, generating European marine data products, and providing discovery and access to these data and data products across seven themes: Bathymetry, Geology, Seabed habitats, Chemistry, Biology, Physics and Human activities. Each theme is developing, maintaining, and operating a gateway (portal), while the EMODnet Central Portal acts as central access point to these portals and provides additional services. EMODnet is making excellent progress in developing added-value services and products which are attracting users from government, research and industry. Just like CMEMS, EMODnet has large interest in additional marine data sets for feeding its workflows for generating its marine data products.

Improving the FAIRness of the RIs in the Marine subdomain will have a major impact on their position as data providers towards CMEMS and EMODnet. The following benefits are expected:

- All five marine RIs will qualify as data providers for CMEMS and EMODnet; currently Euro-Argo, SeaDataNet and LifeWatch (Marine) are delivering data sets, while ICOS (Marine) started to deliver data to CMEMS and not yet to EMODnet and EMSO is not yet delivering to both;
- The successful uptake of EMODnet products, in particular in support of the Marine Strategy Framework Directive (MSFD) implementation, has raised a number of questions and higher requirements from the MSFD community concerning the metadata and data that are delivered by providers. These concern the richness of metadata, presence of documentation on processing and quality control, and the overall quality of data sets. These attributes determine the degree of fitness-for-purpose of the provided data sets for several later applications, such as e.g. the Good Environmental Status (GES) assessments of MSFD and related ecosystems research. The ENVRI-FAIR activities are anticipating these requirements;
- Adoption of controlled vocabularies, implemented as machine services (e.g. SparQL - RDF), by all five RIs for metadata and data mark-up will improve the consistency and machine useability of data entries in each RI; also, it will allow mapping between vocabularies in use at each RI, supporting transformations of metadata and data formats between RIs and towards common formats. The latter will facilitate the transfer and uptake by CMEMS and EMODnet;
- Improving the FAIRness of the machine interface services of each marine RI will improve the interoperability with other infrastructures, such as CMEMS, EMODnet or marine brokerage services such as in use for GEOSS, and will facilitate regular metadata – data exchanges towards these infrastructures;
- Improving the coherence of metadata and data formats within each marine RI and between the five RIs will facilitate aggregation and harmonisation of their in-situ data sets for joint generation of data products. It will also facilitate to streamline and to increase the efficiency of the QA-QC activities that are undertaken as part of the product generation workflow to reach validated data collections. These improvements will also benefit CMEMS and EMODnet;
- The adoption of unique and persistent data object identifiers, supported by established and sustained resolver services, and including richer metadata about processing and QA-QC activities, will support provenance of use of data sets from marine RIs in data products, such as generated by CMEMS and EMODnet. Users of those data products can trace back which original data sets were used from which data providers and which data originators. This will contribute to justification of investments in RIs and to positioning of RIs in the marine research community.

These benefits will be demonstrated for selected EOVs as part of Task 9.8 later in the project.

7. Conclusions and next steps

Task 9.2 "Analysis and priorities to enhance RI data FAIRness" activities has been successfully implemented by MARIS with involvement of all RIs within the Marine subdomain, and while tuning with WP5-WP7 to draft deliverable D9.1 - **Marine subdomain FAIRness roadmap. This roadmap assesses the** FAIRness of each of the marine RIs and gives a list of priorities for enhancing their FAIRness. For the analysis a great support has been received from the NL GO FAIR office with respect to gaining more knowledge on the FAIR Guiding Principles and for using questionnaires as developed by GO FAIR. These questionnaires were tested and later amended and partly combined into one questionnaire that was used for collecting the necessary information from each of the marine RIs.

The WP9 approach has successfully served as a forerunner within ENVRI-FAIR to develop, test and finetune the analysis methodology in close dialogue with WP5-WP7. Following the Marine subdomain, the methodology has also been set-out to the other subdomains for assessing the FAIRness of all ENVRI-FAIR RIs and identifying strengths and weaknesses as well as plans and suggestions for improving weaknesses.

Additional work was done with WP5 to develop an analysis tool to check the current FAIRness using YAML files. This tool was developed, tested and refined with the answers of the marine RIs and is currently in use for analysing the completed surveys of the other RIs. This tool can be used regularly to monitor the progress of activities for improving FAIRness of the RIs over the course of the ENVRI-FAIR project.

This deliverable D9.1 is to be considered as an analysis and assessment providing a major basis for formulating the next deliverable D9.2 - Marine subdomain implementation plan. This will be prepared in the coming months interacting closely with WP5 (Community standards and catalogue of services) and WP7 (Common implementation and support).

Finally, a subset of Essential Ocean Variables (EOVs) has been determined between the five RIs, thereby considering relevance for data delivery to CMEMS and EMODnet. This subset will be adopted in Task 9.8 to develop a demonstrator that will enable to underpin the WP9 FAIRness achievements. The selected subset includes the following EOVs: Temperature, Salinity, and Chlorophyll-A (or alternatively Oxygen). This might be expanded with an EOV for biology and CO_2. Both suggestions will be evaluated in the framework of the planned activities for formulating D9.2 - Marine subdomain implementation plan.

Appendix 1: Glossary

ACDD	Attribute Convention for Data Discovery (for NetCDF)
API	Application Programming Interface
B2HANDLE	EUDAT minting, storing, managing and accessing persistent identifiers
CAS	Central Authentication Service
CC-BY	Creative Commons Attribution License
CDI	Common Data Index (metadata format and data access system by SeaDataNet)
CF	Climate and Forecast (semantics for NetCDF)
CMEMS	Copernicus Marine Environment Monitoring Service
COPERNICUS	A major earth observation programme run by European Commission and European Space Agency
CSR	Cruise Summary Report
CSW	Catalogue Service for the Web
DMP	1) Data Management Plan 2) Data Management Platform (WP9)
DOI	Digital Object Identifier
DSA	Data Seal of Approval
ECV	Essentia Climate Variable
EDIOS	European Directory of ocean Observing Systems
EDMED	European Directory of Marine Environmental Datasets (SeaDataNet)
EDMO	European Directory of Marine Organisations
EDMERP	European Directory of Marine Environmental Research Projects
EML	Election Markup Language
EMODNET	European Marine Observation and Data Network
EMSO	European Multidisciplinary Seafloor and water column Observatory
ENVRI	1) An environmental RI cluster FP7 project 2) Environment research infrastructures (in ESFRI level or upcoming) as a community
ENVRIplus	An environmental RI cluster H2020 project
EOSC	European Open Science Cloud
EOV	Essential Ocean Variable(s)
ERDDAP	NOAA developed science data server technology
ERIC	European Research Infrastructure Consortium (legal entity type)
EUMETNET	Grouping of 31 European National Meteorological Services
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable Accessible Interoperable Reusable
FAQ	Frequently Asked Questions
FORCE11	a community to help facilitate the change toward improved knowledge creation and sharing
GBIF	Global Biodiversity Information Facility
GCMD	Global Change Master Directory
GDAC	Global Data Assembly Center
GEMET	GEneral Multilingual Environmental Thesaurus
GEO	Group on Earth Observation (System of Systems)
GEOSS	Global Earth Observation System of Systems
GOFAIR	An international programme on FAIR implementation
GOOS BGC	Global Ocean Observing System Biogeochemistry Panel
GUI	Graphical User Interface

ICOS	Integrated Carbon Observation System
ICT	Information and Communications Technology
IMIS	Integrated Marine Information System
INSPIRE	Infrastructure for Spatial Information in the European Community
iRODS	Open Source Data Management Software
JCOMM	Joint Technical Commission for Oceanography and Marine Meteorology
LW	LifeWatch
Marine-ID	Registration and authentication services for marine data services
MDA	Marine Data Archive
NetAPP	Hybrid cloud service
NetCDF	Network Common Data Format
NVS	NERC Vocabulary Services
NOAA	US National Oceanic and Atmospheric Administration
OAUTH	Open Authorization (standard)
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OBIS	Ocean Biogeographic Information System
ODIP	Ocean Data Interoperability Platform
OGC	Open Geospatial Consortium
OpenDAP	Open-source Project for a Network Data Access Protocol
ORCID	Open Researcher and Contributor ID
OWL	Web Ontology Language
PID	Persistent Identifiers
PROV-O	Web Ontology Language encoding of the PROV Data Mode
QA/QC	Quality Assurance/Quality Control
RDF	Resource Description Framework
RI	Research Infrastructure
RSS	Really Simple Syndication
SAML	Security Assertion Markup Language
SEADATANET	SeaDataNet pan-European infrastructure for marine data management
SME	Small or medium Enterprise
SparQL	SparQL Protocol And RDF Query Language
SWOT	Analysis on Strengths, Weaknesses, Opportunities and Threats
VRE	Virtual Research Environment
WMO	World Meteorological Organisation
WoRMS	World Registry of Marine Species
WPS	Web Processing Services
YAML	Yet Another Mockup Language

Appendix 2: XLS spreadsheets from RIs

ICOS (Marine)

Question	Response
1.0 Timestamp	4/1/2019 18:05:16
1.2 Email Address	alex.vermeulen@icos-ri.eu
1.1 Contact name	Alex Vermeulen
1.3 Research Infrastructure Name	ICOS
1.4 Please provide the URL of one of the datasets in scope for your answers	https://data.icos-cp.eu/objects/-ffoiHjX5NDN0Vq_fKuVmas0
1.5 Please provide the URL to the discovery portal in which the dataset can be downloaded	https://data.icos-cp.eu/portal
1.6 Which repositories do you use for data?	Local (ICOS servers in Lund), B2SAFE (EUDAT/CDI) at CSC, iRODS at SNIC (Swedish national storage system)
1.7 Which repository software is being used?	HTTP API for B2STAGE access to B2SAFE, iRODS Java client library Jargon 4.3.0.1
1.8 Which repositories do you use for metadata?	local (at ICOS Carbon Portal production server) versioned RDF triple store
1.9 Do your repositories use PIDs? If so which PID systems?	Yes. Handle system generic PIDs and Datacite DOIs. In both cases we have ICOS-specific prefixes.
1.10 Do you assign PIDs manually or automatically?	Handle PIDs automatically, DOIs currently manually (but using custom-build user-friently web application).
1.11 Which PID registration provider do you use?	DataCite for DOIs, Handle.net software server hosted by PDC at KTH.
1.12 Do you use the PID Record to store attributes about the data?	Not for generic Handle PIDs. In case of DOIs we copy relevant attributes into the DataCite metadata catalog (schema version V4.1).
1.13 Are these repositories certified? If so, which methods are used?	The ICOS Carbon Portal itself isn't (yet CTS) certified. However, our certification status might not be relevant, as we have outsourced our storage to data centers (EUDAT and SNIC) that are not certified.
1.14 Are repository policies mentioned at the website? If so, indicate the major ones.	What is a repository policy? Yes the ICOS data Policy is clear and described and available at the website, but does not cover the elements that are part of the B2SAFE policy, although we agreed with them for two replicates.
1.15 Are your repositories registered in a registry? If so which registry?	Yes, re3data.org, working on B2FIND, GEOSS, WMO WDCGG
1.16 Which persistency guaranties are typically given?	By whom? ICOS has 20-25 year longevity goal.
1.17 Which are the most popular data types used?	time series, spatial raster data (3D-5D)
1.18 Which are the preferred data formats?	netcdf, csv
1.19 Do those formats include metadata headers? if so, which?	Yes, netcdf conforms usually to CF-1.4 conventions, CSV file have headers that are community specific and contain general metadata and column names and units etc.
1.20 Do you provide search on data?	ICOS doesn't provide searches inside datasets,

Question	Response
	as these typically contain mostly numbers. Searches for files containing specific variables (column names) are supported via queries to the metadata database.
1.21 Did you register your schemas in a common registry?	No, not formally (e.g. in schemas.org). But ICOS data objects follow schemata hosted by ICOS that links to the data format specification definition in the ICOS ontology, which is openly accessible (in OWL) via the ICOS SPARQL endpoint (https://meta.icos-cp.eu/sparqlclient/).
1.22 Which metadata schemas are mostly used?	rdf, inspire, iso19115, geo-dcat
1.23 Are all categories used in the schemas defined in open registries?	The ICOS-specific ontologies are openly available as linked open URLs, e.g. via the ICOS SPARQL endpoint (https://meta.icos-cp.eu/sparqlclient/).
1.24 How is provenance included?	We follow a simplified PROV-O and track lineage of data objects. The versioned metadata store keeps track of all metadata updates. However, information on data processing steps etc is not yet described in the ICOS data model.
1.25 Are PIDs included in the metadata description?	Yes
1.26 What is the primary storage format for metadata?	As assertions in the form of RDF triples
1.27 Which are the export formats supported?	Metadata export formats include json, xml, turtle, txt, html (all available via content negotiation of dataset landing pages).
1.28 Which metadata exchange/harvesting methods are supported?	SPARQL open endpoint, landing pages contain a subset in content negotiable formats (see previous question).
1.29 Do you have a local search engine?	Yes, see https://data.icos-cp.eu/portal/ (data) and https://meta.icos-cp.eu/sparqlclient/ (metadata).
1.30 Do you support external search engines?	Yes, through access to our open SPARQL endpoint. (We do not operate an OAI-PMH endpoint, though…)
1.31 Do you make statements about access policies in your metadata?	Access and licence information is provided at the Carbon Portal, and specifically to users during data download.
1.32 Is your metadata machine actionable?	In principle yes, as the metadata follows OWL ontology which is exposed in adherence with Linked Data principles.
1.33 How is authentication done?	The Carbon Portal operates its own AAI service (CPAuth), which apart from local username/passwords also allows logins via eduGain and OAuth (ORCID id, Facebook)
1.34 Do you maintain an own user database?	Yes, users can register at the Carbon Portal and store voluntarily-provided profile information including individual settings for services (e.g. data cart contents). We also use logins to control access to specific local services, such as metadata modification, access to VREs and on-demand computation etc.). In support, we provide a trivial username/encrypted password database for users of password authentication.

Question	Response
1.35 Do you use ORCID in your AAI?	Yes, ORCID ids are supported (see above).
1.36 What is the major access technology supported?	HTTP GET method
1.37 How is authorization done?	For data download/access, we use API tokens, or email address/password. For access-controlled services, login via the cpauth service is required, with some operations or functionality requiring the user to be listed in service-specific configuration files.
1.38 Which specific licenses do you use for your data?	In most cases, CC4-BY. Some data however are CC-zero.
1.39 Are metadata openly available?	Yes
1.40 Do you use or provide specific DMP tools?	DMP tools allow to describe the DMP, they do not make them. We used DMPonline, EU template, up to midterm
1.41 Do you apply special data publishing steps?	ICOS is all about curation from data generation through QC, processing to cataloguing and dissemination/publishing.
1.42 Do you apply special data processing steps?	ICOS observation data are processed and quality controlled at our Thematic Centres, before they are uploaded to the Carbon Portal. Here, extreme care is taken to preserve all ingested data objects in a binary-exact form and store indefinitely. To ensure fixity can be proven, data object checksums are evaluated on the fly during ingestion, and the Handle PID suffix is calculated based on this checksum.
1.43 Do you apply workflow frameworks for processing your data?	Yes, mostly custom workflows (carried out at the ICOS Thematic Centres.)
1.44 Do you use distributed workflow tools? if so, which?	Yes and no; we do not use specific workflow tools (like Taverna), but we extensively use pseudo-standardized scripts to e.g. instantiate Virtual Machines for HTC computation and storage.
1.45 Do you offer other type of support or analytics services?	Yes, VREs through Jupyter Lab/Notebooks
1.46 Do you offer data products in your RI?	Yes, see a description at https://www.icos-cp.eu/dataproducts
1.47 Do you use semantic vocabularies from generic vocabularies, ontologies, etc.? If so point to the registries.	Yes, rdf, prov, foaf, … See e.g. http://purl.org/dc/elements/1.1/; http://purl.org/dc/terms/; http://www.w3.org/ns/prov#; http://www.w3.org/2002/07/owl#; http://www.w3.org/2000/01/rdf-schema#
1.48 Do you use discipline specific vocabularies, ontologies etc.? If so point to the registries.	Yes, wds (world data system) and http://www.w3.org/2003/01/geo/wgs84_pos#
1.49 Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	Yes, ICOS (domain) specific, e.g. BADM, WMO GAW. See http://meta.icos-cp.eu/ontologies/cpmeta/ !
1.50 Do you believe that your data is Findable (F)? if not, indicate where you see major gaps.	Yes
1.51 Do you believe that your data is Accessible (A)? if not indicate where you	Yes

Question	Response
see major gaps	
1.52 Do you believe that your data is interoperable (I)? if not indicate where you see major gaps	Starting to be interoperable... Need to work more on our data model, applying relevant standards (including vocabularies) and making sure the definitions and attribute names we use are registered.
1.53 Do you believe that your data is re-usable (R)? if not, indicate where you see major gaps	Mostly, yes - but we need to include more provenance information in the metadata (including more links to ICOS observation and data processing protocols).
2.0 Timestamp	5/2/2019 17:29:55
2.2 Email Address	alex.vermeulen@icos-ri.eu
2.4 Please provide the IRI of the resource to be evaluated	https://hdl.handle.net/11676/-ffoiHjX5NDN0Vq_fKuVmas0
2.5_F1A: Please provide the IRI for a registered identifier schema for your resource's IRI (e.g. DOI, HTTP):	https://handle.net ?
2.6_F1B: Please provide the IRI to the document describing the persistence policy for the identifier of this(meta)data (this may be a document from your service provider, e.g. Zenodo, UniProt, etc.):	N/A, but as the identifier is derived from the checksum of the object it is guaranteed to be unique and persistent
2.7_F2A: Please provide the IRI to a document that contains machine-readable metadata for the digital resource	https://hdl.handle.net/11676/-ffoiHjX5NDN0Vq_fKuVmas0
2.8_F2B: Please provide the IRI for the file format of this metadata:	Content negotiation, File formats: JSON, XML, Turtle, html
2.9_F3: Please provide the IRI of the metadata	https://hdl.handle.net/11676/-ffoiHjX5NDN0Vq_fKuVmas0
2.10_F3: Please provide the IRI of the data described by the metadata above	https://data.icos-cp.eu/objects/-ffoiHjX5NDN0Vq_fKuVmas0
2.11_F4: Please provide the URL to a search engine where the dataset can be found.	https://data.icos-cp.eu/portal
2.12_F4: Search query/terms: the query that will be executed to discover your RESOURCE ID (found in the first page of the search):	https://data.icos-cp.eu/portal/#%7B%22filterCategories%22%3A%7B%22project%22%3A%5B%22icos%22%5D%2C%22theme%22%3A%5B%22atmosphere%22%5D%2C%22level%22%3A%5B2%5D%2C%22type%22%3A%5B%22atcCh4L2DataObject%22%5D%2C%22station%22%3A%5B%22iAS_HPB%22%5D%7D%7D
2.13_A1.1 Please provide a URL to the description of the Access Protocol	https://doi.org/10.17487%2FRFC2616
2.14_A1.1 Is the protocol open?	Yes
2:15_A1.1 Is the protocol (royalty) free?	Yes
2.16_A1.2 Authorization is required to access the content of my RESOURCE ID	No
2.17_A1.2 If "yes" above, please provide a IRI that resolves to a description of the process to obtain access to restricted content	N/A
2.18 A2 Please provide the URL to a metadata longevity plan	https://meta.icos-cp.eu/objects/5stgMTshDNdDe_Y8obQOpe0u
2.19_I1 Please provide the URL to the	https://www.w3.org/TR/owl2-overview/

Question	Response
specification of the language.	
2.20_I2 Please provide one or more (max 3) IRIs from vocabularies used within the (meta)data that is returned by resolving the RESOURCE ID.	http://meta.icos-cp.eu/ontologies/cpmeta/
2.21_I3 Please provide the URL to a formal Linkset (defined at: https://www.w3.org/TR/void/#linkset) or copy/paste the content of a formal linkset that describes at least a portion of the content at RESOURCE ID (you may need to create this, or edit the example below) e.g. :DBpedia2DBLP a void:Linkset;void:target :DBpedia;void:target :DBLP;void:linkPredicate owl:sameAs;void:triples 10000;	Don't see how this is relevant in this context
2.22_R1.1 Please provide the IRI for your usage license regarding the content returned from RESOURCE ID (be that data, or metadata):	https://creativecommons.org/licenses/by/4.0/
2.23_R1.2 Please provide the IRIs (maximum 3) for the vocabularies being used to describe the provenance of the content resolved from RESOURCE ID (be that data, or metadata; e.g. http://purl.org/dc/terms/):	http://www.w3.org/ns/prov/
2:24_R1.2 Please provide the IRIs (maximum 3) for the vocabularies being used to describe the domain information of the content resolved from RESOURCE ID (be that data, or metadata; e.g. http://edamontology.org/data_ will be used for many bioinformatics data/metadata):	http://meta.icos-cp.eu/ontologies/cpmeta/
2.25_R1.3 Please provide the IRI that represents the certification from a recognized authority in your community or domain, indicating that the content of RESOURCE ID is compliant with the standards of your community	N/A
2.3 Research Infrastructure Name	ICOS ERIC

SeaDataNet CDI

Question	Response	Remark
1.0 Timestamp	5/23/2019	The CDI service for data access is being upgraded and the new service will be launched mid July 2019. The given answers concern the new situation
1.2 Email Address	dick@maris.nl	
1.1 Contact name	Dick Schaap	
1.3 Research Infrastructure Name	SeaDataNet	

Question	Response	Remark
1.4 Please provide the URL of one of the datasets in scope for your answers	https://cdi.seadatanet.org/report/15222/xml	Active mid July 2019
1.5 Please provide the URL to the discovery portal in which the dataset can be downloaded	https://cdi.seadatanet.org/search	Active mid July 2019
1.6 Which repositories do you use for data?	SeaDataNet Central Data repository at EUDAT for unrestricted data, complemented with SeaDataNet Local Data repositories at Data Providers for restricted data	
1.7 Which repository software is being used?	EUDAT B2SAFE (based on iRODS)	
1.8 Which repositories do you use for metadata?	Microsoft SQL Server	
1.9 Do your repositories use PIDs? If so which PID systems?	PIDs for data using the EUDAT B2Handle service	
1.10 Do you assign PIDs manually or automatically?	automatically	
1.11 Which PID registration provider do you use?	EUDAT B2Handle	
1.12 Do you use the PID Record to store attributes about the data?	yes	
1.13 Are these repositories certified? If so, which methods are used?	none	
1.14 Are repository policies mentioned at the website? If so, indicate the major ones.	SeaDataNet data policy (https://www.seadatanet.org/Data-Access/Data-policy)	
1.15 Are your repositories registered in a registry? If so which registry?	GEOSS catalogue (http://www.geoportal.org/); IODE Ocean Data Portal (http://www.oceandataportal.org/)	
1.16 Which persistency guaranties are typically given?	long term stewardship (guaranteed by SeaDataNet data centres)	
1.17 Which are the most popular data types used?	ASCII; binary	
1.18 Which are the preferred data formats?	ODV; NetCDF CF; Medatlas	
1.19 Do those formats include metadata headers? if so, which?	CDI metadata reference; Data Provider; Parameter measured; Unit; Cruise; Station; Observation type; Date; Time; Location; Depth;	Most are multiple in use

Question	Response	Remark
	Value; Quality Flag	
1.20 Do you provide search on data?	planned	
1.21 Did you register your schemas in a common registry?	yes	At the SeaDataNet portal with link in the metadata XML
1.22 Which metadata schemas are mostly used?	ISO19115-19139	SeaDataNet community profile
1.23 Are all categories used in the schemas defined in open registries?	partially	Most supported by controlled vocabularies, but a few with free text
1.24 How is provenance included?	partially	Metadata enrichment planned for structured provenance metadata, using linked data principles and SeaDataNet directories and vocabularies
1.25 Are PIDs included in the metadata description?	yes	Active mid July 2019
1.26 What is the primary storage format for metadata?	XML	Extension with RDF - SPARQL planned
1.27 Which are the export formats supported?	XML;HTML	Extension with JSON and RDF planned
1.28 Which metadata exchange/harvesting methods are supported?	OGC CSW;OAI-PMH;OpenSearch	Extension with SPARQL endpoint planned
1.29 Do you have a local search engine?	yes; using ElasticSearch (https://cdi.seadatanet.org/search)	Active mid July 2019
1.30 Do you support external search engines?	GEO portal; IODE Ocean Data Portal (ODP)	
1.31 Do you make statements about access policies in your metadata?	yes	
1.32 Is your metadata machine actionable?	yes	
1.33 How is authentication done?	Marine-ID	
1.34 Do you maintain an own user database?	yes	
1.35 Do you use ORCID in your AAI?	no	
1.36 What is the major access technology supported?	CAS	Shibboleth planned for wider interoperability
1.37 How is authorization done?	CAS;SAML;Oauth	
1.38 Which specific licenses do you use for your data?	SeaDataNet license, derived from SeaDataNet data policy	

Question	Response	Remark
1.39 Are metadata openly available?	yes	
1.40 Do you use or provide specific DMP tools?	yes	SeaDataNet has a documented workflow for populating the CDI service, supported by dedicated software tools and services
1.41 Do you apply special data publishing steps?	yes	The import, validation and publishing of new metadata and data entries follows a strict workflow, which is automated with human interactions
1.42 Do you apply special data processing steps?	yes	Syntax, semantics, and coherence of metadata and data are vaildated before publishing
1.43 Do you apply workflow frameworks for processing your data?	yes	SeaDataNet structured workflow with services
1.44 Do you use distributed workflow tools? if so, which?	yes	New entries for the CDI service are first validated at data provider level, using standard tools such as Octopus, before import workflow can be started
1.45 Do you offer other type of support or analytics services?	no	Additional data validation services are planned
1.46 Do you offer data products in your RI?	yes	Aggregated, validated and harmonised data collections for selected parameters; interpolated maps for sea regions for selected parameters
1.47 Do you use semantic vocabularies from generic vocabularies, ontologies, etc.? If so point to the registries.	ISO countries list; GEMET - INSPIRE themes list;	
1.48 Do you use discipline specific vocabularies, ontologies etc.? If so point to the registries.	SeaDataNet NERC Vocabulary Service (NVS) (https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/)	
1.49 Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	SeaDataNet European directories for organisations (EDMO) (https://edmo.seadatanet.org), projects (EDMERP) (https://edmerp.seadatanet.org), cruises (CSR)(http://csr.seadatanet.org), data collections (EDMED) (https://edmed.seadatanet.org/) and monitoring networks	The https URLS are planned

Question	Response	Remark
	(https://edios.seadatanet.org)	
1.50 Do you believe that your data is Findable (F)? if not, indicate where you see major gaps.	yes, none	
1.51 Do you believe that your data is Accessible (A)? if not indicate where you see major gaps	partially, improved API	API is planned as part of the CDI upgrading
1.52 Do you believe that your data is interoperable (I)? if not indicate where you see major gaps	yes	
1.53 Do you believe that your data is re-usable (R)? if not, indicate where you see major gaps	partially, provenance and quality information	More metadata planned by Linked Data (see also question 1.24)

SeaDataNet Data products catalogue

Question	Response	Remark
1.0 Timestamp	5/23/2019	These answers concern the SeaDataNet discovery and access service for data products
1.2 Email Address	dick@maris.nl	
1.1 Contact name	Dick Schaap	
1.3 Research Infrastructure Name	SeaDataNet	
1.4 Please provide the URL of one of the datasets in scope for your answers	https://doi.org/10.12770/90ae7a06-8b08-4afe-83dd-ca92bc99f5c0	Each SeaDataNet data product has a DOI at DataCite with a landing page hosted at SeaDataNet SEXTANT catalogue
1.5 Please provide the URL to the discovery portal in which the dataset can be downloaded	https://www.seadatanet.org/Products	SeaDataNet SEXTANT products catalogue user interface
1.6 Which repositories do you use for data?	SeaDataNet Central Data Products repository at IFREMER as part of the SEXTANT catalogue	
1.7 Which repository software is being used?	SEXTANT (based on GeoNetwork)	
1.8 Which repositories do you use for metadata?	Oracle relational database	
1.9 Do your repositories use PIDs? If so which PID systems?	DOI, DataCite	

Question	Response	Remark
1.10 Do you assign PIDs manually or automatically?	manually	
1.11 Which PID registration provider do you use?	DataCite	
1.12 Do you use the PID Record to store attributes about the data?	yes	
1.13 Are these repositories certified? If so, which methods are used?	none	
1.14 Are repository policies mentioned at the website? If so, indicate the major ones.	SeaDataNet data policy (https://www.seadatanet.org/Data-Access/Data-policy)	
1.15 Are your repositories registered in a registry? If so which registry?	none	
1.16 Which persistency guaranties are typically given?	long term stewardship (guaranteed by IFREMER)	
1.17 Which are the most popular data types used?	binary	
1.18 Which are the preferred data formats?	NetCDF CF	
1.19 Do those formats include metadata headers? if so, which?	Conventions; Project name; Contact organisation; Contact email; Data access url; OGC WMS URL; Product ID; Author email; Product Date; Product title; File name; Source; Comment; Correlation Length; Climatology Bounds; Number of data in bins; Depth; Location; Signal to noise; Parameters calculated; Values; Date; Time	Part is metadata header and others are multiple in use
1.20 Do you provide search on data?	planned	Planned by means of ERDDAP
1.21 Did you register your schemas in a common registry?	yes	At the SeaDataNet portal with link in the metadata XML
1.22 Which metadata schemas are mostly used?	ISO19115-19139	SeaDataNet community profile for data products
1.23 Are all categories used in the schemas defined in open registries?	partially	Most supported by controlled vocabularies, but a few with free text
1.24 How is provenance included?	partially	Metadata enrichment planned for structured provenance metadata, using linked data principles and

Question	Response	Remark
		SeaDataNet directories and vocabularies
1.25 Are PIDs included in the metadata description?	yes	
1.26 What is the primary storage format for metadata?	XML	Extension with RDF - SPARQL planned
1.27 Which are the export formats supported?	XML;HTML	Extension with JSON and RDF planned
1.28 Which metadata exchange/harvesting methods are supported?	OGC CSW	Provided by GeoNetwork
1.29 Do you have a local search engine?	https://www.seadatanet.org/Products#/search	Built in GeoNetwork
1.30 Do you support external search engines?	none	
1.31 Do you make statements about access policies in your metadata?	yes	
1.32 Is your metadata machine actionable?	partially	Via the use of vocabs in the XML
1.33 How is authentication done?	Marine-ID	
1.34 Do you maintain an own user database?	yes	
1.35 Do you use ORCID in your AAI?	no	
1.36 What is the major access technology supported?	CAS	Shibboleth planned for wider interoperability
1.37 How is authorization done?	CAS;SAML;Oauth	
1.38 Which specific licenses do you use for your data?	SeaDataNet license, derived from SeaDataNet data policy	
1.39 Are metadata openly available?	yes	
1.40 Do you use or provide specific DMP tools?	yes	SeaDataNet has a documented workflow for generating the data products, supported by dedicated software tools and services. Also the population of the Sextant products catalogue follows a workflow, supported by tools and services
1.41 Do you apply special data publishing steps?	yes	The import, validation and publishing of metadata for new data products follows a workflow, which is operated manually

Question	Response	Remark
1.42 Do you apply special data processing steps?	yes	SeaDataNet has a documented workflow for generating the data products, supported by dedicated software tools and services. Developments are underway for a Virtual Research Environment (VRE) for further structuring.
1.43 Do you apply workflow frameworks for processing your data?	yes	SeaDataNet has a documented workflow for generating the data products, supported by dedicated software tools and services. Developments are underway for a Virtual Research Environment (VRE) for further structuring.
1.44 Do you use distributed workflow tools? if so, which?	none	
1.45 Do you offer other type of support or analytics services?	yes	Data products are visualised by OGC WMS - WFS services and by WPS for data plots. Planned is ERDDAP for data searches and visualisation.
1.46 Do you offer data products in your RI?	yes	The SEXTANT products catalogue service is set up for providing discovery and access to the SeaDataNet products
1.47 Do you use semantic vocabularies from generic vocabularies, ontologies, etc.? If so point to the registries.	ISO countries list; GEMET - INSPIRE themes list;	
1.48 Do you use discipline specific vocabularies, ontologies etc.? If so point to the registries.	SeaDataNet NERC Vocabulary Service (NVS) (https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/)	
1.49 Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	SeaDataNet European directory for organisations (EDMO) (https://edmo.seadatanet.org); SeaDataNet CDI data discovery and access service (CDI) (https://cdi.seadatanet.org)	The https URLS are planned
1.50 Do you believe that your data is Findable (F)? if not, indicate where you see major gaps.	yes, none	
1.51 Do you believe that your data is Accessible (A)? if not indicate	partially, improved WPS services for data plots	Planned

Question	Response	Remark
where you see major gaps		
1.52 Do you believe that your data is interoperable (I)? if not indicate where you see major gaps	yes, none	
1.53 Do you believe that your data is re-usable (R)? if not, indicate where you see major gaps	partially, provenance and quality information	More metadata planned by Linked Data (see also question 1.24)

Euro-Argo

Question	Response	Remark
1.0 Timestamp	4/18/2019 10:53	4/18/2019
1.2 Email Address	Thierry.Carval@ifremer.fr	thierry.carval@ifremer.fr
1.1 Contact name	Thierry Carval	Thierry Carval
1.3 Research Infrastructure Name	Euro-Argo ERIC	Euro-Argo ERIC
1.4 Please provide the URL of one of the datasets in scope for your answers	http://doi.org/10.17882/42182	http://doi.org/10.17882/42182
1.5 Please provide the URL to the discovery portal in which the dataset can be downloaded	http://doi.org/10.17882/42182	http://www.argodatamgt.org/Access-to-data/Argo-data-selection
1.6 Which repositories do you use for data?	"The dataset is continuously mirrored on 2 ftp servers (on in USA, the other in Europe) ftp://ftp.ifremer.fr/ifremer/argo ftp://usgodae.org/pub/outgoing/argo "	Euro-Argo Data, central
1.7 Which repository software is being used?	In Europe: a collection of NetCDF file on a high availability fileserver (RAID5 disks), continuous snapshots (one month span), a monthly archive	NetCDF file
1.8 Which repositories do you use for metadata?	"Data and metadata are preserved in NetCDF files with a dedicated Argo NetCDF CF format on Argo GDAC ftp server: ftp://ftp.ifremer.fr/ifremer/argo ftp://usgodae.org/pub/outgoing/argo Metadata are also distributed through JCOMMOPS metadata server: http://www.jcommops.org/board "	central
1.9 Do your repositories use PIDs? If so which PID systems?	The whole dataset has a unique DOI (https://doi.org/10.17882/42182).	DOI

Question	Response	Remark
	Monthly snapshots are preserved, archived and accessible with the DOI and an additional monthly fragment (http://doi.org/10.17882/42182#60707)	
1.10 Do you assign PIDs manually or automatically?	The Argo DOI fragments are assigned automatically	automatically
1.11 Which PID registration provider do you use?	DataCite	SEANOE
1.12 Do you use the PID Record to store attributes about the data?	"Yes for the monthly snapshot (the DOI+ fragment) No otherwise. "	yes
1.13 Are these repositories certified? If so, which methods are used?	"Yes, Ifremer is DSA and IODE certified. Ifremer-Sismer is in certification process as "RDA-Trustworthy repository" "	- Data Seal of Approval
1.14 Are repository policies mentioned at the website? If so, indicate the major ones.	"Yes https://creativecommons.org/licenses/by/4.0/ "	- data access
1.15 Are your repositories registered in a registry? If so which registry?	Yes, GEO registry	GEO
1.16 Which persistency guaranties are typically given?	"The Argo long term archive is managed by US-NCEI. US-NCEI has a Unesco-WMO mandate as world data centre (WDC-A). "	NULL
1.17 Which are the most popular data types used?	The self-describing NetCDF CF format Argo implementation	binary
1.18 Which are the preferred data formats?	"NetCDF file format https://www.unidata.ucar.edu/software/netcdf/docs/ "	NetCDF
1.19 Do those formats include metadata headers? if so, which?	"Yes, NetCDF ACDD metadata https://www.unidata.ucar.edu/software/thredds/current/netcdf-java/metadata/DataDiscoveryAttConvention.html "	- NetCDF ACDD metadata
1.20 Do you provide search on data?	"Yes, among others: • ERDDAP http://www.ifremer.fr/erddap/index.html • Web GUI : http://www.argodatamgt.org/Access-to-data/Argo-data-selection • Thredds : http://www.ifremer.fr/thredds/catalog/CORIOLIS-ARGO-GDAC-OBS/catalog.html "	yes
1.21 Did you register your schemas in a common registry?	No, the format is described online in "Argo user's manual, http://dx.doi.org/10.13155/29825"	no
1.22 Which metadata	"The data and metadata	Argo user manual

Question	Response	Remark
schemas are mostly used?	organization is described in: "Argo user's manual, http://dx.doi.org/10.13155/29825"	
1.23 Are all categories used in the schemas defined in open registries?	No. The Argo vocabulary server is under implementation.	planned
1.24 How is provenance included?	The history section of each data file lists data processing steps, with an action code for each step.	- text only
1.25 Are PIDs included in the metadata description?	Yes, each Argo float has a WMO platform code, each sensors have their serial-ids stored in metadata files	yes
1.26 What is the primary storage format for metadata?	NetCDF CF Argo implementation	NetCDF CF Argo
1.27 Which are the export formats supported?	NetCDF, csv	NetCDF CSV
1.28 Which metadata exchange/harvesting methods are supported?	"OAI-PMH, OpenDAP, OGC-CSW ISO19115-2/19139 schema, in FGDC-STD-001-1998 schema (Federal Geographic Data Committee) "	OAI-PMH OpenDAP OGC CSW
1.29 Do you have a local search engine?	No, investigating Elasticsearh	NULL
1.30 Do you support external search engines?	OAI-PMH, OpenDAP, OGC-CSW	CSW endpoint
1.31 Do you make statements about access policies in your metadata?	"Yes, included in the Argo GDAC DOI metadata. Licence CC-BY Utilisation A user of Argo data is expected to read and understand this manual and the documentation about the data contained in the ""attributes"" of the NetCDF data files, as these contain essential information about data quality and accuracy. A user should acknowledge use of Argo data in all publications and products where such data are used, preferably with the DOI and following standard sentence:""These data were collected and made freely available by the international Argo project and the national programs that contribute to it.""""	yes
1.32 Is your metadata machine actionable?	Yes, through DataCite DOI (license metadata)	yes
1.33 How is authentication done?	No authentication, the dataset to totally and mandatorily open.	VOID
1.34 Do you maintain an own user database?	"No, users are anonymous. But we manage the database of data downloads. "	no

Question	Response	Remark
1.35 Do you use ORCID in your AAI?	"There is no AAI, but we use ORCID to give credit to data provider (in ""contributors"" section) "	no
1.36 What is the major access technology supported?	ftp, opendap, rsync	VOID
1.37 How is authorization done?	No authorization, accesses are anonymous	VOID
1.38 Which specific licenses do you use for your data?	https://creativecommons.org/licenses/by/4.0/	CC-BY4.0
1.39 Are metadata openly available?	"Yes on DataCite for the global dataset Yes on ERDDAP for individual observations (profile files) "	yes
1.40 Do you use or provide specific DMP tools?	No	none
1.41 Do you apply special data publishing steps?	Yes, we have well documented real-time and delayed mode data publishing steps	- manual QA QC
1.42 Do you apply special data processing steps?	Yes, we have well documented real-time and delayed mode data publishing steps	- QA QC
1.43 Do you apply workflow frameworks for processing your data?	"Yes, each data processing step is documented and recorded in the history section The history section of each file list data processing steps , with a time stamped code for each step "	- data processing steps recorded
1.44 Do you use distributed workflow tools? if so, which?	No	none
1.45 Do you offer other type of support or analytics services?	"Yes, we implement Spark, Cassandra, Elasticsearch and Pangeo analytics services. These services are typically accessed through JupyterHub "	Spark Cassandra Elasticsearch
1.46 Do you offer data products in your RI?	Yes: temperature and salinity analysis, climatologies	temperature salinity climatologies
1.47 Do you use semantic vocabularies from generic vocabularies, ontologies, etc.? If so point to the registries.	"Yes, we use NetCDF CF Metadata Conventions (is it a semantic ?). See http://cfconventions.org/ We also comply with NetCDF "ACDD" conventions (Attribute Convention for Dataset Discovery) "http://wiki.esipfed.org/index.php/Category:Attribute_Conventions_Dataset_Discovery "	
1.48 Do you use discipline specific vocabularies, ontologies etc.? If so point to the	"We comply with SeaDataNet P01 and P06 vocabularies for ocean in situ parameters. For platform types, we comply with SeaDataNet C17	codelist

Question	Response	Remark
registries.	vocabulary. For institutions, we use the SeaDataNet EDMO codes. P01: http://seadatanet.maris2.nl/bandit/browse_step.php P06: http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P06 C17: http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=C17 EDMO: http://seadatanet.maris2.nl/v_edmo/welcome.asp All Argo metadata are also registred and served by JCOMMOPS, the WMO-IOC Joint Technical Commission for Oceanography and Marine Meteorology in situ Observations Programme Support Centre http://www.jcommops.org/board "	
1.49 Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	"Yes, we use the Argo vocabulary tables CF convention parameters, P07 vocabulary in NERC server https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/P07/ We also use the Argo vocabulary tables, that are not yet implemented yet in the NERC vocabulary "	
1.50 Do you believe that your data is Findable (F)? if not, indicate where you see major gaps.	"Argo DOI is a huge success in term of findability: on March 24th 2019, there were 4249 publications linked to the GDAC data DOI (… everyday, one more publication). See https://www.seanoe.org/data/00311/42182/relateddoc.htm from: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) https://doi.org/10.17882/42182 To access individual data, Argo ERDDAP server allows machine to machine as well as human to machine queries. See https://www.ifremer.fr/erddap/tabledap/ArgoFloats.html "	yes
1.51 Do you believe that your data is Accessible (A)? if not indicate where you see major gaps	"Yes, data and metadata are accessible through ERDDAP server. Metadata are provided in ISO19115-2/19139 schema, in FGDC-STD-001-1998 schema (Federal Geographic Data Committee). The access protocol is OpenDAP, OGC-WMS (for geospatial observation) and RESTful Web Services. "	yes

Question	Response	Remark
1.52 Do you believe that your data is interoperable (I)? if not indicate where you see major gaps	"NetCDF format contributes to make the Argo data interoperable for criteria I1, but implementing a Argo vocabulary in link with existing community vocabularies is needed for criteria I2 and I3 The ERDDAP server plugged on top of the GDAC qualifies Argo for criteria I1, I2 and I3. "	partially
1.53 Do you believe that your data is re-usable (R)? if not, indicate where you see major gaps	Criteria R1.1 and R1.2 are fulfilled, improvement is needed to implement a Argo vocabulary in link with existing community vocabularies for criteria R1.3	partially
2.0 Timestamp	4/18/2019 11:38	
2.2 Email Address	Thierry.Carval@ifremer.fr	
2.4 Please provide the IRI of the resource to be evaluated	"Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) https://doi.org/10.17882/42182 " IS JUST THIS 10.17882/42182	this must be the metadata record, for automated validation, maybe a URL to start
2.5_F1A: Please provide the IRI for a registered identifier schema for your resource's IRI (e.g. DOI, HTTP):	"https://www.doi.org Within the DOI landing page, we also provide schma.org metadata including data and metadata download services https://schema.org/Dataset <!DOCTYPE html> <html lang=""fr""> <head> <meta charset=""utf-8""> <script type=""application/ld+json""> { ""@context"": ""http://schema.org"", ""@type"": ""Dataset"", ""@id"": ""https://doi.org/10.17882/42182"", ""name"": ""Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)"", ""url"": ""https://www.seanoe.org/data/00311/42182/"", ""sameAs"": ""https://doi.org/10.17882/42182"", ""thumbnailUrl"": ""https://www.seanoe.org/data/00311/42182/thumbnail.gif"", ""datePublished"": ""2019"", ""description"": ""Argo is a global array of 3,000 free-drifting profiling floats ..., via the GDACs within six months of collection. "", ""license"" : ""https://creativecommons.org/licenses/by/4.0/"", ""keywords"" : ""float, Argo, global ocean observing system, ocean circulation, in-situ, ocean pressure, sea water salinity, sea water	http://doi.org/ also: https://fairsharing.org/search/?q=identifier&selected_facets=type_exact:identifier%20schema

Question	Response	Remark
	temperature, multi-year, weather climate and seasonal observation, global-ocean"", ""publisher"" : {""@type"": ""Organization"",""name"": ""SEANOE""}, ""citation"" : ""Argo (2019). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. https://doi.org/10.17882/42182"", ""distribution"":[{ ""@type"": ""DataDownload"", ""encodingFormat"": ""NC, NetCDF"", ""contentUrl"": ""https://www.seanoe.org/data/00311/42182/data/42250.tar.gz"" }, "	
2.6_F1B: Please provide the IRI to the document describing the persistence policy for the identifier of this(meta)data (this may be a document from your service provider, e.g. Zenodo, UniProt, etc.):	"DOIs for ocean data, general principles and selected examples (Argo, French cruises) https://doi.org/10.13155/44515 "	https://doi.org
2.7_F2A: Please provide the IRI to a document that contains machine-readable metadata for the digital resource	"Argo DOI : https://doi.org/10.17882/42182 In the schem.org section, the Argo dataset metadata are machine readable. http://www.ifremer.fr/erddap/tabledap/ArgoFloats.html Argo ERDDAP metadata ISO19115-2: http://www.ifremer.fr/erddap/tabledap/ArgoFloats.iso19115 Argo ERDDAP metadata FGDC : http://www.ifremer.fr/erddap/tabledap/ArgoFloats.fgdc Argo Inspire metadata from Sextant catalogue : https://sextant.ifremer.fr/Donnees/Catalogue#/metadata/3df904de-e47d-4bf9-85a0-7c0942aff8b6"	https://doi.org/10.17882/42182
2.8_F2B: Please provide the IRI for the file format of this metadata:	https://www.w3.org/TR/turtle/. also: https://json-ld.org/ Argo user's manual : https://doi.org/10.13155/29825 NetCDF CF XML vocabulary : http://cfconventions.org/Data/cf-standard-names/64/src/cf-standard-name-table.xml Argo parameters vocabulary : XXX"	https://doi.org/10.13155/29825
2.9_F3: Please provide the IRI of the metadata	"Seanoe : https://doi.org/10.17882/42182 "	When I resolve to metadata from datacite I get this: curl -L -H "Accept: text/turtle"

Question	Response	Remark
		https://doi.org/10.17882/42182 @prefix schema: <http://schema.org/> . <https://doi.org/10.17882/56503> a schema:CreativeWork; schema:isBasedOn <https://doi.org/10.17882/42182> . <https://doi.org/10.17882/57596> a schema:CreativeWork; schema:isBasedOn <https://doi.org/10.17882/42182> . <https://doi.org/10.17882/42182> a schema:Dataset; schema:author [a schema:Person; schema:givenName "Argo"; schema:name ", Argo"]; schema:datePublished "2019"^^schema:Date; schema:description "Argo is a global array of 3,000 free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean. This allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection. The array provides 100,000 temperature/salinity profiles and velocity measurements per year distributed over the global oceans at an average of 3-degree spacing. Some floats provide additional bio-geo parameters such as oxygen or chlorophyll. All data collected by Argo floats are publically available in near real-time via the Global Data Assembly Centers (GDACs) in Brest (France)

Question	Response	Remark
		and Monterey (California) after an automated quality control (QC), and in scientifically quality controlled form, delayed mode data, via the GDACs within six months of collection."; schema:identifier [a schema:PropertyValue; schema:value "https://doi.org/10.17882/42182"]; schema:keywords "float, Argo, global ocean observing system, ocean circulation, in-situ, ocean pressure, sea water salinity, sea water temperature, multi-year, weather climate and seasonal observation, global-ocean"; schema:license <https://creativecommons.org/licenses/by/4.0>; schema:name "Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)"; schema:provider [a schema:Organization; schema:name "DataCite"]; schema:publisher [a schema:Organization; schema:name "SEANOE"]; schema:schemaVersion <http://datacite.org/schema/kernel-3>; schema:url <https://www.seanoe.org/data/00311/42182/> .
2.10_F3: Please provide the IRI of the data described by the metadata above	"In Argo GDAC DOI https://doi.org/10.17882/42182 ""distribution"":[{ ""@type"": ""DataDownload"", ""encodingFormat"": ""NC, NetCDF"", ""contentUrl"": ""https://www.seanoe.org/data/00311/42182/data/42250.tar.gz"" }, { ""@type"": ""DataDownload"", ""encodingFormat"": ""NC, NetCDF"", ""contentUrl"": ""https://www.seanoe.org/data/00311/42182/data/42251.tar.gz"" },"	this is the distribution segment in the DOI that describes the url to the datafile.

Question	Response	Remark
2.11_F4: Please provide the URL to a search engine where the dataset can be found.	"seanoe.org is indexed among others by :	search at the moment via Bing in validator, if the DOI can be found. Very generic test if resource can be found.
2.12_F4: Search query/terms: the query that will be executed to discover your RESOURCE ID (found in the first page of the search):	"Argo floats data" shows Argo GDAC DOI in the first page of the search "Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)"	answer is ok.
2.13_A1.1 Please provide a URL to the description of the Acess Protocol	https://doi.org/10.17882/42182 https://www.w3.org/Protocols/rfc2616/rfc2616.html	https
2.14_A1.1 Is the protocol open?	Yes	yes
2:15_A1.1 Is the protocol (royalty) free?	Yes	yes
2.16_A1.2 Authorization is required to access the content of my resource	No	no
2.17_A1.2 If "yes" above, please provide a IRI that resolves to a description of the process to obtain access to restricted content		VOID
2.18 A2 Please provide the URL to a metadata longevity plan	https://www.nodc.noaa.gov/worlddatacenter/	https://www.nodc.noaa.gov/worlddatacenter/ + policy of the DOI provider
2.19_I1 Please provide the URL to the specification of the language.	http://cfconventions.org/Data/cf-standard-names/64/src/cf-standard-name-table.xml https://www.w3.org/RDF/	this is the code language, in this case RDF, of datacite and of the landing page of the DOI (json LD)
2.20_I2 Please provide one or more (max 3) IRIs from vocabularies used within the(meta)data that is returned by resolving the RESOURCE ID.	"Argo parameters vocabulary : http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P06 http://schema.org, DCAT, PROV	correct answer here is schema.org (both DataCite, and your landing page, contain metadata using schema.org; HOWEVER, if you wanted to be MORE FAIR, you could enhance the metadata in your landing page using other vocabularies. I assume that, in your data, you do use more specific vocabularies?
2.21_I3 Please provide the URL to a formal Linkset (defined at: https://www.w3.org/TR/void/#linkset) or copy/paste the content of a formal linkset that describes at least a portion of the content at RESOURCE ID (you may	"Is the link between an Argo parameter (such as sea water salinity) and its CF community standard name relevant for this question. NetCDF variable standard_name attribute : cf_standard_name : sea_water_salinity SeaDataNet_parameter_uri :	none

Question	Response	Remark
need to create this, or edit the example below) e.g. :DBpedia2DBLP a void:Linkset;void:target :DBpedia;void:target :DBLP;void:linkPredicate owl:sameAs;void:triples 10000;	http://vocab.nerc.ac.uk/collection/P01/current/CNDCST01/ cf_standard_name_uri : from http://cfconventions.org/Data/cf-standard-names/64/src/cf-standard-name-table.xml <entry id=""sea_water_salinity""><canonical_units>1e-3</canonical_units><grib>88</grib><amip>so</amip><description>Sea water salinity is the salt content of sea water, often on the Practical Salinity Scale of 1978. However, the unqualified term 'salinity' is generic and does not necessarily imply any particular method of calculation. The units of salinity are dimensionless and the units attribute should normally be given as 1e-3 or 0.001 i.e. parts per thousand. There are standard names for the more precisely defined salinity quantities: sea_water_knudsen_salinity, S_K (used for salinity observations between 1901 and 1966), sea_water_cox_salinity, S_C (used for salinity observations between 1967 and 1977), sea_water_practical_salinity, S_P (used for salinity observations from 1978 to the present day), sea_water_absolute_salinity, S_A, sea_water_preformed_salinity, S_*, and sea_water_reference_salinity. Practical Salinity is reported on the Practical Salinity Scale of 1978 (PSS-78), and is usually based on the electrical conductivity of sea water in observations since the 1960s. Conversion of data between the observed scales follows: S_P = (S_K - 0.03) * (1.80655 / 1.805) and S_P = S_C, however the accuracy of the latter is dependent on whether chlorinity or conductivity was used to determine the S_C value, with this inconsistency driving the development of PSS-78. The more precise standard names should be used where appropriate for both modelled and observed salinities. In particular, the use of sea_water_salinity to describe	

Question	Response	Remark
	salinity observations made from 1978 onwards is now deprecated in favor of the term sea_water_practical_salinity which is the salinity quantity stored by national data centers for post-1978 observations. The only exception to this is where the observed salinities are definitely known not to be recorded on the Practical Salinity Scale. The unit ""parts per thousand"" was used for sea_water_knudsen_salinity and sea_water_cox_salinity.</description></entry> "	
2.22_R1.1 Please provide the IRI for your usage license regarding the content returned from RESOURCE ID (be that data, or metadata):	Licence CC-BY: https://creativecommons.org/licenses/by/4.0/	https://creativecommons.org/licenses/by/4.0/
2.23_R1.2 Please provide the IRIs (maximum 3) for the vocabularies being used to describe the provenance of the content resolved from RESOURCE ID (be that data, or metadata; e.g.http://purl.org/dc/terms/):	"Argo GDAC DOI https://doi.org/10.17882/42182 <!DOCTYPE html> <html lang=""fr""> <head> <meta charset=""utf-8""> <script type=""application/ld+json""> { ""@context"": ""http://schema.org"", ""@type"": ""Dataset"", ""@id"": ""https://doi.org/10.17882/42182"", ""name"": ""Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)"", ""url"": ""https://www.seanoe.org/data/00311/42182/"", ""sameAs"": ""https://doi.org/10.17882/42182"", ""thumbnailUrl"": ""https://www.seanoe.org/data/00311/42182/thumbnail.gif"", ""datePublished"": ""2019"", ""description"": ""Argo is a global array of 3,000 free-drifting profiling floats t... six months of collection."", ""license"" : ""https://creativecommons.org/licenses/by/4.0/"", ""keywords"" : ""float, Argo, global ocean observing system, ocean circulation, in-situ, ocean pressure, sea water salinity, sea water temperature, multi-year, weather	NULL

Question	Response	Remark
	climate and seasonal observation, global-ocean"", ""publisher"" : {""@type"": ""Organization"",""name"": ""SEANOE""}, ""citation"" : ""Argo (2019). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. https://doi.org/10.17882/42182"", ""distribution"":[{ "	
2:24_R1.2 Please provide the IRIs (maximum 3) for the vocabularies being used to describe the domain information of the content resolved from RESOURCE ID (be that data, or metadata;e.g. http://edamontology.org/data_ will be used for many bioinformatics data/metadata):	cf_standard_name_uri : from http://cfconventions.org/Data/cf-standard-names/64/src/cf-standard-name-table.xml	https://doi.org/10.17882/42182
2.25_R1.3 Please provide the IRI that represents the certification from a recognized authority in your community or domain, indicating that the content of RESOURCE ID is compliant with the standards of your community	"Argo NetCDF file format checker : https://doi.org/10.17882/45538 NetCDF CF compliance checker : http://cfconventions.org/compliance-checker.html"	https://doi.org/10.17882/45538
2.3 Research Infrastructure Name	Euro-Argo ERIC	
2.1 Contact name	Thierry Carval	
2.2 Email		

LifeWatch Marine Data Archive

Question nr.	Question	Response
0	Date of response	4/4/2019
General		
1	Contact name *	Klaas Deneudt, Katrina Exter
2	Email*	klaas.deneudt@vliz.be, katrina.exter@vliz.be
3	Research Infrastructure Name *	LifeWatch (Marine)
4	Please provide the URL of one of the datasets in scope for your answers	http://www.marinedataarchive.org/datasets/imis?module=dataset&dasid=1606
5	Please provide the URL to the	http://www.marinedataarchive.org

Question nr.	Question	Response
	discovery portal in which the dataset can be downloaded	/
Repositories		
6	Which repositories do you use for data?	Marine Data Archive
7	Which repository software is being used?	Home-made (MS SQL)
8	Which repositories do you use for metadata?	IMIS (integrated marine information system; http://www.vliz.be/en/integrated-marine-information-system)
9	Do your repositories use PIDs? If so which PID systems?	Short answer: UUID - Every data file stored in the Marine Data Archive is assigned with an UUID called MDA ID. - DOIs are assigned to datasets that have a data record in IMIS if requested - IMIS PIDs are constructed from the record sequence number, are unique, and are part of the URL.
10	Do you assign PIDs manually or automatically?	- IMIS UUID automatically - DOI manually
11	Which PID registration provider do you use?	-MDA UUIDs are created by the system (local service) - DOI is acquired through DataCite
12	Do you use the PID Record to store attributes about the data?	Relevant metadata - MDA UUIDs: the data have some metadata attached to those data, separately from any IMIS entry, although these are not separately findable. - The DOI is linked to a set of metadata elements specific for the version of the dataset. These elements are exchanged with DataCite.
13	Are these repositories certified? If so, which methods are used?	IODE certified; CoreTrustSeal
14	Are repository policies mentioned at the website? If so, indicate the major ones.	Terms of Use and FAQ available through home page (marinedataarchive.org/mdadatapolicy.pdf)
15	Are your repositories registered in a registry? If so which registry?	No
16	Which persistency guaranties are typically given?	To be written No specific persistence guarantees mentioned, but management of the database and the metadata catalogue is part of the mandate of the VLIZ Data Centre.
Data		
17	Which are the most popular data	No requirements on the data types

Question nr.	Question	Response
	types used?	stored in the MDA, can be anything from ascii to omics data types -> Any
18	Which are the preferred data formats?	See Q17
19	Do those formats include metadata headers? if so, which?	No requirements for this are set by the MDA; probably partially
20	Do you provide search on data?	Yes: search through discovery metadata is based on IMIS APIs. Specific search at data file level within MDA is also possible but requires login and access rights
21	Did you register your schemas in a common registry?	No
Metadata		
22	Which metadata schemas are mostly used?	ISO 19115, xml (eml for a subset) Metadata added to data record directly in MDA is home-made schema
23	Are all categories used in the schemas defined in open registries?	In the metadata descriptions we use Marine Regions for metadata on geographic coverage, WoRMS for information on taxonomic coverage, ASFA for thematic keywords, BODC vocabularies for metadata on measured parameters and parameter groups Most
24	How is provenance included?	No specific provenance metadata schemas, but partial provenance information is covered by metadata fields such as: contact details, publication link, description, data origin, dates. Text only
25	Are PIDs included in the metadata description?	For the discovery metadata, DOI are part of dataset citation, URL download link and/or IMIS record URL are part of the metadata
26	What is the primary storage format for metadata?	Metadata is stored in a relational database. Storage format is ISO19115 compliant.
27	Which are the export formats supported?	For discovery metadata, output available in xml, json, rss, html formats
28	Which metadata exchange/harvesting methods are supported?	Working towards OAI-PMH compliance
29	Do you have a local search engine?	Yes, IMIS search APIs
30	Do you support external search engines?	No, not for data simply archived in the MDA. Data with a record in IMIS may have their metadata exported (e.g. GBIF).

Question nr.	Question	Response
		If a DOI is requested metadata is exchanged with DataCite.
31	Do you make statements about access policies in your metadata?	Usually. Metadata author or data provider can chose from a range of access constraints. Use of CC licensing is promoted.
32	Is your metadata machine actionable?	Where eml is available, yes. Otherwise, we are working towards making all records have eml schema.
Access mechanisms		
33	How is authentication done?	For search and download no authentication required. For advanced editing account login is required. Simple username and passworkd, no machine2machine authentication.
34	Do you maintain an own user database?	Yes
35	Do you use ORCID in your AAI?	No
36	What is the major access technology supported?	Webform using HTTP POST
37	How is authorization done?	For advanced editing, via user profile stored in a SSQL servive database
38	Which specific licenses do you use for your data?	CC mainly
39	Are metadata openly available?	Yes, in case a discovery metadata record has been created
Data Management Plans		
40	Do you use or provide specific DMP tools? If so, which DMP tool are you using or advocating in your community?	No
41	Do you apply special data publishing steps?	No. Webform is available to publish data (DOI) but is up to data owner if they want to do that
Data processing		
42	Do you apply special data [***processing***] steps?	no
43	Do you apply workflow frameworks for processing your data?	no
44	Do you use distributed workflow tools? if so, which?	no
45	Do you offer other type of support or analytics services?	no
46	Do you offer data products in your RI?	no
Semantics		

Question nr.	Question	Response
47	Do you use semantic vocabularies from generic vocabularies, ontologies, etc.? If so point to the registries.	For data files described as dataset in the metadata catalogue: WoRMS (http://www.marinespecies.org/); Marine Regions (http://www.marineregions.org/)
48	Do you use discipline specific vocabularies, ontologies etc.? If so point to the registries.	Yes. WoRMS (http://www.marinespecies.org/); Marine Regions (http://www.marineregions.org/)
49	Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	No
FAIRness		
50	Do you believe that your data is Findable (F)? If not, indicate where you see major gaps.	Yes (if described in the metadata catalogue)
51	Do you believe that your data is Accessible (A)? If not indicate where you see major gaps	Partially: if described in the metadata catalogue but records need cleaning up
52	Do you believe that your data is interoperable (I)? If not indicate where you see major gaps	Partly: most data are in standard file formats but if not for specific data publication creation, no checks are run on the format of the actual data therein
53	Do you believe that your data is re-usable (R)? If not, indicate where you see major gaps	Partly; data with an IMIS record are supposed to indicate reusability but records need cleaning up

LifeWatch - EUROBIS

Question nr.	Question	Response
0	Date of response	4/4/19
General		
1	Contact name *	Klaas Deneudt, Katrina Exter
2	Email*	klaas.deneudt@vliz.be, katrina.exter@vliz.be
3	Research Infrastructure Name *	LifeWatch (Marine)
4	Please provide the URL of one of the datasets in scope for your answers	http://www.eurobis.org/imis?module=dataset&dasid=5514
5	Please provide the URL to the discovery portal in which the dataset can be downloaded	http://www.emodnet-biology.eu/toolbox/en/download/occurrence/explore
Repositories		
6	Which repositories do you use for data?	EUROBIS
7	Which repository software is being used?	EUROBIS
8	Which repositories do you use for metadata?	IMIS (integrated marine information system; http://www.vliz.be/en/integrated-marine-information-system)
9	Do your repositories use PIDs? If so	Short answer: Yes. UUID for data

	which PID systems?	held in the MDA or GBIF, or download using IMIS ID - Every data file stored in the Marine Data Archive is assigned with an UUID called MDA ID. - DOIs are assigned to most, but but all, datasets that have a data record in IMIS, or which are send to GBIF - IMIS PIDs are constructed from the record sequence number, are unique, and are part of the URL.
10	Do you assign PIDs manually or automatically?	- UUIDs Automatically - DOI manually
11	Which PID registration provider do you use?	- UUIDs are created by the MDA or GBIF system - DOI is acquired through DataCite
12	Do you use the PID Record to store attributes about the data?	Relevant metadata - Some LIFEWATCHdata (whether GBIF- or MDA-provided) have metadata separately from any in the IMIS entry, although these are not separately findable. - The DOI is linked to a set of metadata elements specific for the version of the dataset. These elements are exchanged with DataCite.
13	Are these repositories certified? If so, which methods are used?	IODE certified; CoreTrustSeal
14	Are repository policies mentioned at the website? If so, indicate the major ones.	Terms of Use and FAQ available through home page (http://www.eurobis.org/citation)
15	Are your repositories registered in a registry? If so which registry?	EUROBIS is listed in fairsharing.org
16	Which persistency guaranties are typically given?	To be written No specific persistence guarantees mentioned, but management of the database and the metadata catalogue is part of the mandate of the VLIZ Data Centre.
Data		
17	Which are the most popular data types used?	Tracking and monitoring data in structures lists, occurance data, biotic and abiotic associated data (structured lists), sound files, images
18	Which are the preferred data formats?	zip files, spreadsheets, xml, Darwin Core Archive (DwC), OBIS event data scheme
19	Do those formats include metadata headers? if so, which?	Yes if provided as Darwin Core Archive format (not all data are so)
20	Do you provide search on data?	Yes: search through discovery metadata is based on IMIS APIs.

		Specific data-level search is realised through the EMODnet Biology portal (http://www.emodnet-biology.eu/portal/index.php)
21	Did you register your schemas in a common registry?	Yes: OBIS event scheme has been published in the Biodiversity data journal and registered as an extension in the Integrated Publishing Toolkit
Metadata		
22	Which metadata schemas are mostly used?	ISO 19115, eml
23	Are all categories used in the schemas defined in open registries?	In the metadata descriptions we use Marine Regions for metadata on geographic coverage, WoRMS for information on taxonomic coverage, ASFA for thematic keywords, BODC vocabularies for metadata on measured parameters and parameter groups. Most
24	How is provenance included?	No specific provenance metadata schemas, but partial provenance information is covered by metadata fields such as: contact details, publication link, description, data origin, dates Text only
25	Are PIDs included in the metadata description?	For the discovery metadata, DOI are part of dataset citation, URL download link and/or IMIS record URL are part of the metadata
26	What is the primary storage format for metadata?	Metadata is stored in a relational database. Storage format is ISO19115 compliant.
27	Which are the export formats supported?	For discovery metadata, output available in xml (eml), json, rss, html formats
28	Which metadata exchange/harvesting methods are supported?	Working towards OAI-PMH compliance
29	Do you have a local search engine?	Yes, IMIS search APIs
30	Do you support external search engines?	Yes. At metadata level we export subsets to EDMED (https://www.seadatanet.org/Metadata/EDMED-Datasets), GCMD (https://gcmd.nasa.gov/), GBIF (for export to EUROBIS); If a DOI is created, metadata is exchanged with DataCite
31	Do you make statements about access policies in your metadata?	CC (CC BY)
32	Is your metadata machine actionable?	Where eml is available, yes. Otherwise, we are working towards making all records have

		eml schema.
Access mechanisms		
33	How is authentication done?	For search and download no authentication required. For advanced editing account login is required. Simple username and passworkd, no machine2machine authentication.
34	Do you maintain an own user database?	Yes
35	Do you use ORCID in your AAI?	No
36	What is the major access technology supported?	Webform using HTTP POST
37	How is authorization done?	For advanced editing, via user profile stored in a SSQL servive database
38	Which specific licenses do you use for your data?	CC mainly
39	Are metadata openly available?	Yes
Data Management Plans		
40	Do you use or provide specific DMP tools? If so, which DMP tool are you using or advocating in your community?	DMP to be written, will be based on H2020 template and DMPonline
41	Do you apply special data publishing steps?	Yes, for DOI registration. Yes, using a webform to publish the data Yes for uptake into EuROBIS (DwC formatting, WoRMS taxon matching, etc.)
Data processing		
42	Do you apply special data [***processing***] steps?	Yes. QC is applied to the collected data and the data placed in EUROBIS
43	Do you apply workflow frameworks for processing your data?	Yes. Home-made (R) Through LifeWatch e-lab specific data processing procedures (taxon matching, geolocation, qc, ..) can be performed
44	Do you use distributed workflow tools? if so, which?	Yes. Home-made (R) LifeWatch has webservices to access data and specific virtual labs to standardise, analyse and visualise data; Lifewatch also offers an online data explorer; EMODNET provides an online map viewer for the datasets in EurOBIS
45	Do you offer other type of support or analytics services?	Yes. Home-made (R) R packages exist for specific access and processing procedures.
46	Do you offer data products in your RI?	Yes e.g. geospatial datasets, integrated standardised datasets

Semantics		
47	Do you use semantic vocabularies from generic vocabularies, ontologies, etc.? If so point to the registries.	WoRMS (http://www.marinespecies.org/); Marine Regions (http://www.marineregions.org/)
48	Do you use discipline specific vocabularies, ontologies etc.? If so point to the registries.	Yes. WoRMS (http://www.marinespecies.org/); Marine Regions (http://www.marineregions.org/)
49	Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	No
FAIRness		
50	Do you believe that your data is Findable (F)? If not, indicate where you see major gaps.	Yes
51	Do you believe that your data is Accessible (A)? If not indicate where you see major gaps	Yes (to a large majority)
52	Do you believe that your data is interoperable (I)? If not indicate where you see major gaps	Yes
53	Do you believe that your data is re-usable (R)? If not, indicate where you see major gaps	Yes

EMSO

Question	Response	Updated answers (short for YAML)
1.0 Timestamp	3/5/2019 01:27	
1.2 Email Address	ivan.rodero@emso-eu.org	
1.1 Contact name	Ivan Rodero	
1.3 Research Infrastructure Name	EMSO ERIC	
1.4 Please provide the URL of one of the datasets in scope for your answers	http://www.moist.it/sites/western_ionian_sea/2/SMO1#datasets	http://data.emso.eu/files/emso/azores/exif0001/EXIF0001_1944-2019.nc
1.5 Please provide the URL to the discovery portal in which the dataset can be downloaded	http://moist.it	http://data.emso.eu
1.6 Which repositories do you use for data?	EMSO ERIC is a Research Infrastructure distributed across different heterogeneous Regional Facilities and Test Sites. EMSO ERIC does not have a central repository yet. Multiple and distributed data repositories are currently in use, including custom-made internal repositories, cloud (e.g., AWS) repositories, and third-party repositories. Third party	http://data.emso.eu/files/

Question	Response	Updated answers (short for YAML)
	repositories include: EMODnet, PANGAEA, CORIOLIS, OCEANSITES, CMEMS, MOIST.IT, BODC, SEANOE, and SeaDataNet. An interim file-based centralized data repository is available through an web server.	
1.7 Which repository software is being used?	In addition to the repository software deployed by third-party repositories, custom-made (e.g., based on Python and Django,) based on 52-North SOS, THREDDS, OpenDAP and ERDDAP. EMSO ERIC does not use repository software such as Fedora/hyrax or similar. The interim file-based centralized data repository is based on an Apache HTTP server.	Apache HTTP server
1.8 Which repositories do you use for metadata?	Currently, EMSO ERIC does not have a central metadata system but it is distributed across Regional Facilities and integrated into different systems using PostgreSQL, THREDDS (netCDF files), raw netCDF files and custom solutions. Metadata is also stored in third-party repositories (e.g., MOIST.IT, SeaDataNet, EMODNet, DataCite, data.gov.ie, data.marine.ie, BODC, SEANOE).	Apache HTTP server
1.9 Do your repositories use PIDs? If so which PID systems?	The use of PIDs varies across the different regional facilities (e.g., at least two facilities do not use PIDs). PIDs used include: DOIs (e.g., via DataCite and PANGAEA), NetCDF Oceansites ID attribute (.nc file name), and NetCDF EGO ID attribute (.nc file name).	PID
1.10 Do you assign PIDs manually or automatically?	In most of the cases the PIDs are manually (e.g., via PANGAEA: https://wiki.pangaea.de/wiki/DOI).	manually
1.11 Which PID registration provider do you use?	PID registration providers include DataCite, and PANGAEA Data Publisher.	planned
1.12 Do you use the PID Record to store attributes about the data?	Although it varies across regional facilities, in some cases the PID records used to store data attributes are DataCite metadata fields.	planned
1.13 Are these repositories certified? If so, which methods are used?	It is unknown if the third-party repositories are certified.	none
1.14 Are repository policies mentioned at the	The main EMSO ERIC website does not provide repository policies;	https://creativecommons.org/licenses/by/4.0/

Question	Response	Updated answers (short for YAML)
website? If so, indicate the major ones.	however, some regional facilities provide repository policies (e.g., http://data.plocan.eu/thredds/DataPolicy.pdf).	
1.15 Are your repositories registered in a registry? If so which registry?	Repositories registered in a registry include EMODNet registry and data.gov.ie.	none
1.16 Which persistency guaranties are typically given?	The different regional facilities implement their local persistency guaranties policies; however, EMSO ERIC does not currently have unified persistency guaranties. Facilities may implement generic agreements on ocean observing with central and regional governments but there are not explicit persistency guaranties.	none
1.17 Which are the most popular data types used?	Most popular data types include time series, sound, profiles, trajectories, model forecast data, CTD and vessel underway, fisheries effort, bathymetry, image, video.	binary
1.18 Which are the preferred data formats?	Preferred data formats include NetCDF, CSV, WAV, JSON, MSEED, Excel, O&M.	NetCDF CF
1.19 Do those formats include metadata headers? if so, which?	In general data includes metadata headers including NetCDF Oceansites Profile metadata (http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf), NetCDF EGO format metadata (https://archimer.ifremer.fr/doc/00239/34980/), standard NetCDF CF headers, subset of CSV/ASCII are served as SEADATANET ODV with the included semantic header, metadata are included directly in NetCDF e.g., dimensions, variables, global attributes (principal investigator, principal investigator email, principal investigator url, institution, project, network, metadata link, geospatial information, time coverage information, data type, conventions, metadata conventions, netcdf version, publisher information, license, etc.).	NetCDF ACDD
1.20 Do you provide search on data?	It varies across regional facilities, some provide data catalogues and search functionalities (e.g., opensearch) and tools that provide search functionalities such as	planned

Question	Response	Updated answers (short for YAML)
	ERDDAP. In addition, third party repositories provide search on data.	
1.21 Did you register your schemas in a common registry?	EMSO ERIC does not currently register the schemas in a common registry; however, some facilities use schemas/formats already present in a common registry.	yes
1.22 Which metadata schemas are mostly used?	The metadata schemas mostly used in EMSO ERIC include XML, NetCDF (e.g., NetCDF Oceansites, NetCDF EGO), ISO19139, Excel.	NetCDF CF
1.23 Are all categories used in the schemas defined in open registries?	It varies across regional facilities but metadata typically uses well-defined semantic concepts and use standard conventions such as Climate and Forecast (CF) metadata conventions for variables standard names (http://cfconventions.org/Data/cf-standard-names/65/build/cf-standard-name-table.html) and geometries (http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html#appendix-examples-discrete-geometries), and NERC Vocabulary Service (e.g., http://vocab.nerc.ac.uk/collection/P01/current/, https://vocab.nerc.ac.uk/collection/P06/current/ for NetCDF).	planned
1.24 How is provenance included?	It varies across regional facilities but mechanisms for including provenance include NetCDF Oceansites profile metadata attributes (site_code, platform_code, wmo_platform_code, principal_investigator, institution, network) and NetCDF EGO metadata attributes (platform_code, wmo_platform_code, institution, institution_references, SeaDataNet_edmo_code - SeaDataNet EDMO code of the institution-, contact, author, data_assembly_center, principal_investigator, principal_investigator_email, observatory, deployment_code, deployment_label).	partially
1.25 Are PIDs included in the metadata description?	In general PIDs are included in the metadata description where they exist. An example is the	yes

Question	Response	Updated answers (short for YAML)
	"publisher_url" in NetCDF Oceansites.	
1.26 What is the primary storage format for metadata?	The primary storage formats for metadata are heterogeneous across regional facilities and include custom solutions (e.g., human-readable PDF), relational databases, Drupal and GeoNetwork database, WAV files headers for acoustic data, and metadata are also stored in a PostgreSQL database, then used for creating JSON files which feed the production of NetCDF files.	NetCDF CF HTML XML
1.27 Which are the export formats supported?	Export formats include HTML, XML, JSON, RDF, all formats produced by NetCDF conversion toolboxes, all formats compatible with .wav PCM, Excel.	NetCDF CSV
1.28 Which metadata exchange/harvesting methods are supported?	It varies across regional facilities but they include methods compatible with THREDDS, OAI-PMH, CSW (available as standard with GeoNetwork).	OAI-PMH OpenDAP OGC CSW
1.29 Do you have a local search engine?	EMSO ERIC does not currently have a centralized search engine; however, repositories used by the Regional Facilities and third-party systems provide metadata search engines (e.g., http://www.moist.it/search/html and https://data.marine.ie).	planned
1.30 Do you support external search engines?	Yes, EMSO ERIC data and metadata is integrated into third-party search engines such as EMODnet, CMEMS, CORIOLIS, and dataset metadata as Schema.org on both local catalogue and ERDDAP server.	yes
1.31 Do you make statements about access policies in your metadata?	In general EMSO ERIC metadata make statements about access; however, this is not uniform across regional facilities. Examples include free access upon request, a data policy which clearly states access and licencing terms (https://www.marine.ie/Home/site-area/data-services/data-policy/data-policy), NetCDF metadata excerpt: licence = 'Follows FixO3 (FIXed-point Open Ocean Observatory)standards, cf. Data available free of charge. User assumes all risk for use of data. User must display citation in any publication or product using data.	yes

Question	Response	Updated answers (short for YAML)
	User must contact PI prior to any commercial use of data.' -, distribution statement. Distribution_statement = "Follows CLIVAR (Climate Varibility and Predictability) standards, cf. http://www.clivar.org/data/data_policy.php. Data available free of charge. User assumes all risk for use of data. User must display citation in any publication or product using data. User must contact PI prior to any commercial use of data.",	
1.32 Is your metadata machine actionable?	In general EMSO ERIC metadata is machine actionable; however, this is not uniform across regional facilities.	partially
1.33 How is authentication done?	In general data access does not require authentication; however, some datasets are currently provided under demand by the Regional Facilities. The authentication mechanisms are not uniform across Regional Facilities and include HTTP Authentication.	void
1.34 Do you maintain an own user database?	It varies across regional facilities. EMSO ERIC does not have a centralized user database at this point	no
1.35 Do you use ORCID in your AAI?	No.	no
1.36 What is the major access technology supported?	The major access technology supported include HTTP, FTP, THREDDS, ERDDAP.	HTTP FTP THREDDS ERDDAP
1.37 How is authorization done?	EMSO ERIC does not currently implement a centralized authorization system; however it varies across regional facilities and includes web login (user name/password) and web login with token (client_ID, client_secret).	void
1.38 Which specific licenses do you use for your data?	It varies across regional facilities but specific licenses include project-based licenses such as FixO3 (Fixed-point Open Ocean Observatory) standards and Copernicus License (i.e., data available free or charge, user assumes all risk for use of data, user must display citation in any publication or product using data, user must contact PI prior to any	CC-BY-4.0

Question	Response	Updated answers (short for YAML)
	commercial use of data), EGO network data policy (real-time data in EGO format are publicly available through GDAC FTP), Creative common CC BY 4.0 (e.g., https://www.marine.ie/Home/sites/default/files/MIFiles/Docs/DataServices/Marine%20Institute%20Data%20Policy%202017.pdf).	
1.39 Are metadata openly available?	Yes, in general metadata is openly available.	yes
1.40 Do you use or provide specific DMP tools?	It varies across regional facilities, including a DMP template as part of a data management – quality management framework implementation pack and a data policy document where the data management strategy is defined (e.g., http://data.plocan.eu/thredds/DataPolicy.pdf).	planned
1.41 Do you apply special data publishing steps?	EMSO ERIC does not have an integrated data curation process at this time, the data publishing workflow varies across facilities (and dataset), and include basic QA/QC, data format conversion, derived variables calculation, data format standardization, data sync automation and scientific QC. Undergoing efforts include the implementation of QARTOD quality tests.	QA/QC
1.42 Do you apply special data processing steps?	In general no special data processing steps are taken, some additional community standard processing is applied to hydrophone data.	none
1.43 Do you apply workflow frameworks for processing your data?	No.	none
1.44 Do you use distributed workflow tools? if so, which?	No.	none
1.45 Do you offer other type of support or analytics services?	In general no, but it is possible upon request.	planned
1.46 Do you offer data products in your RI?	In general no, but it varies across regional facilities. Some facilities offer data products for near real time and delayed mode, and model output.	planned
1.47 Do you use semantic vocabularies from generic	It varies across regional facilities but they include Oceansites reference table 2.1	ACDD CF NVS

Question	Response	Updated answers (short for YAML)
vocabularies, ontologies, etc.? If so point to the registries.	(http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf), CF Standard names vocabulary (http://cfconventions.org/Data/cf-standard-names/65/build/cf-standard-name-table.html), CF metadata conventions (http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html), NERC Vocabulary Service (e.g., http://vocab.nerc.ac.uk/collection/P01/current/, http://vocab.nerc.ac.uk/collection/P06/current/ for NetCDF).	
1.48 Do you use discipline specific vocabularies, ontologies etc.? If so point to the registries.	It varies across regional facilities but they include NetCDF Oceansites conventions and NetCDF CF conventions (http://cfconventions.org), CF Standard names vocabulary, CF metadata conventions, NERC Vocabulary Server.	ACDD CF NVS
1.49 Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	No.	no, none
1.50 Do you believe that your data is Findable (F)? if not, indicate where you see major gaps.	EMSO ERIC data in general is findable (including third-party systems such as Oceansites, FixO3 portal, local portals, EGO Network portal, EMODNet, and CMEMS); however, data and metadata are not homogenous across regional facilities. The major gap is the integration of the different regional facilities in a way that data findable in a more uniform manner.	yes, none
1.51 Do you believe that your data is Accessible (A)? if not indicate where you see major gaps	EMSO ERIC data is accessible using different mechanisms (e.g., third-party systems such as EMODNet and CMEMS portals). Integrating data access is an undergoing effort to harmonize data access and, as a result, improve its accessibility. The lack of a single point of access (e.g., API) is another gap that EMSO ERIC is currently addressing.	partially, harmonization needs completion
1.52 Do you believe that your data is interoperable (I)? if not indicate where you see major gaps	EMSO ERIC data is interoperable in general (e.g., it is integrated into different third-party repositories); however, integration efforts are required to improve interoperability, including more mature	partially, harmonization needs completion

Question	Response	Updated answers (short for YAML)
	standardization, better semantics and standard metadata.	

Appendix 3: YAML content

ICOS (Marine)

https://github.com/envri-fair/fairness-assessment/blob/master/descriptions/ICOS-20190401.yaml

```
survey:
  date: 2019-04-01
  version: 1
  creator:
    name: Alex Vermeulen
    email: alex.vermeulen@icos-ri.eu
infrastructure:
  acronym: ICOS
  name: Integrated Carbon Observation System
  website: http://www.icos-cp.eu
  domain:
  - marine
  - atmosphere
  - ecosystem
  URL/IRI of dataset: https://hdl.handle.net/11676/-ffoiHjX5NDN0Vq_fKuVmas0
  URL of discovery portal: https://data.icos-cp.eu/portal
  repositories:
  - URL: https://data.icos-cp.eu/portal
    name: Carbon Portal
    kind:
    - metadata repository
    - data repository
    allocation: central
    software:
    - iRODS
    identifier:
    - kind: PID
      system: Handle
      landing page: NULL
      assigned: automatically
      provider: PDC
      includes metadata schema: none
    - kind: PID
      system: DOI
      landing page: yes
      assigned: manually
      provider: DataCite
      includes metadata schema: NULL
    certification methods: none
    policies:
    - usage
    registries:
    - re3data
    persistency-guaranty: 20 years
    access mechanisms:
      authentication method: OAuth paired with eduGain
      access protocol URL: https://doi.org/10.17487%2FRFC2616
      access without costs: yes
```

```
    own user database maintained: yes
    person identification system: ORCID
    major access technology supported: HTTP GET
    authorisation technique: OAUTH
    authorization for accessing content needed: no
    data licenses in use:
    - CC-BY4.0
    data license IRI: https://creativecommons.org/licenses/by/4.0/
    metadata openly available: yes
  data:
  - type name: time series
    preferred formats:
    - format name: NetCDF
      metadata types in data headers:
      - CF-1.4
    - format name: CSV
      metadata types in data headers:
      - parameters measured
      - units
    registered data schema: none
    search on data: no
  - type name: spatial
    preferred formats:
    - format name: raster data
      metadata types in data headers: none
    registered data schema: none
    search on data: planned
  metadata:
    schema:
    - URL: NULL
      name: ISO 19115
      provenance fields included:
      - text only
    - URL: https://semiceu.github.io/GeoDCAT-AP/releases/1.0.1/geodcat-ap_1.0.1.pdf
      name: GeoDCAT
      provenance fields included:
      - simplified PROV-O
    categories defined in registries: yes
    machine readable provenance: no
    PIDs included: yes
    primary storage format: RDF
    export formats supported:
    - JSON
    - XML
    - Turtle
    - HTML
    - txt
    search engine indexing: NULL
    exchange/harvesting methods:
    - SPARQL endpoint
    local search engine URL:  https://meta.icos-cp.eu/sparqlclient/
    external search engine types supported:
    - SPARQL endpoint
    access policy statements included: yes
    metadata longevity plan URL: https://meta.icos-
cp.eu/objects/5stgMTshDNdDe_Y8obQOpe0u
    machine actionable: yes
```

IRI of machine readable metadata of dataset: https://hdl.handle.net/11676/-ffoiHjX5NDN0Vq_fKuVmas0
vocabularies:
- IRI: http://purl.org/dc/elements/1.1/
  name: Darwin Core
  type: vocabulary
  topic: general
  specification language: RDF
- IRI: http://www.w3.org/ns/prov
  name: PROV-O
  type: ontology
  topic: general
  specification language: OWL
- IRI: http://meta.icos-cp.eu/ontologies/cpmeta/
  name: ICOS
  type: ontology
  topic: domain specific
  specification language: OWL
data management plans:
  specific DMP tools used: DMPonline
  data publishing steps applied:
  - QA QC
  compliance validation service: no
data processing:
  special data processing steps applied:
  - checksum
  - Handle PID suffix
  - QA QC
  workflow frameworks applied:
  - custom workflows
  distributed workflows tools used:
  - community standardised scripts
  other analysis services offered:
  - Jupiter Notebooks
  data products offered:
  - near real time observational data
  - final QC observational data
  - elaborated products
fairness:
  data findability:
    data findable: yes
    gaps: none
  data accessibility:
    data accessible: yes
    gaps: none
  data interoperability:
    data interoperable: yes
    gaps:
    - harmonising vocabularies
    - harmonising data models
  data re-usability:
    data reusable: yes
    gaps:
- provenance information

SeaDataNet

https://github.com/envri-fair/fairness-assessment/blob/master/descriptions/SEADATANET-20190523.yaml

```
survey:
  date: 2019-05-23
  version: 1
  creator:
    name: Dick Schaap
    email: dick@maris.nl
infrastructure:
  acronym: SEADATANET
  name: SeaDataNet
  website: http://www.seadatanet.org
  domain:
  - marine
  URL/IRI of dataset: https://doi.org/10.12770/90ae7a06-8b08-4afe-83dd-ca92bc99f5c0
  URL of discovery portal:  https://www.seadatanet.org/Products
  repositories:
  - URL: https://www.seadatanet.org/Products
    name: SeaDataNet Central Data Products
    kind:
    - data repository
    allocation: central
    software:
    - Sextant
    identifier:
    - kind: PID
      system: DOI
      landing page: yes
      assigned: manually
      provider: DataCite
      includes metadata schema: NULL
    certification methods: none
    policies:
    - data usage
    registries: none
    persistency-guaranty: stewardship
    access mechanisms:
      authentication method: Marine-ID
      access protocol URL: http://vocab.nerc.ac.uk/collection/L07/current/
      access without costs: yes
      own user database maintained: yes
      person identification system: none
      major access technology supported: CAS
      authorisation technique: OAUTH
      authorization for accessing content needed: no
      data licenses in use:
      - local license
      data license IRI:  https://www.seadatanet.org/Data-Access/License/1.0
      metadata openly available: yes
    data:
    - type name: binary
      preferred formats:
      - format name: NetCDF CF
        metadata types in data headers:
```

```
      - NetCDF CF
    registered data schema: none
    search on data: planned
  - type name: text
    preferred formats:
    - format name: ODV ASCII
      metadata types in data headers:
      - CDI metadata reference
    - format name: Medatlas ASCII
      metadata types in data headers:
      - CDI metadata reference
    registered data schema: none
    search on data: planned
  metadata:
    schema:
    - URL: NULL
      name: ISO 19115/19139
      provenance fields included:
      - text only
    - URL: NULL
      name: SEADATANET community profile
      provenance fields included: planned
    machine readable provenance: no
    categories defined in registries: yes
    PIDs included: yes
    primary storage format: XML
    export formats supported:
    - XML
    - HTML
    search engine indexing: NULL
    exchange/harvesting methods:
    - OGC CSW
    - OAI-PMH
    local search engine URL: https://cdi.seadatanet.org/search
    external search engine types supported:
    - Geo portal
    - IODE Ocean Data Portal
    access policy statements included: yes
    metadata longevity plan URL: VOID
    machine actionable: yes
    IRI of machine readable metadata of dataset:
http://seadatanet.maris2.nl/v_cdi_v3/print_xml.asp?n_code=2626292
  vocabularies:
  - IRI:
http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codeli
st/ML_gmxCodelists.xml
    name: ISO 19115/19139
    type: codelist
    topic: INSPIRE
    specification language: https://www.w3.org/XML/
  - IRI: http://vocab.nerc.ac.uk/collection/V22/current/
    name: NERC vocabulary service
    type: thesaurus
    topic: domain specific
    specification language: https://www.w3.org/XML/
  - IRI: https://edmo.seadatanet.org
    name: EDMO
```

```
    type: thesaurus
    topic: project specific
    specification language: https://www.w3.org/XML/
  - IRI: https://cdi.seadatanet.org
    name: SeaDataNet CDI
    type: thesaurus
    topic: project specific
    specification language: planned
  data management plans:
    specific DMP tools used: SEADATANET framework
    data publishing steps applied:
    - metadata import
    - metadata validation
    - metadata publication
    - DOI generation
    compliance validation service: no
  data processing:
    special data processing steps applied: planned
    workflow frameworks applied: planned
    distributed workflows tools used: none
    other analysis services offered: planned
    data products offered:
    - interpolations
  fairness:
    data findability:
      data findable: yes
      gaps: none
    data accessibility:
      data accessible: partially
      gaps:
      - improved WPS service
    data interoperability:
      data interoperable: partially
      gaps:
      - harmonising vocabularies
    data re-usability:
      data reusable: partially
      gaps:
      - provenance information
      - quality information
- URL: https://cdi.seadatanet.org/search
  name: SeaDataNet Common DAta Index (CDI)
  kind:
  - data repository
  allocation: central
  software:
  - B2SAFE
  identifier:
  - kind: PID
    system: B2HANDLE
    landing page: no
    assigned: automatically
    provider: EUDAT
    includes metadata schema: NULL
  certification methods: none
  policies:
  - data usage
```

registries:
- GEOSS
- IODE Ocean Data Portal
persistency-guaranty: stewardship
access mechanisms:
authentication method: Marine-ID
access protocol URL: http://vocab.nerc.ac.uk/collection/L07/current/
access without costs: yes
own user database maintained: yes
person identification system: none
major access technology supported: CAS
authorisation technique: OAUTH
authorization for accessing content needed: no
data licenses in use:
- local license
data license IRI: https://www.seadatanet.org/Data-Access/License/1.0
metadata openly available: yes
data:
- type name: binary
preferred formats:
- format name: NetCDF CF
metadata types in data headers:
- NetCDF CF
registered data schema: none
search on data: planned
metadata:
schema:
- URL: NULL
name: ISO 19115/19139
provenance fields included:
- text only
- URL: NULL
name: SEADATANET community profile
provenance fields included: planned
machine readable provenance: no
categories defined in registries: yes
PIDs included: yes
primary storage format: XML
export formats supported:
- XML
- HTML
search engine indexing: NULL
exchange/harvesting methods:
- OGC CSW
local search engine URL: https://www.seadatanet.org/Products#/search
external search engine types supported: none
access policy statements included: yes
metadata longevity plan URL: VOID
machine actionable: partially
IRI of machine readable metadata of dataset: https://doi.org/10.17882/42182
vocabularies:
- IRI:
http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml
name: ISO 19115/19139
type: codelist
topic: INSPIRE

```
    specification language: https://www.w3.org/XML/
  - IRI: http://vocab.nerc.ac.uk/collection/V22/current/
    name: NERC vocabulary service
    type: thesaurus
    topic: domain specific
    specification language: https://www.w3.org/XML/
  - IRI: https://edmo.seadatanet.org
    name: EDMO
    type: thesaurus
    topic: project specific
    specification language: https://www.w3.org/XML/
  - IRI: https://cdi.seadatanet.org
    name: SeaDataNet CDI
    type: thesaurus
    topic: project specific
    specification language: planned
  data management plans:
    specific DMP tools used: SEADATANET framework
    data publishing steps applied:
    - metadata import
    - metadata validation
    - metadata publication
    - DOI generation
    compliance validation service: no
  data processing:
    special data processing steps applied:
    - QA QC
    - ODV aggregation
    workflow frameworks applied: planned
    distributed workflows tools used:
    - Octopus
    other analysis services offered: planned
    data products offered:
    - interpolations
    - aggregations
  fairness:
    data findability:
      data findable: yes
      gaps: none
    data accessibility:
      data accessible: partially
      gaps:
      - improved WPS service
    data interoperability:
      data interoperable: partially
      gaps:
      - harmonising vocabularies
    data re-usability:
      data reusable: partially
      gaps:
      - provenance information
      - quality information
```


Euro-Argo

https://github.com/envri-fair/fairness-assessment/blob/master/descriptions/EuroArgo-20190418.yaml

```
survey:
  date: 2019-04-18
  version: 1
  creator:
    name: Thierry Carval
    email: thierry.carval@ifremer.fr
infrastructure:
  acronym: Euro-Argo
  name: Euro-Argo ERIC
  website: http://www.euro-argo.eu
  domain:
  - marine
  URL/IRI of dataset: http://doi.org/10.17882/42182
  URL of discovery portal: http://www.argodatamgt.org/Access-to-data/Argo-data-selection
  repositories:
  - URL: http://doi.org/10.17882/42182
    name: Euro-Argo Data
    kind:
    - data repository
    allocation: central
    software:
    - NetAPP
    identifier:
    - kind: PID
      system: DOI
      landing page: yes
      assigned: automatically
      provider: SEANOE
      includes metadata schema: NULL
    certification methods:
    - Data Seal of Approval
    - IODE certification
    policies:
    - data access
    registries:
    - GEOSS
    persistency-guaranty: 50 years
    access mechanisms:
      authentication method: VOID
      access protocol URL: https://doi.org/10.17882/42182
      access without costs: yes
      own user database maintained: no
      person identification system: none
      major access technology supported: FTP
      authorisation technique: VOID
      authorization for accessing content needed: no
      data licenses in use:
      - CC-BY4.0
      data license IRI:  https://creativecommons.org/licenses/by/4.0/
      metadata openly available: yes
    data:
    - type name: binary
      preferred formats:
```

```
    - format name: NetCDF
      metadata types in data headers:
      - NetCDF ACDD metadata
    registered data schema: NULL
    search on data: yes
  metadata:
    schema:
    - URL: http://dx.doi.org/10.13155/29825
      name: Argo user manual
      provenance fields included:
      - text only
    - URL: NULL
      name: NetCDF CF checker
      provenance fields included:
      - codes
    machine readable provenance:
    categories defined in registries: planned
    PIDs included: yes
    primary storage format: NetCDF CF Argo
    export formats supported:
    - NetCDF
    - CSV
    search engine indexing: NULL
    exchange/harvesting methods:
    - OAI-PMH
    - OpenDAP
    - OGC CSW
    local search engine URL: NULL
    external search engine types supported:
    - CSW endpoint
    - Google dataset search
    access policy statements included: yes
    metadata longevity plan URL: https://www.nodc.noaa.gov/worlddatacenter/
    machine actionable: yes
    IRI of machine readable metadata of dataset: https://doi.org/10.17882/42182
  vocabularies:
  - IRI: http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P06
    name: NVS P06
    type: codelist
    topic: project specific
    specification language: http://cfconventions.org/Data/cf-standard-names/64/src/cf-
standard-name-table.xml
  - IRI: http://vocab.nerc.ac.uk/collection/P01/current/PSALST01/
    name: NVS P01
    type: codelist
    topic: domain specific
    specification language: http://cfconventions.org/Data/cf-standard-names/64/src/cf-
standard-name-table.xml
  data management plans:
    specific DMP tools used: quality management framework
    data publishing steps applied:
    - subsetting monthly snapshots
    - DOI generation
    compliance validation service: yes
  data processing:
    special data processing steps applied:
    - QA QC
```

```
    workflow frameworks applied:
    - data processing steps recorded
    distributed workflows tools used: none
    other analysis services offered:
    - Spark
    - Cassandra
    - Elasticsearch
    - Pangeo
    data products offered:
    - gridded fields
    - temperature
    - salinity
    - climatologies
  fairness:
    data findability:
      data findable: yes
      gaps: none
    data accessibility:
      data accessible: yes
      gaps:
      - impovements in subsetting access
      - WMS service
    data interoperability:
      data interoperable: partially
      gaps:
      - harmonising vocabularies
    data re-usability:
      data reusable: partially
      gaps:
      - harmonising vocabularies
```


LifeWatch (Marine)

https://github.com/envri-fair/fairness-assessment/blob/master/descriptions/LWmarine-20190430.yaml


```
survey:
  date: 2019-04-30
  version: 1
  creator:
    name: Klaas Deneudt
    email: klaas.deneudt@vliz.be
infrastructure:
  acronym: LW (Marine)
  name: LifeWatch (Marine)
  website: http://www.euro-argo.eu
  domain:
  - marine
  URL/IRI of dataset: http://www.vliz.be/en/imis?module=dataset&dasid=1606
  URL of discovery portal: http://www.marinedataarchive.org/
  repositories:
  - URL: http://www.marinedataarchive.org/
    name: Marine Data Archive
    kind:
    - data repository
    allocation: central
```

software:
- iRODS
identifier:
- kind: local ID
  system: none
  landing page: no
  assigned: automatically
  provider: local service
  includes metadata schema: none
- kind: PID
  system: DOI
  landing page: yes
  assigned: manually
  provider: DataCite
  includes metadata schema: NULL
certification methods:
- IODE certification
- CoreTrustSeal
policies:
- Terms of Use
- FAQ
registries: none
persistency-guaranty: stewardship
access mechanisms:
  authentication method: account manually
  access protocol URL: https://doi.org/10.17487%2FRFC2616
  access without costs: yes
  own user database maintained: yes
  person identification system: none
  major access technology supported: HTTP POST
  authorisation technique: SSQL service
  authorization for accessing content needed: no
  data licenses in use:
  - CC BY.CC 0
  data license IRI: NULL
  metadata openly available: yes
data:
- type name: text
  preferred formats:
  - format name: ASCII
    metadata types in data headers: none
  registered data schema: none
  search on data: yes
metadata:
  schema:
  - URL: NULL
    name: ISO 19115
    provenance fields included:
    - text only
  - URL: http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language
    name: EML
    provenance fields included:
    - text only
  categories defined in registries: yes
  machine readable provenance: no
  PIDs included: yes

```
      primary storage format: relational database
      export formats supported:
      - JSON
      - XML
      - RSS
      - HTML
      search engine indexing: NULL
      exchange/harvesting methods: planned
      local search engine URL: http://www.vliz.be/en/integrated-marine-information-system
      external search engine types supported: planned
      access policy statements included: partially
      metadata longevity plan URL: VOID
      machine actionable: partially
      IRI of machine readable metadata of dataset:
http://www.lifewatch.be/en/imis?module=dataset&dasid=1841&show=eml
    vocabularies:
    - IRI: http://www.marinespecies.org/
      name: WoRMS
      type: taxonomy
      topic: domain specific
      specification language: NULL
    - IRI: http://www.marineregions.org/
      name: Marine Regions
      type: gazetteer
      topic: domain specific
      specification language: NULL
    - IRI: http://www.marinespecies.org/
      name: Marine species
      type: gazetteer
      topic: domain specific
      specification language: NULL
    data management plans:
      specific DMP tools used: none
      data publishing steps applied: none
      compliance validation service: no
    data processing:
      special data processing steps applied: none
      workflow frameworks applied: none
      distributed workflows tools used: none
      other analysis services offered: none
      data products offered: none
    fairness:
      data findability:
        data findable: yes
        gaps: none
      data accessibility:
        data accessible: partially
        gaps:
        - records cleaning
      data interoperability:
        data interoperable: partially
        gaps:
        - format checking
      data re-usability:
        data reusable: partially
        gaps:
        - provenance information
```

```
    - records cleaning
- URL: http://www.eurobis.org/
  name: EUROBIS
  kind:
  - data repository
  allocation: central
  software:
  - relational database
  identifier:
  - kind: local ID
    system: none
    landing page: no
    assigned: automatically
    provider: local service
    includes metadata schema: none
  - kind: PID
    system: DOI
    landing page: yes
    assigned: manually
    provider: DataCite
    includes metadata schema: NULL
  certification methods:
  - IODE certification
  - CoreTrustSeal
  policies:
  - Terms of Use
  - FAQ
  registries:
  - fairsharing.org
  persistency-guaranty: stewardship
  access mechanisms:
    authentication method: account manually
    access protocol URL: https://doi.org/10.17487%2FRFC2616
    access without costs: yes
    own user database maintained: yes
    person identification system: none
    major access technology supported: HTTP POST
    authorisation technique: SSQL service
    authorization for accessing content needed: no
    data licenses in use:
    - CC BY.CC 0
    data license IRI: NULL
    metadata openly available: yes
  data:
  - type name: text
    preferred formats:
    - format name: ASCII
      metadata types in data headers: none
    - format name: DwC
      metadata types in data headers:
      - DwC
    registered data schema: Darwin Core Archive
    search on data: yes
  - type name: sound files
    preferred formats:
    - format name: NULL
      metadata types in data headers: none
```

registered data schema: none
search on data: yes
- type name: occurance data
preferred formats:
- format name: OBIS
metadata types in data headers:
- OBIS event scheme
registered data schema: OBIS event scheme
search on data: yes
metadata:
schema:
- URL: NULL
name: ISO 19115
provenance fields included:
- text only
- URL: http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language
name: EML
provenance fields included:
- text only
categories defined in registries: yes
machine readable provenance: no
PIDs included: yes
primary storage format: relational database
export formats supported:
- JSON
- XML
- RSS
- HTML
search engine indexing: NULL
exchange/harvesting methods: planned
local search engine URL: http://www.vliz.be/en/integrated-marine-information-system
external search engine types supported:
- EDMED
- GBIF
- GCMD
- DataCite
access policy statements included: partially
metadata longevity plan URL: VOID
machine actionable: partially
IRI of machine readable metadata of dataset: http://www.lifewatch.be/en/imis?module=dataset&dasid=1841&show=eml
vocabularies:
- IRI: http://www.marinespecies.org/
name: WoRMS
type: taxonomy
topic: domain specific
specification language: NULL
- IRI: http://www.marineregions.org/
name: Marine Regions
type: gazetteer
topic: domain specific
specification language: NULL
- IRI: http://www.marinespecies.org/
name: Marine species
type: gazetteer
topic: domain specific

specification language: NULL
data management plans:
specific DMP tools used:
- DOI generation
- DwC formatting
- WoRMS taxon matching
data publishing steps applied: none
compliance validation service: no
data processing:
special data processing steps applied:
- QA QC
workflow frameworks applied:
- WoRMS taxon matching
- geolocation
distributed workflows tools used:
- LifeWatch virtual lab
- map viewer
other analysis services offered:
- R packages
data products offered:
- geospatial datasets
- integrated standardised datasets
fairness:
data findability:
data findable: yes
gaps: none
data accessibility:
data accessible: partially
gaps: none
data interoperability:
data interoperable: yes
gaps: none
data re-usability:
data reusable: yes
gaps: none

EMSO

https://github.com/envri-fair/fairness-assessment/blob/master/descriptions/EMSO-20190503.yaml

survey:
date: 2019-05-03
version: 1
creator:
name: Ivan Rodero
email: ivan.rodero@emso-eu.org
infrastructure:
acronym: EMSO
name: EMSO ERIC
website: http://www.emso.eu
domain:
- marine
URL/IRI of dataset: http://data.emso.eu/files/emso/azores/exif0001/EXIF0001_1944-2019.nc
URL of discovery portal: http://data.emso.eu

repositories:
- URL: http://data.emso.eu/files/
  name: EMSO data
  kind:
  - data repository
  allocation: central
  software:
  - Apache HTTP
  identifier:
  - kind: planned
    system: planned
    landing page: no
    assigned: manually
    provider: planned
    includes metadata schema: NULL
  certification methods: none
  policies: none
  registries: none
  persistency-guaranty: none
  access mechanisms:
    authentication method: VOID
    access protocol URL: VOID
    access without costs: yes
    own user database maintained: yes
    person identification system: none
    major access technology supported: HTTP
    authorisation technique: VOID
    authorization for accessing content needed: no
    data licenses in use:
    - CC-BY4.0
    data license IRI: NULL
    metadata openly available: yes
  data:
  - type name: binary
    preferred formats:
    - format name: NetCDF CF
      metadata types in data headers:
      - NetCDF ACDD metadata
    registered data schema: planned
    search on data: planned
  - type name: text
    preferred formats:
    - format name: ODV ASCII
      metadata types in data headers:
      - NetCDF metadata
    registered data schema: none
    search on data: planned
  metadata:
    schema:
    - URL: NULL
      name: ISO 19115/19139
      provenance fields included:
      - text only
    - URL: NULL
      name: SEADATANET community profile
      provenance fields included: planned
    machine readable provenance: no

```
      categories defined in registries: yes
      PIDs included: yes
      primary storage format: XML
      export formats supported:
      - XML
      - HTML
      search engine indexing: NULL
      exchange/harvesting methods:
      - OGC CSW
      - OAI-PMH
      local search engine URL: https://cdi.seadatanet.org/search
      external search engine types supported:
      - Geo portal
      - IODE Ocean Data Portal
      access policy statements included: yes
      metadata longevity plan URL: VOID
      machine actionable: yes
      IRI of machine readable metadata of dataset:
http://seadatanet.maris2.nl/v_cdi_v3/print_xml.asp?n_code=2626292
    vocabularies:
    - IRI: http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery_1-3
      name: ACDD
      type: thesaurus
      topic: domain specific
      specification language: https://www.w3.org/RDF/
    data management plans:
      specific DMP tools used: planned
      data publishing steps applied:
      - QA QC
      compliance validation service: no
    data processing:
      special data processing steps applied: none
      workflow frameworks applied: none
      distributed workflows tools used: none
      other analysis services offered: planned
      data products offered: planned
    fairness:
      data findability:
        data findable: yes
        gaps: none
      data accessibility:
        data accessible: partially
        gaps:
        - improved WPS service
      data interoperability:
        data interoperable: partially
        gaps:
        - harmonization
      data re-usability:
        data reusable: partially
        gaps:
        - quality information
        - provenance information
```