EarthScope Oceans Data Plan

Key Aspects

- Data Management Plan
    - Identification of Key Elements
        - Proprietary Period
            - Examples
                - OBSIP
                - PASSCAL
                - SEIS-UK
                - SeisMob (_FR)
        - Formats
            - Distinct aspects of EarthScope Oceans
                - Changing position, sample rate, etc.
        - Digital Object Identifiers and the FDSN
        - Data Access and Data Centers
            - Central Data Center
            - Federated Model
            - Some solutions

8.2. OBSIP Data Policy

http://www.obsip.org/data/

http://www.obsip.org/experiment-planning/obsip-instrument-use-policies-and-procedures/

All data collected on research projects by OBSIP will be archived at the IRIS Data Management Center and at the OBSIP IIC that provided the instruments. The PI must provide all necessary ancillary data (instrument locations, shot times, etc) to the OBSIP technicians prior to the end of the recovery cruise.

Upon completion of a cruise, OBSIP will provide one data set to the project PI. Depending on cruise logistics and quantity of data collected, this may occur several weeks after the ship has reached port. Requests for additional copies should be made at the time the instruments are requested in order to allow for the additional costs to be budgeted.

Short-deployment data will be archived in PASSCAL-SEGY format, and long-deployment data will be archived in SEED format. Data will be submitted to IRIS within six months of the end of the last recovery cruise.

In accordance with NSF requirements, the IRIS DMC data can be restricted to the PIs for 24 months from the date of the end of the instrument recovery cruise. However, for long-term broadband experiments, data from one instrument (selected by the PI) will be made publicly immediately as part of the Oceans Observing Initiative. After this 24 month period, all data will be available to any interested investigator.

PASSCAL Data Delivery Policy

https://www.passcal.nmt.edu/content/general-information/policy/data-delivery-policy

October 9, 2015

This policy outlines the guidelines and archiving requirements for data collected with IRIS PASSCAL equipment. IRIS's policy is that delivery of data to the IRIS Data Management Center (DMC) is an obligation of the Principal Investigator (PI). It is important to IRIS that the PI acknowledges this obligation and meets it within the required time frame. Failure to complete this requirement not only deprives the community of a valuable data resource, but also may jeopardize future requests to borrow IRIS equipment. IRIS expects data delivery while the experiment is in the field (for long term deployments), or immediately at the conclusion of the field experiment. At the request of the PI, the data will remain confidential for a period of 2 years after the end of the fieldwork, but will then become part of the open data archive to benefit the entire seismological community. The Data report and all QC metrics will be immediately open and available to ensure data sharing after the moratorium period is facilitated.

The equipment in the PASSCAL facility represents a significant community resource. In addition, the data collected by the PI community using the PASSCAL equipment has tremendous value past the specific scientific goals of the original experiment. Abiding by this data policy ensures the continued contribution of the IRIS PASSCAL facility to seismological studies.

Data Evaluation Report and Data Object Identifiers (DOI)

IRIS has worked with the editors of the Seismological Research Letters to allow publishing data reports on experimental data sets to introduce the dataset and experiment to the community, and to provide a citable reference for others who will eventually use your data. Guidelines for these reports can be found at www.iris.edu/TO_BE_DETERMINED

IRIS Data Services has worked with the FDSN to develop a procedure to create DOI's for data from temporary networks. DOIs promote the citation of a PI's work as well as the facility support for each experiment. More information on DOI's can be found at: http://www.fdsn.org/services/doi/ and temporary experiments can request a DOI from the same page where they request a temporary network code. Look for the section titled Network Citation.

The Data Evaluation Report is due immediately after the completion of the experiment. This form can be found from the PI's Home Page, or while viewing an experiment; the PI must be logged in to access this form.

Data

The actual format of the data and the amount of data depend upon the type of experiment. Most PASSCAL experiments are classified as passive source, controlled source, or more recently some combination of controlled and passive recording (mixed-mode).

Passive Source Experiment Data

PI's conducting a passive source experiment will provide all of the data and appropriate metadata from their experiment to the IRIS DMC for archive in SEED or stationXML format. It is expected that the PI will coordinate with the PASSCAL Instrument Center to arrange delivery of the data to the DMC on a continuing basis during the experiment (after service runs, as appropriate, or in real time if available). The final data shall be delivered to the DMC immediately after the experiment is over. The DMC will make the data available only to the PI or his

designated representative for a period of up to two years after the completion of the experiment. This exclusive access can be waived by the PI if open data can be immediately made available. After the proscribed moratorium (no more than 2 years after the conclusion of the experiment), the data will be made freely and openly available to the public.

Controlled Source Experiment Data

Controlled source experiments are more frequently recording continuous data encompassing the controlled source time windows. To best preserve the entire data record the accepted data format for continuously recorded data is PH5, SEED, or stationXML. For conventional, shot windowed data , data should be delivered in SEG-Y format. In addition it may be delivered in PH5 format as well.The data should include all of the necessary information on the geometry of the experiment (metadata) and they should be corrected for all known timing problems.

Mixed-Mode Experiment Data

More recently, experiments are being proposed and deployed that involve both controlled source and passive source experiment components. In these cases, we are working towards a single format (PH5) to allow ingestion, reformatting and distribution of the particular components of the data in the appropriate formats as required by the data users and this may be used in its proto-form. However, at the time of this report, the operational release of PH5 is not complete, so it may be used intermittently. Until this becomes a standard, or a new standard is found, it is acceptable to provide continuous data in SEED and shot data in SEG-Y, as appropriate.

Non-Standard

There will always be some experiments that do not fit directly into one of the above categories. In those cases the exact form of the data delivery will be negotiated between the PI, Data Services and PASSCAL.

Proprietary Data

Data of all types should be delivered to the DMC, in the appropriate format with complete metadata, as soon as possible and well before the general release of the data. The DMC will only allow access to the waveforms to the PI and others designated by the PI during the moratorium (not more than 2 years from the completion of the field experiment). Access will be by password that will be provided by the DMC to the PI. The PI can share the password with anyone he/she wishes. The PI will be notified when anyone registers for access to a proprietary dataset. Information about the experiment contained in the metadata and data quality characteristics will be made publicly available during the experiment, only waveform data will be limited in distribution during the proprietary period.

All passive experiments with five or more stations will designate at least one station as an “open station”. The data from the “open station/s” will be made available to the public immediately upon being archived.

Support Available from IRIS

Every PI and their field staff need to attend training at the IRIS PASSCAL Instrument Center (PIC) prior to their experiment where they will receive training on instrumentation and installation techniques, and the handling of the data and development of the station metadata. Teams will be taught how to archive their data using PASSCAL field computers and software that is provided as a part of the experiment equipment set. Every field computer has the software necessary to accomplish the data delivery task and the PASSCAL Instrument Center has personnel who can provide assistance to the PI during and after the experiment. We encourage PI's to interact with PIC staff before and after service runs and immediately after the experiment conclusion to allow refreshers on data handling and ensure immediate

ingestion of the data to the DMC. This will not only get help PI's with data handling training that may have occurred months to years ago, but also assure that the data delivery requirement to the DMC is met in a timely manner. Please note that many new tools for assessing data quality are now available for experiment datasets that can only be accessed once the data are uploaded.
The PI is encouraged to utilize these resources at all stages of the work. In all cases, however, the ultimate responsibility for delivery of the data rests with the Principal Investigator. The PI must ensure that adequate resources are budgeted to accomplish this task.

Data Quality Assessment

All data submitted to the IRIS DMC will be run through the data quality assessment tools (MUSTANG) and the data quality metrics will be made open without delay for data quality evaluation - ie quality metrics are not subject to the data moratorium described above. MUSTANG data may be accessed through standard web services tools or through the use of the MUSTANG client interfaces (like MUSTANGDataBrowser or LASSO).

Experiment Completion and Delinquent Datasets

A PASSCAL data submission is not considered complete until both the PASSCAL Program Manager and the Director of Data Services certify that the information contained in the report is sufficient to allow other members of the community to utilize the data. IRIS will not certify that it has received data from any PI until the data submission is deemed useable.
PIC staff will remind PI's that data are due beginning 6 months prior to the data becoming delinquent (ie, more than 90 days after completion of the experiment) and will offer guidance on how to ensure your data are properly uploaded to the DMC. If your dataset becomes delinquent, the Program Manager and PASSCAL Standing committee chair will work with you to determine how best to make your data available to the community to ensure the program resources are being adequately allocated. We certainly hope that this escalation will not be required, but it is in the best interest of the entire IRIS community to ensure that when IRIS PASSCAL equipment is used, that the data are archived in a useable format as required.

Exemptions for Strictly Educational Use Datasets

If you are using equipment from the PASSCAL facility for strictly educational training on the instrumentation and are not using the equipment for observing a science target, PASSCAL management may grant an exemption from the data distribution policy if there is no useful information to be shared among the seismological community. This is at the discretion of the PASSCAL program manager. Of course, if you are willing to share the data, all data sets are welcome. All exemptions must be request of the PASSCAL Program Manager and are to be considered due to IRIS unless the Program Manager grants and exemption.
There may be other needs for exemptions, including complete failure of the experiment to capture data, experiment cancellation, or no useable data were captured. However, all of these need to be approved by management with no exceptions.

This policy is effective as of Sept 29, 2015 and is subject to change and revision as needs dictate. For updated versions of the policy and additional information on data delivery see the PASSCAL and DS pages on the IRIS web site. There is also an IRIS document detailing the policies regarding release of restricted data.

SEIS-UK

http://seis-uk.le.ac.uk/equipment/downloads/data_management/seisuk_data_manage2.pdf

2 Data Archiving with IRIS DMC and Network Codes

All data obtained with SEIS-UK equipment is archived with the IRIS DMC at the end of an experiment as soon as it has been converted to miniseed (passive) or SEGY (controlled source). The IRIS DMC archive has restricted access for 3 years after the fieldwork completion date. The PI will be notified of the access procedures by IRIS. After 3 years the data becomes publicly available unless the PI has reasonable grounds to delay the release.

In order for data to be archived with IRIS all experiments require a unique network code, assigned by FDSN for passive arrays or IRIS for controlled source experiments. It is the responsibility of the PI to request a network code from the appropriate source:

• Passive: http://www.fdsn.org/getcode.html

• Controlled source: http://www.iris.washington.edu/data/assembled_ID.htm
The PI should also fill out the mobilisation form, preferably before or soon after deployment:

• http://www.iris.washington.edu/stations/mob.htm

(where the data shipment date will be approximately 6 months after the end of the fieldwork period).

After the fieldwork has been completed a demobilisation form should also be completed:

• http://www.iris.washington.edu/stations/demob.htm

Details on these forms as to volume sizes etc do not need to be precise, they are used to give IRIS an idea of how much data they will need to prepare for and can be updated when precise figures are known. However as a rough estimate, sampling at 100sps will generate approximately 60Mb per day per sensor depending on the noise conditions of the site.

2.1 Once your data have been sent to IRIS

IRIS DMC have created a virtual network of SEIS-UK stations archived there. This allows users rapid access to network and station meta-data available at IRIS, including network maps, data availability and instrument responses:

• http://www.iris.washington.edu/vnets?vnet=_SEIS-UK

A map of all SEIS-UK stations with data archived at IRIS can be found here:

• http://www.iris.edu/gmap/_SEIS-UK

A list of ALL temporary deployment network codes can be found here (Select ”Short Description Query - View Results”):

• http://www.iris.washington.edu/SeismiQuery/tempNets.phtml

You can easily get a GOAT plot of your data holdings at IRIS here:

• http://www.iris.edu/data_available

- or try the following for a SEIS-UK example from Afar (takes a few seconds) then manipulate the URL for your network:

• http://www.iris.edu/data_ava
ilable/ZF/*/*/BHZ?timewindow=2007-2009&both=on&byday=off

SEISMOB-FRhttps://sismob.resif.fr/spip.php?article24&lang=en

or in this landing page

https://sismob.resif.fr/spip.php?article21&lang=en

Question

What is Restricted Data?

Answer

An important function of the DMC is to collect data, sort it, and make it readily available to a wider research community, especially with short-term temporary experiments (such as from PASSCAL, SEIS-UK, OBSIP, SISMOB-FR, and others) that have multiple partners conducting research, and don't want to be burdened with distributing data while the research is underway. For this reason, there are data at the DMC that appear to be available in the holdings, but are restricted.

Data from PASSCAL (and some other temporary deployments) are subject to a two-year moratorium, allowing the Principal Investigators (PI's) who carried out the scientific study the opportunity to report and publish results with exclusive access to their data. As a result, access is allowed only to designated investigators, where the experiment PI's have sole authority to grant such access. At the end of the moratorium, barring extenuating circumstances, the data are released to open public access.

There are four generalized policies for what data are restricted:

A) IRIS/PASSCAL

The policy is that 2 (two) years after the last instrument is pulled from the field, data should become open.

B) OBSIP

The policy is that 2 (two) years after the last instrument is pulled from the field, data should become open.

C) SEIS-UK

The policy is that 3 (three) years after the last instrument is pulled from the field, data should become open.

D) SISMOB-FR

The policy is that 3 (three) years after the last instrument is pulled from the field, data should become open.

IRIS Data Management Center Policies

https://www.iris.edu/hq/programs/ds/#policies

Board-Approved Policies

- Large Data Temporary Deployments (LDTD) (v 1.2)
- Data Acceptance Policy for Seismological Data from Permanent Networks (v 3.3)

Other IRIS DMC Policies or Practices

- Data Provider Agreement (v 1.3)
- Redistribution of IRIS Data (v 2.0)
- Release of Restricted Data (v 1.0)