Grant agreement no: 654310

Ocean Data Interoperability Platform

Deliverable D3.1: Definition of ODIP Prototypes 1

Work package	WP3	ODIP Prototypes
Author (s)	Dick M.A. Schaap	MARIS
Author (s)		
Author (s)		
Author (s)		
Authorized by	Helen Glaves	NERC
Reviewer		
Doc Id	ODIP II_D3.1	
Dissemination Level	PUBLIC	
Issue	2.0	
Date	30 October 2016	

Abstract:

The Ocean Data Interoperability Platform project is developing a series of prototype solutions that establish interoperability between existing regional and relevant global marine data infrastructures. These interoperability solutions aim to demonstrate effective sharing of data between these systems and support development of a common global framework for marine data management.

Initial formulation and development of these interoperability solutions formed part of the activities in the previous ODIP project. This document describes the continued development and expansion of the three ODIP prototype interoperability solutions as part of the current ODIP II project.

Document History				
Version	**Author(s)**	**Status**	**Date**	**Comments**
1.0	Dick M.A. Schaap (MARIS)	DRAFT	20 September 2016	First draft
1.1	Dick M.A. Schaap (MARIS)	DRAFT	3 October 2016	Table of Terminology added
1.2	Helen Glaves (NERC)	DRAFT	21 October 2016	Editing/review
2.0	Helen Glaves (NERC)	FINAL	30 October 2016	Final review and sign-off

Contents

Executive Summary

A key objective of the ODIP II project is the further development of the three existing prototype interoperability solutions created during the previous ODIP project, and the formulation of further additional prototype interoperability solutions. The purpose of these prototype solutions is to evaluate and test selected potential common standards and to establish and/or improve interoperability between the participating regional marine data infrastructures and also with selected global infrastructures.

The final outcomes of the three ODIP prototype development tasks were presented and evaluated during the 1st ODIP II Workshop which took place during September 2015. As a result of this evaluation it was decided to continue and expand these prototype development tasks to incorporate additional aspects as part of the ODIP II project. In the seven-month period between the 1st and 2nd ODIP II workshops further analysis of the options for expanding the prototype development tasks was carried out. The resulting specifications for the extensions to the three prototype interoperability solutions were presented, discussed and agreed during the 2nd ODIP II workshop which took place in Boulder, Colorado, USA on 2 to 6 May 2016.

Deliverable *D3.1 Definition of ODIP Prototypes* 1 gives a concise description of the status of the three existing prototype interoperability solutions and also outlines the specifications and associated activities for their continued development and expansion which will be implemented in the ODIP II project.

1 Introduction

The ODIP II: Extending the Ocean Data Interoperability Platform project is the follow-on from the previous ODIP project (Establishing an Ocean Data Interoperability Platform) which ran from 1st October 2012 to 30th September 2015. ODIP II is promoting the development of a common global framework for marine data management which is being supported by establishing interoperability between existing regional marine e-infrastructures in Europe, USA and Australia and also with global infrastructures such as GEOSS, IOC-IODE and POGO.

To establish interoperability between the different data systems four workshops are being organised during the three years of the project that bring together relevant domain experts and other technical specialists to provide insights into commonalities and differences between the individual marine data systems. They will also discuss and compare their respective approaches, best practices and standards, and also identify opportunities for the development of common standards and interoperability solutions.

During the previous ODIP project these workshops resulted in the formulation and successful delivery of a suite of prototypes to evaluate and test selected potential common standards and solutions for establishing and demonstrating improved interoperability between the regional and global data infrastructures. The three prototype development tasks successfully undertaken during the initial ODIP project were:

- **ODIP 1**: Establishing interoperability between the SeaDataNet CDI, US NODC, and IMOS MCP data discovery and access services using a brokerage service, and moving towards interacting with the global IODE-ODP and GEOSS portals, led by SeaDataNet (Europe);
- **ODIP 2**: Establishing a common deployment and interoperability between Cruise Summary Reporting systems in Europe, the USA and Australia by making use of GeoNetWork, and moving towards interaction with the global POGO portal, led by R2R (USA);
- **ODIP 3**: Establishing a prototype for a Sensor Observation Service (SOS) and formulating common O&M and SensorML profiles for selected sensors (SWE), installed at research vessels and in real-time monitoring systems, led by AODN (Australia).

During the 1st ODIP II workshop which took place in Paris, France, from 28th September to 1st October 2015, the final results of the three ODIP prototype development tasks were presented and discussed. As a direct result of this evaluation, it was decided to continue and to expand these prototype development tasks to include additional features as part of the ODIP II project. In the seven-month period between the 1st and the 2nd ODIP II workshops further analysis of the extended prototype solutions took place and the resulting specifications for the three ODIP II prototype development tasks were presented, evaluated and approved during the 2nd ODIP II workshop which took place in Boulder, USA, from 2nd May to 5th May 2016.

This deliverable *D3.1 Definition of ODIP Prototypes 1* begins with a concise description of the status of the three ODIP prototype projects resulting from the previous ODIP project. This is followed by the specifications and associated actions for expanding and further developing these prototype interoperability solutions as part of the ODIP II project.

Within the framework of the previous ODIP a number of cross-cutting topics were also addressed which included vocabularies, data publishing and citation, and persistent unique identifiers.

The term ‘cross-cutting’ has been used because these topics are relevant to all of the prototype development tasks and also to marine data management in general. These topics address various aspects of harmonising metadata and data descriptions, achieving semantic interoperability between different regional systems using mappings and ontologies, and encouraging researchers to release and publish their data sets in scientific literature and gaining academic credit through citation of their data by others.

At the 1st ODIP II Workshop it was decided to continue also activities on these topics, partly by monitoring and discussing at the Workshops on-going developments in the regions and relevant developments at global level, and partly by developing solutions contributing to the three ODIP II prototype projects.

In addition to the three ODIP II prototype projects new ODIP II prototype projects have been formulated at the 2nd ODIP II Workshop. The associated specifications and action plan will be worked out and documented in ODIP II Deliverable D3.2.

A significant challenge for the ODIP II project is that unlike the previous project only the European partners have successfully obtained further funding for participation in ODIP II. Unfortunately partners in the USA and Australia are now reliant on their very limited organizational funding for participation in the ODIP II project. This will potentially significantly curtail these partners’ ability to contribute to the ODIP II activities. This situation further emphasizes the need for an approach where the ODIP II prototype development activities are largely leveraging the activities of current regional projects and initiatives being undertaken by the ODIP II partners. The ODIP II prototype development tasks must therefore be formulated taking these constraints into account.

2 Existing prototypes

The ODIP II project is the successor to the previous ODIP project ("Establishing an Ocean Data Interoperability Platform") which ran from 1st October 2012 till 30th September 2015. The development activities for the previous ODIP project included creating three prototype solutions to establish interoperability between the existing regional and global data systems.

Below are concise descriptions of the three prototype development tasks that were successfully delivered during the previous ODIP project.

2.1 ODIP 1 prototype development task

The objective of the ODIP 1 prototype development task has been to establish interoperability between the SeaDataNet CDI, US NODC and IMOS MCP data discovery and access services, and to expose metadata records from the regional marine data systems in the global IODE-ODP and GEOSS portals. The GEO-DAB brokering service has been used to establish interoperability between these regional data systems and also with global data systems at the metadata level use. The GEO-DAB broker technologies are largely based upon the GI-Cat software that is maintained by partner CNR (http://essi-lab.eu/do/view/GIcat). This task was led by MARIS on behalf of SeaDataNet (Europe).

Europe, USA and Australia each operate and manage their own data discovery and access services, that include use of local metadata formats, vocabularies, access control, data formats and data policies.

The three participating regional data infrastructures are:

- SeaDataNet CDI service in Europe
- US NODC (now NCEI) service in USA
- AODN MCP service in Australia

The GEO-DAB Brokerage Service harvests XML entries from the regional data systems and converts them to the Generic Brokerage Reference Schema, including the adoption of the existing vocabularies. Using the two global portals as a guide and also taking into account the potential performance issues, vertical interoperability of the three regional infrastructures with the global portals (IODE-ODP and GEOSS) has been established at collections level using the following approaches:

SeaDataNet: delivers the required XML via a REST / CS-W service at the collections level, using the CDI ISO 19139 schema and supporting vocabularies. Aggregations of metadata records for individual datasets into 'collections' has been created using criteria such as: data centre (EDMO-code), discipline (P08 vocabulary terms) and geometry type (points / tracks / polygons). This has resulted in approximately 480 CDI collections, each with a URL to the CDI service for details of the individual datasets within an individual collection.

The SeaDataNet XML output from the REST web service is taken up by the GEO-DAB brokerage and converted to the Brokerage Common Reference Model, thereby maintaining the use of specific SeaDataNet vocabulary terms alongside the ISO standard terms. The GEO-

DAB brokerage service provides the SeaDataNet collections as common XML via an OAI-PMH service:

http://seadatanet.essi-lab.eu/gi-cat/services/oaipmh?verb=ListIdentifiers&metadataPrefix=ISO19139&set=SEADATANET

and by CS-W version 2.0.2 Service – HTTP POST method:

http://seadatanet.essi-lab.eu/gi-cat/services/cswiso

using SEADATANET as parameter for retrieving the SeaDataNet output.

The GEO-DAB brokerage output of SeaDataNet collections has been harvested and imported into the GEOSS portal using the CS-W service, while the harvesting and loading into the IODE-ODP portal takes place using the OAI-PMH service and the jOAI harvester. Triggers have been configured for fully automatic and the most efficient synchronisation between the different services in the chain.

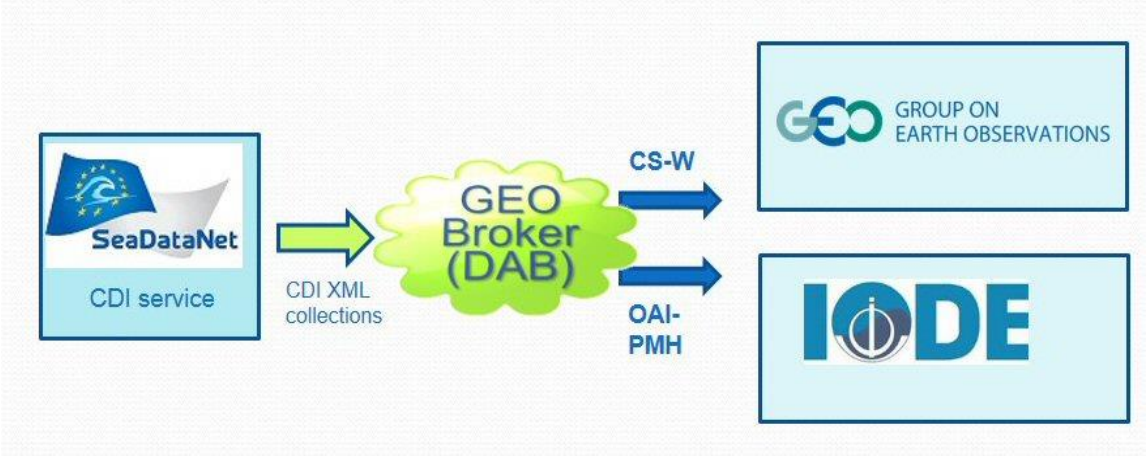

Figure 1 Establishing metadata interoperability between SeaDataNet and the GEOSS and IODE-ODP portal

US-NODC (now NCEI - National Centers for Environmental Information): following a similar to that approach used for SeaDataNet a comparable analysis and procedure has been applied by MARIS and CNR for the US NODC portal service. US NODC provides CSW and OAI-PMH services at collection level. The resulting XML output contains a latitude – longitude box which can be used to display the spatial coverage of the collection at the receiving portal and also a URL to the web page for the collection with the title: '*Navigate directly to the URL for a descriptive web page with download links'.* The GEO-DAB Brokerage Service has harvested the US NODC CSW for collections (around 28,000 entries) and converted the XML to the Generic Brokerage Reference Schema, adopting the US NODC vocabularies, for supplying harmonised CSW and OAI-PMH to the GEOSS and ODP portals via an OAI-PMH interface, at:

http://seadatanet.essi-lab.eu/gi-cat/services/oaipmh?verb=ListIdentifiers&metadataPrefix=ISO19139&set=NODC

OGC Catalogue Service for the Web (CSW) Version 2.0.2 Service – HTTP POST method:

http://seadatanet.essi-lab.eu/gi-cat/services/cswiso

with 'apiso:parentIdentifier' = NODC

An individual US NODC collection record after brokerage can be seen at:

http://seadatanet.essi-lab.eu/gi-cat/services/cswisogeo?service=CSW&request=GetRecordById&id=gov.noaa.nodc:0069000&outputschema=http://www.isotc211.org/2005/gmi&elementSetName=full

The US NODC collections have been harvested and loaded into the GEOSS and IODE-ODP portals in comparable way to that used for the SeaDataNet collections.

AODN (MCP) service: the Australian Ocean Data Network (AODN) makes use of GeoNetWork and provides web services at collections level as CSW and OAI-PMH services. The collections are defined somewhat differently from those delivered by SeaDataNet and US NODC, with the use of IMOS facilities (similar to data originator), geometric object, timeliness (real-time or delayed mode) as aggregators. The XML output contains a latitude – longitude box which can be used to display the spatial coverage of the collection at the receiving portal and also a URL to the web page for the collection with the title: '*Point of truth URL of this metadata record'.* AODN collections are available via public web services as provided by the GEO-DAB Brokerage service. However, in this instance the brokerage has to make use of the Dublin Core metadata profile in order to ensure that all links are included in the exchange. The OAI-PMH interface is available at:

http://seadatanet.essi-lab.eu/gi-cat/services/oaipmh?verb=ListIdentifiers&metadataPrefix=ISO19139&set=AODNCSWCORE

OGC Catalogue Service for the Web (CSW) Version 2.0.2 Service – HTTP POST method:

http://seadatanet.essi-lab.eu/gi-cat/services/cswiso

with 'apiso:parentIdentifier' = AODNCSWCORE

An individual collection record of AODN after brokerage can be seen at:

http://seadatanet.essi-lab.eu/gi-cat/services/cswisogeo?service=CSW&request=GetRecordById&id=63db5801-cc19-40ef-83b3-85ccba884cf7&outputschema=http://www.isotc211.org/2005/gmi&elementSetName=full

The AODN collections have also been harvested and loaded into the GEOSS IODE-ODP portal in a comparable way to that used for the SeaDataNet and US-NODC (now NCEI) collections.

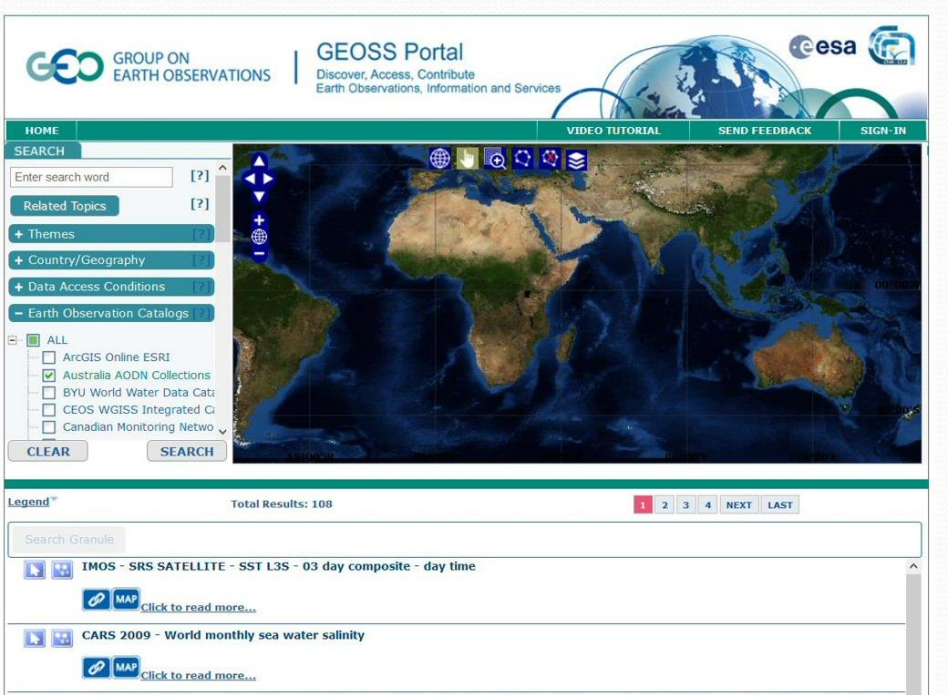

Figure 2 SeaDataNet, AODN and US NODC data 'collections' exposed in the GEOSS portal (http://www.geoportal.org)

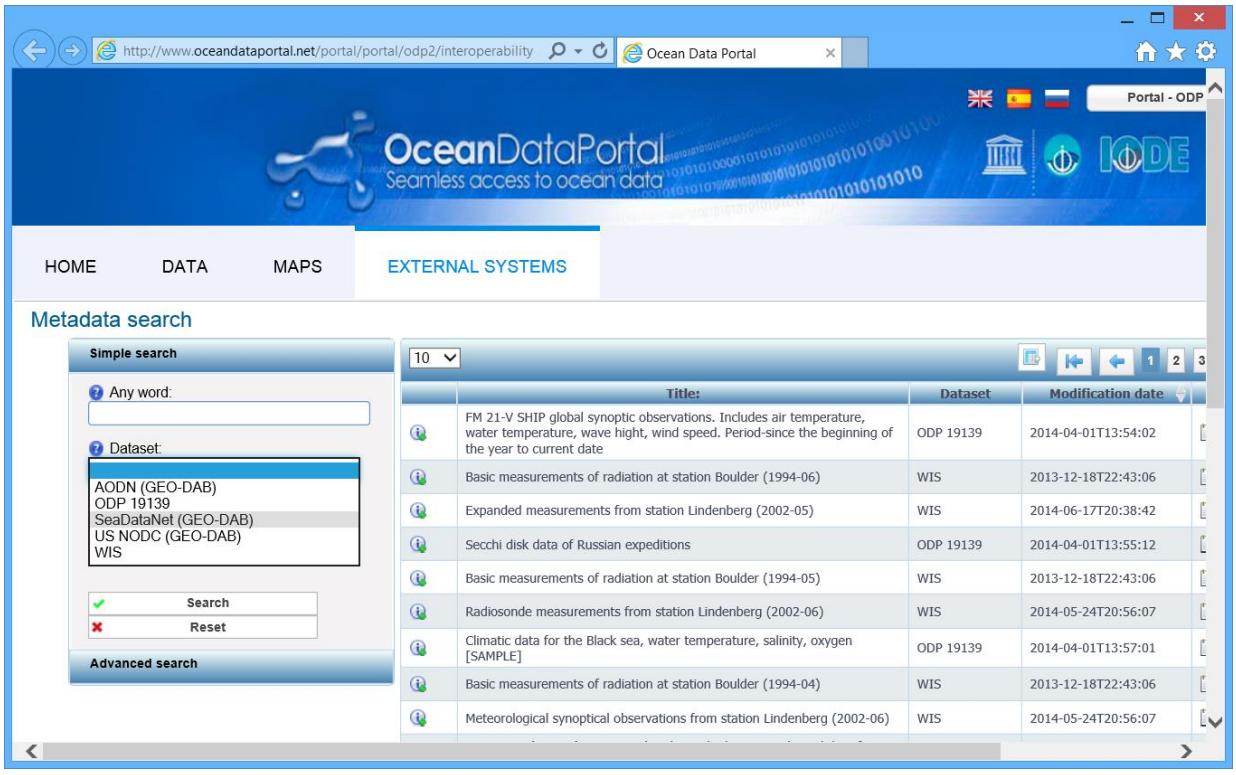

Figure 3 SeaDataNet, AODN and US NODC data collections exposed in the IODE-ODP (*http://www.oceandataportal.net/portal/portal/odp2/interoperability*)

Project partners responsible for the ODIP 1 prototype development task were: MARIS on behalf of European SeaDataNet consortium, US NODC (now NCEI) on behalf of NOAA, UTAS on behalf of AODN consortium, CNR and IEEE on behalf of GEO-DAB brokerage service development team, and RIHMI-WDC and DFO-BIO on behalf of IODE ODP development team.

2.2 ODIP 2 prototype development task

The ODIP 2 prototype development task is focused on a developing a common deployment and establishing interoperability between Cruise Summary Reporting (CSR) systems in Europe, the USA and Australia and moving towards interaction with the global POGO portal through the use of GeoNetWork. This task has been led by LDEO on behalf of the Rolling Deck to Repository (R2R) program in the USA.

Efforts are being undertaken in Europe, USA and Australia for administering cruises and related data acquisition activities undertaken by their research vessels, as well as for developing new innovations for the related information systems.

The three regional infrastructures engaged in the administration of these CSR systems are are:

- SeaDataNet CSR metadata service in Europe
- R2R Cruise Information service in USA
- Marine National Facility (MNF) Cruise metadata in Australia

All three of these regions are also participating in the Partnership for Observation of the Global Oceans (POGO) which maintains a global catalogue of planned and past cruises that provides information on the coverage of past cruises and raises awareness of potential opportunities to improve the cost-effectiveness of future cruises.

For the development of the ODIP 2 prototype interoperability solution the focus has been on three regional CSR systems contributing to the POGO portal which already makes use of the European SeaDataNet CSR metadata service. The ODIP partners contributing the development of the ODIP 2 prototype interoperability solution have been: LDEO, SIO, FSU, and WHOI on behalf of R2R (USA); BSH, BODC, IFREMER, RBINS, CSIC, and MARIS on behalf of SeaDataNet, Eurofleets and the POGO consortia; CSIRO and UTAS on behalf of Australian MNF consortium.

SeaDataNet CSR service: the Cruise Summary Report (CSR) service provides a low level inventory of research cruises with reference to observations and also provides a tool for tracking.

Cruise Summary Reports (CSR) is a component of the SeaDataNet metadata infrastructure which, at present, contains almost 48,000 entries from more than 2000 research vessels dating back to 1873. It also includes historic European cruise summary reports from the ICES database from 1960 onwards. The cruise summary report format and schema were upgraded to ISO 19139 towards the end of 2013 and make use of the SeaDataNet vocabularies which

have been extended to include several additional attributes such as ports (C38) and platforms (C17).

EurOcean, (European Centre for Information on Marine Science and Technology, (http://www.eurocean.org/) maintains a global directory of research vessels, which are linked by ICES Ship codes (C17). The CSR schema is based upon the ISO 19115 and 19115-2 standards (for measurements and samples part) content models. The format and interface also supports linking to track charts (image / WMS) and station lists.

The operational Cruise Summary Reports WMS service displays CSRs that have been supplied with their navigation tracks in GML. Furthermore, CSR XML entries are also provided as a web service and used in the SeaDataNet CDI service to facilitate retrieval of data sets from specific cruises. SeaDataNet partner BSH from Germany manages the central SeaDataNet CSR directory and services which can be queried through the SeaDataNet portal.

The CSR service is also used for the POGO research vessel portal that is managed by SeaDataNet partners (BODC, MARIS and BSH). It focuses on ocean-going research vessels with a length greater than 60metres. The POGO portal provides the Global Research Vessels directory, the CSR directory and a directory of Cruise Programmes (CP).

Details of planned cruises are gathered by BODC through dialogue with global research vessel operators including those in the USA and Australia, and documented using a simple Cruise Programmes format, supported by an online content management system (CMS) or provided as CSV or XML. The Cruise Programme format is also supported by the SeaDataNet common vocabularies.

For the purposes of documenting the cruise summary reports, each data centre includes its own individual unique identifier, the Local_CSR_ID in the CSR XML. The central master CSR database, managed by partner BSH, then uses a combination of the Local_CSR_ID and the EDMO_code for the organisation that authored the cruise summary report to manage unique CSR entries in the central portal. A Central_CSR_ID is then also added to the CSRs for the public delivery services. These internal Local_CSR_IDs and external Central_CSR_IDs. IDs are also maintained and used for identifying updates to existing CSRs and the addition of new ones.

R2R Cruise Information service: the Rolling Deck to Repository (R2R) project provides uniform stewardship of routinely-collected environmental sensor data from the US academic research fleet. At present it includes more than 6000 cruise IDs from 38 research vessels and covers 30 device classes from 60 manufacturers. Furthermore, it has more than 3000 unique IDs for people from over 400 organisations and documents information from in excess of 200 portals and over 25 programmes.

R2R publishes cruise records in multiple formats: ISO 19115-2 Metadata records; OGC Web Services; W3C “Linked Open Data” with SPARQL endpoint; DOI-DataCite Metadata records; all of which use controlled vocabularies and best practices where available. For ISO 19115-2 it uses the vanilla schema, as provided by the NOAA/R2R Template. The NOAA/R2R records are published in a GeoNetwork (GN) 2.8 based web portal.

Marine National Facility (MNF) cruise metadata: Australian research vessel cruise metadata is stored in the CSIRO Marine and Atmospheric Research metadata system, MarLIN, along with metadata on datasets from the Marine National Facility (MNF) and other

sources. For each research vessel it provides the survey details, cruise plan, track and data which are loosely structured.

The majority of the development of this prototype interoperability solution has been undertaken by the EU and USA partners, while the Australian partners have been working on preparatory mappings. This is largely due to the fact that the current Australian system does not manage a comparable cruise summaries catalogue service, and the MNF has not been a partner in ODIP.

BSH and R2R each publish ISO Cruise Summary Reports using the SeaDataNet CSR 3.0 schema and associated vocabularies at their regional nodes. A GeoNetwork catalogue has also been deployed by each of the regional services, providing both a graphical user interface (GUI) in the form of a web portal and an API (CSW service). These GeoNetwork nodes are harvested into the POGO global catalogue for the purpose of providing an integrated search.

The GeoNetwork software has been adapted by IFREMER to support the SeaDataNet CSR schema and to facilitate harvesting of CSR XML entries, and it is freely available, including the supporting documentation. BSH has provided a mechanism for the harvested CSR entries, once validated, to be published in both the SeaDataNet CSR Catalogue service for all research vessels and the CSR Catalogue service at the POGO portal for larger ocean-going research vessels.

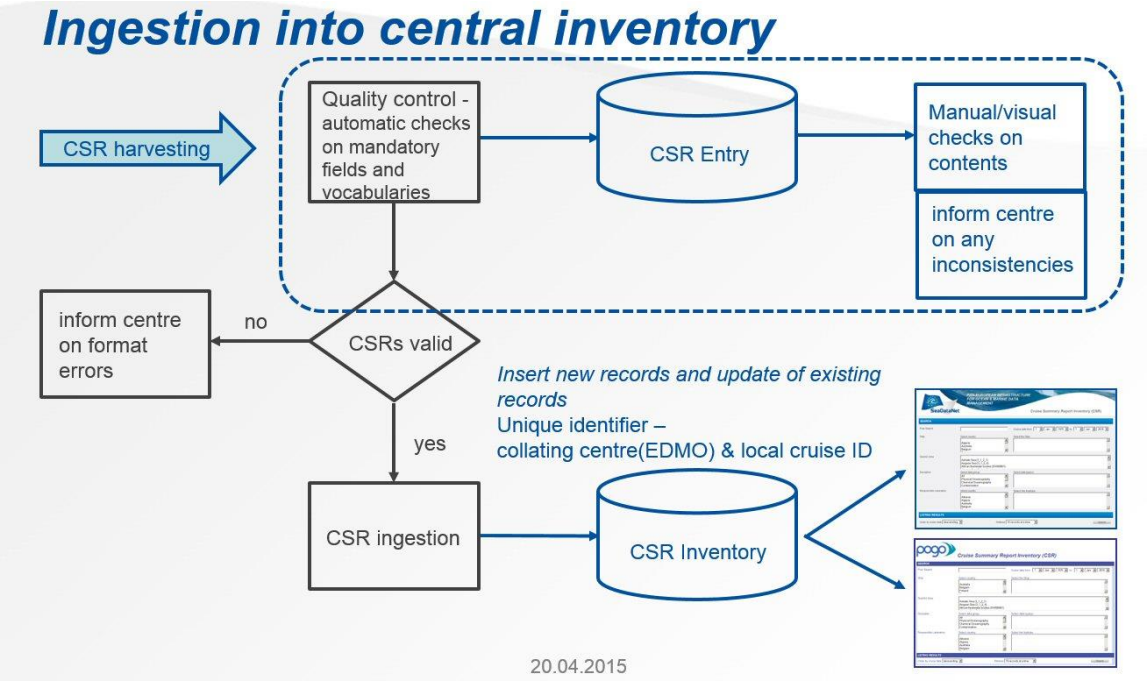

Figure 4 Cruise summary report (CSR) ingestion process at BSH

R2R has also adopted the GeoNetwork software solution and undertaken efforts to convert their USA cruise entries to the SeaDataNet CSR schema. US and Australian partners have invested considerable effort in mapping between their local vocabularies and those used SeaDataNet. They have also added additional missing terms to the SeaDataNet vocabularies with support of NERC-BODC that manages the SeaDataNet vocabularies. The European Directory of Marine Organizations (EDMO) has also been extended with several hundred

additional entries for institutes in the USA and Australia. To facilitate on-going maintenance of the entries for organizations in the USA and Australia user accounts have been set-up in the EDMO CMS system for the responsible partners in each region.

The R2R team has prepared and published an initial set of 130 CSRs for the USA research vessels *Kilo Moana* and *Falkor* via its GeoNetWork node, which have been harvested and successfully integrated into the POGO CSR Catalogue service. This initial submission of CSRs marks the successful completion of the ODIP 2 prototype development task.

Since this milestone further population of the POGO catalogue has continued with more than 1700 CSRs from US research vessels now available.

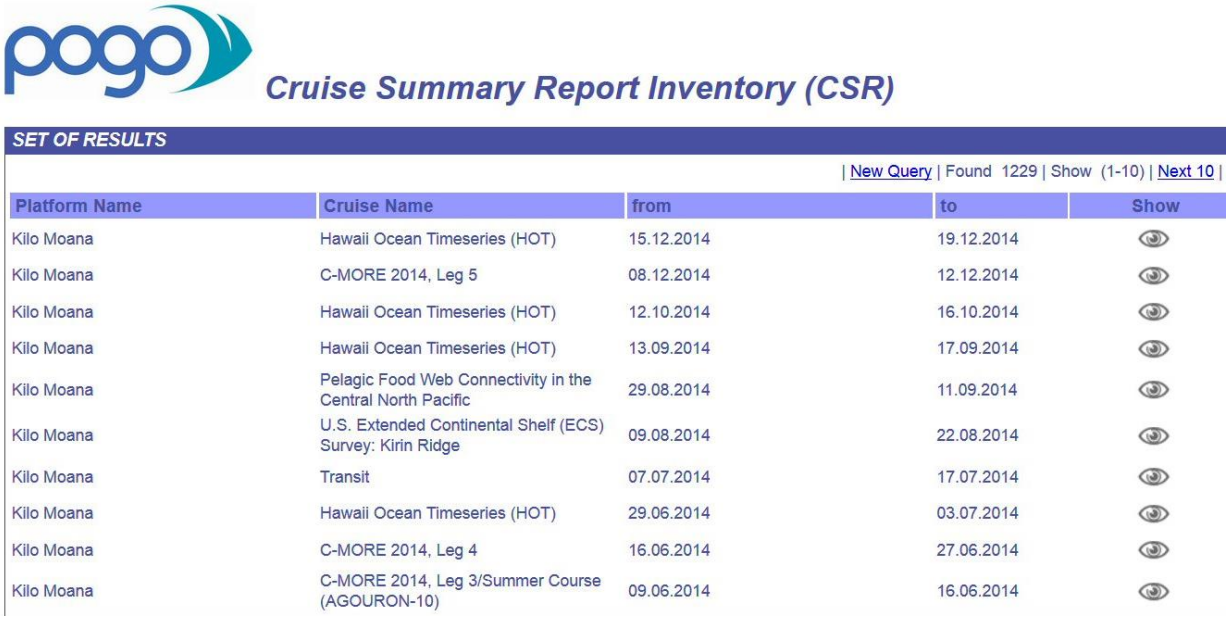

Figure 5 Cruise reports from the USA in the POGO portal (http://www.pogo-oceancruises.org)

2.3 ODIP 3 prototype development task

This task has focused on establishing a prototype for a Sensor Observation Service (SOS) and formulating common O&M and SensorML profiles for selected sensors (SWE), installed at research vessels and in real-time monitoring systems. It was led by IMOS on behalf of AODN (Australia).

Europe, USA and Australia are taking steps towards adopting Sensor Web Enablement' (SWE) and developing standards that can be applied by operators of operational marine observation systems to describe their observations in more detail and to provide standardized access to these observations using the SOS service protocol. This can provide a mechanism for direct access to the related data streams from operational sensor systems, such as real-time ocean monitoring networks and underway data from systems onboard research vessels.

The SWE architecture comprises two major building blocks: 1) the information which consists of the underlying conceptual models and encodings, and 2) the service model that provides the specification of services. SWE comprises SensorML profiles, Observation & Measurement profiles and Sensor Observation Services.

The ODIP directly involved partners directly involved in this task are: UTAS (leading), AIMS and BOM on behalf of AODN; CSIC, OGS, IFREMER, MARIS, CNR, BODC, and AWI on behalf of SeaDataNet – Eurofleets and RITMARE projects; NOAA on behalf of US-IOOS; and 52North.

The ODIP 3 prototype development task had originally aimed to establish a set of common standards for Sensor Web Enablement (SWE) applicable to a sensor observation service (SOS) for selected sensors (SWE) installed on research vessels and in real-time monitoring systems. However, as the ODIP project has progressed it has become evident that it is not currently feasible to define common SWE standards. There are multiple SWE initiatives in the marine domain as part of several European research projects for observing systems (EuroFleets 1 and 2, JERICO and JERICO-NEXT, FixO3, AtlantOS, GROOM), data management (SeaDataNet 2), and instruments and sensors (Sense OCEAN, NEXOS, and SCHeMA). There are also major implementations of SWE in the USA as part of the US-IOOS programme, and several initiatives working with SWE in Australia. It has therefore been decided to realign the aims of the ODIP 3 prototype development task to instead follow these SWE developments in the aforementioned projects, in which ODIP partners are involved.

During the ODIP workshops a range of presentations were given by a number of these SWE initiatives. Specific presentations have also been dedicated to SensorML and O&M profiles, SOS services (52North SOS service, ncSOS), vocabulary mappings (instruments, manufacturers, etc.), SWE editors, and performance of SOS services. These presentations and the resulting discussions have led to synergies and cross-fertilisation between the SWE initiatives, but as yet not to common standards

Steps have also been taken to set up a testbed for different SOS services and adding further services to it. A Github (https://github.com/aodn/ODIP) collaboration tool for compiling an inventory of services and endpoints, vocabularies and registries has also been established.

3 Expanding the existing prototypes

The final results of the three ODIP prototype development tasks (see previous chapter) were presented and discussed as part of the 1st ODIP II workshop which took place in Paris, France, from 28th September to 1st October 2015. Following a discussion, it was decided to continue and expand these prototype development tasks by adding extra elements as part of the ODIP II project. In the seven-month period between the 1st and 2nd ODIP II workshops further analysis took place and the resulting specifications for the three expanded ODIP II prototype development tasks were presented, discussed and agreed at the 2nd ODIP II workshop which took place in Boulder, USA from 2nd May to 5th May 2016. The details of the new features, planned activities and the partners involved in each of the three ODIP II prototype development tasks are described below.

The activities related to the cross-cutting topics addressing vocabularies, data publishing and citation, and unique identifiers for researchers will be continued, partly by monitoring and discussing on-going developments in the regions and relevant developments at global level during the workshops, and partly by developing solutions contributing to the three ODIP II prototype development task.

3.1 ODIP 1 prototype development task expansion

The ODIP 1 prototype development task addresses interoperability of regional discovery services (SeaDataNet, NCEI and AODN) with the global GEOSS and ODP portals. The current solution makes use of metadata brokerage at the collections level and entries are included in GEOSS and ODP with return links at the granule level to the regional portals and their respective data access options.

The expanded ODIP 1 prototype development task will be led by MARIS with contributions by European, USA, and Australian partners. In addition to the existing regional data systems (SeaDataNet, NCEI and AODN) Ocean Networks Canada may also be included in this activity.

A number of development plans are being considered at the metadata and data level for the ODIP 1 prototype development task.

Development plans at **metadata** level:

- Making the service more operational and fully dynamic so that new entries propagate automatically through the system to the global portals (Partners responsible: MARIS and CNR)

- Work on semantic interoperability (Partners responsible: BODC (lead), CNR, MARIS, NCEI, UTAS, CSIRO, RIHMI-WDC and ONC): the GEO-DAB brokerage service converts all metadata entries from the regional systems to a Generic Brokerage Reference Schema. However, each region may be using their own vocabularies or free text to populate common metadata fields which are then incorporated as provided in the GEOSS and ODP user interfaces (searching and publishing).

  It is not practical for the regional managers and connected data providers to rewrite their metadata model population and apply it to all legacy data holdings. A more efficient system is to provide a translation service that will interact with the GEO-DAB brokerage service. The translation service will capture knowledge of which terms are identical, similar or specialisations of other terms used across the regional systems

(SeaDataNet, NCEI and AODN) and publish it in a machine to machine system. This service could then automate the translation on ingestion of regional metadata records into the GEO-DAB brokerage service and following exposure of converted metadata records for inclusion in the GEOSS and ODP portals and their catalogue search user interfaces.

Analysis to be undertaken includes the following actions:

- Identify vocabularies used, if any, to populate comparable metadata fields by the regional systems
- Determine mappings between vocabularies being used in comparable metadata fields
- Identify differences in the levels of granularity of the published metadata and how they could be mapped with each other
- Identify and feedback to regional system managers where vocabularies are being used by other systems in place of free text entries
- Look at options for the publication of the mappings in human and machine readable formats. An “ODIP Rosetta Stone” vocabulary service might be provided and maintained. One or more vocabulary services host all mappings that are required for catalogue harmonisation.
- Analyse how the ODIP Rosetta Stone vocabulary service will interact with the GEO-DAB brokerage service and implement this solution in order to expose XML metadata from the regional systems including harmonised vocabularies as CSW and OAI-PMH services.

- Integrate the ODIP 1 and ODIP 3 prototype solutions to combine discovery metadata and SWE metadata for operational oceanography systems that go directly from data discovery to access and visualisation (Activity led by: CSIC and IFREMER)
-

Development plans at **data format** level:

- Regional systems use and deliver their own data formats. In Europe the INSPIRE Directive dictates the need for a more harmonized approach using specific data formats etc. Both EMODnet and SeaDataCloud (the successor to theSeaDataNet project) are working towards full INSPIRE compliance at the data level. The contributions made from the various initiatives participating in the ODIP II project will also potentially support these objectives. These activities include:

  - Analysis of how the SeaDataNet ODV format can be used as the basis for an INSPIRE compliant data format, following the INSPIRE Observations & Measurements (O&M) data model and guidelines. Potentially relevant are the new RDF/OWL and JSON implementations, which have resulted in proposals for a minor revision to version 2.1 of O&M.
  - Consider any extensions to the existing SeaDataNet NetCDF (CF) data format that may be required

- Development and deployment of central transformation services for converting SeaDataNet ODV datasets to other required output formats such as SeaDataNet NetCDF (CF) and relevant INSPIRE application schemas.

The planned interaction between the SeaDataCloud developments and other relevant regional initiatives around the world within the framework of ODIP II will be led by CNR, MI and IFREMER with involvement of other European ODIP II partners.

Development plans for **horizontal interoperability**:

- The present ODIP 1 prototype interoperability solution addresses vertical interoperability between the three regional systems and the GEOSS and ODP portals using the brokerage and CS-W and OAI-PMH services. It has been proposed to also analyse and implement, where possible, horizontal interoperability between the three regional systems, using OpenSearch and WMS-WFS. **OpenSearch** facilitates remote queries between different data portals. It is supported by Google, Yahoo, as well as other search engines and also an OGC standard. OpenSearch requests are simple and consist of HTTP GET with parameters: Free search, Geospatial (area or point + radius), Temporal (from to). OpenSearch can be used to go from collections to granules. **OGC WMS** facilitates sharing and exchanging map layers between each of the regional portals. Web Feature Service (**WFS**) allows interrogation of objects on maps for retrieving feature information. The feature information can include URLs to the associated portal to retrieve further metadata and options to request access and delivery of the data sets. WMS – WFS can be applied at collections and granules level. This activity will be led by MARIS with contributions from NCEI, UTAS, BODC and possibly others.

  The aanalyses to be undertaken include the following activities:

  - Identify existing OpenSearch and WMS-WFS services at the regional systems level and analyze these in detail: e.g. the metadata level being supported
  - Build a common OpenSearch user interface in the ODIP portal that provides search options for the connected regional systems at the metadata level
  - Build a common WMS – WFS user interface in the ODIP portal giving search options for the connected regional systems at the metadata level
  - Possibly integrate the “ODIP Rosetta Stone” vocabulary service for harmonisation of the vocabulary terms in use
  - Analyze and arrange dynamic maintenance of the user interfaces.

3.2 ODIP 2 prototype development task expansion

The ODIP 2 prototype solution addresses interoperability of regional cruise summary reporting (CSR) systems in Europe (SeaDataNet), the USA (R2R) and Australia (MNF) with the global POGO portal. At present the SeaDataNet CSR V3.0 schema is used for cruise summary reporting in Europe and USA while relevant bodies in Australia are currently undertaking preparatory mappings.

The ODIP 2 prototype development task will be led by BSH with contributions by the SeaDataNet (Europe), R2R (USA) and MNF (Australia) partners. A number of development plans for the expansion of this prototype solution are under consideration.

- Further population of the POGO Cruise Summary Report service:
  - Ingesting new and updated CSR entries from the SeaDataNet partners in Europe that are received through regular submissions of CSR XML files and increasingly through weekly harvesting from SeaDataNet partners that have set up GeoNetwork nodes. Task will be led by BSH with the SeaDataNet partners.
  - R2R partners are converting additional US cruise reports to the SeaDataNet CSR 3.0 format and making these available for harvesting by BSH through the R2R GeoNetwork node in preparation for ingestion into the POGO portal. Task will be led by R2R with contributions from BSH.
  - Australian partners are exploring further options for providing cruise reports for Australian research vessels, such as the new RV Investigator, in the SeaDataNet CSR V3.0 format. Task will be led by CSIRO with contributions from BSH.
- Identification and upgrading of legacy CSR records for USA, Canada and Australia received through ICES (Task led by BSH with R2R, Australian and Canadian partners)

Upgrading of the CSR schema following suggestions from the R2R project and leveraging the new SeaDataCloud project (the successor to SeaDataNet):

- The linked data concept will be worked out for directory services such as CSR as part of the SeaDataCloud project. This will optimise the SeaDataNet directories as resources for the semantic web. The metadata formats will be reviewed and possibly amended to achieve optimal data linking between cruises, observation platforms, projects, actors and resulting data resources. The CSR schema will be upgraded in response to suggestions from the R2R project within the framework of the ODIP project for example by:
  - Introducing gmx:Anchor to support links to vocabularies
  - Introducing ORCiDs to provide unique persistent identifiers for researchers involved in cruises
  - Consider introducing DOIs for data sets associated with the CSR records
  - Ensuring compliance with the INSPIRE standards.

  The upgrading of the CSR schema will be led by CNR with input from several of the SeaDataNet and R2R partners

Upgrading the CSR schema will have implications for the various metadata services for harvesting, import, search, and publishing as well as for the editing tools. Upgrading of the SeaDataNet and POGO services will be undertaken by SeaDataNet partners leveraging the activities of the SeaDataCloud project. Task led by IFREMER with support from BSH, MARIS and BODC. Upgrading at US nodes will be undertaken by the R2R partners.

The CSR services will also be provided as a SPARQL endpoint. This involves publishing CSR records within Resource Description Framework (RDF) representations. RDF is the encoding language of linked data and the semantic web, and has an associated query language (SPARQL) which allows interrogation of links between datasets published in RDF. Once the CSR records have been published as RDF, they can be harvested into SPARQL endpoints. This will be undertaken by SeaDataNet partners. Task led by BSH with support from MI, MARIS and BODC.

3.3 ODIP 3 prototype development task expansion

The ODIP 3 prototype development task deals with the further development of sensor web enablement (SWE) standards and services in concertation with many on-going regional projects as well as the OGC standards body. It will build upon the results of the previous ODIP project and explore further enhancements and new technological approaches for SWE. This task will also focus on establishing synergy and encouraging alignment between the ongoing regional projects that are addressing aspects of SWE.

The ODIP 3 prototype development task will be led by 52°North, a leading specialist in SWE for the marine domain, with contributions from several ODIP II partners that are involved in a range of regional SWE initiatives.

A number of development plans are being considered for the expansion of ODIP 3:

Establish OGC Sensor Web Enablement (SWE) standards for facilitating the interoperable sharing of oceanographic observation data and metadata including the development of marine profiles of relevant OGC SWE standards:

- OGC SWE standards will be evaluated for the interoperable sharing of oceanographic observation data and metadata. Recommendations and guidance will be derived for how SWE can be used in practice and special attention will be given to the different approaches for applying the SWE standards used by different members of the ODIP II consortium. Based on these findings, ODIP II intends to contribute to the development of marine profiles of relevant OGC SWE standards (i.e. OGC Sensor Observation Service (SOS), OGC Observations and Measurements (O&M), and OGC Sensor Model Language (SensorML)). This also includes the use of appropriate controlled vocabularies for consistent marking-up of the metadata profiles. There are already several controlled vocabularies in the marine domain. These will be applied and augmented with extra terms and lists where needed. The overall aim of defining these profiles is to further improve the interoperability between SWE systems deployed by different organization around the globe. This approach will be similar to that used in other domains such as hydrology, which have resulted in domain-specific SWE profiles (e.g. the OGC SOS 2.0 Hydrology Profile which was an output of the European GEOWOW project). This activity will be led by 52°North with contributions from many of the ODIP II partners.

Handling large volumes of data within SWE-based infrastructures:

- Analysing and optimising the ability to handle large volumes of data are closely related to SWE best practices. An assessment will be made of which data modelling and data management challenges need to be addressed in order to use SWE for serving the content of very large, potentially heterogeneous data archives. The result of this activity is planned to be a prototypical SOS implementation by AWI and 52°North.

Use of lightweight technologies such as JSON and REST as complementary solutions:

- Complementary to the XML-based SWE standards, lightweight technologies such as JSON and REST are becoming more and more relevant. The advantages they provide such as reduced data volumes as well as easier data processing on (potentially resource constrained) client platforms are becoming of great interest. Examples of implementations of lightweight technologies and Sensor Web concepts (e.g. Sensor Cloud, the 52°North Sensor Web Client REST-API or the ESRI GeoServices REST

Specification) will therefore be collected and compared in order to share experiences and derive common best practices. A possible output could be a JSON encoding for O&M and SensorML. This activity will be led by CSIRO (Australia) with contributions from those ODIP II partners that have an interest in this topic.

RDF-based approaches for supporting the discovery of marine sensors and data sets:

- Approaches such as RDF allow a better description of the semantics of marine observation data and metadata. Using linked data concepts which are based on RDF could become a valuable complementary solution for enabling faster and better discovery of marine observation data sets and sensors. To evaluate the benefits of this technology, the development of a lightweight RDF layer will be considered. This layer could be put on top of SOS server databases. The idea behind this approach is to expose the most important data fields such as observed property, procedure, feature of interest, etc. as linked data so that marine observations and sensors can be more flexibly discovered across multiple data providers and organizations. A possible way to achieve this goal it to work with the D2RQ framework - a system for accessing relational databases as virtual, read-only RDF graphs. As a result, existing SOS servers would still be used for accessing marine sensor data but the additional linked data implementation could be used as a facilitator for the discovery of content offered by these SOS servers. Task led by 52°North with contributions from other ODIP II partners

Linking between sensor web services and global infrastructures such as GEOSS:

- The linkages between sensor web services and global infrastructures such as GEOSS will be analysed. An approach similar to that developed in the GEOWOW project for the hydrology domain could be followed. Based on a marine Sensor Web profile (see above) it would be possible to offer a common interface that may be used by GEOSS Common Infrastructure components such as the GEOSS Discovery and Access Broker in order to harvest metadata from marine SWE services and to make these resources discoverable and accessible through GEOSS. Activity led by CNR and 52°North.

Synchronise efforts for metadata/SensorML Editors:

- Implementation of SWE profiles requires the availability of dedicated editors to efficiently generate high-quality SensorML files. A number of existing projects are already developing such editors as stand-alone software (e.g. SeaDataNet MIKADO editor software) and as online services (e.g. RITMARE starter kit). Activities will be undertaken to seek synchronisation of these efforts in the different projects. Task led by IFREMER, CNR and CSIC.

3.1 Cross-cutting topics

The activities relating to the so-called “cross-cutting topics” will be continued in the ODIP II project. The topics being addressed are:

- Vocabularies (Led by BODC)
- Data publishing and citation (Led by IFREMER)
- Persistent identifiers (Led by IFREMER)

The activities will be continued as follows by:

- monitoring and discussing relevant developments in these areas both at the regional and at the global level during ODIP II workshops
- developing appropriate solutions as required by the ODIP II prototype development tasks.

Annex A: Figures and Tables

Annex B: Terminology

Term	Definition
AODN	Australian Ocean Data Network
API	Application Programming Interface (API): a set of routine definitions, protocols, and tools for building software and applications
CDI	Common Data Index metadata schema and catalogue developed by the SeaDataNet project
CF	Climate and Forecast conventions: metadata conventions for the description of Earth sciences data, intended to promote the processing and sharing of data files http://cfconventions.org/
CSR	Cruise Summary Reports is a directory of research cruises.
CSW	Catalog Service for the Web (CSW): OGC standard for exposing a catalogue of geospatial records in XML on the Internet
D2RQ	A system for accessing relational databases as virtual, read-only RDF graphs. It offers RDF-based access to the content of relational databases without having to replicate it into an RDF store
DataCite	Global non-profit organisation that provides persistent identifiers (DOIs) for research data to support improved citation https://www.datacite.org/
DOI	Digital Object Identifier (DOI): a unique persistent identifier for objects which takes the form of a unique alphanumeric string assigned by a registration agency
EDMO	European Directory of Marine Organisations
EMODnet	EU-funded intiative to develop and implement a web portal delivering marine data, data products and metadata from diverse sources within Europe in a uniform way. http://www.emodnet.eu/

GEO	Group on Earth Observations: a voluntary partnership of governments and organizations supporting a coordinated approach to Earth observation and information for policy making
GEO-DAB	Brokering framework developed and implemented by GEO for interconnecting heterogeneous and autonomous data systems http://www.geodab.net/
GeoNetwork	An open source catalogue application for managing spatially referenced resources. It provides a metadata editing tool and search functions as well as providing embedded interactive web map viewer
GEOSS	Global Earth Observation System of Systems: international initiative linking together existing and planned observing systems around the world http://www.earthobservations.org/geoss.php
GitHub	Distributed revision control and source code web-based Git repository hosting service.
GML	Geography Markup Language (GML): XML grammar defined by the OGC to express geographical features
ICES	International Council for the Exploration of the Sea http://www.ices.dk/
IMOS	Integrated Marine Observing System: Australian monitoring system; providing open access to marine research data http://imos.org.au/
INSPIRE	EU Directive (May 2007), establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment.
IOC	Intergovernmental Oceanographic Commission of UNESCO (IOC/UNESCO).
IODE	International Oceanographic Data and Information Exchange" (IODE) of the "Intergovernmental Oceanographic Commission" (IOC) of UNESCO

IOOS	US Integrated Ocean Observing System https://ioos.noaa.gov/
ISO	International Organization for Standardization http://www.iso.org
jOAI	Java-based OAI software that supports the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), version 2.0 http://www.dlese.org/oai/
JSON	JavaScript Object Notation: an open-standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs. It is the most common data format used for asynchronous browser/server communication.
MarineID	Registration and authentication services for selected marine data services including SeaDataNet and EMODnet
MCP	Marine Community Profile: ISO19115 profile developed by Australian Ocean Data Centre Joint Facility (AODCJF) for marine data
MIKADO	Java-based software tool, for creating XML metadata records for the SeaDataNet directories EDMED, CSR, EDMERP, CDI and EDIOS.
MNF	Marine National Facility is owned and operated by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) http://mnf.csiro.au/
NetCDF	Network Common Data Form (NetCDF): a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.
NCEI	NOAA's National Centers for Environmental Information https://www.ncei.noaa.gov/
O&M	Observations and Measurements: OGC standard defining XML schemas for observations, and for features involved in sampling when making observations

ODP	Ocean Data Portal: data discovery and access service, part of the IODE network http://www.oceandataportal.net/portal/
ODV	Ocean Data View: a software package for the interactive exploration, analysis and visualization of oceanographic and other geo-referenced profile, time-series, trajectory or sequence data
OGC	Open Geospatial Consortium: international voluntary consensus standards organization http://www.opengeospatial.org/
OIA-PMH	Open Archives Initiative Protocol for Metadata Harvesting https://www.openarchives.org/pmh/
OpenDAP	Open-source Project for a Network Data Access Protocol: a data transport architecture and protocol widely used by earth scientists https://www.opendap.org/
OpenSearch	Collection of technologies that allow publishing of search results in a format suitable for syndication and aggregation http://www.opensearch.org/Home
ORCID	Open Researcher and Contributor ID: a non-proprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors http://orcid.org/
POGO	The Partnership for Observation of the Global Oceans: a forum created by the major oceanographic institutions around the world to promote global oceanography. http://www.ocean-partners.org/
R2R	Rolling Deck to Repository: a US project responsible for the cataloguing and delivery of data acquired by the US research fleet.
RDF	Resource Description Framework (RDF): family of W3C specifications for conceptual description or modeling of information that is implemented in web resources https://www.w3.org/RDF/

REST	REpresentational State Transfer (REST): an architectural style, and an approach to communications often used in the development of web services
SensorML	OGC standard providing models and an XML encoding for describing sensors and process lineage
SOS	Sensor Observation Service: a web service to query real-time sensor data and sensor data time series. Part of the Sensor Web
SPARQL	SPARQL Protocol and RDF Query Language: a semantic query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (RDF) format http://www.w3.org/TR/rdf-sparql-query/
SWE	Sensor Web Enablement: OGC standards enabling developers to make all types of sensors, transducers and sensor data repositories discoverable, accessible and useable via the web
US-NODC	US National Oceanographic Data Centre (now the NOAA National Centres for Environmental Information) https://www.nodc.noaa.gov/
W3C	World Wide Web Consortium: main international standards organization for the World Wide Web http://www.w3.org/
WCS	Web Coverage Service Interface Standard: OGC standard defining Web-based retrieval of coverages i.e. digital geospatial information representing space/time-varying phenomena http://www.opengeospatial.org/standards/wcs
WFS	Web Feature Service: standards allowing requests for geographical features across the web using platform-independent calls
WMS	Web Map Service: standard protocol for serving geo-referenced map images over the Internet

XML	Extensible Markup Language: a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable http://www.w3.org/XML/