



Australian Ocean Data Network



Integrated **Marine**
Observing System



Australian Ocean Data Network

What eMII needs to achieve

IMOS from the Users Point of View

Australian Ocean Data Network

Data Discoverable:

- Uniform metadata and intelligent searching semantics

Data Accessible :

- Web services interfaces linked from metadata to data

Data Meaningful :

- Metadata / QAQC / Ancillary Data / Standardised Processing

Data 'Integratable' :

- Single high-level framework for IMOS data

IMOS from eMII's point of view

Australian Ocean Data Network

- Centralised metadata
 - Central catalogue harvested from the custodian metadata systems
- Data Versioning
 - Directory of what data has been lodged with eMII, the version and status
- Data Context
 - Linking of the data to the range of ancillary data that provide context
- Data Integration
 - Single over-arching framework in which all data sits
- Data Accessible
 - Oceans Portal using Web Services delivery but other mechanisms may be used

IMOS from the data providers point of view

Australian Ocean Data Network

- Data Described
 - Metadata systems for each data custodian / creator
- Quality Data
 - Set of uniform QA/QC systems and flags
- Data Stored
 - Local Custodian storage + TPAC Storage
- Data Processed
 - Range of mechanisms from TPAC processing to custodian processing
- Data Available
 - Via web sites that allow for searching and retrieval of data

What IMOS needs to deliver...

Australian Ocean Data Network

- Is about a 'sea-change' in data collection, management and use;
- Is based on freely available data for use by a community of people;
- Has data integration as a fundamental component.

IMOS demands a new way of thinking about data and, as a result, means new models for data management – there are 'strings' attached!

Data Frameworks

Frameworks - Introduction

What is a Data Framework?

- A data framework can be described as all the components, including hardware and software, standards, systems, tools, dictionaries, defined values and terms, processing workflows, data storage schemas (such as table definitions), data exchange formats and schemas, processing rules and systems that together meet the needs of ingesting (either virtually or for real) data generated outside the framework, storing, processing and transforming the data using set systems, rules and workflows, describing the data using terms defined by the user communities, linking the data to other external and internal supporting data and making the data discoverable and accessible to a range of users.

What is a Data Framework?

Australian Ocean Data Network

- What is a Data Framework?
 - Standards;
 - Tools;
 - Resources (people, etc);
 - Dictionaries;
 - Defined values and terms;
 - Processing work-flows;
 - Data storage schemas (such as table definitions);
 - Data exchange formats and schemas;
 - Processing rules;
 - Hardware and software components.

Function of a Framework

Australian Ocean Data Network

- Is the fundamental way that data integration is delivered;
- Ensures conformity amongst data from the community of providers;
- Facilitates use of the IMOS data by other communities;
- Gives data providers definitions and standards to work to;
- Works to raise all providers to a similar standard of data management.

One Framework or Many Frameworks?

Australian Ocean Data Network

- The goal of data integration across IMOS means one higher level framework (but this could be simple);
- Can accommodate a number of more complex sub-frameworks for equivalent data types;
- MUST be standards compliant (whatever this means)

Many disparate frameworks will not deliver data integration outcomes to IMOS.

The importance of a Framework in a machine dominated system

Australian Ocean Data Network

- We are moving towards systems targeted at machines interacting with data not just people;
- We need to design systems to accommodate machine interfaces as well as human interfaces;
- Frameworks allow us to design in some of the interfaces to meet this emerging need while still focusing on the people part!

What type of Framework?

Australian Ocean Data Network

- Database or record centric frameworks
- XML or object centric frameworks
- Agency versus Community frameworks
- National versus International frameworks
- Human orientated versus Machine orientated frameworks

Components of the Framework

Australian Ocean Data Network

- Metadata – standards, tools, processes
- Data Schemas – how the data is stored
- Standard nomenclature, definitions and mapping between these where multiple 'namespaces' need to be supported
- Data exchange standards;
- Data processing standards, in particular QA/QC standards;
- Mappings between the IMOS framework and other frameworks as required;
- Tools, documentation and people to support the Framework.

ISO/OGC have been busy...

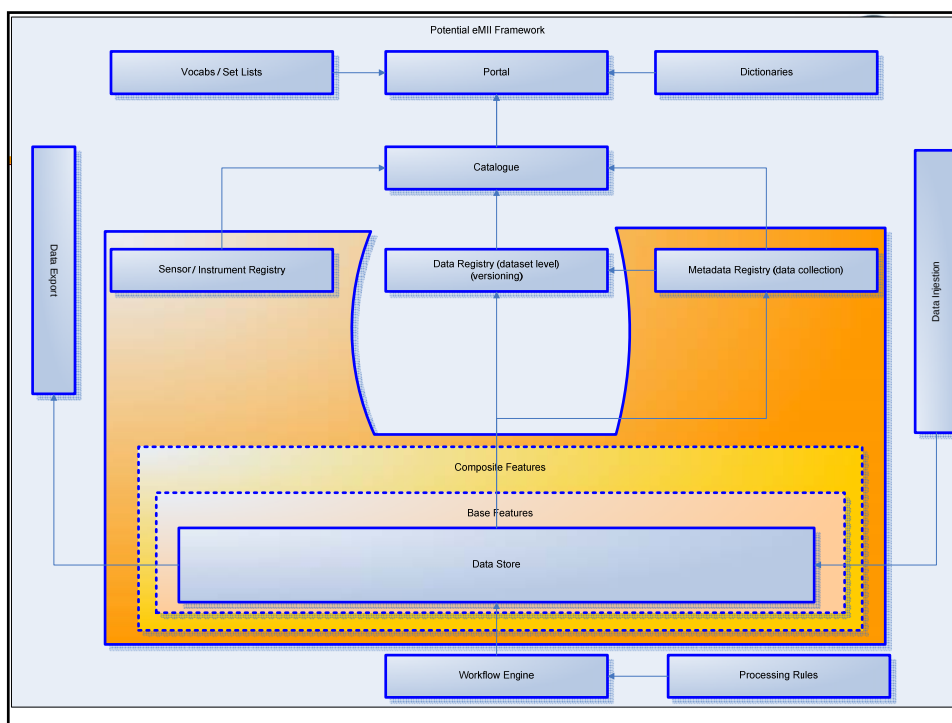
Australian Ocean Data Network

```
<sensorDefinitions>
  <definition type="airtemp" href="urn:x-ogc:def:phenomenon:OGC:AirTemperature"/>
  <definition type="windspd" href="urn:x-ogc:def:phenomenon:OGC:WindSpeed"/>
  <definition type="winddir" href="urn:x-ogc:def:phenomenon:OGC:WindDirection"/>
  <definition type="pressure" href="urn:x-ogc:def:phenomenon:OGC:AtmosphericPressure"/>
  <definition type="wtemp1 and wtemp2" href="urn:x-ogc:def:phenomenon:OGC:WaterTemperature"/>
</sensorDefinitions>
<uomDefinitions>
  <definition type="deg C" href="urn:ogc:unit:celsius"/>
  <definition type="dir deg" href="urn:ogc:unit:degree"/>
  <definition type="hPa" href="urn:x-ogc:def:uom:hPa"/>
  <definition type="knots" href="urn:x-ogc:def:uom:knots"/>
</uomDefinitions>
```

Describing the Framework

Australian Ocean Data Network

- Use of Data Standards Documents (DSD's)
- Use of Data Flow Diagrams
- Use of Standard Operational Procedures (SOP's)
- Via mappings to other frameworks and standards





AQDN
Australian Ocean Data Network

Standards



IMOS
Integrated Marine
Observing System

Australian Ocean Data Network

Why use Standards?

- Big picture versus small picture...
- Only way to realise the goals of IMOS

Where Standards are going...

- Geographic standards are evolving fast;
- Many of the issues / problems with environmental data are being addressed at the International level (e.g. Taxonomy);
- Most standards are generic with individual communities left to define profiles of the standards;
- XML representations are the base of most new standards;
- ALL international efforts are standards based;
- Standards are the 'soldiers' of the sea-change currently underway in how we treat, manage and use data.

Metadata Standards

Australian Ocean Data Network

- ISO 19115 vs ANZLIC II vs FGDC
 - ANZLIC II is the existing standard (since 1998??)
 - FGDC is used by ESRI and others
 - ISO is the new standard and adopted by the Commonwealth
- ISO 19115 and ISO 19139
 - ISO 19115 is the standard
 - Purposely vague and general
 - Idea of communities developing profiles
 - ISO 19139 is the validation schema (XSD) but it is lacking...
- The Australian and Marine Profiles
 - The Australian Profile is just the ISO standard with one element changed from conditional to mandatory
 - The Marine Profile is cut down to include elements that are only meaningful to marine data;
 - The Marine Profile includes pick lists, vocabs, etc to guide
 - The Marine Profile belongs to the Marine Community.

Metadata Implementation

Australian Ocean Data Network

- One MEST or Many MEST's ??
- Keeping it all in sync, harvesting
- Human versus machine readable systems
 - data level metadata vs service level metadata
- Metadata granularity
 - Fundamental issue not easily resolved
- Parent and child records – what this may mean
- Metadata and versioning
- What is metadata and what is ancillary data?
- Metadata and marine catalogues and portals

Implementing Metadata

Australian Ocean Data Network

- Steps in implementing metadata for your project
 - Adapting the Marine Profile
 - Installing or using a MEST
 - Developing a template
 - Developing a validation schema
 - Using it!

Data Standards Documents

Australian Ocean Data Network

- The role of DSD's
 - Documents the components of the data management for a data type;
 - Details the data flow, processing, quality control, storage and what standards exist or will be used;
 - Details the roles of eMII and the data provider.
- Who develops these?
 - Developed jointly by eMII and the Data Provider
 - Published as reference documents

Ancillary Data

Australian Ocean Data Network

- Why this is important – adding value for the User
- What this may include:
 - Sensor registries
 - SOP's
 - Reports and outcomes
- Ancillary data vs metadata
- Ancillary data is more important than most people think!

Quality Assurance / Quality Control

Australian Ocean Data Network

- Focus on Quality Control / Quality Assurance
 - We need to have a greater focus on quality control
 - Part of the value of the data;
 - Need to present a single framework for QC
- Lossy versus conservative QA/QC
 - Many QC systems are 'lossy' – only accept 'good' data
 - One idea is to move to a conservative system that conserves data but 'tags' it so Users know how to use the data
- Data set level versus record set level QA/QC
 - Much of the QA will be at the data set level
 - Do we need to do value level QA/QC?
 - Best practice says we do need value level QA/QC

QA/QC

Australian Ocean Data Network

- Need for single uniform QA/QC standard and representation
 - Makes it easy for User to understand the data;
 - Essential for data integration
- eMII standards versus other standards versus no standards
 - Recognition that there are a number of existing standards;
 - Need to develop a single eMII/IMOS standard
 - Understanding that some projects will need help with this including tools, software, etc;
- Real-mode versus delayed mode data
 - Differing needs for real and delayed mode data;
 - Met. data example where real-time data is not as QA/QC'd as delayed mode data;
 - Does all data have real-mode and delayed mode data??

Example QA/QC Framework for eMII

Australian Ocean Data Network

- Principles:
 - All data has error, we need to understand and document this so that Users of the data can decide for themselves how useful the data will be;
 - All QA/QC procedures should preserve data, even if it is deemed to be of low usability;
 - The end user is the ultimate determiner of the usability of the data and any QA/QC systems should make it easy for the user to discover the lineage, chain-of-custody and history of the data at all levels from the entire dataset or series down to the individual data point;
 - As a result the QA/QC should preserve all data at all stages of the QA/QC process along with enough information for the User to understand the exact process where by the final 'corrected' data value has been derived;

Australian Ocean Data Network

- Principles (cont):
 - The QA/QC system should allow the user to fully deconstruct a data value back to the original reading, that is at any level of processing it should be possible to back-track to the original reading and forward-track to the final corrected value;
 - The QA/QC system should follow an agreed set of nomenclature standards so that it is easy for users to understand the data they are using – the QA/QC information has to be easy to understand and use.

QA/QC Implementation

Australian Ocean Data Network

- Implementation:
 - Idea of QA/QC levels
 - Real-mode versus delayed mode data
 - Record level QA/QC
 - Standard set of QA/QC 'flags'
 - Standard Processing toolbox
 - Mapping of eMII QA/QC to other systems / standards

QA/QC issues

Australian Ocean Data Network

– Issues:

- What level of QA/QC do we need to do?
- Do we need record level QA/QC?
- The cost of doing QA/QC – is this prohibitive?
- What about existing QA/QC standards?
- What about non simple non numeric data (e.g. video)?
- What about the other components to Quality, such as calibration of equipment, etc;
- Where is the QA/QC information stored?
- Does anyone care? Does it matter?

AIMS Sensor QA/QC

Australian Ocean Data Network

- Record or value level QA/QC
- Rules based
- Users can see what rules a data value has passed or failed and filter or accept data based on the rules outcome;
- QA/QC data is stored and presented with the data, not separate to the data;
- Hierarchal level of 'flags' to allow mapping to other systems yet present more information;
- Based on QA/QC 'levels';
- May be a template for what eMII does...

Versioning

Australian Ocean Data Network

– Versioning

- What does this mean?
- Real mode versus delayed mode
- Availability of data
- Simple or complex – what do we need to do?
- One or many copies of the data?
- How would it work?
- Matching data versions to metadata and other framework components
- What data do we want to expose to the world?
- Issues and Implementation

Versioning - what it Means

Australian Ocean Data Network

- Recognizes that most data goes through processing stages and from this a number of differing versions of the data will exist;
- Also recognizes that at a future point a data provider may want to replace one data set with another to accommodate changes in processing, quality control or to correct errors;
- Users of the data need to know if the data they are using has been altered or updated, so there needs to be a link between the data and the users;
- Versioning is one area that can be as simple or as complex as you want to make it!

Versioning - real mode vs delayed mode

Australian Ocean Data Network

- We need to make data available ASAP even if it is 'incomplete' – if we leave too long a period between collection and availability we may never get it;
- So most or all data will have a 'real' mode version that will be made available ASAP to an agreed to schedule;
- Many data will also have a delayed mode version and we need to identify this as a separate data product to the real mode;
- The delayed mode product will normally be the product that the scientist will want to advertise;
- The idea of real vs delayed mode data is problematic to many data providers.

Versioning - how it may work

Australian Ocean Data Network

- Differing version of the data can be viewed as differing data sets and treated as such;
- Differing versions can be hidden by 'replacing' one version with another;
- We can use a separate data registry to implement a mixed mode – only a small number of fundamentally different versions of a data set are presented (e.g. real vs processed) and sub-versions are managed by a registry system.

Versioning - what we want to expose

Australian Ocean Data Network

- For most scientists they only want to expose the processed 'final' data;
- Users may have different expectations...
- We need to increase our sophistication as to what we are prepared to expose and what caveats we place on the data;
- Maybe we also need to educate our Users.

Versioning - issues...

Australian Ocean Data Network

- What we as a community want to do with regard to real mode vs delayed mode data;
- How do we educate the Users as to what the data is and what it can be used for?
- How sophisticated with versioning do we need to be?
- Does anyone care??

Data Flows

Australian Ocean Data Network

– Data frameworks and data flow

- The data flow describes how data moves between the components of the frameworks and, optionally, the standards, tools, systems and processes attached to each stage of the flow;
- So the framework is the bits, the flow is how the bits work together;
- Understanding the flow helps with the framework – that is the framework should make possible the required data flow.

Key Framework Issues

Australian Ocean Data Network

– Key issues

- How do we start?
- Where are we going?
- How to start simple and move towards complex?
- How to integrate what we do with the various standards and other systems out there, don't want to re-invent the wheel or be an orphan.
- Who does what, who is responsible for what?
- Role of eMII and role of the providers.

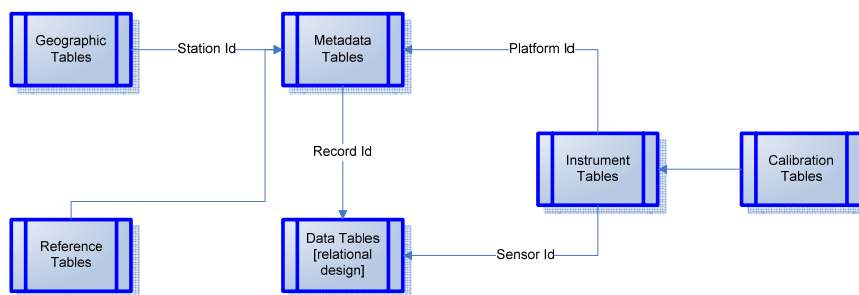
Questions...

Australian Ocean Data Network

Types of Frameworks

Australian Ocean Data Network

Traditional Data Framework



Feature Based Frameworks

Australian Ocean Data Network

- New way of managing data based on 'objects' or 'features' which reflect real world phenomena;
- Still being developed (bleeding edge) but potentially is a new way of looking at the world;
- Gives a direction for the future we need to engage with but not to distract us from delivering outcomes now.

Feature Based...

Australian Ocean Data Network

- All data is resolved to a small number (13 in CSML) of features and so potentially complex sets of data are reduced to a manageable and well defined set of features, for IMOS this would bring 30+ data types down to 10-13 features;
- Data from the same feature type by definition will be conformal and so, if appropriate, can by definition be integrated;
- Data integration occurs by creating composite features, that is over-lays of base features into more complex ones, once this mapping has been done this can be used for any data belonging to those base features – this is how true full data integration is achieved

Feature Based

Australian Ocean Data Network

- The feature defines what data is collected and so we can ensure that all the information we need to collect is collected and so data collection is complete;
- Features represent real-world phenomena or structures and so the data by definition has real-world significance that is easy to understand;
- Helps with the issue of granularity – by definition a dataset is all the data associated with a single feature - e.g. all the data associated with one profile is a dataset, the data for a set (collection) of features is a collection of datasets and so on;
- All of the data is inherently spatial and so doing spatial operations are made easy.

Feature vs traditional

Australian Ocean Data Network

- Walk before we can crawl;
- Too early – bleeding edge;
- Not everyone is at the same level;
- Important base standards still to be defined;
- Let someone else make the mistakes but keep this in mind and plan for it.

Summary

Australian Ocean Data Network

- We need to implement a single over-arching framework for IMOS, this needs to come from the community of data providers facilitated by eMII;
- The framework can be complex and sophisticated, issues such as metadata, QA/QC, versioning and standards will be important;
- Data Standards Documents (DSD's) will be one way that the framework is described and developed;
- The framework will evolve from one that is common to what most of us do today to future feature based object type frameworks – we have to go with the tide;
- The will to develop a single common framework is fundamental to achieving the outcomes for eMII and for IMOS.

Questions...

Australian Ocean Data Network

Data Flows

Australian Ocean Data Network

– The use of Data Flow models

- Data flow models are a powerful way of describing and defining how the data creators in each Investment Area will work with the eMII Data Management component. They detail who does what, what services, infra-structure and systems eMII needs to provide and also they highlight potential issues.
- eMII will use data flow models as a fundamental information gathering exercise to design the overall frameworks.

Data Flows...

Australian Ocean Data Network

- The eMII project is looking to use the idea of Data Flows to set the standards, processes and protocols that together make up the Framework.
- The idea is to start at the back end and work forward; that is from the eMII component (Portals, Catalogue, data storage, etc) back towards the point of data collection.
- For some data types this will only be appropriate to the point where eMII is responsible for the data (hence the concept of 'conformal' data where it is the responsibility of the Data Provider to ensure that the data is conformal).
- For other data types we will want to push further forward and work to set standards and protocols for the data processing (such as QA/QC standards) and for some even the data collection.

Data Flows...

Australian Ocean Data Network

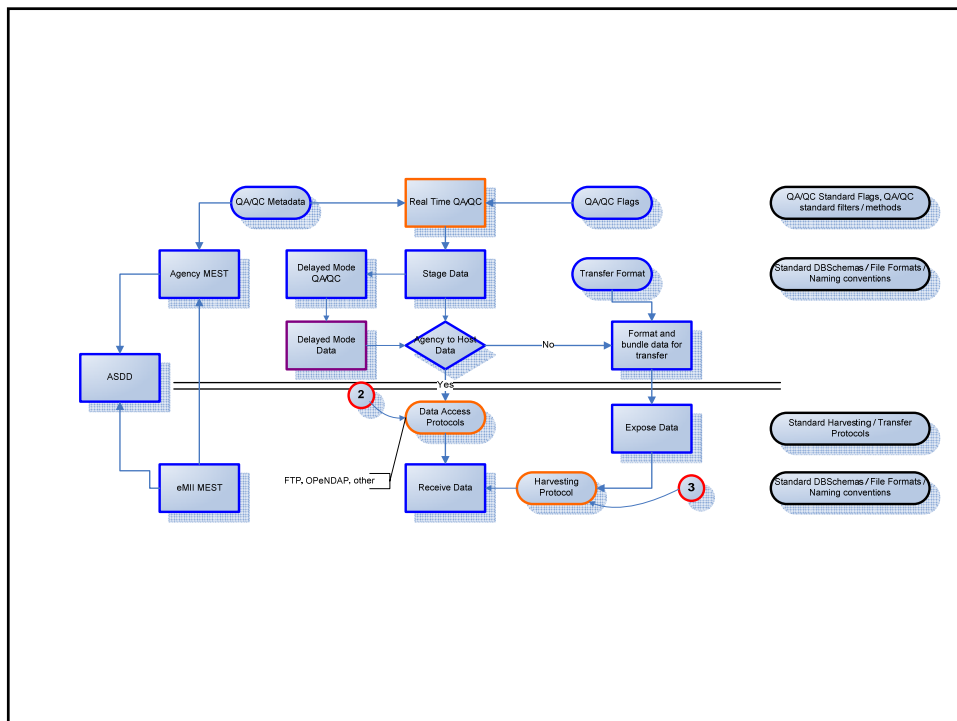
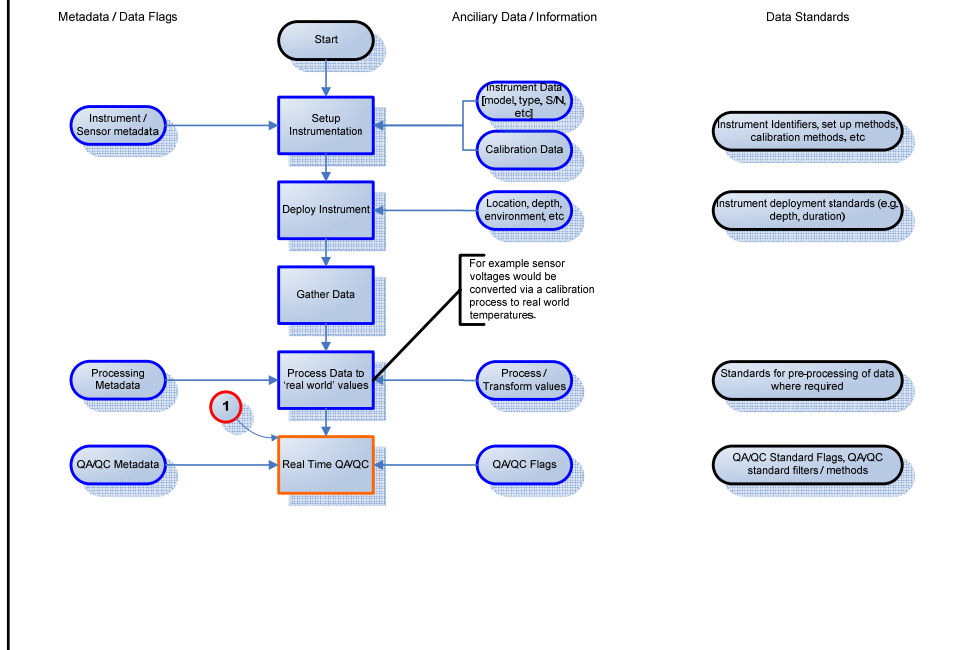
- By developing a data flow model for each data type we can better understand what eMII has to provide to each area;
- The data flow model also defines where eMII takes responsibility for the data flow and how these boundaries are defined and implemented.
- The idea is to come up with a small number of base flows that are supported - not an unlimited number.

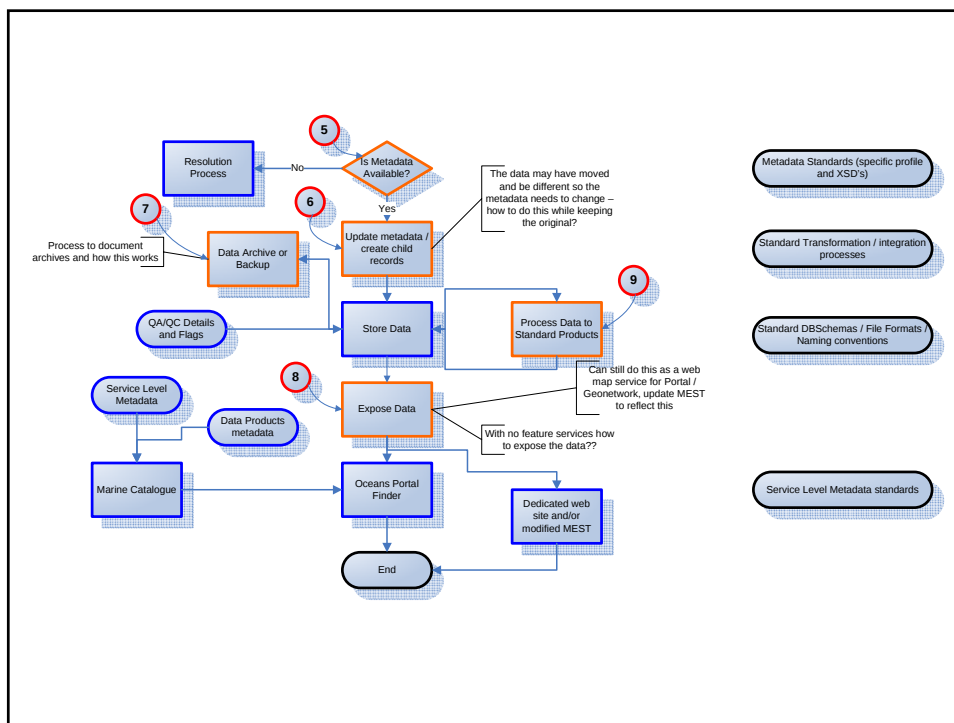
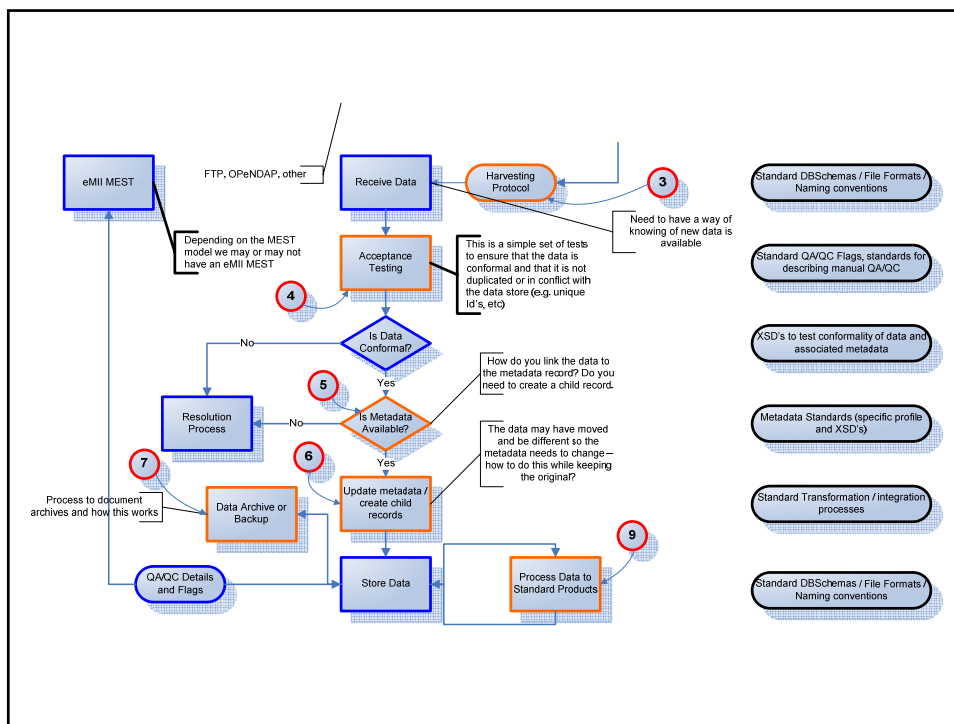
Generic Flow Diagram -1

Australian Ocean Data Network

- Is a generic flow of data with most of the complexities we expect to see;
- Has a number of issues that still need to be resolved;
- Breaks down to a small number of real world flows;
- Shows the relationships between the main parts of the framework.

Figure One: Draft Proposed eMII Data and Logic Flow (V1.0)





Issues with this flow...

Australian Ocean Data Network

1. How do we do real time QA/QC;
2. What protocols do we use to transfer data to eMII or at least let eMII know that new data is available?
3. If we 'harvest' data from another data store how do we actually do this – how many copies of the data, where does the 'original' lie?
4. When we 'ingest' data how do we ensure it is conformal and not duplicate?

More issues...

Australian Ocean Data Network

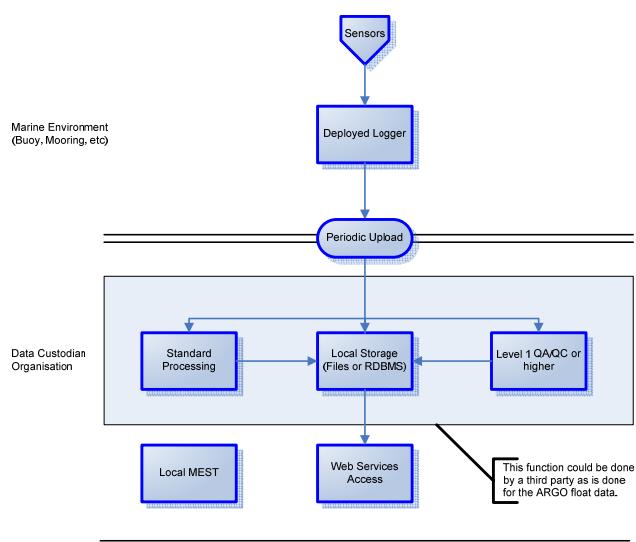
5. How do you link data to metadata, especially where you have real and delayed mode data and differing versions?
6. Do you need to update the metadata if the data has changed due to processing or by being delayed mode not real mode?
7. How do we back up data, how do we archive it?? Who does what??
8. What standards are used to expose the data, that is do we use WFS or web services or other standards?
9. If we create processed data products who does the processing, how are processing resources allocated??

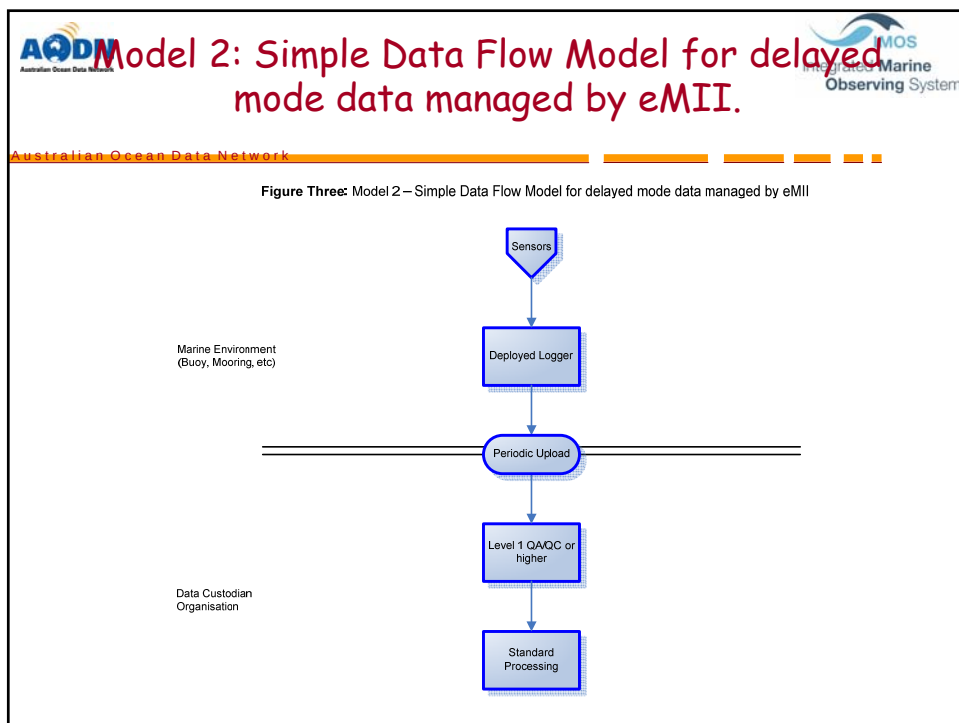
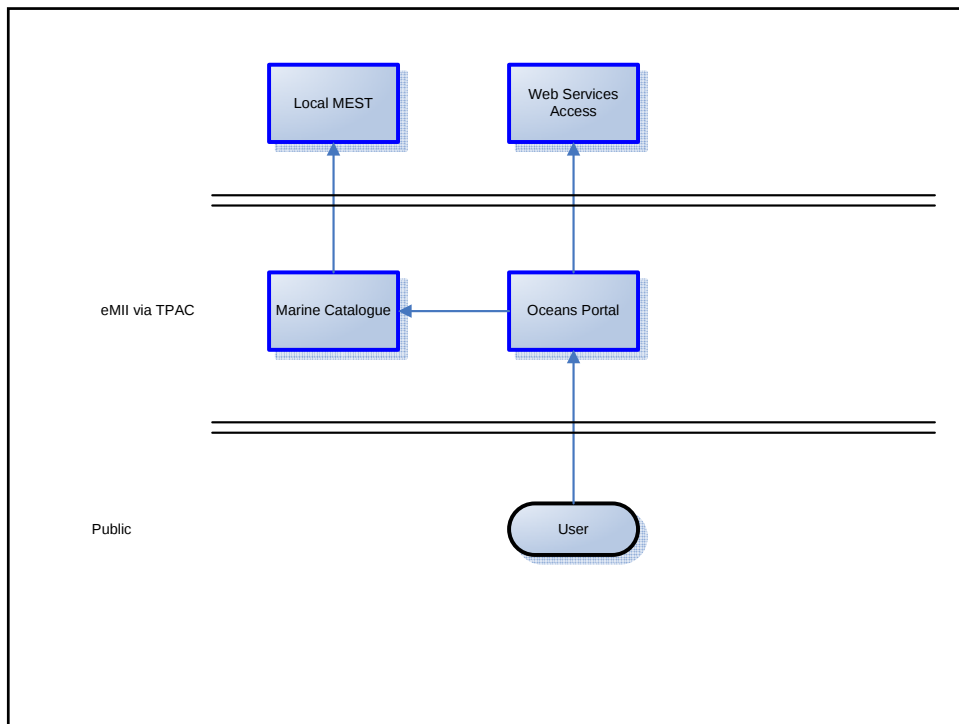
Example Flows

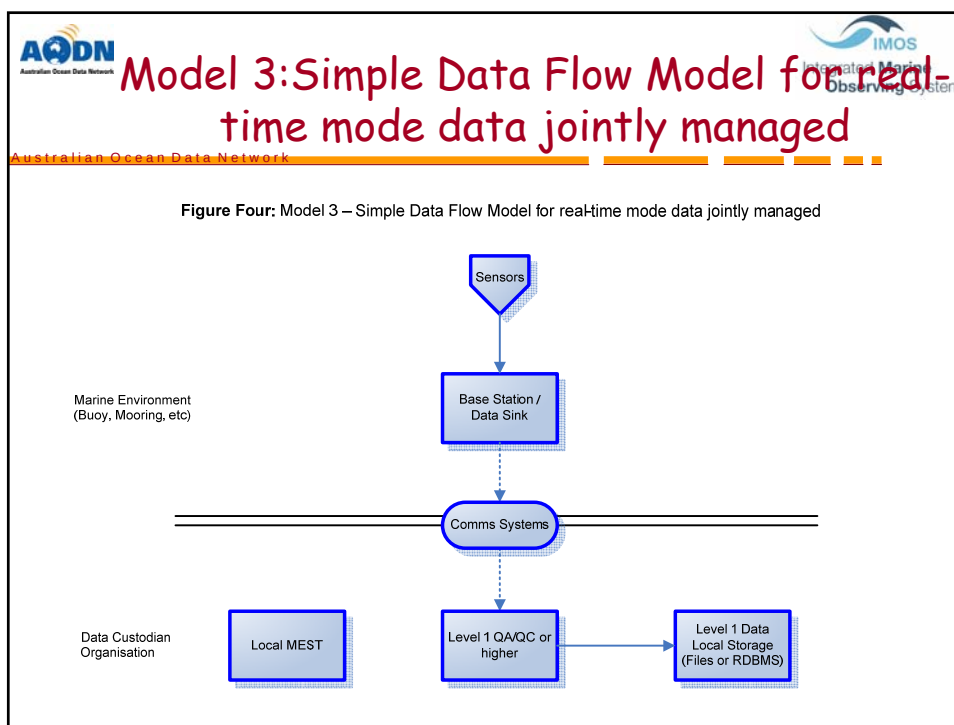
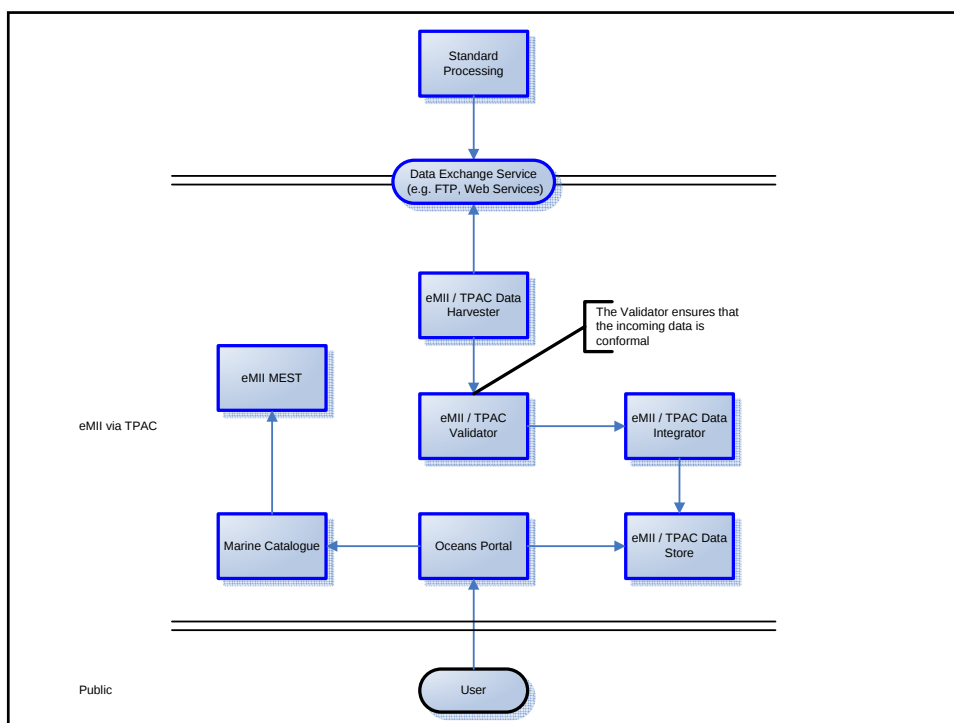
Australian Ocean Data Network

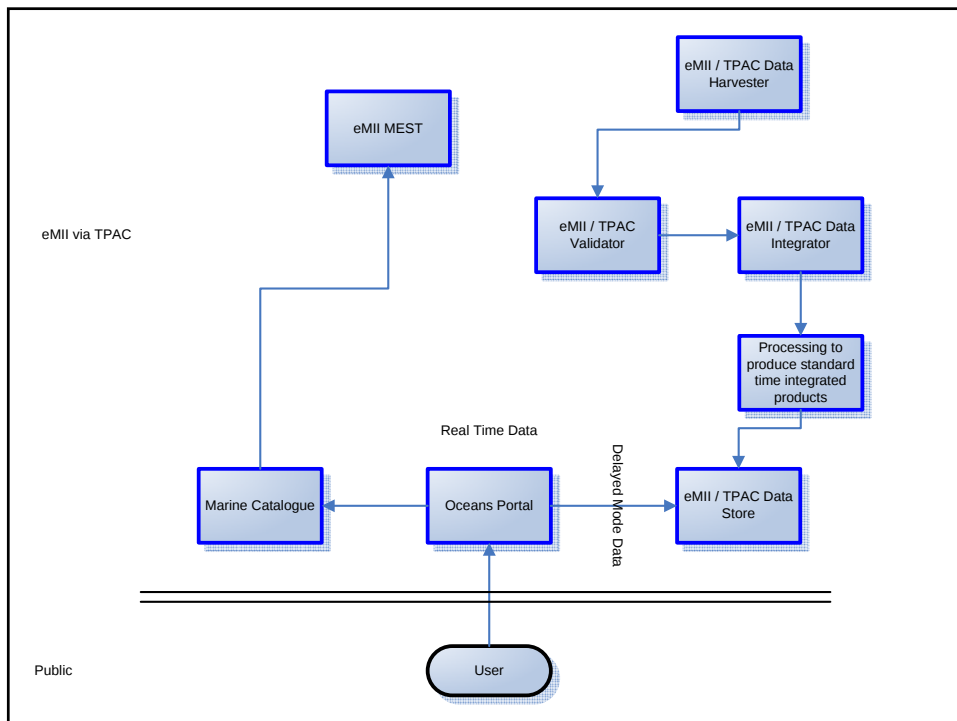
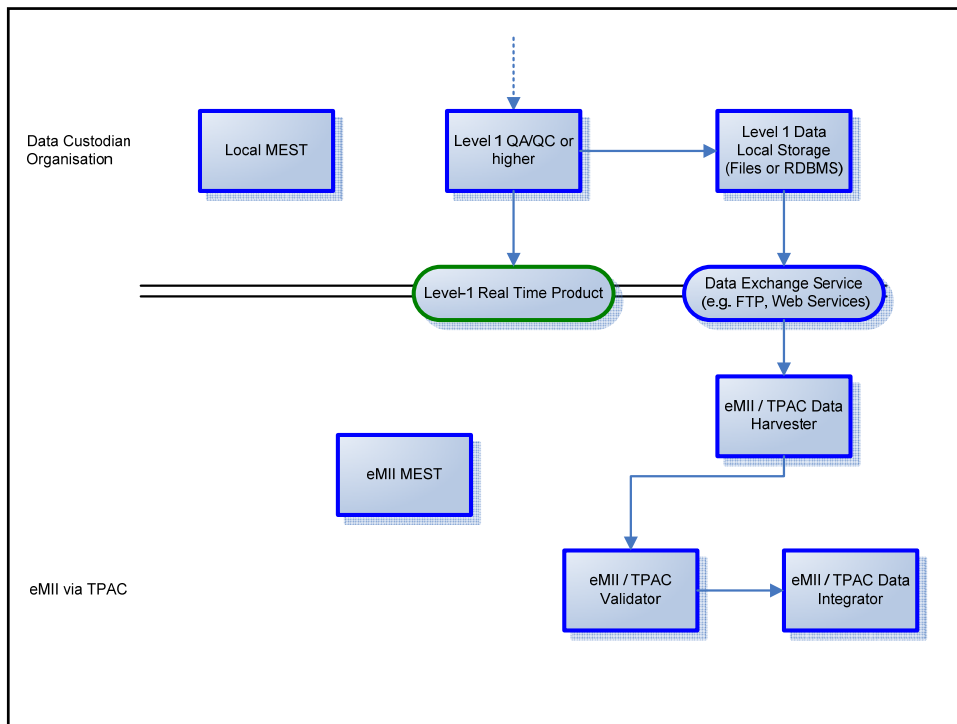
Model 1: Simple Data Flow Model for delayed mode data managed by the custodian.

Figure Two: Model 1 – Simple Data Flow Model for delayed mode data managed by the custodian









Issues...

Australian Ocean Data Network

– Key Issues

- The cost of best practice – who wears this and why?
- Metadata and ancillary data, how to tag datasets, data series and data points with all of the meta and ancillary data that is required to maximise the value of the data;
- Quality Control , how to implement rigorous quality control frameworks for the data to be collected at a range of scales from the dataset to the data point;
- Versioning – how to ensure that we are able to make changes to the data (such as a scientist correcting data) and have these changes flow through the system.
- Tools and processes
- Making it work

More Issues!

Australian Ocean Data Network

– Other Issues:

- Do we want to implement best-practice data management, if not what level do we want to implement?
- How will resources be allocated in meeting the cost of the level of data management required – do we just do the back end or do we actively get involved at the 'front-end' with those projects that do not currently have data management systems in place?
- How do we deal with those projects that have well defined and often set data management systems in place especially if these do not line up with the eMII framework?

And more...

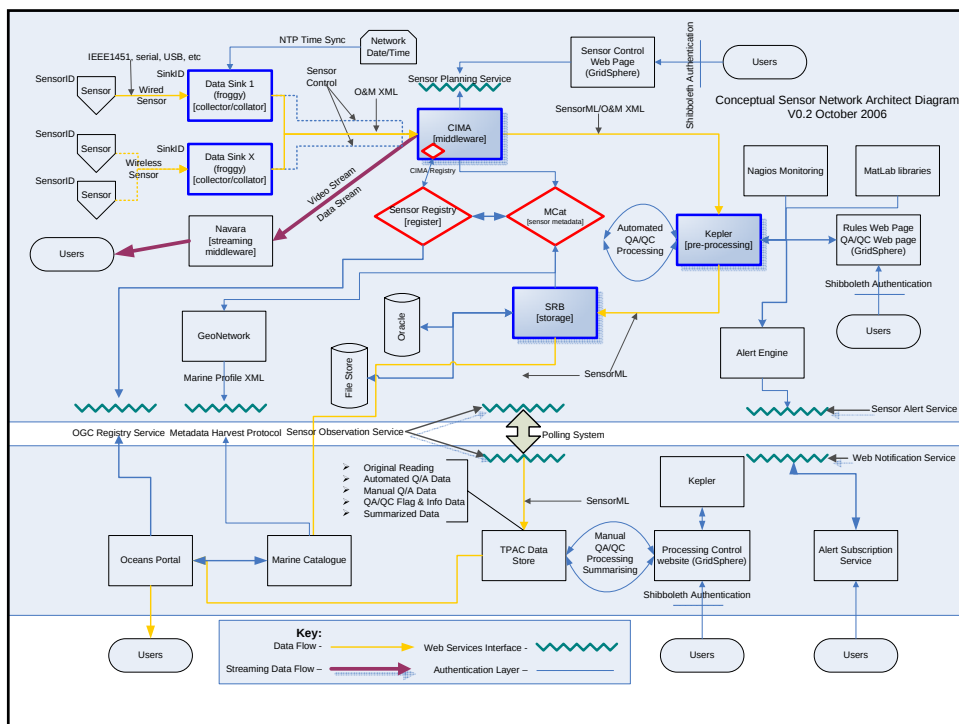
Australian Ocean Data Network

– Other Issues:

- How do we deal with the risk and uncertainty to do with the use of emerging and cutting edge technologies and approaches?
- How do we build better capability and capacity in eMII to deliver given that cooperative models, while they do deliver, are slow and often ineffective?
- How do we survive politically given the level of risk and the dependence on eMII and the Data Providers?
- How to we add value to the science work and to the outcomes of IMOS?

An example using the Sensor Network

Australian Ocean Data Network





Australian Ocean Data Network

Questions???



Integrated Marine Observing System

Australian Ocean Data Network

Data Standards Documents

Australian Ocean Data Network

- The role of DSD's
 - Documents the components of the data management for a data type;
 - Details the data flow, processing, quality control, storage and what standards exist or will be used;
 - Details the roles of eMII and the data provider.
- Who develops these?
 - Developed jointly by eMII and the Data Provider
 - Published as reference documents

Content of a DSD...

Australian Ocean Data Network

- Introduction
- Purpose of this document
- Data Flow
- Data Flow Standards
 - Instrumentation
 - Deployment
 - Data Collection
 - Data pre-processing
 - Data QA/QC
 - Data Storage / Staging
 - Data Transfer
 - Data Harvesting
 - Data Validation
 - Data Storage
 - Data Exposure

Content of a DSD... continued...

Australian Ocean Data Network

- Metadata Standards
 - Marine Profile
 - Metadata Validation / XSD
 - Service level metadata
- QA/QC Standards
 - Level 1 QA/QC Standards
 - Level 2 QA/QC Standards
 - QA/QC Flag Standards
- Data Integration Standards
 - Data Transfer Standards
 - Standard Naming and Formatting
- Appendices
 - Database Schema
 - Metadata Profile Elements
 - Metadata XSD
 - File Identifier Standards

Role of 'conformal' data

Australian Ocean Data Network

- eMII has been built on the idea of conformal data and that only conformal data will be accepted;
- The definition of conformal therefore determines the minimum standards for contributing data to eMII and IMOS;
- Compliancy is non-negotiable, these are community set standards that eMII will use to determine the quality of data in IMOS

Conformal data

Australian Ocean Data Network

- Conformal data is data that:
 - Has had basic QA/QC applied;
 - Has real-world values;
 - Has a corresponding metadata record available;
 - Uses standard values for units of measure, location, time and depth;
 - Meets the criteria set out in the DSD;
 - Is in the correct format for data exchange;
 - Passes an 'XSD' set up for this data.

'Base' QA/QC means:

Australian Ocean Data Network

- The lowest level of QA/QC includes:
 - All locations are 'reasonable' (i.e. wet);
 - Times are in the correct era;
 - At the record level data is checked for:
 - Missing values;
 - Out of bounds values;
 - Illogical or illegal values;
 - Logic tests where available.

Conformal Data

Australian Ocean Data Network

- Data that fails the 'conformancy' test will not be accepted by eMII/IMOS;
- Need to have ways to resolve this or at least have unambiguous ways of demonstrating conformance or lack of this;
- Use of XML and XSD's can do this.

Examples

Australian Ocean Data Network

- XBT data
- Argo data
- The Template

Questions...

Australian Ocean Data Network

Issues and Solutions

Australian Ocean Data Network

- The task we are trying to do is non-trivial;
- This is a community based effort not an eMII effort;
- There will be much give and take;
- Need for a 'can-do' attitude for it to work.

Issue: Metadata

Australian Ocean Data Network

– Metadata

- What if there are existing standards for my data?
 - Document these – do they follow an existing standard and if so which one?
 - Is it possible to use ISO19115 as a replacement for the standard currently in place?
 - If no then work with eMII to map your metadata to ISO19115/Marine Profile and look at the data flow to work out where to extract the eMII metadata;
 - Work out what data is missing and how this gets added

metadata

Australian Ocean Data Network

- I don't use a particular standard or set of tools, what do I need to do?
 - Download the Marine Profile from www.aodc.org.au and see what elements it has and how it is represented;
 - See how the data you collect maps to the profile
 - With eMII go through the Marine Profile and see how it works and how it can be adapted to your needs;
 - Look at the granularity of the metadata required;
 - Look at installing a local MEST or using a central one
 - Develop a MEST template to suit
 - Use it!

Issue: QA/QC

Australian Ocean Data Network

– QA/QC

- What framework will eMII use:
 - One task is to develop the eMII framework using best practice
- How do I relate my current QA/QC to what eMII wants?
 - One task is to map the eMII framework to other commonly used frameworks (e.g. Argo).
- What tools will be provided to do QA/QC on my data?
 - eMII will be tasked to provide tools for QA/QC but not determined what this will look like;
- How do I describe the QA/QC I currently do?
 - The DSD's will help with this;
- How will the eMII QA/QC map to other standards?
 - Yet to decide philosophically how we deal with other standards
 - map or compliment?

QA/QC

Australian Ocean Data Network

- What if I have a current QA/QC system?
 - Please document this and let us know what standards it uses and what linkages there are;
 - Talk with eMII about the framework that eMII is implementing and how your framework can map to this (or vice versa);
 - Look to see what is the best solution to move forward with.

QA/QC...

Australian Ocean Data Network

- What if I don't have any formal QC in place?
 - Examine your data and determine what bits are at the dataset level and what bits are at the record or value level;
 - For bits at the dataset level do the following:
 - Manually check all positions are where you expect them (e.g. on Google Earth);
 - Manually check all times;
 - Check units, geodes and names.

QA/QC

Australian Ocean Data Network

- For the record level bits do the following:
 - Adopt a 'flagging' system such as the IODE system

IODE QA Flags

Australian Ocean Data Network

Flag Code	Meaning
0	No QA/QC Applied
1	Good Data
2	Probably Good Data
3	Bad Data that are potentially correctable
4	Bad Data
5	Value Changed
6	Not used
7	Not used
8	Interpolated value
9	Missing value

Reference: MANUAL OF QUALITY CONTROL PROCEDURES FOR VALIDATION OF OCEANOGRAPHIC DATA – IOC/IODE

If I don't have a formal QA/QC system

Australian Ocean Data Network

- Modify your data format to include a QA/QC field / element:

Cast_Id	Pressure	Pressure_QA	Temp	Temp_QA	Salinity	Salinity_QA
2701	1200	1	16.5	1	32	1
2701		9	16.6	1		9
2701	-500	4		9	21	1

- Write a program / scripts to fill in the QA values based on the rules you specify.

QA/QC

Australian Ocean Data Network

- Update table set temp_qc = 9 where temp is null;
- Update table set temp_qc = 1 where temp between 5 and 25;
- Update table set temp_qc = 4 where temp >50 or temp<-50;
- Update table set temp_qc = 3 where temp > 25 and < 50;

QA/QC

Australian Ocean Data Network

- Document this!

Issues: Versioning

Australian Ocean Data Network

– Versioning

- Examine your data and identify what short-term products can be delivered and what longer term products – that is what real mode and delayed mode data there is;
- Work out what other versions will be produced, are these data products or data versions?
- Work out what caveats you need to place on your various data versions / products.
- Talk to eMII about where you get to.

Issues: DSD's

Australian Ocean Data Network

– Data Standards Documents

- Who will develop these?
 - Data Provider with help from eMII
- What do they have to contain?
 - As per the template
- What about data that is the same as other data?
 - We can and should create common documents (e.g. CTD);
- At what level do these need to be done?
 - As detailed as possible but could be links to other document.

Issue: Moving Forward

Australian Ocean Data Network

- Do the data flows for major data streams;
- Do DSD's for each data type generically and then with amendments as required;
- Development of the higher level standards to support this – metadata, QA/QC, Versioning, data access formats, etc;
- Work with the community to develop some of the framework systems – QA/QC, versioning, etc;
- Working with each data provider to put in place systems to deliver the required outcomes;