**IRIS Data Services**

**Tim Ahern**

**Director of Data Services**

The central component of IRIS Data Services (DS) is the IRIS Data Management Center in Seattle, Washington. The DMC relies on other DS components in Albuquerque, La Jolla, University of Washington, LLNL, and Almaty, Kazakhstan to realize its full functionally but the heart of the DS is the DMC. The major CI components are in place at the DMC. We run a fully functional Auxiliary Data Center that is unmanned at LLNL.

1. ***A brief description of the facility, its science mission, and the community (including size, make up – number of individual, number of institutions, etc.). Please include a URL for more information.***

The IRIS DMC is a domain specific facility that meets the needs of the seismological community both within and outside the US. The DMC facilitates science within our domain but does not DO any science. Our science mission can be found in our strategic plan: <http://www.iris.edu/hq/files/programs/data_services/policies/Strategic_Plan_v7.pdf>

Our science community numbers in the thousands worldwide.

**Mission:** *To provide reliable and efficient access to high quality seismological and related geophysical data, generated by IRIS and its domestic and international partners, and to enable all parties interested in using these data to do so in a straightforward and efficient manner*.

IRIS is university consortium with approximately 125 members (US academic institutions with graduate degrees in seismology) and roughly the same number of foreign affiliates scattered all over the globe. We are a 501c3 Delaware corporation. We distribute primary data to roughly 25,000 (3rd level IP address) distinct users or IP addresses per quarter from roughly 12,000 distinct organizations (2nd level IP address). IRIS ingests roughly 75 terabytes of new observable data per year and we project we will more than one petabyte in 2017.

1. ***A description of the key products/services of the facility (data, software, services, etc.)?***

IRIS’ primary products are (Level 0, raw and Level 1 quality controlled) time series data. The time series come from roughly 30 types of sensors deployed on/in the ground, in the water column or water bottom, and in the atmosphere. IRIS also produces Level 2 derived products, and manages community developed Level 2 and higher products. (See http://ds.iris.edu/spud/ ). Level 0 and 1 products are fully documented (metadata) time series data from geophysical sensors distributed globally generated form NSF and other national and international sources. We distribute roughly one petabyte of level 0 and 1 data per year.

Figure 1 shows volume of time series data shipped from the IRIS DMC to end users and or monitoring agencies since 2001. Major types of shipments include legacy requests in the blue, real time data distribution in the red, and web service distribution in the purple.

IRIS also produces a great deal of community software and offers both IRIS developed and community developed software and tools in Redmine and GitHub repositories. IRIS develops and maintains specific client applications for accessing and working with IRIS data.

All IRIS data assets (Level 0-3) are available through service APIs. Some of the APIs have been adopted internationally (FDSN web services) and other APIs are IRIS developed and maintained and not yet adopted internationally. (see <http://service.iris.edu>). IRIS also maintains comprehensive documentation and is also the source of documentation for the SEED format, which is the international seismological domain format. (www.fdsn.org)

 ***3. A brief description (including a figure) of the facility CI (e.g., its architecture, key services/components, underlying infrastructure), how it is deployed/distributed, and its operation. What is the median age since deployment of the key CI components***.

The figure could be provided at a later date, it is very complex obviously and difficult to provide something at a high enough level as to be useful.

The IRIS DMC operates a primary data center in Seattle as well as an unmanned, fully functional Auxiliary Data Center (ADC) in Livermore California. Major components of CI at the DMC and ADC consist of the following

* **Storage** – IRIS operates large volume Hitachi RAID systems that emphasis storage over performance. We improve performance by indexing the RAID contents in a PostgreSql DBMS. We have roughly 700 terabytes of storage RAID at both the DMC and the ADC. We also operate high performance RAID systems made by NetApp both for reception of real time data and PostgreSql database transactions.
* **Servers**- IRIS runs virtual servers on physica Dell Servers. Virtualization software is VMWare.

IRS operates Forcepoint Firewalls and A10 Load Balancers. Load Balancers are configured so that a failure at the DMC or the ADC does not remove outsides user’s access to services,

* **LANs-** We run 10 gigabit/second LANs sometimes in parallel to form a data backbone internal to the DMC and ADC. We connect to the Internet through the University of Washington.

Storage access to observational data has been abstracted through web services for both internal and external use. Access to data is transitioning from direct SQL access to abstractions thorugh web services. We are very close to running a SOA for both internal and external access.

Our goal is to refresh all major computational and storage hardware infrastructure every four years. Budget pressues sometimes pushes this to 5 years.

We are currently testing operating our software in XSEDE and AWS to see if this is viable.

.