AODN Underway Data Publication Plan

EIF023 Publication of Data from National Research Vessels

May 2010

Miles Jordan, Australian Antarctic Division
Siddeswara Guru, CSIRO Marine & Atmospheric Research

Contents

1. Introduction

This technical specification describes the processes that will be used by the Australian Antarctic Division (AAD) and CMAR (CSIRO Marine and Atmospheric Research) to publish underway voyage data in near-real time to the Australian Oceans Data Network (AODN) and the Australian Research Data Commons (ARDC). **Error! Reference source not found.** provides a conceptual diagram to give a high level overview of the implementation.

The specification has been developed specifically for use by the AAD and CMAR vessels RV Aurora Australis and RV Southern Surveyor, respectively. This forms the deliverable for Milestone 2 of EIF023 Publication of Data from National Research vessels.

2. System Overview

The basic system involves data being captured on-board a research vessel. This data is then transmitted via satellite (or in the case of CMAR, 3G if available) to the headquarters of the relative organisation (AAD or CMAR) which acts as a facility node to the AODN.

The facility publishes metadata for the voyage, which is harvested by the AODN Metadata Entry and Search Tool (MEST), and also makes the data available via a WMS.

The AODN Portal is configured to access WMS layers from the facility, which enables a portal user to view the vessel track from the underway dataset on the portal map. If the user clicks the vessel track, a popup is displayed that shows information relating to the specific point on the map, as well as providing links to the full underway dataset (ie. all points for the voyage) in various formats, and links to the metadata records that relate to the voyage.

After the metadata has been published to the AODN MEST, the ARDC can harvest the metadata and in turn provide access to the data in the formats that are provided by the facilities.

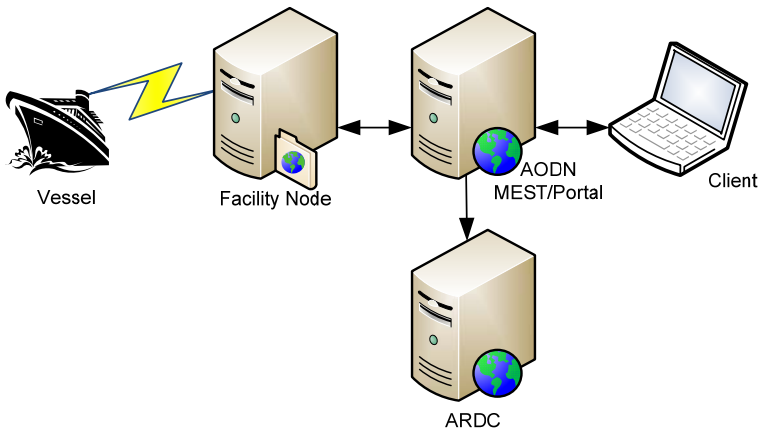

FIGURE 1. A HIGH-LEVEL OVERVIEW OF THE SYSTEM DATA FLOW

The various elements that make up this system are outlined in more detail below.

3. The AODN Portal

The AODN portal system mainly consists of a web application, a MEST metadata catalogue, and various WMS servers. In order to publish data to the AODN portal, the facility nodes must provide at a minimum harvestable metadata records (see section 5, Metadata) and appropriately configured, public WMS services.

The following diagram shows the relationship between the AODN portal and the facility nodes.

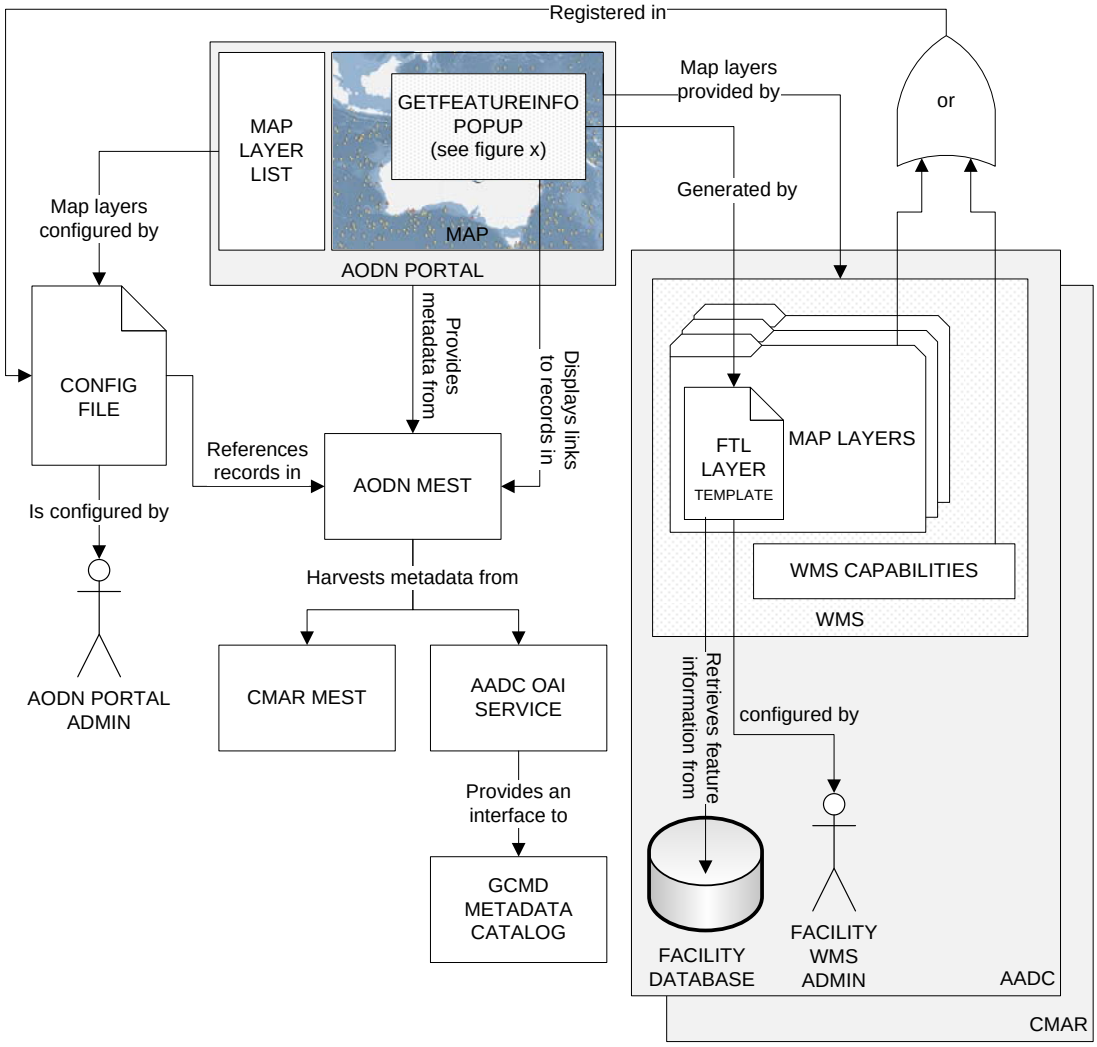

FIGURE 2. THE RELATIONSHIP BETWEEN THE AODN PORTAL AND FACILITY DATA AND METADATA

4. Configuration

In order for facility data and metadata to integrate into the AODN, various elements of the system, both at the facility and AODN, must be configured by technical personnel. Configuration of these elements must be maintained over time in order for the system to function.

4.1 AODN Portal Configuration

The AODN portal must be configured by the AODN portal administrator in order for underway voyage datasets to be added to the portal as they occur. Each voyage should be published as an individual layer. There are two ways in which the portal can be configured for these layers; either by configuring each new voyage as a layer individually in the configuration file, or by allowing the Portal to parse a linked WMS capabilities document to discover available layers, which are automatically added to the portal configuration. These methods are discussed in section 0,
Service Configuration.

NOTE: Either method will enable each underway voyage to have its real-time data provided via the portal; however, ideally these layers should be able to be configured directly via an OGC CSW service query directed at the AODN MEST, which would enable a level of automation with regards to adding layers to the map and enable the Portal and MEST to function in a true service-oriented architecture. The current AODN-MEST architecture is not SOE architecture.

4.2 AODN MEST Configuration

The AODN MEST should be configured by the AODN portal administrator to harvest metadata from the individual metadata systems that are managed by both the AAD and CMAR.

4.2.1 Harvesting CMAR Metadata

CMAR will instantiate a local MEST in which to store metadata. This MEST instance should be configured to be harvestable by the AODN MEST at least every three hours, as this is the rate of data and metadata updates received from the CMAR vessel. The harvest configuration needs to be carried out by the AODN MEST administrator.

4.2.2 Harvesting AAD Metadata

The AODN MEST should also be configured to harvest AAD metadata from the OAI service that is provided by the AADC as an interface to the GCMD metadata catalogue that it maintains. Underway voyage metadata could be updated as frequently as every fifteen

minutes, so this should also be the rate at which the AODN MEST should query the AAD OAI service for new or updated, harvestable records. The harvest configuration needs to be carried out by the AODN MEST administrator.

4.3 Facility WMS Configuration

4.3.1 WMS Software

It is a requirement of the AODN portal that the facility nodes publish mappable data using a WMS that supports the *getFeatureInfo* request. The getFeatureInfo request is designed to return information relating to a feature based on where the user clicks on the map that makes the request. The AODN portal sends this request to a WMS server to retrieve and display feature information for a feature in a pop-up window.

The WMS specification states that the getFeatureInfo request is an optional WMS component, so it may not be included in all varieties of WMS software. With this in mind, Geoserver 2.0, Mapserver 5.6 or Deegree 2 are considerable options; however the WMS software that is recommended by the AODN portal administrators is Geoserver because the portal has been tested with it as a WMS server and it is known to work. Geoserver also supports using the FreeMarker Template Language (FTL) to generate compiled HTML getFeatureInfo responses to getFeatureInfo requests (see section 4.3.2.2 for more information on FTL templates), and AODN portal administrators can provide support for generating FTL templates.

It is this generated HTML that is shown inside of the pop-up window when a user clicks a feature on the map. The AODN portal administrators can provide support for creating FTL templates, and Geoserver is considered to be mature and stable software, which makes it the preferred choice.

4.3.2 Map Data and Feature Pop-ups (getFeatureInfo)

4.3.2.1. Mapping the Underway Voyage Datasets

Underway datasets will be represented on the AODN portal map by their voyage track. This track will initially be point data that represents locations at which observations have been measured from sensors. As underway voyage data is ingested into the facility databases that are sourced by the facility's WMS, it will be automatically available to the AODN portal in real-time.

When a point along the voyage track is clicked by the user, a getFeatureInfo request will be sent to the facility's WMS, requesting information about that point in the track.

4.3.2.2. FTL Templates

FreeMarker Template Language is used by Geoserver to compile HTML from feature attributes, based on a template that is provided. This gives the WMS administrator the

control to decide on the formatting of a getFeatureInfo response, and what information is displayed within it.

Not all feature attributes need to be displayed, however everything that is displayed in a getFeatureInfo response must either be hard-coded into the FTL template or be derived from an attribute of the feature that it is requested for.

4.3.3 getFeatureInfo Responses

The getFeatureInfo response will contain basic voyage-related information as well as basic feature information and URLs to resources. At a minimum, the getFeatureInfo response should show:

- the voyage name
- the coordinates of the feature
- a URL to the full metadata record in the AODN MEST
- a URL to a downloadable NetCDF file for the entire voyage dataset
- a URL to a downloadable CSV file for the entire voyage dataset

Because *all* information that is contained in the getFeatureInfo response must either come from a static template or a single feature, *everything* that displayed in a feature pop-up must be stored as attributes to that feature. For example, the URL (or at least the non-static component of it) to the full metadata record *must* be stored as an attribute to *every* feature in the facility database.

The following diagram shows the relationship between the getFeatureInfo popup and the various components of the rest of the system.

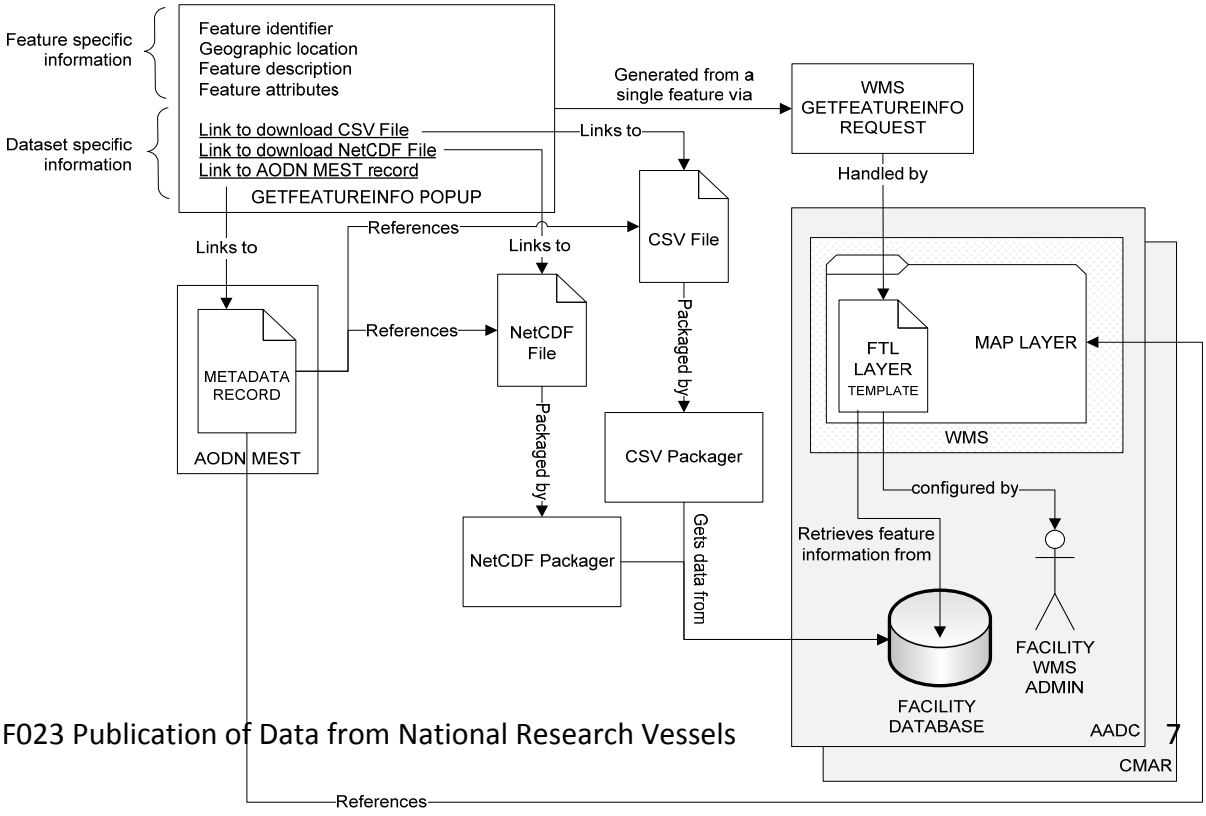

FIGURE 3. THE RELATIONSHIP BETWEEN GETFEATUREINFO POPUPS, DATA AND METADATA

4.3.4 Service Configuration

Voyage layers can be added to the portal configuration individually, or the URL to a WMS capabilities document can be added to the configuration, which would publish every layer that is described in the capabilities document (for a particular GeoServer instance) into the AODN portal.

Using the *capabilities* method is the preferred approach, as it doesn't require intervention by the AODN portal administrator to add a new voyage every time there is one available in the facility WMS, particularly given the high frequency of update for this real-time data stream (see 4.3.1, WMS Software). However, the default capabilities response from the WMS will include all layers that are available from an individual server instance via the WMS, and would not be limited to underway voyage datasets if other data layers were available at the facility node.

One of two approaches should be used to rectify this:

- An individual instance of Geoserver WMS can be set up by the facility to serve underway voyage datasets only, or
- The AODN portal can be configured with a URL to a non-default WMS capabilities document that should describe only underway voyage layers that the facility is publishing.

The former approach will work however it requires more resources to achieve (assuming the facility publishes other WMS services also).

The latter can be accomplished by a facility administrator manually creating web-accessible capabilities documents; however a more efficient approach would be to point the AODN portal to a capabilities document that is created by a web service, which can dynamically create a capabilities document based on database values (i.e. which voyages are available).

Once this service in instantiated, it would operate entirely autonomously.

4.3.5 Layer Styling

The AODN portal will request tiled images of layers from the facility WMS servers using WMS getMap requests. WMS getMap requests can be crafted to include styling information for the layers that are requested, however the AODN portal does not request any specific styles for layers that are rendered, so the styles that are set as default for the underway datasets on the facility WMS servers are used.

This means that *any WMS* request to a facility WMS server for images containing underway voyage data will by default be styled to suit the AODN portal. This has positive and negative aspects.

The main drawback of this method is that facility-specific applications which would usually draw on the use of default styles for facility WMS layers would be forced to use te same layer styles (by default) as the AODN portal

The main benefit of this approach is that tile-caching becomes much more efficient.

4.3.6 Tile Caching

In order for WMS getMap tiles to be served up efficiently, the facility should implement a tile caching service. Geoserver 2.0 ships with GeoWebCache built into it, which is configured by default to serve cached WMS tiles of layers that are published within Geoserver, however it doesn't handle OGC Filters on layer features. Cached versions of basic layers can be accessed via GeoWebCache simply by changing the url in the WMS request to the GeoWebCache module of Geoserver, using the same WMS request parameters.

The default GeoWebCache configuration is only very basic. The AAD implements a separate instance of GeoWebCache to cache WMS tiles for Geoserver. This enables finer configuration of GeoWebCache and can improve performance and scalability.

NOTE: Geowebcache should be configured to expire cached tiles every 15 minutes so that new data is displayed on the AODN portal map as it arrives. 15 minutes is the shortest interval at which new underway voyage data can arrive to a facility.

4.3.7 Java Environment

Geoserver 2.0 can run under any Sun Java Runtime Environment (JRE) version 5 and above, however there are significant speed benefits in configuring Geoserver to use the Sun JRE 6. As Geoserver will be deployed at both facilities running under an instance of the Tomcat servlet container, Tomcat should be configured to use the Sun JRE 6, as this in the JRE that Geoserver will use.

Above all, it is important for Geoserver to be configured for use with a JRE developed by Sun Microsystems, because they have sufficient support for two-dimensional rendering.

For more information, see http://docs.geoserver.org/2.0.1/user/production/java.html

4.3.8 Native Java Advanced Imaging

For best performance, native Java Advanced Imaging and Image I/O extensions should be installed by the facility to the JRE that Geoserver uses. This improves the speed at which tiles are rendered.

5. Metadata

5.1 Recording Metadata

AAD and CMAR should record the same metadata attributes for underway datasets.

Underway datasets should be described by multiple metadata records. A parent metadata record should describe the voyage, whilst sensor-level metadata (almost analogous to parameter level metadata) should describe sensor specific metadata, (e.g. for the Temperature Sensor), such as sensor downtime, calibration times and configuration, and levels of erroneous data. The sensor-level metadata records should be child records of the voyage metadata record.

The sensor-level metadata should consist of detailed information about the sensor that was used to take measurements during the voyage. A metadata record should be created for each sensor on every voyage. The information that is recorded in the sensor-level metadata record should contain, at a minimum:

- sensor ID
- sensor name
- period of operation for the sensor
- calibration details
- phenomena measured
- unit of measurement
- maximum and minimum thresholds that are used to identify good quality data
- history of the sensor

5.1.1 General Voyage-Level Metadata Considerations

Each layer on the AODN portal must relate to a single AODN MEST metadata record. As the underway voyage datasets are represented by a voyage track on the AODN portal, the voyage-level metadata record for a voyage should be flagged as being related to the voyage track that can be displayed on the map. This is achieved by setting the *distribution* tag in the MEST metadata record. The *distribution* tag should not be set for sensor-level metadata, unless it is desired that observations from individual sensors appear on the AODN portal as individual layers, however this is not a current requirement.

5.1.2 Recording Voyage-level Metadata (AAD)

Voyage level metadata relating to AAD voyages should be created for each voyage before the departure of the voyage. This metadata record should be created in the Global Change Master Directory, and be harvestable by the AADC OAI service. This process involves the AAD Metadata officer exporting the metadata record from the GCMD into ISO 19115 format after it has been created, and placing the exported file into a directory that is harvestable by the OAI service.

5.1.3 Recording Voyage-level Metadata (CMAR)

CMAR voyage metadata should also be created prior to the departure of each voyage, using the CMAR MEST instance. Once instantiated, metadata in the CMAR MEST will be harvestable by the AODN MEST. See section 4.2.1, Harvesting CMAR Metadata, for more information.

5.1.4 Recording Sensor-level Metadata (AAD)

Sensor-level metadata is sometimes recorded in the ship's log during AAD voyages, however that only occurs if there is a marine science programmer on-board the voyage (usually one voyage per year), and often this is still not the case if there is a marine science programmer on-board.

The AAD should create a metadata record for data obtained from each sensor, prior to each voyage. Also, an on-board events database should be used by AAD personnel to log events that occur with sensors, such as sensors being swapped out for spares (noting the serial number or sensor ID of each sensor), errors or problems that sensors encounter, and sensor cleaning information.

Metadata should be derived from the events database each time an update of underway data is transmitted from the vessel to shore, and it should be included in the transmission and later included as a part of the sensor-level metadata.

Furthermore, an identical events database containing historical sensor events also should be maintained at the Australian Antarctic Data Centre (AADC).

5.1.5 Recording Sensor-level Metadata (RV-SS)

The sensor-level metadata is fed into the RV-SS on-board data management system, TECHSAS, at the start of the voyage. TECHSAS creates a NetCDF file of the underway data every hour. The metadata can be extracted from the NetCDF file and transmitted to the CMAR facility at a regular interval. A cost-benefit analysis will be conducted to decide the frequency of metadata transmission. Metadata will be created in a format that adheres to the Marine Community Profile (MCP). This will enable seamless translation of metadata to the MEST format because the MEST supports the MCP.

5.1.6 Sensor Calibration Metadata

Sensor calibration metadata must be included with the sensor-level metadata that is published to the AODN MEST. This metadata is input into the TECHSAS System on-board the RV Southern Surveyor for CMAR, however no such system exists for the RV Aurora Australis for the AAD.

The AAD maintains two identical sensors for every sensor that is needed for a voyage. Calibration is performed yearly on one of each sensor by AAD Science Technical Support, so each sensor is calibrated every two years. Identical sensors are swapped every year, meaning that if one sensor is due for calibration then the other is set in operation on the vessel.

AAD Sensors are calibrated during winter, so the sensor that is due for calibration in one year can also be taken on-board voyages that year as a backup sensor in case the primary sensor fails, as AAD vessel operations only occur during the warmer times of the year.

Calibration information will be stored in a web accessible location and referenced by the underway dataset sensor-level metadata.

6. Publishing to the Australian Research Data Commons

Metadata that is published to the AODN MEST will be automatically harvested by the ARDC. The ARDC is currently set up to harvest metadata from the IMOS MEST instance, however IMOS will ensure that the ARDC harvests from the AODN MEST instance also. This process requires no further configuration by AAD or CMAR in order for underway voyage dataset metadata to be published to the ARDC, however the metadata will not be available to the ARDC until the configuration is carried out by IMOS.

The following diagram shows the hierarchy of metadata catalogues, with respect to metadata harvesting.

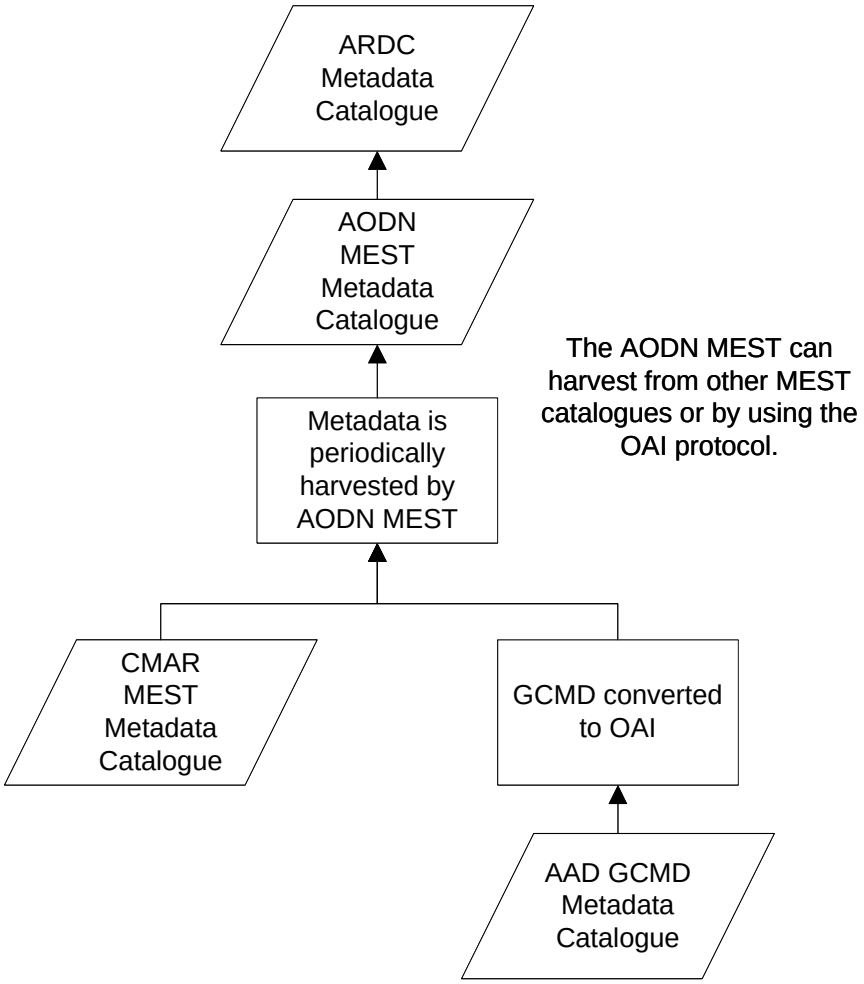

FIGURE 4. HARVESTING METADATA TO THE ARDC

7. Data

7.1 Retrieving Data

Data will be received by the AAD from the RV Aurora Australis via email every 15 minutes. Similarly, CMAR data will be received by the RV Southern Surveyor every 3 hours. In both cases the received data will be ingested into the respective facility's database as it arrives.

7.2 Quality Control

Several issues need to be considered to publish quality data in near-real time. Firstly, the metadata should be published to improve the usability of data. Secondly, data should be at least Level-1 quality controlled, this will remove some of the corrupt data and finally, user should have an option of downloading data for further use and visualise it. Figure 6 gives some of the steps that data will undergo to publish to the AODN.

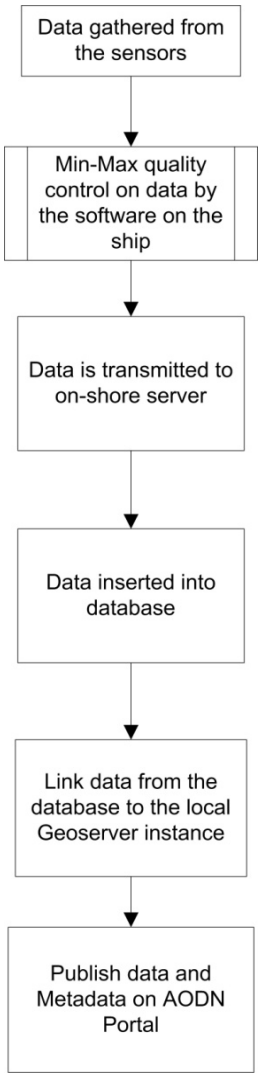

FIGURE 6. FLOWCHART OF THE STEPS INVOLVED TO PUBLISH DATA TO THE AODN PORTAL

The data should be added to an existing table of real-time underway voyage data that is referenced by the Geoserver WMS as the source of underway voyage data.

No on-shore quality control is performed on received data until the voyage has completed and the status of the data has changed to “historical” from “real-time”. The complete quality controlled (level-2) underway data will also be published on the AODN portal in a downloadable format.

A common, defined set of quality control flags will be established by the AAD and CMAR. A possible outcome of this will be to use the SeaDataNet Measurand Qualifier Flags. Alternatively, the CMAR data quality flagging system could be used by both organisations, however the impact of either organisation adopting each system will need to be further investigated before a system is chosen.

7.3 Storage Format

Data is to be stored in a relational database format. This is preferred by both AAD and CMAR as it provides a versatile interface to the data, including one in which a WMS service can source the data.

The only restriction to the database format that is used is that it must be compatible with the Geoserver WMS. With this in regard, the AAD will use PostGIS for the storage of underway voyage data and CMAR use spatial referencing enabled oracle database.

7.4 Dissemination Formats

At a minimum, underway data should be made available by both facilities as NetCDF files, CSV files and ship tracks in the form of WMS tiles.

A set of parameters from the British Oceanographic Data Centre (BODC) parameter vocabulary will be defined, that will be used by both AAD and CMAR for naming and describing datasets and their contents.

Downloadable NetCDF and CSV files will be generated on-the-fly by web services that are to be instantiated by each facility. By using a service to create a downloadable file in real-time, the user will have the benefit of being able to take advantage of the usability of NetCDF and CSV, without having the need for static files to be created at regular intervals as data arrives from the vessels.

Downloadable datasets will contain data for every sensor for the voyage requested. Data pertaining to individual sensors may be available as individual downloads at a later stage, however as a preliminary measure it will only be available as an entire dataset.

All downloadable files will contain all relevant metadata that relates to the voyage requested as well as the sensor-level metadata for each sensor that is included in the dataset.

Optionally, each facility may choose to expose the underway datasets in other formats, however this will not affect the overall operation of the AODN portal.

8. Preparing Data for the AODN Portal Map

8.1 Database Preparation

Databases that are used to store data that is published to the AODN portal should be prepared appropriately. This entails that the relevant database tables should maintain appropriate:

- spatial indexes
- primary keys
- referential integrity, and
- spatial metadata.

9. Historical data Vs Live data

The best possible quality data for a voyage should always be referenced by a single metadata record. This may mean that data has not undergone quality control procedures and yet is still available via the portal, however it is still the best *available* data so should be published. The state of quality control on the dataset should always be stated in the metadata for the dataset.

Dataset versioning may occur at the facility as quality control occurs on a dataset, however only the *best possible* version of the data should be made available via the AODN portal and the associated metadata records.

10. Conclusion

This is report provides the specifications for publishing underway data from RV-SS and RV-AA to the AODN portal. The voyage and sensor-level metadata of the corresponding data will also be published on the AODN MEST.

The implementation will adopt and leverage most of the features and technology used in the AODN portal. However, some of the technologies that the portal is designed to be used with should be replaced with a more client-centric approach where the client has more of an influence on how, where and what data is delivered to the portal. This may require the use of additional OGC web services like the Web Feature Service (WFS) and Sensor

Observation Service (SOS) to maintain the service-oriented architecture principle and abstract data sources from the portal, if the portal were further developed.