A project supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB).

Report for Collaborative Oceanography and Monitoring for Protected Areas and Species (IVA5015)

D2.4.2
Report of Data Processes Review Workshop

23rd March 2020

Report by: Marine Institute, Ireland
Report author: Adam Leadbetter

Executive Summary

This report summarises the Data Process Review workshop held for the COMPASS project at the National Ocenography Centre, Liverpool, UK on 19th and February 2020. The workshop was a part of the project's task on data management to document, review and revise data management processes at partner organisations and to promote data management best practices through a data management quality management framework approach based on ISO 9001 and accredited by the International Oceanographic Data and Information Exchange of UNESCO's Intergovernmental Oceanographic Commission.

The workshop was also an opportunity to share data management best practices and align data management processes with the related SeaMonitor and MarPAMM projects, also funded in the same Interreg scheme. Input to the workshop was also received from the British Oceanographic Data Centre (BODC) and the UK's Marine Environmental Data and Information Network (MEDIN). BODC kindly hosted the event for the COMPASS project and the workshop was supported by MEDIN.

This document introduces the aims of the workshop, reports the workshop activities and gives a list of actions arising from the discussion throughout the workshop.

Contents

1. Introduction

This report forms part of the work for Task 2.2 "Documentation, review and revision of data management processes" of the COMPASS project. The task will first involve an inward look at the current state of the data management processes in all partner organisations (that is to assess the processes of the implementation communities in the parlance of the IOC report). Following this short review phase, a collaborative workshop will be held where areas of common practice can be identified and areas for development, in light of the requirements and standards choices in Activity A.T2.1, can be highlighted. This workshop should include representatives from the relevant NODCs in order to provide guidance where necessary, creating an expert team from both within and external to the project. The final step in this task will be to revise and document the data management processes. This documentation should be of high standard and suitable for submission to the International Oceanographic Data and Information Exchange's Quality Management Framework. All work carried out will be reviewed at a second collaborative data management workshop. The review and revision of the processes should also address the sustainability of the data management network and any applications following the lifetime of the project. Through making appropriate reuse of existing data management pathways and integrating a common suite of tools this sustainability should be achievable.

This report describes the "Data processes review workshop" which was held at the National Oceanography Centre, Liverpool UK on 19th and 20th February 2020 to fulfil Deliverable 2.2.4 of the COMPASS project. The statement of work for the project describes this workshop as follows: "Areas of common practice can be identified and areas for development, in light of the requirements and standards choices". At the COMPASS Advisory Group meeting in October 2019, the policy leads for the regions represented by organisations in the project requested expanding the scope of the meeting to discuss data management approaches with the other projects in the same funding call, SeaMonitor and MarPAMM. The British Oceanographic Data Centre (BODC); Marine Environmental Data and Information Network (MEDIN); and the UK Integrated Marine Observing Network were also represented at the meeting.

2. Acknowledgments

The meeting was kindly hosted by BODC at the National Oceanography Centre, Liverpool. MEDIN generously sponsored the meeting through teas/coffees and lunch on day 2.

3. Aims

The objectives of the workshop were to share the work undertaken within the COMPASS project on data processes, and also to discuss data integration with respect to the management of marine protected areas and potential connections for the COMPASS, SeaMonitor and MarPAMM projects with national and international data management infrastructures.

4. Day 1

Adam Leadbetter provided an overview of the Data Management work package for the COMPASS project and provided an update on the progress made in the work package since the start of the project. The work package is run in a manner which is aware of the proliferation of data standards (such as the European Commission's INSPIRE spatial data infrastructure; the UK's Marine Environmental Data and Information Network; the SeaDataNet infrastructure for marine data in Europe; and the European Marine Observation and Data Network) in existence for marine and environmental data without trying to impose any one of these standards, or indeed a new standard, on the project. This approach is illustrated through the

"Reduce-Reuse-Recycle" paradigm. The deliverables for the work package were reviewed, with the ISO9000 quality management system as a lens through which to view the organisation of the deliverables. The Erddap data broker was introduced, which has been suggested as an approach to delivery of data to a central data portal for the COMPASS project, and with a connection to a ncWMS server may be solution to integrating presentation of the model outputs from the project.

A follow-up discussion was held to the data requirements gathering exercise conducted at the 2019 COMPASS annual meeting. Glider data is a key requirement for the SeaMonitor project. The Passive Acoustic Monitoring data will be catalogued in an instance of Tethys database, which has been connected to Erddap in the US. Tethys can be used to create appropriate metadata and clips from data which has been appropriately quality controlled by researchers. Having this evidence that the data exist will be a key output from the project. It was suggested that the delivery of the contractually obligated data for the project should be scheduled for ingestion into the prototype portal for September 2020 ahead of the next annual meeting. This may include handheld CTD data from mooring locations; and water samples taken for nutrient analysis at mooring locations. Non-COMPASS data can also be incorporated, such as SeaMonitor and MarPAMM data. Other streams which may be of use include

- more fish tag positions
- more glider missions
- behavioural model outputs
- research vessel survey data (underway; CTD profiles)
- CEFAS WaveNet buoy data
- maps of residual currents
- data from the MikNet project
- and remotely sensed ocean colour data

A suggestion was made that data in the portal should link back to a metadata record which also includes links to such information as journal papers written from the dataset.

5. Day 2

The day was introduced by Adam Leadbetter, who gave an overview of the aims and objectives of the day and outlined the agenda. BODC and MEDIN were thanked for their support and hospitality. A brief description of the connection between the Interreg Va Marine Protected Area projects was given.

The participants in the meeting each gave a short introduction to themselves and their work.

5.1 Interreg Va Marine Protected Areas Projects

The first session of the day provided an overview of the three Interreg Va projects with a focus on Marine Protected Areas: COMPASS; SeaMonitor; and MarPAMM.

Adam Mellor introduced the COMPASS project, giving an overview of the instrument arrays which have been deployed in the water; and highlighting the approach which is being taken to take data from the project to produce information and impact, through the development of knowledge on the science affecting Marine Protected Areas. COMPASS has involved over 50 scientists across Ireland; Northern Ireland and Scotland; has deployed 90% of its monitoring array; has tagged 140 salmonids; and has undertaken glider missions. The data infrastructure for the COMPASS project has been defined and is

being established with the implementation of data and metadata standards. Modelling outputs from the project have been used in the assessment of beaked whale strandings. For outreach, three short-films; joint e-zines and a new website have been created. A CEFAS waverider has been deployed at the Afbi 38A mooring location. Time series and spatial data products are being developed. The Data Management work package is trying to aid on both standardisation and delivery, and the next challenges to overcome is integrating the outputs from the acoustic database being used within the cetaceans work package and the presentation of modelling outputs from different domains and resolutions. Cultural change has been an issue which is slow to resolve and can be frustrating, but which is important in addressing the project's goals.

Caroline Finlay introduced the SeaMonitor project and the consortium which includes partners from Ireland; Northern Ireland; Scotland and North America. The project is developing five spatial models and three management plans or management recommendations. University College Cork have undertaken tagging of abandoned, human reared seal pups with GPS transmitters. Basking shark and skate tagging is being scoped, and hotspots are being identified. Salmon receiver arrays are being deployed, and receiver hits will be compared with hydrodynamic model outputs. Cetecean monitoring using Passive Acousitic Monitoring systems are also incorporated into the salmon arrays and other infrastructure in the North Channel. The Loughs Agency are adopting the Data Management model developed by the COMPASS project, with an Erddap server installed and metadata collation is about to begin. Again, data sharing is a cultural issue with a reluctance to open data up as this has not happened historically.

Jay Calvert introduced the MarPAMM project which is developing models and management plans for a number of regions within the Interreg Va area. These regional management plans are in contrast with the localised, small geographic areas covered more traditionally. There will be six Marine Protected Area plans developed: four regional and two site-specific. This work is being led by Scottish natural Heritage. There are four models being developed, concerning: seabirds; benthic or seabed habitats; marine mammals and underwater noise; and coastal processes. The seabirds modelling is led by Marine Scotland Science and is based on GPS fixes of tagged seabirds giving distribution models and can be linked with both climate variables and by-catch estimates to understand the pressures of these on seabird populations. Afbi are leading on the benthic and seabed habitats work focussed on the connectivity between the different regions and utilising backscatter data from multi-beam bathymetry, some from Remotely Operated Vehicles, to categorise the habitats. High-resolution underwater laser scanning data is also being used where possible to give up to millimetre resolution. The marine mammals and underwater noise modelling aims to develop seal foraging and underwater noise models through mapping of seal density through a network of moorings to record underwater noise and clicks from marine mammal species. The coastal processes modelling activity, led by the University of Ulster, examines historic changes in coastal use through mapping from 1834-2014 and shorter timescale assessments of coastal erosion. The SWAN model is being used for sediment transport simulations. Data issues for the project revolve around both spatial and temporal scales of available data, which extends into the resolution and coverage of available datasets which can create a sampling bias. Awareness and availability of existing datasets is an issue, and this feeds a need for an inventory of metadata for the various projects which can be shared. No data management funding was included in the project proposal, and so assistance is required from the COMPASS project.

5.2 Data Infrastructures

This session aimed to outline the marine data infrastructures available in the UK, Ireland and in Europe.

Clare Postlethwaite described the Marine Environmental Data and Information Network (MEDIN), an open partnership of UK organisations for the sharing of data and metadata. The slogan of MEDIN is "measure

once, use many times" which encourages the sharing and reuse of data for purposes which may differ from the original reason for data collection or creation. As such, MEDIN aims to improve access to and the management of the UK's marine data and information holdings. MEDIN works through providing an operational network of marine-specific data archive centres with standards and tools to connect their metadata to a central portal. Over 600 organisations contribute to the portal which describes over 15,000 datasets. The data archive centres manage the data and provide the dataset descriptions to the central portal. Local data holders also send data descriptions to MEDIN. The MEDIN metadata standard is INSPIRE compliant as it is a specialisation of the GEMINI-2 profile which the UK uses; and as such is also ISO19115 compliant. A recent cost-benefit analysis showed MEDIN works well, with a benefit-cost ratio of 8. MEDIN can support the COMPASS project, and related projects, through advertising project data within the portal; through possible submissions of data to the data archive centres; through metadata standards and guidelines; and through targeted data management workshops. A short discussion on passive acoustic monitoring data followed. MEDIN currently does not deal with these data through an archival centre due to the large volume of the data, but liaison with the Joint Nature Conservation Committee Noise Register was suggested.

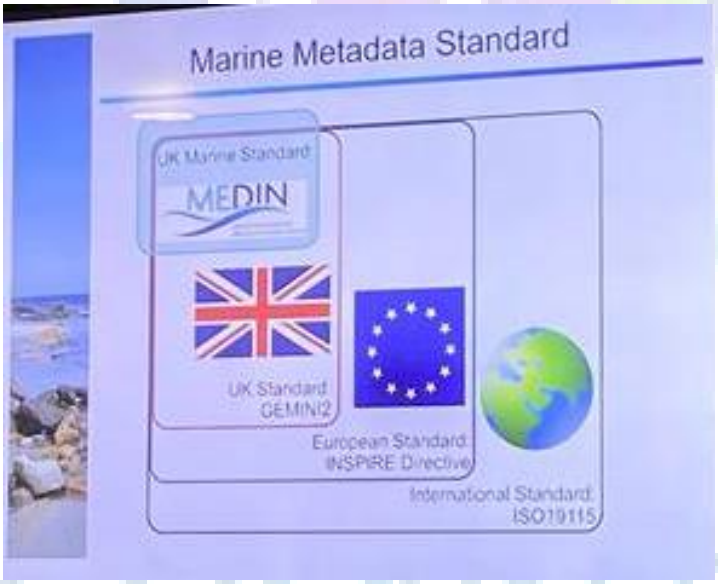

Figure 1. The MEDIN Marine Metadata Standard

Adam Ledbetter presented the work of the Marine Institute, Ireland on national data management. First the different drivers for marine data management were summarised, including supporting the growth of the Blue Economy and national marine strategy. A number of tools developed at the Marine Institute were highlighted including: the Digital Ocean platform; the Galway Bay Subsea Observatory dashboard; the Adrift tool used in search and rescue scenarios; and the aquaculture licensing decision support tool. The Marine Institute's approach to data cataloguing was introduced including standards compliant metadata and data citation through digital object identifiers. The idea of an Application Programming Interface layer to the various data and metadata endpoints, harmonised through consistent documentation was also discussed.

Mark Hebden presented a number of European and global initiatives which aggregate marine data. These included the

- SeaDataNet – the pan-European infrastructure for ocean and marine data management

- The European Marine Observation and Data Network (EMODnet) and associated real-time and downstream data services from the European Global Ocean Observing System and Copernicus.
- International Oceanographic Data and Information Exchange of UNESCO's Intergovernmental Oceanographic Commission
- The International Council for the Exploration of the Seas
- The International Hydrographic Office
- The World Meteorological Organisation
- The Research Data Alliance

In subsequent discussion it was decide that the Marine Institute, Afbi and BODC would work together under the EMODnet ingestion contract, to provide a sample of data from the COMPASS project to EMODnet as a proof of concept. It was also decided that the Marine Institute and BODC would collaborate to ensure that COMPASS was properly represented in the various catalogues of the SeaDataNet infrastructure.

5.3 Data Requirements for the Interreg Va Marine Protected Areas Projects

Andrew Conway introduced the Marine Institute's Data Management Quality Management Framework and in particular the work that has been done through this framework in the COMPASS project to document data processes through visual flows in the Business Process Model and Notation.

Adam Leadbetter introduced a discussion session focussed on the data requirements of COMPASS, SeaMonitor and MarPAMM.

Mark Hebden responded to a question on the management of data from sub-surface gliders. Metadata flows more easily now than it did when these platforms were first deployed. Building relationship with glider technicians and pilots is key to the successful management of data from these platforms. Real-time data exchange can be an issue, so it is important to get information on which sensors are mounted on a platform prior to its deployment. Gliders do not always get deployed by operational infrastructures or teams, often it is a lone researcher so their support structures may be limited. Data should be targeted at the Everyone's Gliding Observatories netCDF profile with terms from the NERC Vocabulary Server and World Meteorological Office and International Council for the Exploration of the Seas platform codes. Data discoverability is the minimum standard to aim for, and it was agreed that an entry in the European Directory of Marine Environmental Datasets (EDMED) would be created for the glider missions SAMS have undertaken for the COMPASS project. The Marine Autonomous Robotics Systems laboratory at the UK National Oceanography Centre offered support to project partners on glider deployments and infrastructures. Matthew Palmer also noted that other autonomous platforms than gliders are also important, such as surface vehicles, and should be considered alongside gliders in data infrastructures.

Andrew Conway chaired a discussion session on data from model outputs. The MarPAMM project wish to dynamically adjust and scale hydrodynamic and other parameters on map visualisations of model outputs, which would require a number of model runs to be available to the visualisation tool. A potential mismatch with species modelling was also identified. For SeaMonitor, the focus of the modelling effort is on specific areas of interest: in particular the Malin Channel and areas targeted by their management recommendations. Again, a desire to dynamically adjust variables for the modelled ranges within the visualisation tool was expressed. COMPASS and SeaMonitor will look to produce migratory routes for salmonids based on some decision parameterisation. This is something the projects will have to do

together with the amount of salmon data SeaMonitor will have from their North Channel and estuary arrays. Discussions on this between the projects are ongoing. It was felt that common disclaimers should be added for these data.

Adam Leadbetter closed the meeting.

6. Actions Arising

Owner	*Action*
Marine Institute	• Forward Erddap Google Group details to SAMS • Contact CEFAS with respect to republishing of the Wavenet data from Isaly / West Hebrides • Investigate Copernicus ocean colour / phytoplankton satellite imagery or other sources of relevant ocean colour data (PML, EOS Data and Analysis service, CEOS, Portugese data) • Test the MEDIN metadata schematron for version 3 • Develop a MEDIN metadata profile from Erddap • Liase with SeaMonitor staff on the incorporation of data management infrastructure into their project, including on-site visits and presentations.
Afbi	• Follow-up on data from Corystes and Miknet • Contact CEFAS re: Rockall survey
Marine Institute / Afbi	• Follow-up on the COMPASS legacy plan, emphasising the cultural change / human impact aspect • Follow-up with data providers to provide a timeline of delivery with a target of September • Arrange for a presentation of the Marine Institue's Data Management Quality Management Framework at Afbi • Develop resourcing models for an Afbi ocean data manager • Finalise COMPASS data sharing agreement and pass on to SeaMonitor and MarPAMM
Marine Institute / SAMS	• Produce Web Map Services to mesh the two COMPASS modelling domains
Marine Institute / BODC	• Develop EDIOS, EDMED and EDMERP entries for COMPASS, which will also support UKDMOS
Marine Institute / Afbi / BODC	• Develop a prototype workflow for the 38A surface deployment to European data infrastructures through the EMODnet ingestion contract

7. Agenda

7.1 Wednesday 19th February 2020

- 14:00 Welcome and local arrangements (Lesley Rickards, BODC)
- 14:05 Introduction & T2 Overview / Recap (Adam Leadbetter, Marine Institute)

- 14:20 Marine Institute Data Management Quality Management Framework (Andrew Conway, Marine Institute)
- 14:45 Round table review of data process (All)
    - Requires reading the deliverable beforehand
    - Please identify
        - Areas of overlap
        - Areas of divergence
        - Areas where help / assistance required
- 15:30 Break
- 15:45 Summary of discussion (Adam Leadbetter, Marine Institute)
- 15:55 Review of data requirements emerging from annual seminar / synergies (Andrew Conway, Marine Institute, All)
- 16:30 Next steps and timelines (All)
- 16:55 Wrap-up and plans for tomorrow (Adam Leadbetter, Marine Institute)
- 17:00 Close

7.2 Thursday 20th February 2020

09:15 Welcome and local arrangements (BODC)

7.2.1 Interreg Va Marine Protected Areas Projects

09:20 Session Introduction (Adam Leadbetter)

09:25 COMPASS Project Overview (Adam Mellor)

09:45 SeaMonitor Project Overview (Caroline Finlay, Loughs Agency)

10:05 MarPAMM Overview (Jay Calvert, Afbi)

10:25 Interreg Va Marine Protected Areas - Data Management Overview (Andrew Conway)

10:35 Discussion

10:45 Coffee (Kindly provided by MEDIN)

7.2.2 Data Infrastructures

11:15 Session Introduction (Adam Leadbetter)

11:20 UK National - Marine Environmental Data and Information Network (Clare Postlethwaite)

11:40 Ireland National (Adam Leadbetter)

12:00 European / Global (Mark Hebden)

12:20 Discussion

12:30 Lunch (Kindly provided by MEDIN)

7.2.3 Data Integration Requirements

13:30 Session Introduction (Adam Leadbetter)

14:00 Open discussion

- Data requirements relevant to management plans for Marine Protected Areas
- Data process documentations and sharing
- What would national data infrastructures like to see from the project?
- What other datasets are available and might be combined to meet the project goals?

15:00 Coffee (Kindly provided by MEDIN)

7.2.4 Wrap-up session

15:30 Session Introduction (Adam Leadbetter)

1535 Open discussion

- Next steps and agreed actions

16:50 Wrap-up (Adam Leadbetter)

17:00 Close

8. Attendees

Organisation	*Name*
Marine Institute, Ireland	Adam Leadbetter Andrew Conway
Afbi	Adam Mellor Colleen Ward Jay Calvert Victoria Dews
British Oceanographic Data Centre	Mark Hebden Robin McCandliss Lesley Rickards
Loughs Agency	Caroline Finlay
Marine Environmental Data and Information Exchange	Clare Postlethwaite
Scottish Association for Marine Science	Paola Arce
National Oceanography Centre, UK	Matthew Palmer