Data Acceptance Policy for Datasets from Temporary Deployments[1]
Policy Version 2.0
BoD Approval on 18 April 2019

Synopsis

IRIS and IRIS Data Services are fully committed to serving the needs of the seismological community by archiving data from temporary deployments of seismometers, whether these deployments are National Science Foundation (NSF) funded or otherwise. This policy lists the data formats that can be accepted, defines data volumes that will be accepted cost-free, and establishes a fee structure for accepting large datasets that depends on the size of the dataset, the provider, and the archive resources needed. For the vast majority of temporary deployments there will be no cost to the data submitters. For experiments that generate data in excess of the volumes that are accepted cost-free, the fee structure ensures that the datasets are archived and managed in perpetuity. In conjunction with this policy, IRIS provides procedures for data submission and identifies resources to Principal Investigators (PIs) who wish to have their seismic datasets archived at the IRIS Data Management Center (DMC).

Formats

The IRIS DMC accepts data in SEED, PH5, SEG-Y, SEG-D, and GeoCSV[2]. Use of the SEED and PH5 formats is strongly encouraged as these formats are supported by powerful DMC tools for discovery and access through web services.

- **SEED data:** SEED data are supported by IRIS DMC web services for discovery and access. Metadata must be supplied in the FDSN SEED metadata format (either dataless SEED or StationXML)[3].

- **PH5 data:** PH5 data are supported by IRIS DMC web services for discovery and access. The PH5 format is particularly suited to data with a large number of nodes and consequently larger volumes (several tens of terabytes).

- **SEG-Y and SEG-D data:** SEG-Y and SEG-D data are _not_ supported by IRIS DMC web services for discovery and access. To take advantage of these tools, data in SEG-Y or SEG-D data may be converted to PH5 for archival, but some information may be lost[4].

- **GeoCSV data:** GeoCSV data are _not_ supported by IRIS DMC web services for discovery and access. This format is intended for low volume tabular data to provide a richer set of metadata that is understandable by humans and machines. These data will be managed as an assembled dataset. The user must provide additional information that adds utility to the data set.

---

[1] This document will be reviewed annually at spring meetings of the IRIS Data Services Standing Committee.

[2] http://geows.ds.iris.edu/documents/GeoCSV.pdf

[3] Knowledgeable consultants are available to provide this service for a fee. The PI is responsible for contracting with an external consultant directly.

[4] Conversion is available for a fee.

Support for SEED and PH5 Data Archiving

To help PIs produce data in the formats that the IRIS DMC can ingest, up-to-date resources are provided on the IRIS webpages. The IRIS DMC staff are also available to provide archiving information and advice on a limited basis.

- Submitting data in SEED format:
  - https://ds.iris.edu/ds/nodes/dmc/data/submitting/

- Building metadata for the SEED format:
  - https://ds.iris.edu/ds/nodes/dmc/software/downloads/PDCC/3-8-1/
  - https://www.passcal.nmt.edu/content/passcal-announces-software-release-nexus-simple-tool-creating-seed-meta-data

- Generating metadata and archiving data in PH5
  - https://www.passcal.nmt.edu/content/data-archiving/documentation/active-source

- DMC Contact and Email for archiving SEED or PH5 formatted data:
  - engine_room@iris.washington.edu

Procedure for Data Provider

Early planning can save PIs time and effort when large datasets from their experiments are expected to be archived at the IRIS DMC. The procedures begin during proposal preparation so that the PI can avoid unexpected costs for data archival.

1. Use the form provided on the IRIS website and coordinate with the IRIS DMC to estimate costs (if any) of data archival as per this policy. These costs should be included in proposals to funding agencies.

   https://ds.iris.edu/ds/nodes/dmc/tools/archive-pricing/

2. Sign a *Data Provider Agreement* with IRIS that outlines responsibilities for each partner and details the services provided by the DMC. The agreement can be found on the IRIS web pages:

   https://www.iris.edu/hq/files/programs/data_services/policies/Data_Provider_Agreement-V1.5.pdf

3. Pay any archival fees to IRIS prior to sending data.

4. Prepare the data for archival by generating metadata, converting formats, etc.

5. Send data to the IRIS DMC for archival.

Large Dataset Fees

The cost to manage large datasets from temporary experiments is not included in the core funding the DMC receives through the cooperative agreement between IRIS and the NSF. For this reason, the DMC charges fees for the initial costs that will be incurred at the DMC for archiving large datasets and for some portion of refreshing hardware in the future. The fee is paid only once and is based on the size of the dataset, the format and the data provider. The DMC assumes the costs for the on-going management of these data within its existing budget for the out years.

IRIS Data Services incurs costs for its primary and backup facilities that depend on the data format. The storage systems for SEED data are different from the storage systems used for data formatted in PH5, SEG-Y, SEG-D and GeoCSV (which are particularly suited to managing data with very large data volumes from larger deployments).

IRIS provides significant discounts for data that are supplied by NSF-funded research and US non-profit educational/research institution[5] or U.S. federal, state, or similar international agencies. These data providers pay the initial hardware costs only for data in excess of a data volume threshold; for-profit companies such as US or international oil companies, mining companies, utility companies or geotechnical companies pay all of the initial hardware costs. As well, the discounted rate includes half of the cost of the first subsequent hardware renewal, while for-profit companies pay for the first two hardware renewals.

The following fees apply according to the various data storage systems, dataset volumes and data provider:

Fees[6] per Terabyte for Archiving Temporary Deployment Experiment Data[7]

	Not-for-profit Orgs & Gvmts	For-Profit Companies
The first 10 terabytes of SEED data	$0	$2,250
SEED data in excess of 10 terabytes	$1,125	$2,250
The first 40 terabytes of PH5 and other data	$0	$480
PH5 and other data in excess of 40 terabytes	$240	$480

The effort to create SEED metadata or to convert SEG-Y or SEG-D data to PH5 can be time consuming. Consultants provide these services for a fee and PIs who wish to pay for these services must contract directly with the consultant. Typical rates are $2,500 per 100 channels for generating SEED metadata and $1,000 per terabyte to convert SEG-Y or SEG-D data to PH5. While IRIS may recommend consultants that they believe are capable and willing to do this work, IRIS does not guarantee the quality of work performed by any consultant hired by the PI.

Examples

An NSF-funded temporary deployment that produces 11 terabytes of SEED data would be assessed a fee of $1,125 (10 terabytes at no cost and 1 terabyte at $1,125 per terabyte).

A company provides 50 terabytes of SEG-Y data would pay $24,000 (50 terabytes at $480 per terabyte).

A government-funded experiment that generates 100 terabytes of SEED data would be assessed a fee of $101,250 (10 terabytes at no cost and 90 terabytes at $1,125 per terabyte).

A government-funded experiment that generates 100 terabytes of PH5 data would be assessed a fee of $14,400 (40 terabytes at no cost and 60 terabytes at $240 per terabyte).

---

[5] IRIS membership is *not* required.

[6] Fees may vary depending on the price of hardware.

[7] An experiment could consist of several deployments.