



# T2: DATA MANAGEMENT

## Data processes for the Compass project\*

November, 2019

\*The views and opinions expressed in this document do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB).

# Contents

1. Purpose and Scope.....	1
2. Paper Structure .....	1
Oceanography .....	2
Salmonids .....	2
Cetaceans .....	2
Modelling.....	3
Data Management.....	3
3. Funding and Governance .....	4
Partners .....	4
Governance .....	4
4. Data Management Plan.....	5
5. Process Flows .....	6
6. Conclusion.....	7
7. References.....	7
Appendix 1: Data Management Plans .....	8
Marine Institute: Fixed Mooring Mace Head.....	8
Data Collection .....	8
Documentation and Metadata .....	8
Ethics and Legal Compliance .....	8
Storage and Backup .....	8
Selection and Preservation .....	9
Data Sharing .....	9
Responsibilities and Resources.....	9
SAMS: Mooring, Receiver and Glider .....	10
Data Collection .....	10
Documentation and Metadata .....	10
Ethics and Legal Compliance .....	10
Storage and Backup .....	10
Selection and Preservation .....	10
Data Sharing .....	11

Responsibilities and Resources .....	11
AFBI: Belfast Pile 8 Moored EXO2 .....	12
Data Collection .....	12
Documentation and Metadata .....	12
Ethics and Legal Compliance .....	13
Storage and Backup .....	13
Selection and Preservation .....	13
Data Sharing .....	13
Responsibilities and Resources .....	14
AFBI: Belfast Pile 8 Moored CTD .....	15
Data Collection .....	15
Documentation and Metadata .....	15
Ethics and Legal Compliance .....	16
Storage and Backup .....	16
Selection and Preservation .....	16
Data Sharing .....	16
Responsibilities and Resources .....	16
MSS: Loch Ewe .....	18
Data Collection .....	18
Documentation and Metadata .....	19
Ethics and Legal Compliance .....	19
Storage and Backup .....	19
Selection and Preservation .....	20
Data Sharing .....	21
Responsibilities and Resources .....	21
MSS: Passive Acoustic Monitoring (PAM) .....	23
Data Collection .....	23
Documentation and Metadata .....	23
Ethics and Legal Compliance .....	23
Storage and Backup .....	23
Selection and Preservation .....	24
Data Sharing .....	25

Responsibilities and Resources .....	25
Appendix 2: Process Flows .....	27
Mooring Overview .....	27
Mooring Operations and Maintenance .....	28
Acoustic Sensor Calibration .....	29
Acoustic Sensor Deployment and Retrieval .....	30
Acoustic Data Processing .....	31
Glider Overview .....	32
Glider Processing .....	33
Modelling Overview .....	34
ADCP Processing .....	35
ERDDAP Publishing .....	36

# 1. Purpose and Scope

The Marine Institute (MI), in its role as data management work package leader for the Compass project, has organised several wide ranging meetings between scientific and project personnel from partners on each work package theme in the project. The purpose of those meetings was to:

1. Scope and flesh out the data management processes of each partner entity for each work package they are involved with.
2. Document those processes and procedures in a readable, concise and consistent manner, using developed and peer reviewed documentary systems.
3. Highlight synergies present in the processes between not only the different work packages but indeed
4. Identify any obvious gaps or opportunities for process improvement within those procedures, noting them to be revisited in the future.

The outputs of those meetings form the basis of this paper.

This paper has not been developed in isolation and builds on the significant contributions which have been made by members of staff at each partner organisation and indeed from each work package.

The focus of this report therefore is to introduce the practices which the partners have for processing Compass data as well as describe the factors which have influenced the design of same. Some of the topics discussed herein can have specific dependencies and application outside of the scope of the Compass project, but where possible, the scope has been limited to the project only.

## 2. Paper Structure

This paper describes the data processes employed by the organisations involved in the Compass project by using the established methodology designed by staff at the National Oceanographic Data Centre (NODC), based out of the Marine Institute.

As detailed by Leadbetter et al. (in press), there are several fundamentals which an organisation must complete to establish a quality data management system. Once in one's possession, a competent individual should be able to comprehend the intricacies of a data generating system which can better help one understand their place within it as well as identify areas open to improvement. Two of those fundamentals, the data management plan and the process flow, are expanded upon in the sections below.

To ensure the highest level of clarity, brevity and succinctness in this paper, a sample completed data management plan and process flow for each operational work package will follow in subsequent sections.

As stated in the Compass project application, the work package themes can be summarised as follows:

## Oceanography

The Oceanography work-package delivers an integrated network of buoys for oceanography in regional seas by creating new moored observation stations at key locations and integrating these with established monitoring stations already within the region. Activities co-developing the skills to deliver new parameters of significance to marine protected areas complements activities designed to co-ordinate monitoring across the three jurisdictions (Northern Ireland, Republic of Ireland, Scotland) in the region. These activities implement common practices and telemetered data interfacing directly with the Data Management work-package.

The theme delivers a reinforced and enhanced monitoring capacity for the region, and acts as a platform for delivering baseline regional scale data, information on key species, and process specific studies. The logistical activity of the Oceanography work package also serves as a platform for other studies, as it is supported by established work programmes with long term provenance and data sets. The delivery of observational data through a co-developed data management infrastructure will facilitate the work-programmes assessments of salmonids, cetaceans and seals, as well as reinforce and extend the validated domain of the regional hydrodynamic models.

The work package design includes periodic grab sampling to validate in situ sensor data and provide data additional essential parameters to ensure consistent quality assured data is generated across partners.

## Salmonids

This work package defines the habitats used by outward migrating salmon and resident marine phase sea trout through a network of moored acoustic receivers. The critical issue for both species is population decline, without full knowledge of cause. The species are closely related and are present for some of their lifecycle in the same rivers and coastal waters, but there is an emerging need for very different population management strategies due to differing life history strategies and ranges used in the marine environment. Separation of legislative provisions for the two species which have been closely linked in the UK and Ireland for almost a century is now an obvious management objective.

## Cetaceans

This work package develops and implements a future-proof passive acoustic monitoring (PAM) program for marine mammals (cetaceans and pinnipeds) in relation to a cross border (Scotland, Northern Ireland and the Republic Ireland) network of Marine Protected Areas (MPAs). This is underpinned by harmonised cross-border methods and partners using consistent methods to ensure project legacy and the ability for large-scale, future data sharing. Data on the occurrence of marine mammals is integrated with oceanographic, visual and ambient noise data sampled at similar spatiotemporal scales to increase understanding of habitat associations and environmental drivers for observed seasonal occurrence of mobile species.

## Modelling

The modelling work package addresses the project objective of developing cross border capacity by drawing together two currently unrelated national hydrodynamic models, the Irish Northeast Atlantic Model and the Scottish Shelf Model (SSM), or an alternative west coast domain, into an integrated model system for the target region. This interfaced hydrodynamic model itself addresses one of the required model outputs, providing simulations of the evolving marine environment, but also the underpinning for two further models. These consider biological connectivity between sites across the region and hydrodynamic habitat type, both of key importance in understanding, developing and managing marine protected area networks.

## Data Management

The data management work package has the responsibility of bringing the outputs from each organisation and work package together and delivering them to the public. That delivery mechanism would take the form of a bespoke portal on the pre-existing Digital Ocean platform.

The framework to facilitate data dissemination and data acquisition addresses the needs of the other work package themes along with the other work performed by this work package theme is detailed at length in this paper.

## 3. Funding and Governance

The Compass project is funded by the Special EU Programmes Body (SEUPB), a body which was established in part to facilitate the positive impact that European Regional Development Funding has on the people living across Northern Ireland, the border region of Ireland and Western Scotland.

More specifically, funding has been drawn from the INTERREG VA Programme, a programme designed to help overcome the issues which arise from the existence of a border, thus providing an opportunity to create a more prosperous and sustainable cross-border region.

### Partners

- Agri-Food and Biosciences Institute (AFBI) – Northern Ireland
- Inland Fisheries Ireland (IFI) – Republic of Ireland
- Marine Institute (MI) – Republic of Ireland
- Marine Scotland Science (MSS) – Scotland
- Scottish Association for Marine Science (SAMS) - Scotland

### Governance

The Marine Institute (MI), as lead Data partner, has implemented a series of data governance procedures and guidelines dictating how data should be handled and how it can be reused. The MI's Data Policy, approved since 2017, clearly states the requirements for and appropriate use of the MI's data and information resources.

This policy identifies who may access data and for what purposes as well as guidelines relating to data access, classification, protection, retention and the effects of relevant legislation (including freedom of information and general data protection regulation) on each. A separate deliverable on the data standards used in the Compass Project is also available.

Metadata created as part of the Compass project from all partners will be included in the National Oceanographic Data Center (NODC)'s data catalogue based at the Marine Institute. This will be a means of firstly attributing generated data to the correct partner but secondly as a way of ensuring that all data is compliant with statutory metadata standards, including the [Climate and Forecast Conventions](#), [SeaDataNet](#), [EMODnet](#) and [MEDIN](#).

## 4. Data Management Plan

The first port of call when understanding how a data process operates is the data management plan (DMP). It describes what data are created during the course of a project, how they will be stored, how it will be archived after the conclusion of the project and how access should be granted to same. In an ideal scenario, a DMP is prepared before the outset of a project, however, it can be reviewed and fundamentally changed throughout as well as after the conclusion of a project so that it remains relevant.

The format taken for the Compass project asked work package participant to answer the following questions under these headings:

### Data Collection

- What data will be collected or created?
- How will the data be collected or created?

### Documentation and Metadata

- What documentation and metadata will accompany the data?

### Ethics and Legal Compliance

- How will ethical issues be managed?
- How will copyright and intellectual property rights (IPR) issues be managed?

### Storage and Backup

- How will the data be stored and backed up during the research process?
- How will access and security be managed?

### Selection and Preservation

- Which data are of long-term value or should be retained, shared and/or preserved?
- What is the long-term preservation plan for the dataset?

### Data Sharing

- How will data be shared?

- Are there any requirements for the restriction of data sharing?

## Responsibilities and Resources

- Who will be responsible for data management?
- What resources are required to deliver this data management plan?

Together, the answers to these questions form the data management plans for the entire Compass project. These can be reviewed in Appendix 1 of this paper.

## 5. Process Flows

Once data management plans have been completed and an organisation has been made aware of all of their data operations, the logical next step is to complete what are known as “Process Flows”. As stated in Leadbetter et al. (In Press), a Process Flow is a visual representation of an activity illustrating the relationship between major components and demonstrating the logical sequence of events. A Process Flow describes ‘the what’ of an activity. While a Process Flow can be split across multiple levels, at its bare minimum, it should illustrate the complete lifecycle of the data process. Creating flows involves gathering all actors involved together and determining the inputs, outputs, steps and process time, all informed by the completed data management plan.

Process flows themselves have multiple advantages. The first and foremost being that it is a visual representation of a process in an easily understandable style. Thus, by enforcing a relatively simple visual design language to a flow, it reduces the time it takes for a person to understand a process at a high level. It can also act as a literal sign-post for those completed a task to follow the already created process flow.

Another advantage involves the introspective examination of a process which occurs while a flow is being drawn. Obvious problems, duplication in work or inconsistency between processes can be easily identified and solutions put in place to improve overall efficiency within an organisation.

Process flows completed with the involvement of each of the work package themes can be found in Appendix 2 accompanying this paper.

## 6. Conclusion

This document and its supporting appendices dictates the frameworks of data processing present at each partner institution in the Compass project. This includes but is not limited to reasoning and ways of collecting different types of data in each work package theme as well as the ethical and legal concerns which exist therein.

It provides significant insight on the detail of the operations of each institution's individual processes, the resource requirements, the methodology behind selection, preservation and sharing of project data outputs.

While there are commonalities between methods, there are also divergences. These are not viewed as obstacles to integration or the success of the Compass project. To the contrary, the differences identified are considered fantastic opportunities for each institution to learn from each other and raise the standards for everyone involved.

## 7. References

Leadbetter, A., Carr, R., Flynn, S., Meaney, W., Moran, S., Brophy, L., Lyons, K., Stokes, D., Thomas, R. in press. Implementing a Data Management Quality Management Framework at the Marine Institute, Ireland. Earth Science Informatics

# Appendix 1: Data Management Plans

## Marine Institute: Fixed Mooring Mace Head

### Data Collection

#### **What data will you collect or create?**

Text files aggregated by a Campbell logger on board a remote vehicle (buoy) based at Mace Head, Co. Galway. Data collection activity is funded by the Compass Project (INTERREG) under the auspices of SEUPBB in association with AFBI, IFI, MSS and SAMS.

#### **How will the data be collected or created?**

Collected by the sensors on-board the buoy, logged and timestamped by a Campbell logger. Transmitted via SMTP to several different email addresses for redundancy (MI mail, Gmail, Outlook). That data is parsed and updates an ever increasing table in PostgreSQL.

### Documentation and Metadata

#### **What documentation and metadata will accompany the data?**

No named metadata standards are enforced at data collection but there are plans to adapt Inspire standards and the P01 vocabulary to the data at ingestion to database (pending other project's work).

### Ethics and Legal Compliance

#### **How will you manage any ethical issues?**

A strict data sharing agreement has been drafted and is under legal review by the different Compass partners.

#### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

The data is owned by the originating organisation. Usage restrictions are explicitly covered in the data sharing agreement.

### Storage and Backup

#### **How will the data be stored and backed up during the research?**

Data is stored on-board the buoy in standard FAT-32 memory but also on an internal shared mail server. It is also backed up in two external places, as Azure blob storage (using the MI's Azure account) and an external mail client (Gmail). The data processor is charged with ensuring that these backups are current and maintained. In the case of a major system failure resulting in a loss of data, services can be rebuilt from backups or directly from source.

#### **How will you manage access and security?**

Access to the data is managed by the data steward in conjunction with IT ops.

### Selection and Preservation

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

Presently, a backup of the full database is retained in both backup locations indefinitely. This poses a security risk for personal information. The data in the viewer is truncated each time a new update is received.

#### **What is the long-term preservation plan for the dataset?**

The intention is that in the event of the project ending without a plan of continuation, that the legacy data will be added to the archive of all Marine Institute buoy data.

### Data Sharing

#### **How will you share the data?**

Internally, the data is shared via an internal ERDDAP server and also through direct access to the email server.

The data will be shared publicly via an ERDDAP portal (likely to be the main MI ERDDAP) along with a project specific ERDDAP instance which will federate data from all partners in the project. This data will be used to create separate visualisations on the Compass portal.

#### **Are any restrictions on data sharing required?**

No explicit restrictions exist though time lapses are expected to guarantee the highest levels of data quality are maintained.

### Responsibilities and Resources

#### **Who will be responsible for data management?**

The data management team, namely the Compass data manager and Data team lead hold end responsibility for the Compass data.

#### **What resources will you require to deliver your plan?**

Staff to carry out data processing. DBMS software to manage data. Server space for storage and publishing. DNS and certificates for public facing sites.

## SAMS: Mooring, Receiver and Glider

### Data Collection

#### **What data will you collect or create?**

Text files from broadband recorders and sensors in different moorings and gliders. Netcdf files with QC glider missions and model output data.

#### **How will the data be collected or created?**

Real time data collected by the gliders is recovered via Iridium, acoustic data is collected when sensors are recovered and then the data is copied to a hard drive.

### Documentation and Metadata

#### **What documentation and metadata will accompany the data?**

Netcdf files will contain metadata. Tethys database also provides a way to add metadata to acoustic data but it's not implemented yet. At the moment, a spreadsheet is used to record metadata and it's saved along the data in an internal network drive.

### Ethics and Legal Compliance

#### **How will you manage any ethical issues?**

A strict data sharing agreement has been drafted and is under legal review by the different Compass partners.

#### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

The data is owned by the originating organisation. Usage restrictions are explicitly covered in the data sharing agreement.

### Storage and Backup

#### **How will the data be stored and backed up during the research?**

Glider data is stored in a base station, preprocessed and backed up in two internal data servers (Windows and Linux based). These data servers also back up acoustic data and modelling outputs. Both data servers are weekly and incrementally backed up using VEEAM in an onsite and offsite backup VEEAM archive.

#### **How will you manage access and security?**

All access to data stored on SAMS servers and in the cloud (One Drive) is managed by Access Control Lists (ACLs) whose membership is managed through set procedures and only by permission of the data owners by requests using the Service Desk.

### Selection and Preservation

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

Most of the generated data has a long-term value, therefore their backups and original data is retained, shared and preserved indefinitely in our servers.

**What is the long-term preservation plan for the dataset?**

In the case of acoustic data, long-term archiving is not planned since the amount of data is big. A clipped record using tethys database will reduce the size of the files and this could be archived indefinitely. Glider data is going to be archived in BODC.

**Data Sharing**

**How will you share the data?**

The data is shared internally in network drives and via THREDDS and ERDDAP with partners. We have a catalogue that gathers COMPAS project outputs.

**Are any restrictions on data sharing required?**

There are no restrictions on the data required by COMPASS.

**Responsibilities and Resources**

**Who will be responsible for data management?**

The SAMS Data manager and ICT department at SAMS will be responsible for COMPASS data management.

**What resources will you require to deliver your plan?**

Staff to carry out data processing. Server space for storage and publishing. DNS and certificates for public facing sites.

## AFBI: Belfast Pile 8 Moored EXO2

### Data Collection

#### What data will you collect or create?

Primary Latitude (54.66348°) and Longitude (-5.83922°) of the moored instrument, system generated 10 digit Unix timestamp, date (mm/dd/yyyy) and time (hh:mm:ss) of received telemetered data, along with EXO2 instrument data: date (ddmmyy), time(hhmmss), system battery (volts) and oceanographic parameters originating from its sensors: sea temperature (degrees centigrade to two decimal places), Depth (meters), Conductivity (uS/cm), Specific Conductivity (uS/cm), Salinity (psu), Turbidity (NTU), Chlorophyll ( $\mu\text{g/L}$ ), Fluorescence (percentage FS), Dissolved Oxygen Concentration (mg/l) and Dissolved Oxygen % Saturation, and Blue-green algae pigment from BGA\_PE sensor (RFU).

The sampling (scan rate) and data transmission interval is once every 20 minutes, with the first reading being 12:00:00 on 15<sup>th</sup> October 2019, programmed just after the instrument was deployed.

Note there are four pre-deployment records in the download, dated 14/10/2019, and should not be included as oceanographic data.

It is expected the telemetered data volume to be less than 5 MB per year but requiring 8 MB per year storage on SQL Server.

#### How will the data be collected or created?

The data is collected by an EXO2 Sonde in moored mode (rather than profiling mode) and set to broadcast every twenty minutes via mobile broadband through a Storm-3-02 data Logger with GSM-3G Cellular Modem to a commercial platform (StormCentral.Waterlog.com). From that data, other measurements are calculated by the data processing software: Practical Salinity, Dissolved Oxygen Concentration, Fluorescence, Density and Depth of the sea water.

From the StormCentral website, the data is manually downloaded to the Afbi file system, to be cleaned and transformed using stored procedures within a new conversion SQL Server database and output to the new AFBI AquaticMeasurement, a SQL Server database being developed as part of the COMPASS project.

Note: The EXO2 was a replacement for the SBE19plus that was recovered in September 2019 and may be swapped back at some stage. The data from the SBE was being telemetered to TriskeData (not StormCentral). For more information on the SBE, see the DMP Moored Instrument Data - Belfast Lough Pile8 SBE CTD.

### Documentation and Metadata

#### What documentation and metadata will accompany the data?

It is intended that information, or links to information, accompanying the data will be stored as metadata within the Data Catalogue. The referenced documents may be located on the internet (such as manufacturer specific information) or in the Afbi file system (such as imagery and certificates).

The documentation and metadata to accompany the data is determined by the MEDIN Metadata standards and BODC Metadata submission requirements. In this case, for the recording of moored

instrument data, the MEDIN\_Moored\_Ocean data guideline from BODC is applicable, requiring: General Metadata - project information and survey report;  
Detail Metadata – sampling methods; moored instrument details including location, deployment, service, calibration, maintenance and recovery details; SOPs: No. Marism3 “Deployment of Moored Equipment from Navigation Pile Platforms (Belfast Lough and Lough Foyle)”, No. Marism5 “Deployment of Moored Equipment of from small Boats”, No Marism2 “Procedures For Small Boat Operation In Lakes And Rivers”;  
Data - Moored instrument profile data consisting of the parameters described in this DMP under Data Collection.

In addition, Discovery Metadata will be provided as per the MEDIN Discovery Metadata format – aimed at allowing the non-informed user to discover datasets and will contain a large range of data types that are in turn covered by a range of data guidelines.

### Ethics and Legal Compliance

#### **How will you manage any ethical issues?**

Not applicable.

#### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

Not Applicable – Public Access

### Storage and Backup

#### **How will the data be stored and backed up during the research?**

Data is stored in a SQL Server database and regularly backed up to tape.

#### **How will you manage access and security?**

The data is made accessible via bespoke and off the shelf software. The data falls under the AFBI security umbrella.

### Selection and Preservation

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

All data downloaded from StormCentral is retained and made available.

#### **What is the long-term preservation plan for the dataset?**

The data will be retained indefinitely.

### Data Sharing

#### **How will you share the data?**

Users can find out about this data via the AFBI data catalogue. The data is available to anyone who requests it and is available as soon as it has been processed. It will also be discoverable on BODC through discoverable metadata and through an ERDDAP gateway being implemented as part of the INTERREG COMPASS project.

**Are any restrictions on data sharing required?**

There are no restrictions on sharing this data.

**Responsibilities and Resources**

**Who will be responsible for data management?**

Moored Instrument Data (Oceanographic Services) Data Steward.

**What resources will you require to deliver your plan?**

Once the Data Catalogue is in production, data owners will be responsible for creating and maintaining their own metadata and DMPs.

Disk storage is the primary resource requirement to cope with the expanding database over time.

## AFBI: Belfast Pile 8 Moored CTD

### Data Collection

#### **What data will you collect or create?**

Date, time, latitude, longitude, volts and oceanographic parameters originating from sensors on the Seabird SBE 19plus instrument coupled with atmospheric parameters from a Li-COR instrument and Airmar instrument, collected every 20 minutes between 01/01/2017 and 17/09/2019 from the Pile 8 Moorings in Belfast Lough.

Data collected is from the Seabird CTD, Turbidity, Chlorophyll and Dissolved Oxygen sensors. The following data is from calculations using the parameters generated from the CTD and other sensors: Practical Salinity, Dissolved Oxygen Concentration, Sound Velocity, Density and Depth of the sea water.

The data volume is small: a manual download from the Triskeldata website of six years data produced a 23mb Excel Spreadsheet requiring less than 8 MB per year storage on SQL Server.

#### **How will the data be collected or created?**

The data is collected by a Seabird SBE19plus in moored mode (rather than profiling mode) and telemetered through a DBT-3 GPRS data logger / modem to a commercial platform (triskeldata.com). From that data, other measurements are calculated by the Seabird Data Processing software: Practical Salinity, Dissolved Oxygen Concentration, Sound Velocity, Density and Depth of the sea water. The sampling (scan rate) and data transmission interval is once every 20 minutes.

From the Triskeldata website, the data is manually downloaded to the Afbi file system, to be cleaned and transformed using stored procedures within a new conversion SQL Server database and output to the new AFBI AquaticMeasurement, a SQL Server database being developed as part of the COMPASS project.

Note: The SBE19plus has been recovered and replaced by an EXO. The data from that instrument is being telemetered to StormCentral and not TriskelData. For more information on them, see the DMP Moored Instrument Data - Belfast Lough Pile8 EXO CTD. The Li-COR and Airmar instruments are in-situ with their parameters still being transmitted to the TriskelData website.

### Documentation and Metadata

#### **What documentation and metadata will accompany the data?**

It is intended that information, or links to information, accompanying the data will be stored as metadata within the Data Catalogue. The referenced documents may be located on the internet (such as manufacturer specific information) or in the Afbi file system (such as imagery and certificates).

The documentation and metadata to accompany the data is determined by the MEDIN Metadata standards and BODC Metadata submission requirements. In this case, for the recording of moored instrument data, the MEDIN\_Moored\_Ocean data guideline from BODC is applicable, requiring: General Metadata - project information and survey report;

Detail Metadata – sampling methods and moored instrument details including location, deployment, service, calibration, maintenance and recovery details;

Data - Moored instrument profile data consisting of the parameters described in this DMP under Data Collection.

In addition, Discovery Metadata will be provided as per the MEDIN Discovery Metadata format – aimed at allowing the non-informed user to discover datasets and will contain a large range of data types that are in turn covered by a range of data guidelines.

### Ethics and Legal Compliance

#### **How will you manage any ethical issues?**

Not applicable.

#### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

Not Applicable – Public Access

### Storage and Backup

#### **How will the data be stored and backed up during the research?**

Data is stored in a SQL Server database and regularly backed up to tape.

#### **How will you manage access and security?**

The data is made accessible via bespoke and off the shelf software. The data falls under the AFBI security umbrella.

### Selection and Preservation

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

All data downloaded from triskeldata is retained and made available.

#### **What is the long-term preservation plan for the dataset?**

The data will be retained indefinitely.

### Data Sharing

#### **How will you share the data?**

Users can find out about this data via the AFBI data catalogue. The data is available to anyone who requests it and is available as soon as it has been processed. It will be discoverable on BODC through discoverable metadata and through an ERDDAP, implemented as part of INTERREG COMPASS project.

#### **Are any restrictions on data sharing required?**

There are no restrictions on sharing this data.

### Responsibilities and Resources

#### **Who will be responsible for data management?**

Moored Instrument Data (Oceanographic Services) Data Steward.

**What resources will you require to deliver your plan?**

Once the Data Catalogue is in production, data owners will be responsible for creating and maintaining their own metadata and DMPs.

Disk storage is the primary resource requirement to cope with the expanding database over time.

## MSS: Loch Ewe

### Data Collection

#### **What data will you collect or create?**

Weekly water samples are collected for temperature salinity nutrients, phytoplankton and zoo-plankton are collected as part of Marine Scotland's Scottish Coastal Observatory.

Specifically for the COMPASS project, a mooring is deployed in Loch Ewe as well. The strong mooring consists of a CTD sensor (MicroCat), and current meters (SeaGuard) at two depth levels, near sea surface, and close to seabed.

Two meteorological weather stations are deployed onshore nearby the mooring to collect information on local weather conditions.

In the future parts of the mooring are intended to be replaced with a data buoy to facilitate transmission or near real-time data.

A Contros HydroC PCO<sub>2</sub> system on a separate mooring is deployed in Loch ewe as well. The deployment is close to the standard sampling site and oceanographic mooring.

#### **How will the data be collected or created?**

Weekly water sampling at the Scottish Coastal observatory site follows the data collection descriptions provided in a set of decadal review reports for the coastal monitoring sites operated by Marine Scotland<sup>1</sup>

The instrumentation deployed for the COMPASS project (PCO<sub>2</sub> system, mooring and weather stations) are visited and data are downloaded every 3 months. Data are transferred onto portable media for return to the institute.

Data are subsequently subject to different QC processes at different intervals.

The weekly Scottish Coastal Observatory water sample data are collated in a one year cycle which is then subjected to a QC process based on a series of Matlab scripts the utilise the prevailing conditions and patterns from previous years in the time series to identify data points to be queried. After investigation of all raised points in conjunction with all available data sources, data are assigned quality flags (aligned with SEADATANET measured qualiter flags L20 vocabulary) before publication.

The mooring data undergo a similar operation at each 3 month interval in terms of using scripts to sense check the data combined with visual plotting of the data. Quality flags are not currently added to individual mooring data points.

---

<sup>1</sup> E Bresnan, K Cook, J Hindson, S Hughes, J-P Lacaze, P Walsham, L Webster and W R Turrell (2016) The Scottish Coastal Observatory 1997-2013. Part 1-3. Scottish Marine and Freshwater Science Vol 7 No 26. DOI: 10.7489/1881-1 <https://doi.org/10.7489/1881-1>

The weather station data undergo a basic sense check, but the work package is developing a more detailed QC plan for this data, so the data management plan will likely be updated to reflect this over the course of the project.

The PCO<sub>2</sub> system data are currently plotted and undergo basic sense checks. A more detailed QC plan is being developed, but the device encountered a number of data logging reliability issues after a firmware upgrade, so there are limited useful data, and the process will be developed further.

## Documentation and Metadata

### **What documentation and metadata will accompany the data?**

Overall datasets for monitoring data in Loch Ewe are described in metadata records following UK MEDIN standard, housed in Marine Scotland's internal metadata catalogue. The MEDIN metadata standard is a superset of the UK GEMINI standard, so it INSPIRE and ISO 19115 compliant.

The Loch Ewe Scottish Coastal Observatory Site metadata record will likely be updated to reflect all available data at the site, but separate records may be created for each component if they are to be published independently or on separate platforms.

## Ethics and Legal Compliance

### **How will you manage any ethical issues?**

A strict data sharing agreement has been drafted and is under legal review by the different Compass partners.

### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

The data are owned by Marine Scotland, and are as such under Crown copyright. Data will be available on request under the UK Open Government license. Published data sets or products associated with this data source will be released under UK OGL as well.

## Storage and Backup

### **How will the data be stored and backed up during the research?**

Water sample data undergo well documented and mostly UKAS accredited processes for analysis, and data are stored in a dedicated Laboratory Information Management System (LIMS)

The LIMS system is housed within the Scottish Government Data Centre, which features a high degree of business continuity planning and full data Centre duplication. Regular backups and server replications are performed to safeguard the database that houses most of Marine Scotland's laboratory based analysis data.

The mooring data are stored in the existing Oceanographic data archive location on Marine Scotland's internal storage network.

The weather station data are currently stored in a dedicated COMPASS project share.

The storage array housing the mooring and weather station data is fully duplicated in a separate physical location. The primary storage array for passive acoustic monitoring data is in Aberdeen with overnight replication to Marine Scotland's site in Pitlochry acting as reciprocal disaster recovery locations. Each storage array further maintains a rolling snapshot policy to allow restoring changes of data on a 1 hour basis for 24 hours, daily basis for 2 weeks, weekly basis for 5 weeks, and monthly basis for 3 months. The array is also design with failure protection in mind, with full disk striping and hot spare capability to reduce the impact of a failing disk. The storage array is on a top tier support contract with a 4 hour disk replacement and engineer call out guarantee.

The PCO<sub>2</sub> system data are currently stored in a file share on the central Scottish Government IT system. It is subject to an automatic rolling retention policy of 6 years, but the data will be transferred and archived in a long term location as soon as the QC processes are finalised.

### **How will you manage access and security?**

Access to the network in general is managed through the Marine Scotland IT group. Only Scottish Government employees and security vetted contractors or placements have access to the network. In addition, each share within the network is limiting modification and writing access to be limited to the relevant business area. In the case of passive acoustic monitoring data, a dedicate share is established allowing access for modification only to the named custodians and Network Administrators. The LIMS system is a closed system where only users with recognised roles and function have access to the database.

### **Selection and Preservation**

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

All the data collected from instrumentation described in this data management plan or resultant from the analysis of collected water samples are considered of long-term value and retention for Marine Scotland. Forming part of the Scottish Coastal Observatory data, this data will be retained and published by Marine Scotland long term.

Other material developed as part of the analysis – such as derived data products, documents, presentations etc. will be retained for the duration of the project. Once the project closes, the content will be reviewed for reuse value. Data products created will be added to Marine Scotland's data archive, covered by metadata, and where practical it will also be published. Content from the project work area not deemed to have high reuse value will be moved to a project archive, where it is retained for a further 3 years before disposal.

#### **What is the long-term preservation plan for the dataset?**

The water sample results and the mooring data already have long term data preservation locations in place. As soon as they are committed to the system, they become part of operational long term data archives and systems.

For the weather station data, there is already an indication that this data will also be of long-term value, and it likely to be committed to the existing oceanographic data archive location once the QC steps and process is finalised.

Likewise, the data from the PCO<sub>2</sub> system will also be considered of long-term value both from a coastal monitoring and climate monitoring perspective, and will be committed to a long term archive location once the QC process is finalised.

## Data Sharing

### How will you share the data?

Internally, the data are accessible to analysts via network shares and via the LIMS systems.

Marine Scotland also submits oceanographic data to British Oceanographic Data Centre for long term archival on an annual cycle. Thus, the mooring data will be submitted to BODC at the end of each financial year (March).

Full resolution, quality flagged data and derived data products of a manageable volume (e.g. less than 10GB/dataset) will be made publicly available, either via a cloud deployed ERDDAP server, or via Marine Scotland's existing open data portal (<https://data.marine.gov.scot>).

### Are any restrictions on data sharing required?

No restrictions are anticipated for sharing this data.

## Responsibilities and Resources

### Who will be responsible for data management?

Marine Scotland as part of Scottish Government follows a tiered information management structure in terms of accountability. These roles are typically associated with specific posts in the Scottish Government, and stay with the role even if the person changes:

The Accountable Officer (AO) for Scottish Government is the permanent Secretary of Scottish Government. This is the role with overall responsibility for ensuring information risks are addressed and mitigated to an acceptable level

The Senior Information Risk Owner (SIRO) is the Directory General – Organisational Development and Operations. This role owns the overall information risk policy and assessment process.

Information Asset Owner for this project is the Deputy Director of Marine Scotland Science. This is the person who manages individual business areas in terms of data and information holdings risks and use, and communication with the senior risk owner.

On a day to day basis, the data management, long term storage, and metadata will be facilitated by the Marine Scotland Data Manager.

Finally, each inventoried and catalogued dataset in Marine Scotland will have assigned custodians. This is typically 2-3 persons from the business are with the right expertise to conduct the QC and analyse the data. They are responsible for completion, analysis, and publication of the data in collaboration with the data manager.

### What resources will you require to deliver your plan?

The deployment, logging and download of data will have been outlined in the work package for oceanographic data collection.

Storage capacity in Marine Scotland's storage of maximum 1-2TB for the COMPASS project's data logger outputs.

The existing resources for Loch Ewe sampling are already covered as part of the routine operation of the Scottish Coastal Observatory and require no additional or dedicated data management resource.

Business area and data management staff to complete inventory and metadata for the assets.

Public facing data publication mechanism – ERDDAP or Marine Scotland's Open data portal - or most likely a mixture of the two (e.g. landing pages via the Marine Scotland data portal, with option to dive deeper into larger datasets via ERDDAP).

Data management staff to register dataset on relevant open data access point(s), metadata preparation, and minting of DOI for data sets and products.

## MSS: Passive Acoustic Monitoring (PAM)

### Data Collection

#### **What data will you collect or create?**

Data loggers are deployed on moorings following the agreed sampling plan for coverage to the extent weather and ship time will allow. Each logger is recovered and downloaded every 4 months and subsequently redeployed. There are two types of data loggers for passive acoustic monitoring: broadband sound recorders and CPOD units.

#### **How will the data be collected or created?**

Raw acoustic recording data are downloaded from loggers and stored on portable drives or SD cards for transport to shore. Once back in the institute, data are stored on a high capacity storage array inside the institute's network for long term keeping. Raw data from passive acoustic recordings are high volume. And average year of deployments consume approximately 8TB capacity.

Following initial storage, the data will undergo QC in the form of validation of CPOD detections against broadband noise detections to test for false positives and negatives. Data files are also trimmed to exclude periods where the data loggers were not in the sea etc. Observations of any data quality issues will be flagged and noted in metadata.

The initial QC process is deliberately light-touch to avoid narrowing the potential reuse of the data for other analytical purposes.

### Documentation and Metadata

#### **What documentation and metadata will accompany the data?**

The overall dataset of passive acoustic monitoring is described in a metadata record following UK MEDIN standard, housed in Marine Scotland's internal metadata catalogue. The MEDIN metadata standard is a superset of the UK GEMINI standard, so it INSPIRE and ISO 19115 compliant.

In addition, a more detailed level of metadata for individual loggers is maintained in an Access database to allow analysts quick retrieval of relevant station information etc.

### Ethics and Legal Compliance

#### **How will you manage any ethical issues?**

A strict data sharing agreement has been drafted and is under legal review by the different Compass partners.

#### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

The data are owned by Marine Scotland, and are as such under Crown copyright. Data will be available on request under the UK Open Government license. Published data sets or products associated with this data source will be released under UK OGL as well.

### Storage and Backup

#### **How will the data be stored and backed up during the research?**

A full backup of the raw data and subsequently QC'ed data will be stored in Marine Scotland's dedicated storage array. Additional copies will be utilised for analysis by copying the master data to portal drives or local computers hard drives to facilitate speedier analysis and the avoidance of corruption of the master data.

The storage array is fully duplicated in a separate physical location. The primary storage array for passive acoustic monitoring data is in Aberdeen with overnight replication to Marine Scotland's site in Pitlochry acting as reciprocal disaster recovery locations. Each storage array further maintains a rolling snapshot policy to allow restoring changes of data on a 1 hour basis for 24 hours, daily basis for 2 weeks, weekly basis for 5 weeks, and monthly basis for 3 months. The array is also design with failure protection in mind, with full disk striping and hot spare capability to reduce the impact of a failing disk. The storage array is on a top tier support contract with a 4 hour disk replacement and engineer call out guarantee.

### **How will you manage access and security?**

Access to the network in general is managed through the Marine Scotland IT group. Only Scottish Government employees and security vetted contractors or placements have access to the network. In addition, each share within the network is limiting modification and writing access to be limited to the relevant business area. In the case of passive acoustic monitoring data, a dedicate share is established allowing access for modification only to the named custodians and Network Administrators.

### **Selection and Preservation**

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

The raw data, once QC'ed/trimmed is of long term value and reuse potential

The work package for passive acoustic monitoring is currently discussion potential cloud deployment of a shared database for compiling the project specific analysis of marine mammal detections. Once the outcome and plan of these discussions are completed, the data management plan will need to be updated to reflect this.

Other material developed as part of the analysis – such as derived data products, documents, presentations etc. will be retained for the duration of the project. Once the project closes, the content will be reviewed for reuse value. Data products created will be added to Marine Scotland's data archive, covered by metadata, and where practical it will also be published.

Content from the project work area not deemed to have high reuse value will be moved to a project archive, where it is retained for a further 3 years before disposal.

#### **What is the long-term preservation plan for the dataset?**

The raw data for passive acoustic monitoring is likely to have high reuse potential given that the range of acoustic recordings can support analytical work wider than the scope for this project. As such, it is highly likely that the Quality controlled data will be retained for the long term as part of an overall archive of passive acoustic monitoring. It is possible that data at some point could be placed in a "cold" storage location, where it would require time and money to bring it back to full operational status. Potential deep archiving like this will not occur until a minimum of 3 years after the conclusion of the project, when it will be subject to a disposal review. Typically, there is a recognised value in retaining the data for longer, especially for data that have high collection cost (e.g. ocean based sampling and equipment

deployment). The dataset is subsequently reviewed on a rolling 3 year basis after that. Currently, the disposal review is only binary (e.g. Keep in archive, or delete), but Marine Scotland is actively developing a roadmap for extension of data archives to cloud – which would facilitate deep archiving to be included in future disposal reviews.

## Data Sharing

### How will you share the data?

Internally, the data are accessible to analysts via network shared – one for the large scale data archive of data loggers, and one for a project work area.

Marine Scotland is also providing copies of the passive acoustic monitoring data to SAMS. Currently this is facilitated by transfer on disk. This is a pragmatic solution given the large volume of the involved in the transfer.

Derived data products of a manageable volume (e.g. less than 10GB) will be made publically available, either via a cloud deployed erddap server, or via Marine Scotland's existing open data portal (<https://data.marine.gov.scot>).

Raw data from passive acoustic monitoring is unlikely to be made directly available for download online. The balance between the cost of server space and the size of the audience of specialists that would want to retrieve the data is likely to make this cost-prohibitive in the long term. However, at the very minimum, metadata and a catalogue list of the data available will be published.

### Are any restrictions on data sharing required?

No restrictions are anticipated for sharing this data.

## Responsibilities and Resources

### Who will be responsible for data management?

Marine Scotland as part of Scottish Government follows a tiered information management structure in terms of accountability. These roles are typically associated with specific posts in the Scottish Government, and stay with the role even if the person changes:

The Accountable Officer (AO) for Scottish Government is the permanent Secretary of Scottish Government. This is the role with overall responsibility for ensuring information risks are addressed and mitigated to an acceptable level

The Senior Information Risk Owner (SIRO) is the Directory General – Organisational Development and Operations. This role owns the overall information risk policy and assessment process.

Information Asset Owner for this project is the Deputy Director of Marine Scotland Science. This is the person who manages individual business areas in terms of data and information holdings risks and use, and communication with the senior risk owner.

On a day to day basis, the data management, long term storage, and metadata will be facilitated by the Marine Scotland Data Manager.

Finally, each inventoried and catalogued dataset in Marine Scotland will have assigned custodians. This is typically 2-3 persons from the business are with the right expertise to conduct the QC and analyse the

data. They are responsible for completion, analysis, and publication of the data in collaboration with the data manager.

**What resources will you require to deliver your plan?**

The deployment, logging and download of data require ships operations as will have been outlined in the work package for passive acoustic monitoring.

Storage capacity in Marine Scotland's storage of up to approximately 20TB for the COMPASS project's passive acoustic monitoring.

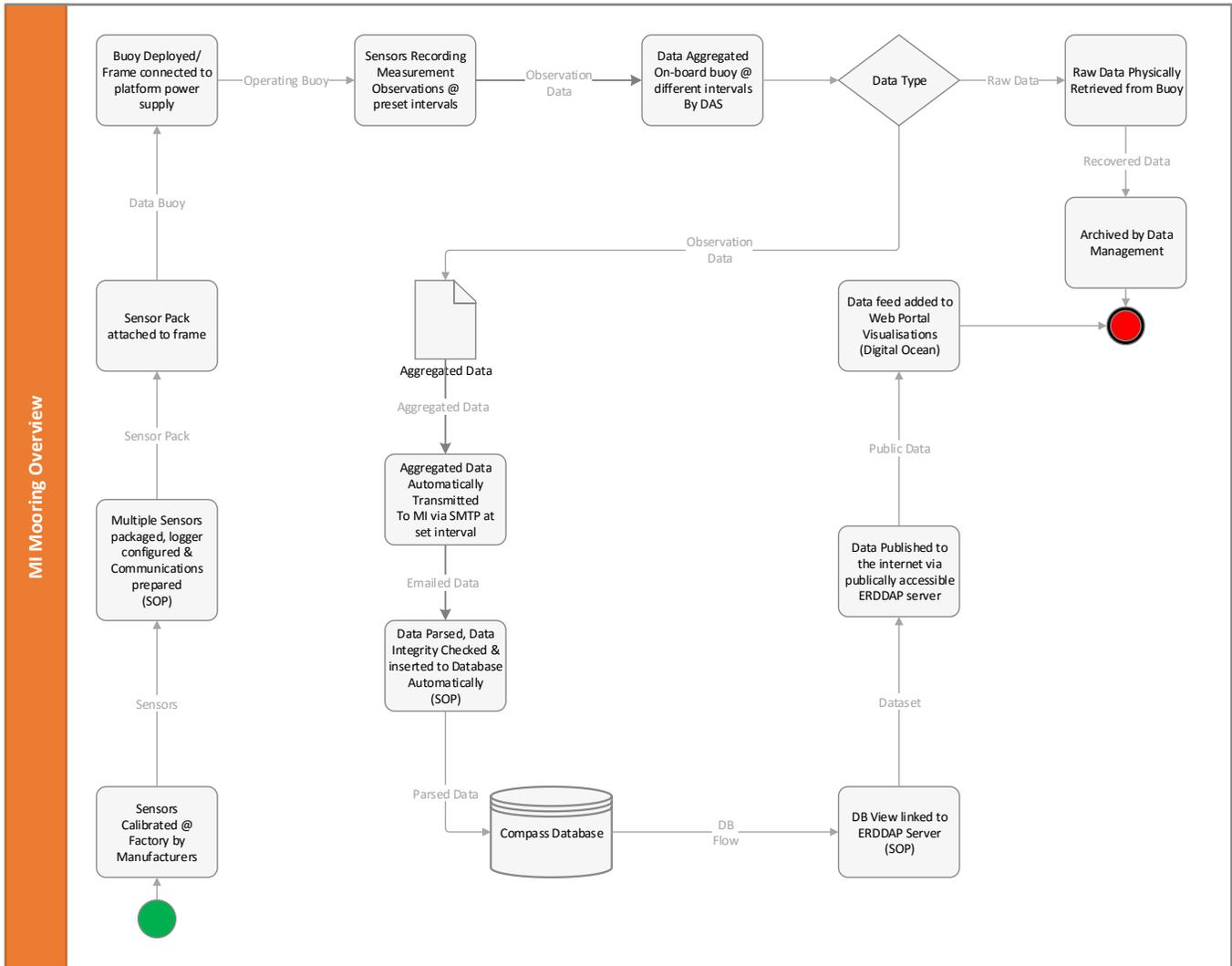
Business area and data management staff to complete inventory and metadata for the asset

Public facing data publication mechanism –erddap, Tethys database, or Marine Scotland's Open data portal - to be determined as the work package plans for Tethys database are finalized.

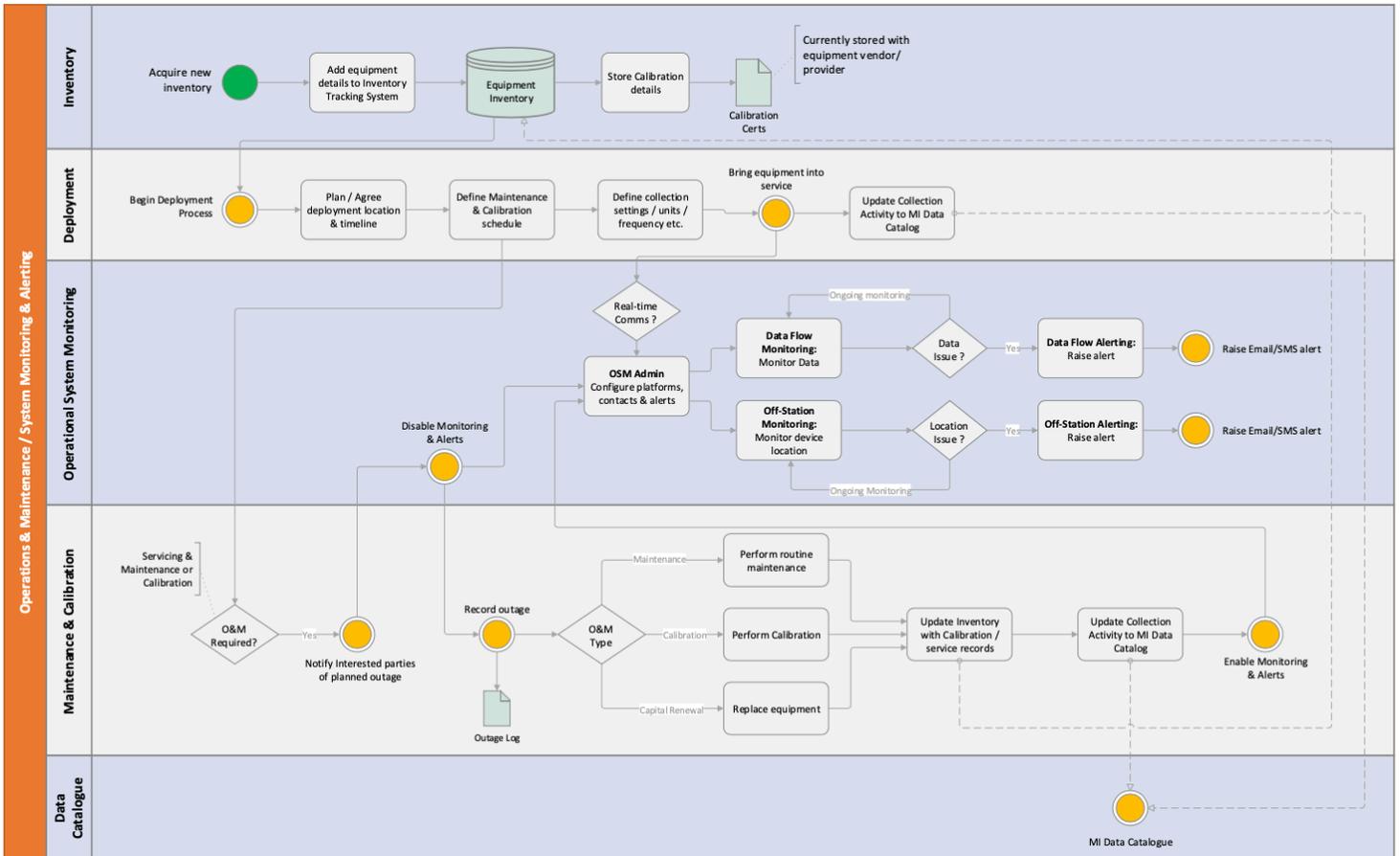
Data management staff to register dataset on relevant open data access point(s), metadata preparation, and minting of doi for data sets and products.

# Appendix 2: Process Flows

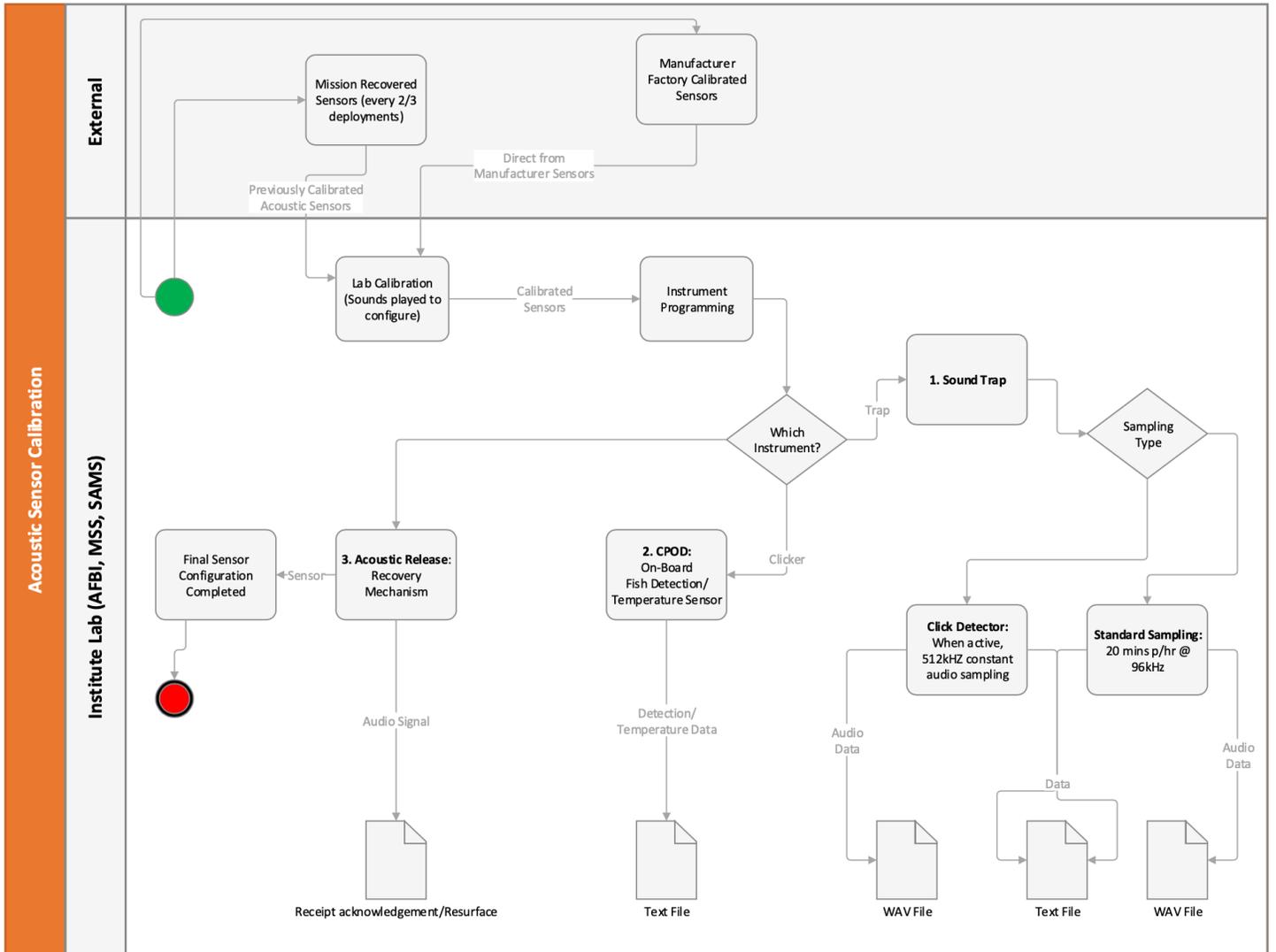
## Mooring Overview



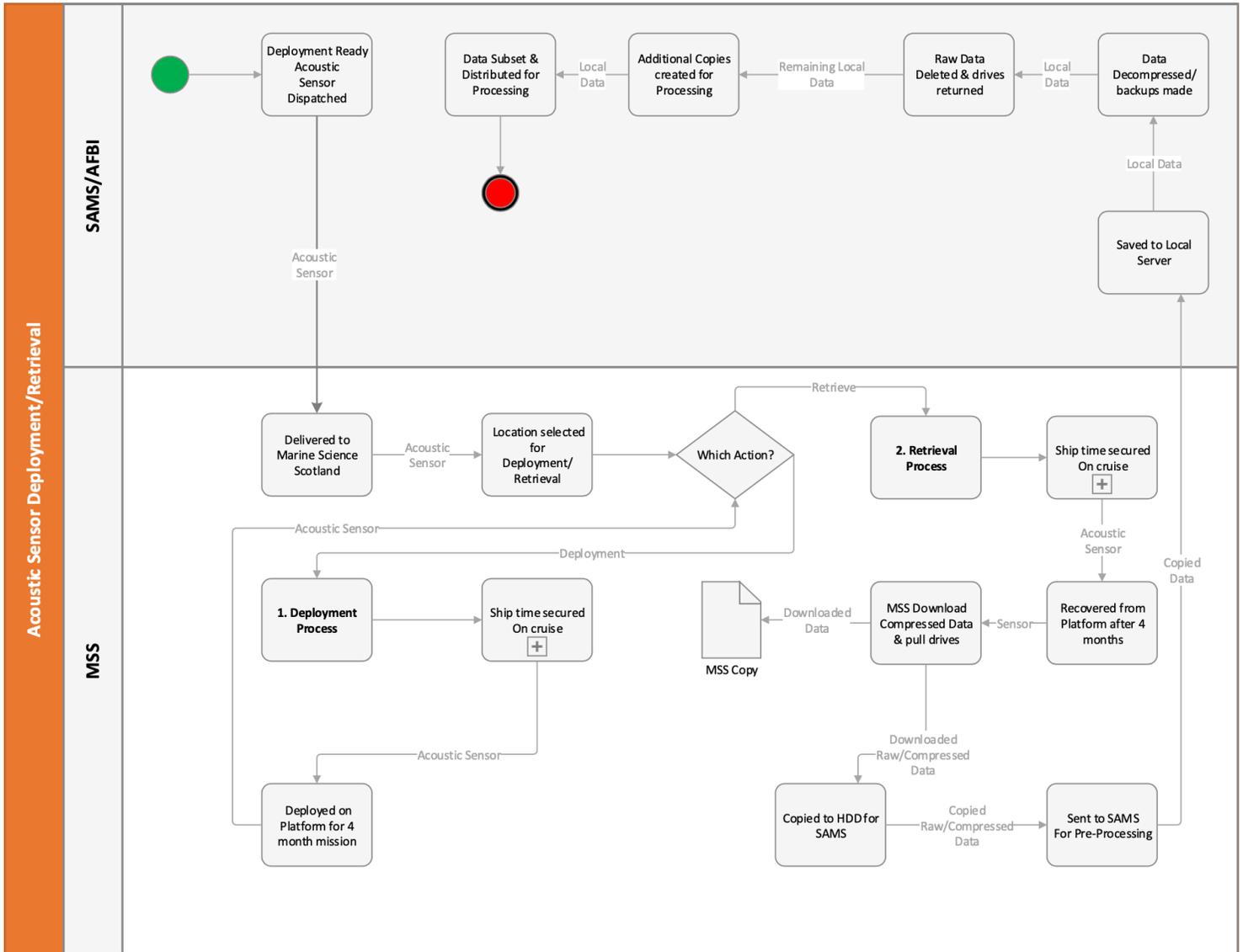
# Mooring Operations and Maintenance



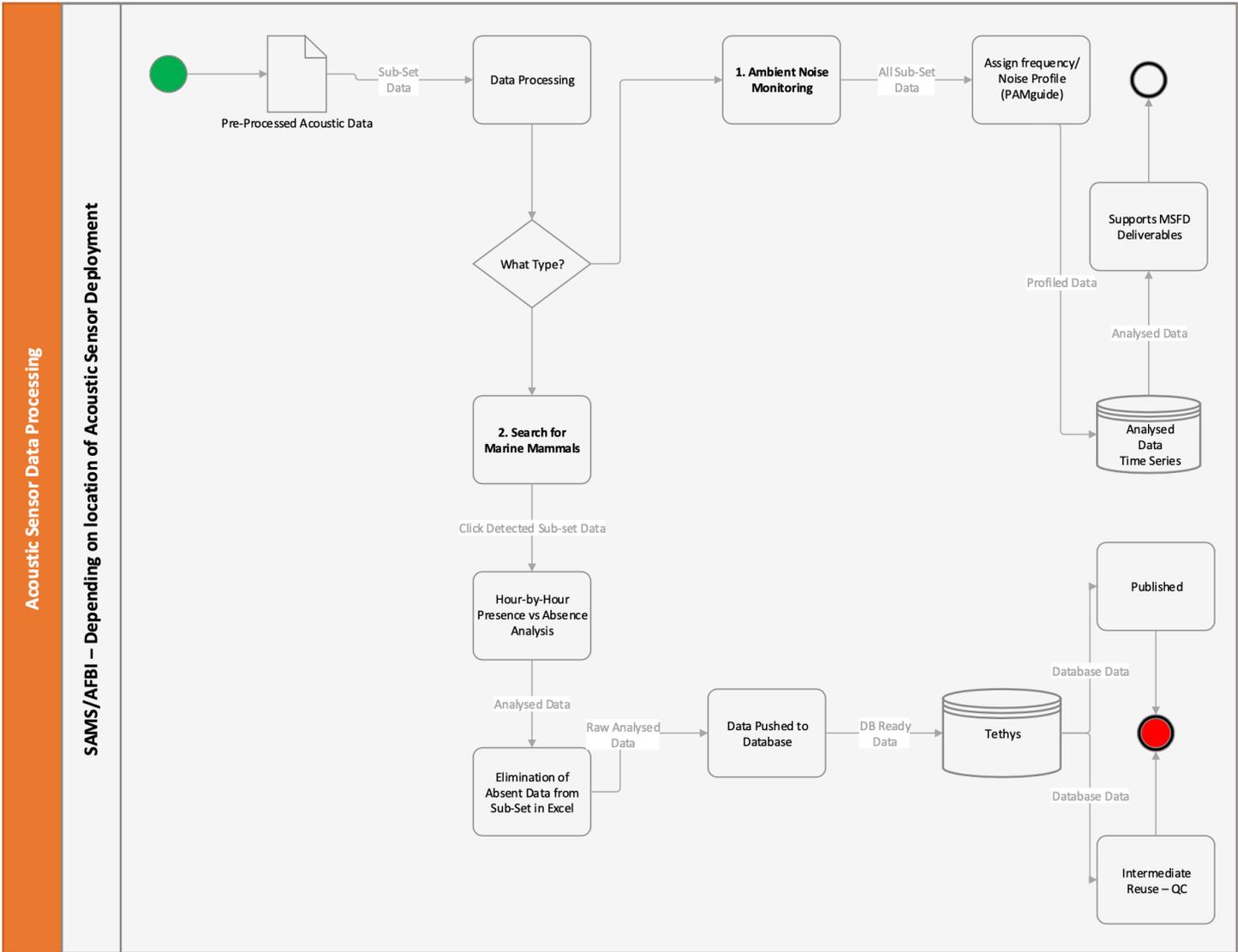
# Acoustic Sensor Calibration



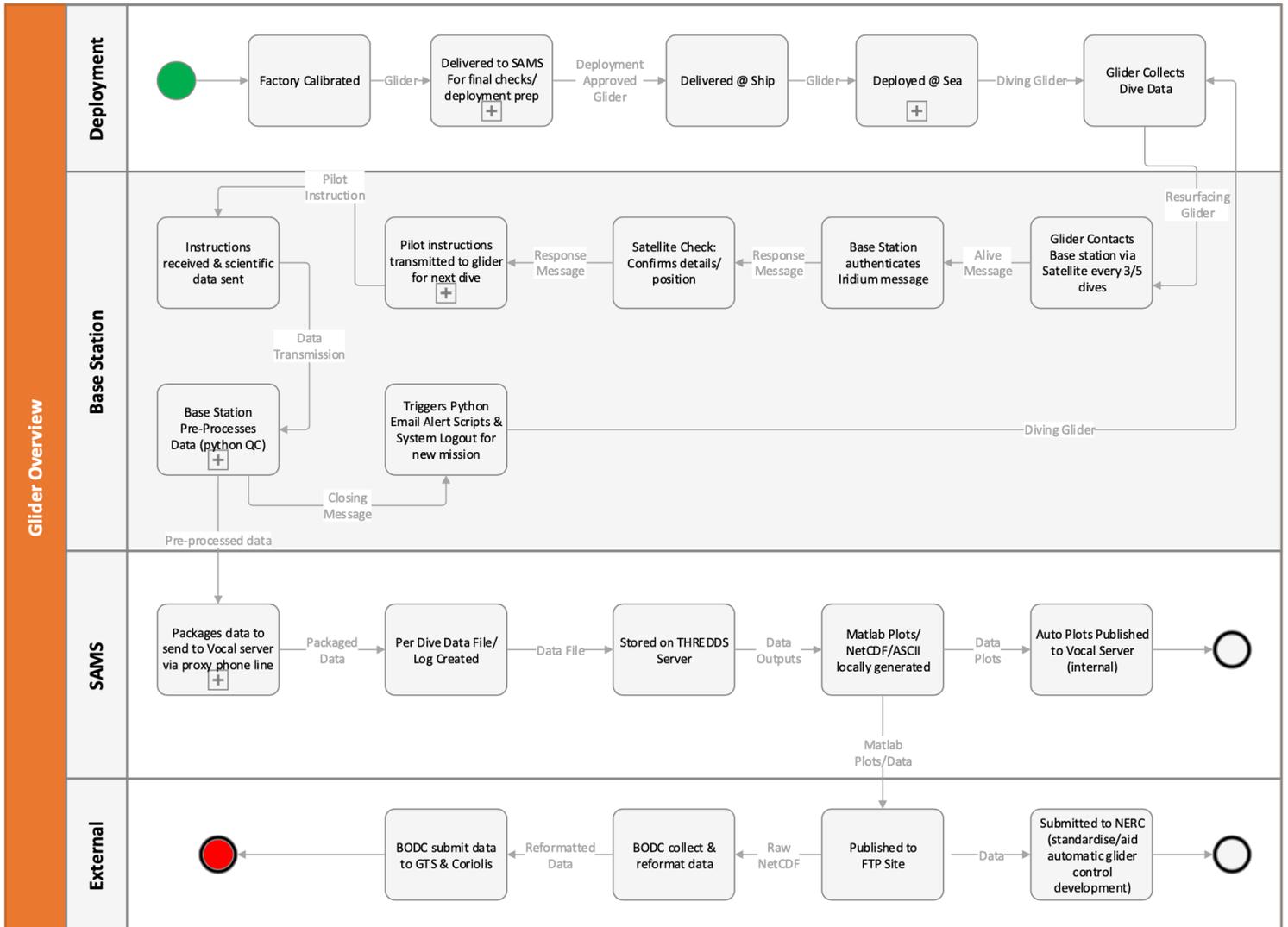
# Acoustic Sensor Deployment and Retrieval



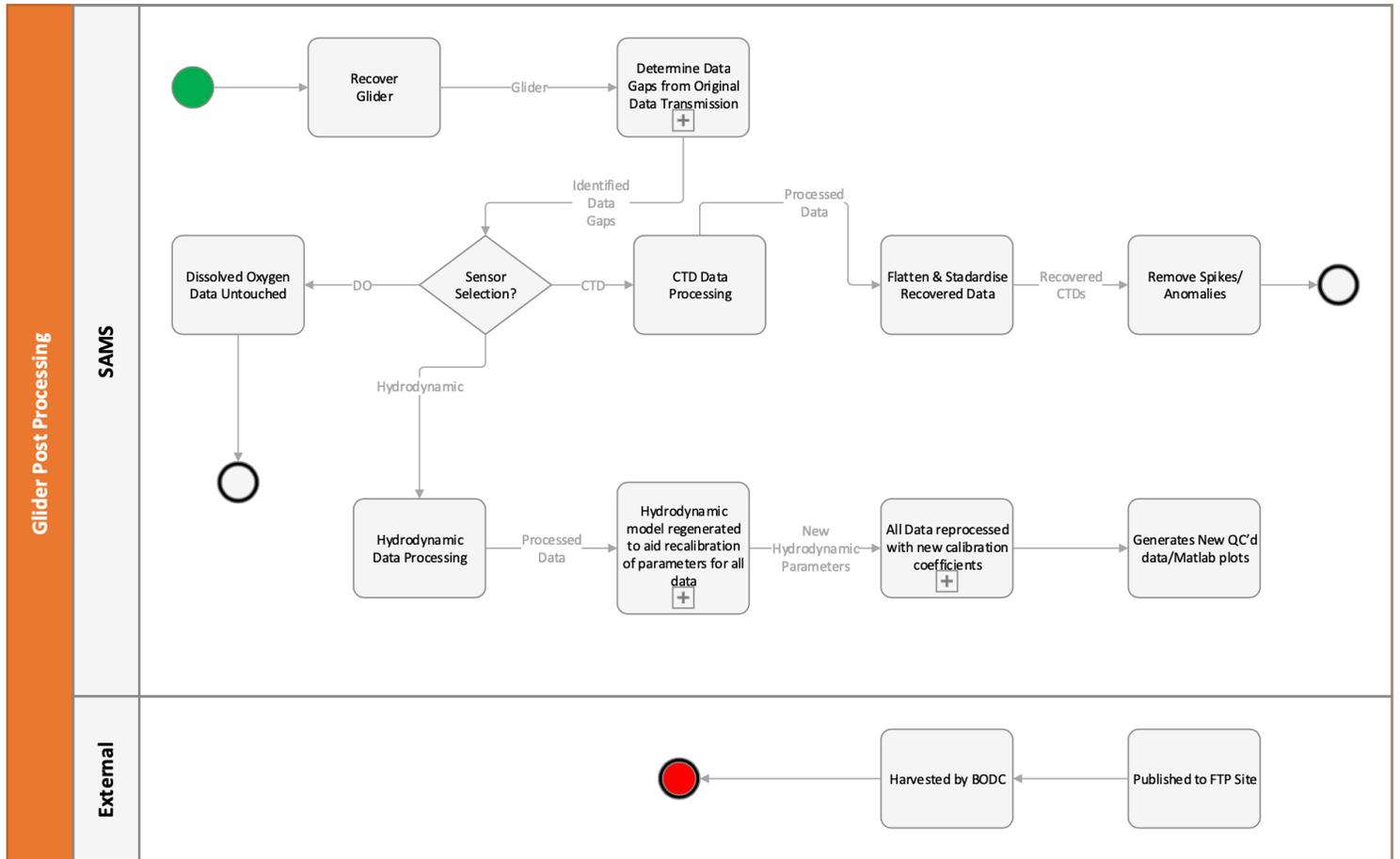
# Acoustic Data Processing



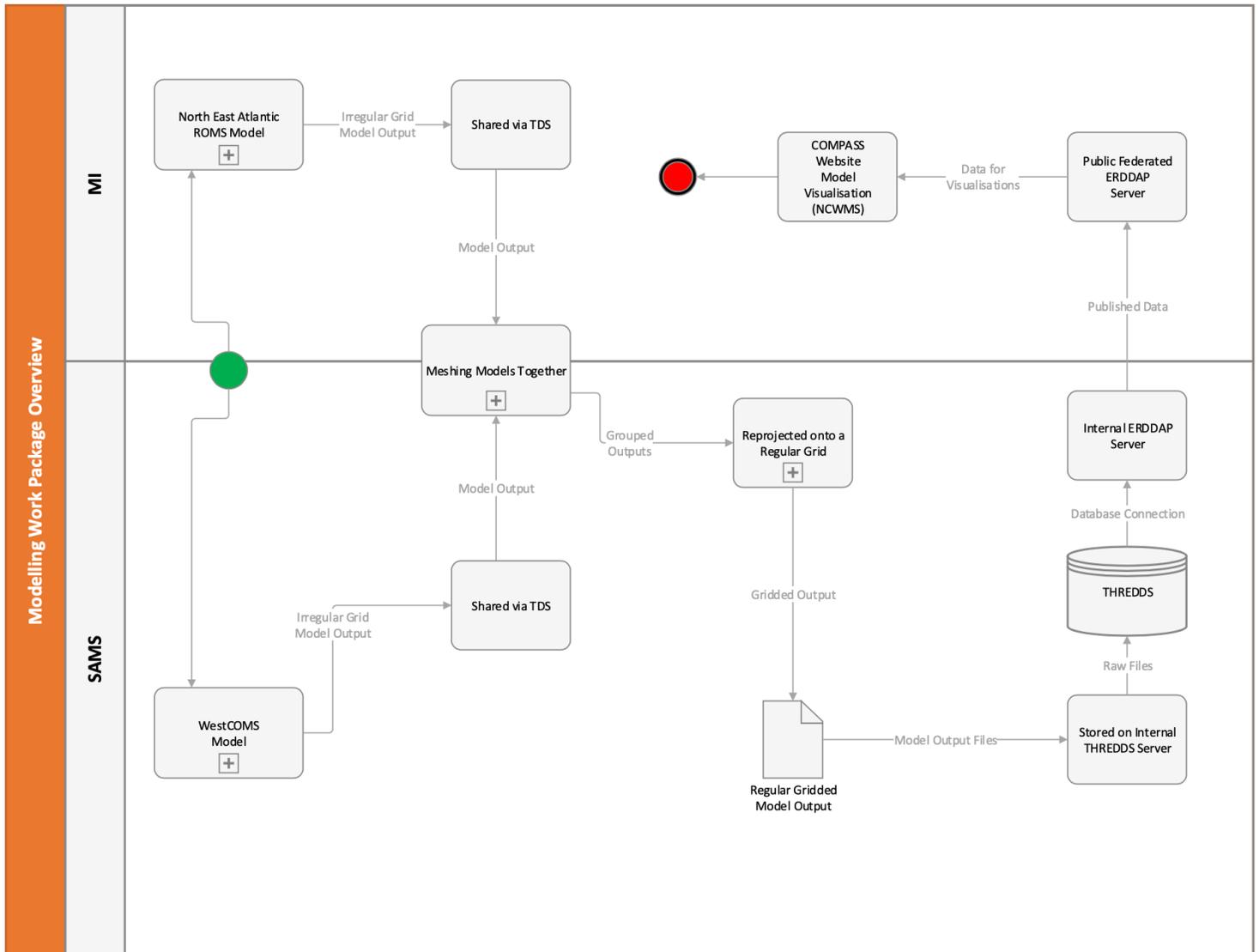
# Glider Overview



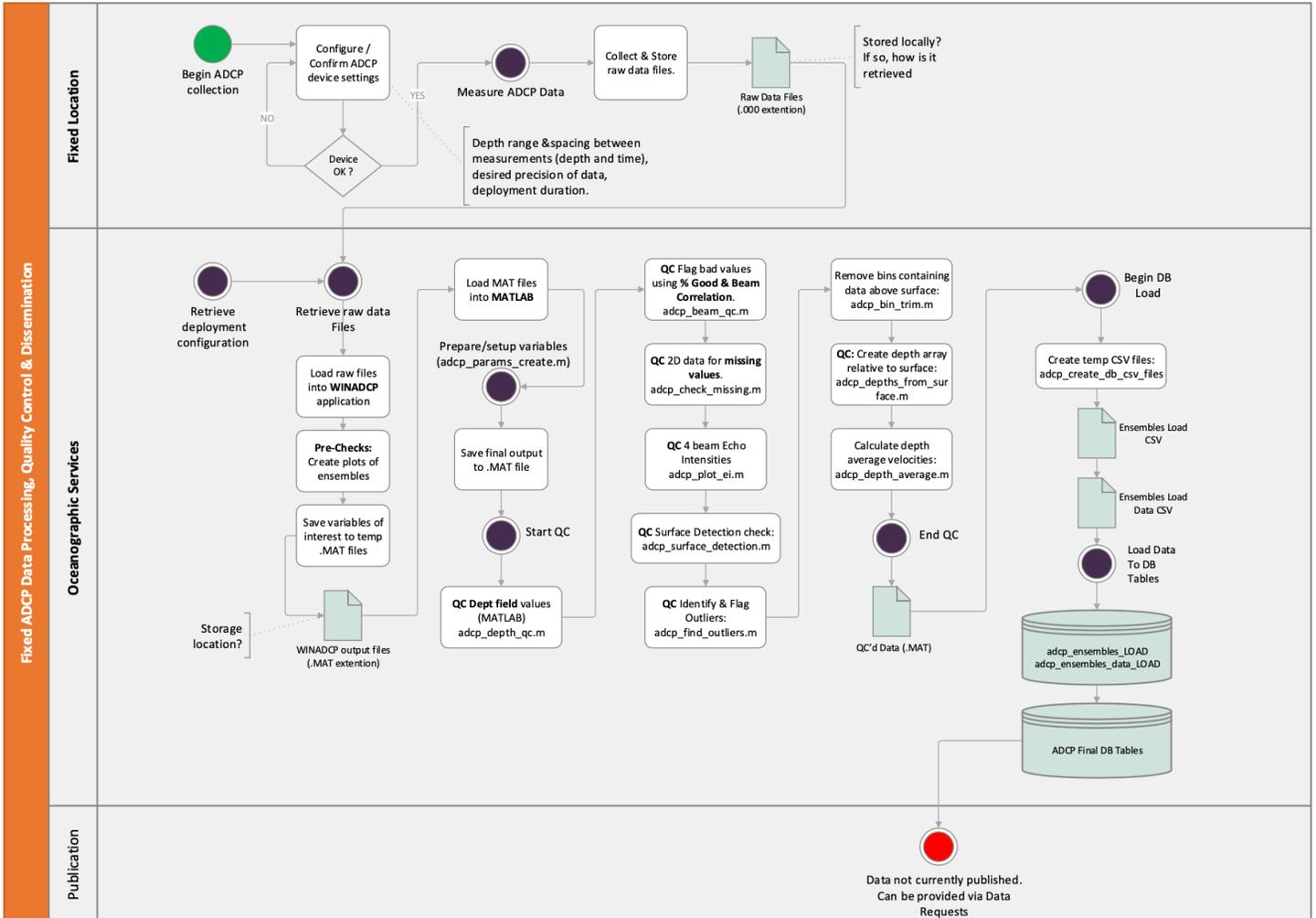
# Glider Processing



# Modelling Overview



# ADCP Processing



# ERDDAP Publishing

