Importance of data stewardship for marine research

Greg Reed

Executive Officer, Australian Ocean Data Centre Joint Facility

Past Co-Chair, IODE Committee

International Symposium for the Integration of marine-related Data and Information
Tokyo, Japan, 8th December 2011

Presentation

- What is data stewardship: why do we need it
- Case study: Australian Ocean Data Network

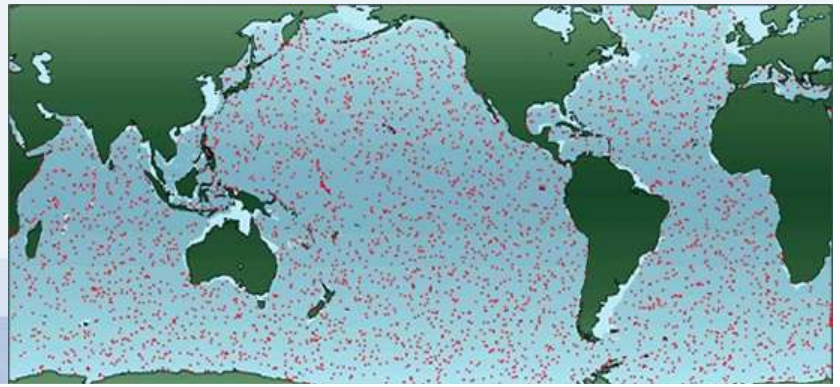
Marine data are important

- Marine data are fundamental to understand processes that control the environment
- Marine data are a key requirement for effective strategic decision making
  - play an important role in promoting the development of economic activities
- Underpin many of our activities, such as:
  - navigation
  - sea transportation
  - fisheries
  - marine disaster mitigation
  - environmental monitoring

Marine data are unique

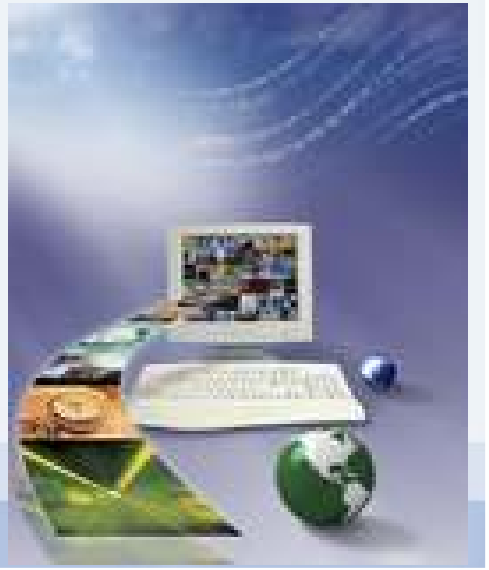
- Marine data are expensive to collect
- Marine data are unique and unrepeatable
  - the environment is constantly changing
- Spatial and temporal coverage is quite sparse
  - Research vessels and moorings are small dots on a map
- Important to ensure that maximum benefit derived from data
  - Share data !

"capture once – use many times"



Data Stewardship

- The *acquisition, processing, preservation, quality assurance and dissemination* of marine data are key elements of curation and archiving data
- This is known as ***Data Stewardship***
- A community approach
- Stewardship ≠ ownership

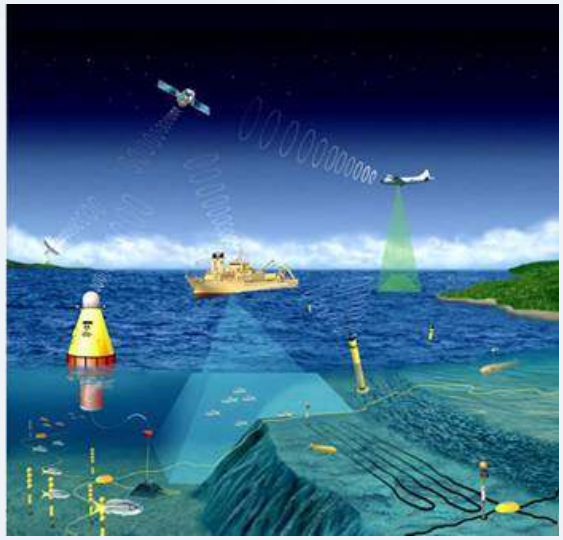


Components of Data Stewardship

- Key components include
  - Preserving detailed information about observed variables
  - Details of observation instruments used
  - Techniques and calibrations
  - Comprehensive metadata
  - Creation of products derived from archived data

End-to-end data management

- End-to-end data management system handles data from the point of collection, through processing and quality control, to archival and dissemination.
- End-to-end management of data facilitates the ability to integrate data from multiple sources and sensors (i.e. satellites, *in situ*, and model data)



Elements of end-to-end data systems

1. Standardized data collection
   - The lack of standardized data collection efforts can hamper long-term value of datasets
   - Data collection must be standardized to allow data sets from a variety of sources to be integrated.

Elements of end-to-end data systems

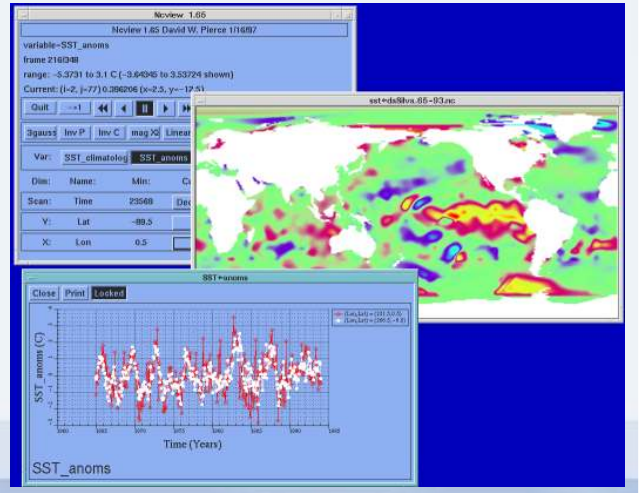
2. Common vocabularies

- Use of common vocabularies is an important prerequisite towards consistency and interoperability.
- Common vocabularies consist of lists of standardized terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community.
- Using standardised sets of terms reduces ambiguities and enables records to be interpreted by computers.

Elements of end-to-end data systems

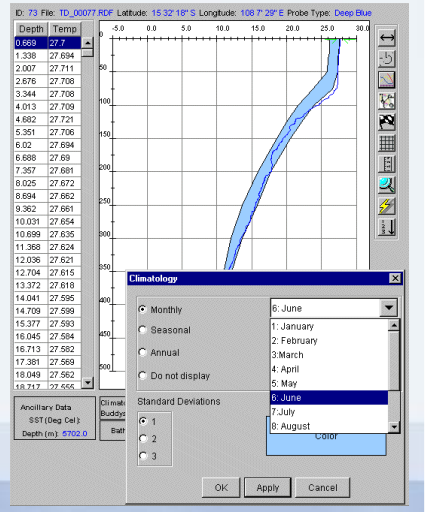
3. Standard data formats

- The selection and adoption of a small number of standardized data formats is essential to ensure effective data stewardship
- The use of just a few formats can enhance the ability of data stewards to preserve information over the long term
- netCDF (network Common Data Form) is emerging as a *de facto* standard



Elements of end-to-end data systems

4. Quality assurance and quality control (QA/QC)
   - Quality Assurance (QA) - procedures performed prior to instrument deployment to support the return of best possible quality data
   - Quality Control (QC) - procedures/processes applied to the data returned from the instrument
   - Standards for QA/QC should be well documented
   - The preservation of original values, even if they appear wrong, is important for possible future re-processing



Elements of end-to-end data systems

5. Data archive
   - To ensure the long-term preservation and dissemination of data, the data producers and data archives need to work together to generate the information needed to be able to understand and re-use data. This includes:
     - Well-defined file naming conventions and format descriptions
     - Related descriptive metadata
     - Information to facilitate data dissemination

Researchers can assist . . .

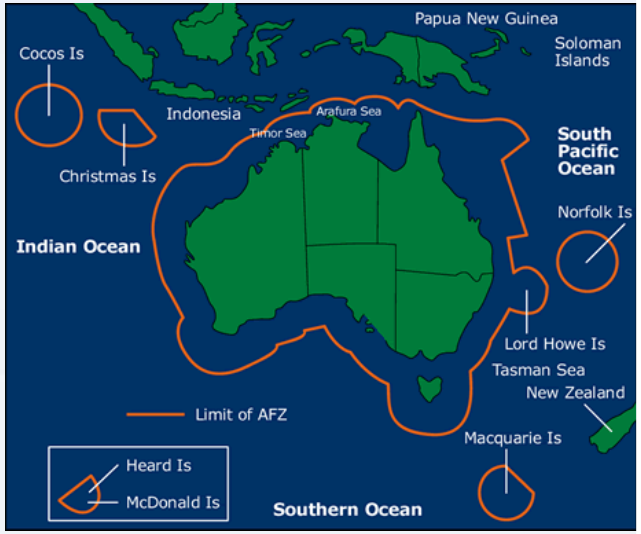
- Data derived from publicly-funded research should be available for public use
- The value of data increases if aggregated into collections and are available for re-use
- Researchers can work collaboratively with the wider community to develop a data stewardship framework
- It is important that researchers receive appropriate credit for their work

Case Study.

Data Stewardship in Australia: The Australian Ocean Data Network

Setting the scene . . .

- Australia has one of the largest marine areas in the world
  - More than 8 million square kilometres
  - EEZ more than double the land area
  - >80% population lives near coast
- Tropical through temperate to sub-Antarctic



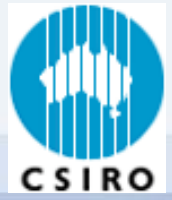
AODCJF

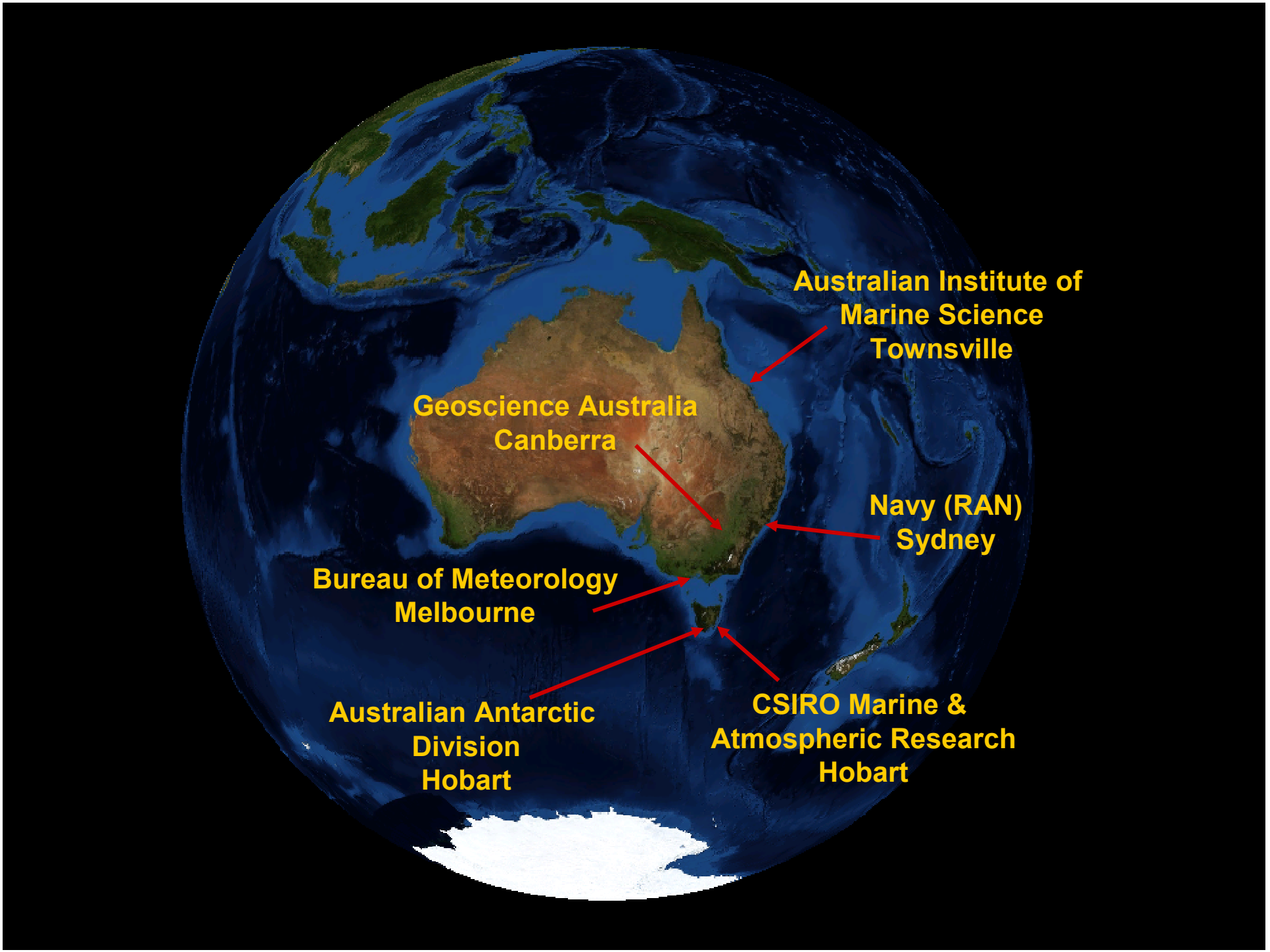
- Australian Ocean Data Centre Joint Facility
  - AODCJF operates as a distributed system
  - Promotes the discovery, transfer and implementation of knowledge about marine environment by facilitating the exchange and management of, and ready access to, marine data
- Established in 2005

Australian Ocean Data Centre
JOINT FACILITY

AODCJF Partners

- Six Australian government marine agencies as initial partners
  - Australian Antarctic Division (AAD)
  - Australian Institute of Marine Science (AIMS)
  - Bureau of Meteorology (BoM)
  - CSIRO Marine & Atmospheric Research (CMAR)
  - Geoscience Australia (GA)
  - Navy Hydrography and Metoc Branch (RAN)




Australian Institute of Marine Science Townsville
Geoscience Australia Canberra
Navy (RAN) Sydney
Bureau of Meteorology Melbourne
Australian Antarctic Division Hobart
CSIRO Marine & Atmospheric Research Hobart

Australian Ocean Data Network

- The AODCJF has a shared vision of developing and delivering an Australian Ocean Data Network (AODN) to manage national ocean data resources
- The AODN provides a national infrastructure to access the distributed archives of data held in the partner institutions across Australia.
- Distributed network
  - Data will reside on agency servers
  - QC and management responsibility of agency

Publishing interoperable data

- Achievements of the AODN
  - automated real-time Research Vessel Underway Data from AODCJF agencies, funded by the Australian National Data Service (ANDS)
  - identifying requirements for interoperability (Marine Community Profile, CTD use case)
  - facilitation and promotion of common cross-sector infrastructure in marine research

Research Vessels Aurora Australis (l) and Southern Surveyor (r); image acknowledgements: ABC Australia, CSIRO Australia

Integrated Marine Observing System

- Nation-wide collaborative program designed to observe the oceans around Australia
- Assemble and provide free, open and timely access to data that support research on
  - Role of the oceans in the climate system
  - Interaction between major boundary currents, shelf environments and ecosystems
  - Preservation of biodiversity

IMOS Integrated **Marine Observing** System

IMOS Nodes and Facilities

- 6 nodes
- 13 facilities

Data diversity in the AODN

Marine research sectors

- government, universities, private industry, not-for-profit organisations

Marine science disciplines

- Physical and chemical oceanography, biology, biogeochemistry, modelling

Marine data sources

- observation data, model output, data products

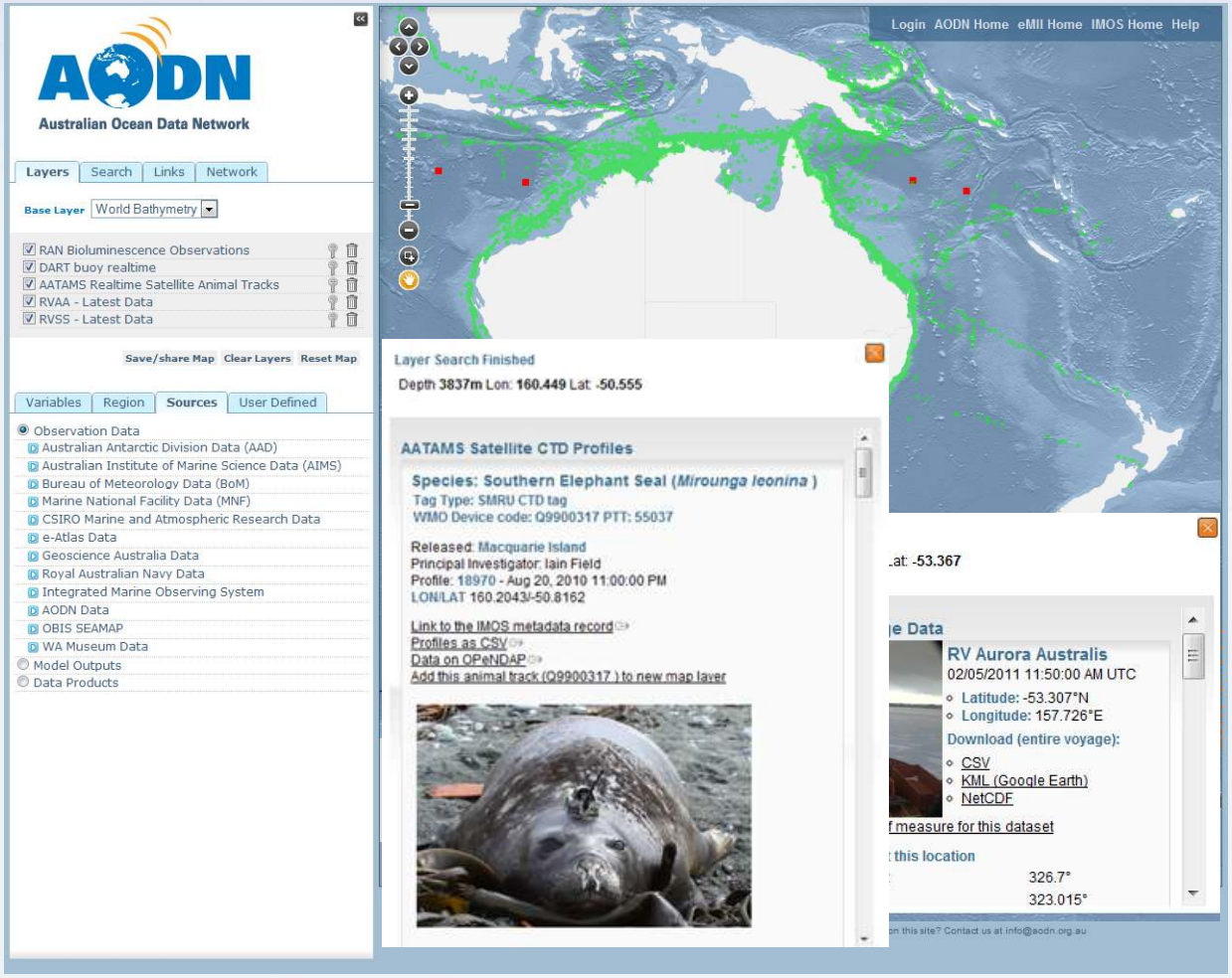
Marine data delivery

- real-time (e.g. tide gauges, tsunami buoys), delayed mode, archived / historic

Marine data formats

- gridded data (e.g. satellites, HF radar), image data, time series

Data discovery: AODN Portal



snapshot from http://portal.aodn.org.au

current functionality:

- map-based interface
- zoom & pan tools
- scale bar
- optional overview map
- base layers

- list of active layers
- visibility options
- time series animation

- list of data sources
- drop-down menu of available layers
- text-based search
- relevant info & links

AODN Data Policy

- Data provided to AODN is unencumbered, i.e. it is freely accessible
- Data is documented with metadata
  - ISO 19115 standard metadata
  - Metadata hosted by participating agencies
- All data is publicly available and licenced through Creative Commons BY licence

- Result: Increase discoverability, access and re-use of marine data

Summary and Conclusions.

Summary

- Benefits of data stewardship:
  - Provision of good quality and properly archived data
  - Timely distributions of data, as well as supplementary metadata and derived product
  - Easy discovery and access to data and products
  - Elimination of major barriers to efficient use and re-use of data

Conclusion

- Data derived from publicly-funded research should be available for public use
- Every research project should have a data management component in the budget
- A framework for stewardship of research data should involve all parties ensure the needs of researchers are met
- This will ensure that the maximum benefit is derived from research data

Thank you for your attention . . .