

The Implementation of SCEC Data Management

Thanakij Pechprasarn, Patrick Small, Scott Callaghan, Kevin Milner, and Thomas H. Jordan

Southern California Earthquake Center
University of Southern California
3651 Trousdale Parkway
Los Angeles, CA 90089
USA

A common problem facing researchers is promoting accessibility and distribution of research data products to the large scientific community. The SCEC Data Management system (SDM) has been created to improve accessibility of the earthquake simulation results. The SDM has two components: the CyberShake API and the CyberShake Seismogram Retrieval Tool.

The CyberShake API provides a Python API to the CyberShake site database. Besides the API for developers, the CyberShake API also provides a collection of command-line scripts to perform commonly used queries. The CyberShake API can be used to find information about sites, given an interested map. The CyberShake API also can query a maximum peak amplitude in the given map or site. The returned data includes source, rupture and rupture variations. Moreover, the CyberShake API can construct the names of seismogram and SRF files from source, rupture, and rupture variations information. The names will be used by the CyberShake Seismogram Retrieval Tool to identify the physical locations of them.

The CyberShake Seismogram Retrieval Tool (SRT) allows seismologists to recover the synthetic seismogram file generated for any rupture variation used in the computation of the peak amplitudes for a site. Using the CyberShake SRT, the scientist can identify a set of rupture variations that are of interest for one or more sites, then employ the SRT to retrieve those seismograms and save them on a local filesystem. The output seismograms can be saved in a variety of formats, including SAC and KBO. To perform this transfer, the SRT first queries a Globus Replica Location Services (RLS) at SCEC to determine which data collection(s) contain the desired seismograms. Then the SRT extracts them from that repository by a remote job submitted to that host via the Grid Resource Allocation and Management (GRAM) and transfers the set of files to the local host with GridFTP.