

INTRODUCTION TO DATA MANAGEMENT PLAN

Harrison Ong'anda

INTRODUCTION

- A data management plan should be implemented
- Data should be moved to public domain as soon as possible
 - *via* global archives
 - These archives support certain analysis
- Data acquisition, assembly, processing and dissemination often involve different groups

Definitions

- Acquisition: - data is collected in-situ or remotely
- Assembly: - combining data into a single data set. Metadata is created
- Processing: quality checks, duplicate resolution, unit conversions
- Archiving: ensuring long term storage and preservation
- Dissemination: - on-line offline data access

What is a data management plan

- describes the data that will be collected,
- the data management practices that will be used,
- the policies that will apply to the data,
- who will have custody of the data,
- who will have access to the data, and
- who will be responsible for the preservation of the data.

Initial planning

- What kind of data will be collected and how
- Is data coming in through a PI or a system
- Is data processing and availability time-bound
- Are there standards for metadata, processing and delivery
- Are there reasons for use restriction of the data
- What metadata must accompany the data

Data system coordination and planning

- Planners of the research activity
- Data management team
- Document a data management system
- Roles and activities of participants identified
- All project participants should be party to the system

Components of a data management plan

- Data to be collected and managed
 - description of the expected types of data, samples, collections and other materials that will be produced
 - Conductivity and temperature from moorings and shipboard CTD surveys
 - Currents from ADCP and moorings during the project

- *Documenting the data that will be collected:Metadata*
 - enable data discovery
 - Discover errors that were due to, for example, systemic instrument faults
 - generation of a cruise report (e.g. Cruise Summary Report form) and a sampling event log recording all instrument activities and deployments
 - “Marine Metadata Interoperability” web site (<http://www.marinemetadata.org>)
 - Vocabularies, ontologies etc

NAME OF DATA FIELD	REPORTING UNITS OR CODE	METHODS OF OBSERVATION AND INSTRUMENTS USED (SPECIFY TYPE AND MODEL)	ANALYTICAL METHODS (INCLUDING MODIFICATIONS) AND LABORATORY PROCEDURES	DATA PROCESSING TECHNIQUES WITH FILTERING AND AVERAGING
Salinity	% or	Nansen bottles	Inductive salinometer (Hytech moel S510)	N/A (Not applicable)
		STD Bisset-Berman model 9006	N/A	Values averaged over 5-meter intervals
Water color	Forel scale	Visual comparison with forel bottles	N/A	N/A
Sediments size	Ø units and percent by weight	Ewing corer	Standard sieves. Carbonate fraction removed by acid treatment	Same as "sedimentary rock manual," folk '65

- Data Formats

- What standards (e.g. vocabularies, units, analysis, etc.) will be used?
- Description of any non-standard formats and how to convert to standard formats
- Will the IODE/JCOMM recommended standards and best practice be used? (see:
<http://www.oceandatastandards.org> or
<http://www.oceandatapactices.org>)

- Data assembly and processing
 - complete set of operations that are applied to incoming data to transform it into a form suitable for the existing or planned archives
 - Who is responsible
 - Do all of the data arrive at once
 - Will all of the required metadata accompany the data
 - Are there real-time data
 - What are the processing steps (e.g. reformatting, unit conversions, etc)
 - What file naming conventions will be used (including standard use of vocabulary, date style, punctuation and numbers)

- Data archival and preservation
 - While data access may be restricted for some time to allow analysis and publication of results, all data should ultimately be archived in permanent archives
 - IODE National Oceanographic Data Centres (NODCs)
 - ICSU World Data System centres
 - following questions will need to be addressed
 - archive update procedures
 - disaster recovery
 - Where data is to be submitted (NODC, ???)
 - data retention period and any plans for disposal

- Data dissemination
 - Once data have been processed and inserted in the appropriate archive, they should be made available, preferably online
 - Need for data exchange policy
 - All data should be discoverable through metadata
 - The ultimate goal should be to make all data openly available as soon as possible and ideally within two years of acquisition
 - A new way of making data available is through data publishing/data citation
 - DOI -data sets can be referred to in a unique and consistent manner

Online data resources

- Argo profiles - (<http://www.coriolis.eu.org/>)
- Ocean color - <http://oceancolor.gsfc.nasa.gov/cms/>
- Altimetry - <http://www.aviso.altimetry.fr/en/home.html>
- Currents - <http://www.aoml.noaa.gov/phod/goos.php>
- Sea Surface Temperature - <http://iobis.org/node/639>

Example (SWIOFP/ASCLME)

- A data manager to be on-board familiar with data capture on-board systems
- must have back up capture systems and must ensure that the real-time capture of the data is not compromised
- must also ensure that the Chief Scientist or Cruise Leader has a backup copy of the data
- At the end of a cruise the data manager and chief scientist must cross check the data and go through a verification process
- Thereafter the Chief Scientist must ensure the data are given to the particular scientific component leaders and that **these data are also sent to the RMU for archiving.**