≋USGS

science for a changing world

An Overview of the National Earthquake Information Center Acquisition Software System, Edge/Continuous Waveform Buffer

By John M. Patton, David C. Ketchum, and Michelle R. Guy

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015

Contents

Figures

Abbreviations

ANSS	Advanced National Seismic System
AQMS	ANSS Earthquake Monitoring System
ASL	Albuquerque Seismic Labs
CD1.1	CTBTO International Data Center Continuous Data Protocol Version 1.1
COSMOS	Consortium of Organizations for Strong Motion Observation Systems
CTBTO	Comprehensive Nuclear-Test Ban Treaty Organization
CWB	Continuous Waveform Buffer
DAS	data acquisition system
DMC	data management center
EW	Earthworm
FTP	file transfer protocol
GB	gigabyte
GSE	Group of Scientific Experts
GSN	Global Seismographic Network
IDA	International Deployment of Accelerometers
IRIS	Incorporated Research Institutions for Seismology
ISI	IDA System Interface
JDK	Java Development Kit
LISS	Live Internet Seismic Server
NEIC	National Earthquake Information Center
NGP	USGS National Geomagnetism Program
no-db	no-database
OEM	original equipment manufacturer
RRP	ring replication protocol
RTPD	Ref Tek Protocol Server
SAC	Seismic Analysis Code
SCREAM	Seismometer Configuration, Real-time Acquisition and Monitoring
TCP	transmission control protocol
UDP	user datagram protocol
USGS	U.S. Geological Survey
WFDISC	Waveform Disc

An Overview of the National Earthquake Information Center Acquisition Software System, Edge/Continuous Waveform Buffer

By John M. Patton,[1] David C. Ketchum,[2] and Michelle R. Guy[1]

Abstract

This document provides an overview of the capabilities, design, and use cases of the data acquisition and archiving subsystem at the U.S. Geological Survey National Earthquake Information Center. The Edge and Continuous Waveform Buffer software supports the National Earthquake Information Center's worldwide earthquake monitoring mission in direct station data acquisition, data import, short- and long-term data archiving, data distribution, query services, and playback, among other capabilities. The software design and architecture can be configured to support acquisition and (or) archiving use cases. The software continues to be developed in order to expand the acquisition, storage, and distribution capabilities.

Introduction

The Edge and Continuous Waveform Buffer (CWB) is a software system that handles the acquisition, storage, and distribution of seismic waveform data when those data reach the edge of the U.S. Geological Survey (USGS) National Earthquake Information Center (NEIC) system infrastructure.

The Edge software acquires data from many sources by using multiple protocols and many different formats. It supports robust, automated, and routine data acquisition, including the processing of late-arriving data and the capability to fill gaps. All acquired seismic data are immediately archived for a fixed number of days in an organized and uniform format. The Edge also includes efficient distribution capabilities that forward any acquired seismic data to cooperating networks, the Incorporated Research Institutions for Seismology (IRIS) data center, and NEIC real-time processing.

A CWB is a specific configuration of the Edge software that provides an infinitely deep archive of seismic data, bound by available disk space. The CWB configuration includes robust and numerous query capabilities to support research and processing activities.

---

[1]U.S. Geological Survey
[2]STW Software Inc.

Background and Motivation

The NEIC is the Nation's primary source for rapid and historical earthquake information on felt earthquakes in the United States and significant earthquakes worldwide. Federal and State government agencies, the national and international scientific communities, news media, and the general public all rely on information and products produced by the NEIC.

This mission is fulfilled, in part, by acquiring seismic waveform data from a well-distributed set of high-quality seismograph stations. The NEIC acquires these data from nearly 100 seismic networks, including the Global Seismographic Network (GSN), U.S. Advanced National Seismic System (ANSS), European Geofon, International Deployment of Accelerometers (IDA), MedNet, and many national networks. This waveform data acquisition spans many vendors and software collection systems from around the world.

The Edge/CWB project started as an effort to encompass all of the legacy software for data acquisition written by the NEIC and replace it with a single software package. This same software package would also act as the archival system for all waveform data; use of the software in this configuration is termed a "CWB." The project has the following additional software requirements:

- Portability via Java or other platform-independent programming languages
- Modular data acquisition that is independent of data format
- Immediate storage upon acquisition
- Organized and uniform internal storage format
- Support for out-of-order data and automated gap filling
- Ability to retrieve stored data in a robust and efficient manner
- Minimal interdependencies with other systems
- Real-time distribution of data to cooperating organizations
- Ease of writing new data acquisition modules

Typically, the Edge/CWB software is used as the acquisition and (or) archiving solution within an organization, as figure 1 shows. However, the CWB component of the system can be used without enabling the Edge acquisition software.
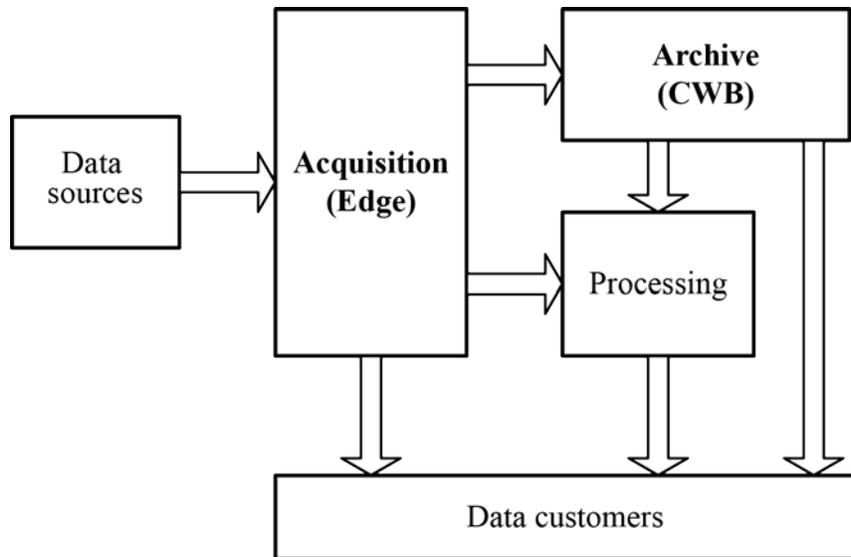
Figure 1. Diagram showing the location of the Edge/Continuous Waveform Buffer (CWB) software within the National Earthquake Information Center seismic analysis system. The Edge/CWB is used for acquisition, as well as import, distribution, and archiving.

Capabilities

The Edge/CWB software has extensive capabilities in the areas of direct station acquisition, data import, short- and long-term archiving, data distribution, query services, and playback.

Acquisition

The Edge acquires data directly from seismic stations via several data acquisition systems (DASs) by using a mixture of locally developed and vendor-provided software. The NEIC uses two different DASs when acquiring seismic data critical to the NEIC mission, the Quanterra Q330 system (for ANSS backbone stations) and the Albuquerque Seismic Labs (ASL) Slate/Q330 system (for GSN network stations). Other supported DASs include Reftek RT–130; Kinemetrics Basalt; Kinemetrics K2; Guralp Seismometer Configuration, Real-time Acquisition and Monitoring (SCREAM); and NetQuakes. Some of these DASs are used primarily for temporary seismic network deployments. Other DASs systems are used to acquire triggered strong-motion data.

Import

In addition to directly acquiring data from stations, the Edge software supports importing data from other organizations by using a wide range of protocols including Earthworm Import and Earthworm user datagram protocol (UDP) methods, Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) CD1.1, IDA System Interface (ISI), Miniseed ring replication protocol (RRP), IRIS Seedlink, Ref Tek Protocol Server (RTPD), and Guralp SCREAM 4.5.

Archiving

The Edge/CWB software provides short- and long-term data archiving. The archive supports out-of-order, back-filling, and gap-filling data processes to produce a comprehensive and complete archive. All seismic data acquired or imported are converted to the Miniseed format and stored in daily data files with corresponding index files. The total amount of waveforms a data/index file can support is up to 8,000 discrete channels, with sizes on disk exceeding 25 gigabytes (GB).

Distribution

Edge/CWB supports robust distribution by using many of the common protocols in use for seismic data exchange. The following export protocols are available: Earthworm Export, Earthworm UDP, IRIS Seedlink, Live Internet Seismic Server (LISS), EdgeBlockServer, CD1.1, and RRP.

Playback

The software supports playing any set or subset of data in the Edge/CWB archive out to other systems via Earthworm UDP for purposes of testing, evaluation, or portable deployment processing. The data can be played back at rates from real time to many times real time, depending on network and system capacity.

Query Services

When configured appropriately, the Edge/CWB software offers extensive services for requesting data from the data storage archives. Additionally, query services allow requesting of data from the Edge memory buffer to support near-real-time data processing. By using the provided CWBQuery Java client software, waveform data can be written to files in the Miniseed, Seismic Analysis Code (SAC), Waveform Disc (WFDISC), Group of Scientific Experts (GSE), and Consortium of Organizations for Strong Motion Observation Systems (COSMOS) formats. The software also supports the Earthworm and Winston Waveserver interfaces.

Configuration

The configuration and maintenance of an Edge/CWB installation is designed for simplicity. The Edge/CWB software stores all configuration files/information in a centralized location, either a MySQL database or a no-database (no-db) flat-file system. All configuration maintenance is performed by using a robust suite of platform-independent graphical user interfaces written in Java.

Platform Independence/Portability

With the exception of a few original equipment manufacturer (OEM)–provided libraries, all of the Edge/CWB software is designed to support platform independence. This is accomplished by writing the software in the Oracle Java programming language (Java

Development Kit [JDK] 6.0 or higher). Additionally, the software supports using open-source database management systems, including MySQL and Postgres, as well as a no-db option.

System Architecture and Design

Edge Node Configuration

An Edge installation is built around one or more Edge nodes. An Edge node is a set of Edge functions grouped according to operational needs. A simplified diagram of an Edge node is depicted in figure 2. Every Edge node consists of an instance of the EdgeMom application, which manages the Edge functions, and an associated configuration file.
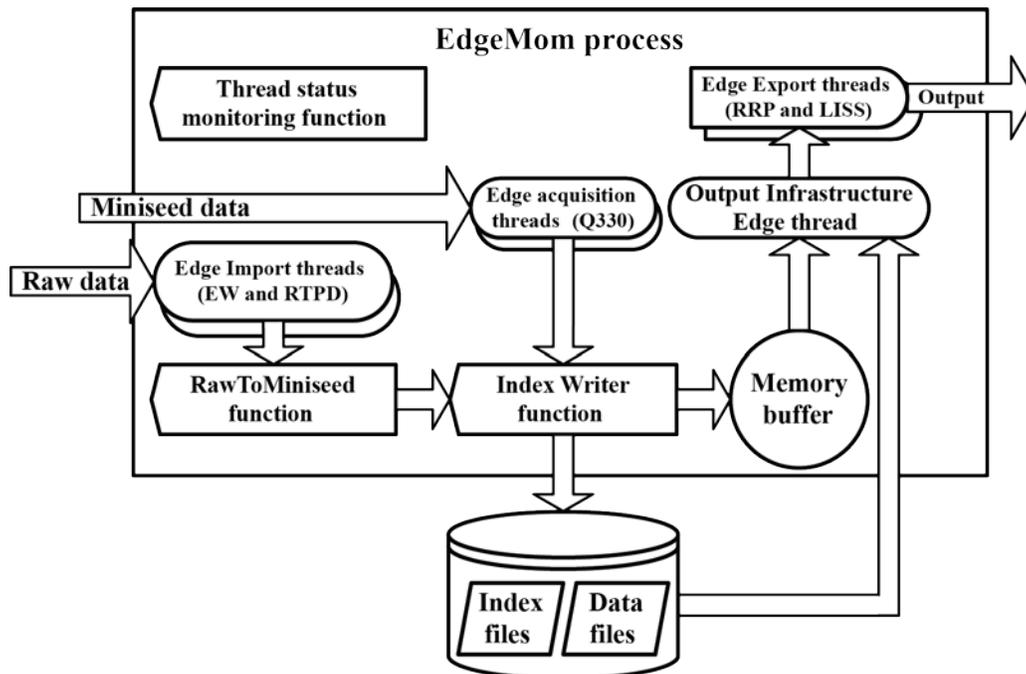

Figure 2. Diagram showing the configuration design of an Edge node, including the main EdgeMom process, the various Edge threads used for input and output, and the functions used for data conversion and archiving. (EW, Earthworm; LISS, Live Internet Seismic Server; RRP, ring replication protocol; RTPD, Ref Tek Protocol Server)

By design, the EdgeMom application starts and manages processing threads called Edge threads that implement the desired functionality. The EdgeMom application also hosts software functions to perform logging and health monitoring of the processing threads. The EdgeMom application also hosts functions for data conversion, memory management, and file writing. Edge threads are threads managed and monitored by EdgeMom. Edge threads handle all the input processing, whether acquisition directly from stations, or data received via any of the various supported import methods. All input Edge threads call the EdgeMom functions to write the received data to the data/index files as well as to the EdgeMom memory buffer. Function examples include the RawToMiniseed function for handling non-Miniseed input formats, such as

Earthworm or RTPD, and the Index Writer function for input that is already in Miniseed. This use of EdgeMom ensures proper data conversion and file synchronization.

An Edge node limits the amount of data stored in the data/index files. An Edge installation may customize this limit in order to support functionality such as back-filling gaps and allowing out-of-order data. The default value for this option is 10 days, which at the NEIC is a reasonable timeframe to ensure 90-percent or greater data completeness.

The Edge design includes an Output Infrastructure thread that supports specific output paths and methods. This thread is an Edge thread managed by EdgeMom and receives acquired data from the memory buffer. The thread dispatches the data to subthreads that perform the actual output. Various subthreads exist to handle protocols such as RRP, LISS, Earthworm Export, and others.

Continuous Waveform Buffer Configuration

In addition to the Edge node installation, the Edge software can also be configured as a long-term archive, or CWB. A simplified diagram of a CWB configuration is shown in figure 3. A CWB typically omits the input Edge threads found in an Edge node configuration, in favor of a single input method, usually the Edge RRP (although any of the supported Edge input methods can be used).
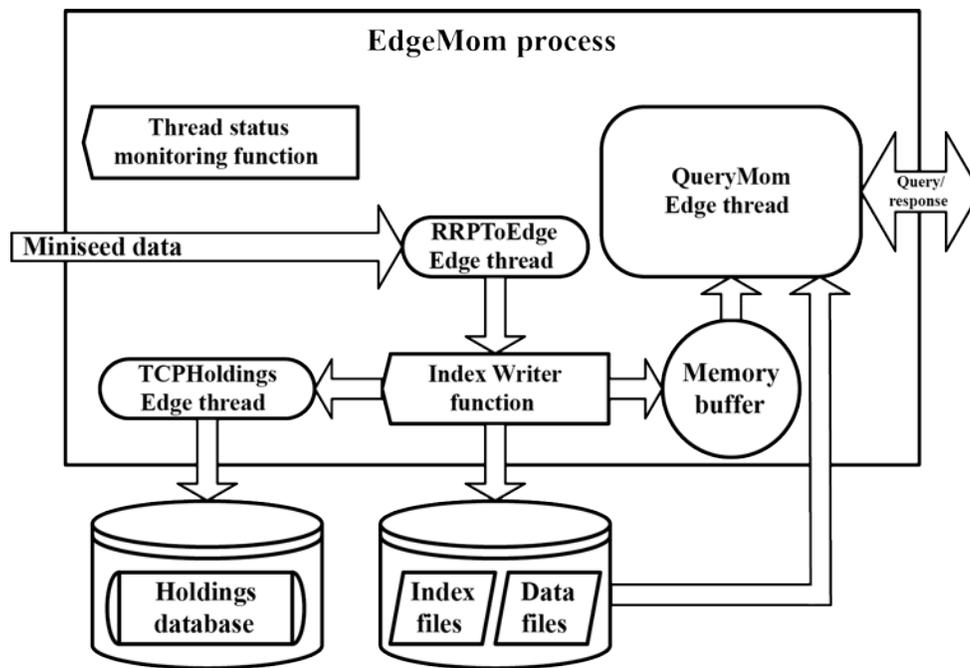

Figure 3. Diagram showing the configuration design of a Continuous Waveform Buffer (CWB) node, including the main EdgeMom process, the Edge thread used for input, the output QueryMom thread, and the holdings database. (RRP, ring replication protocol; TCP, transmission control protocol)

Additionally, a CWB, unlike an Edge node, is configured to store as much data in the data/index files as the hosting server's resources allow. It is simple to load the CWB with all historical data from the IRIS Data Management Center (DMC) or another archival waveform data source and create a single local source for waveform data query services. For example, a

CWB instance can be filled with data from many sources to provide a unified interface for seismic research problems. The CWB also maintains a holdings database that tracks information about the channels and time intervals for which waveform data is available in the CWB.

The QueryMom thread manages all the data query services provided by the CWB. As with the Output Infrastructure thread, QueryMom subthreads handle connections and specific output formats and methods, including QueryServices, Earthworm Waveserver, Trinet Waveserver, and Winston Waveserver formats and protocols, all in one server program. The QueryMom can also be configured to cache recent data in memory to speed up queries of near-real-time data.

Please note that a perfectly valid configuration for the Edge software combines the Edge node and CWB configurations into a single installation. This is particularly attractive for small organizations.

Use Cases

Typical Use Case

Figure 4 shows a typical installation of the Edge/CWB software. A single Edge node handles the data acquisition from stations and also imports data from other contributing networks. It provides a near-real-time export service for interested clients for tasks such as phase arrival picking in real time, as well as for sending real-time data to a seismic processing system.
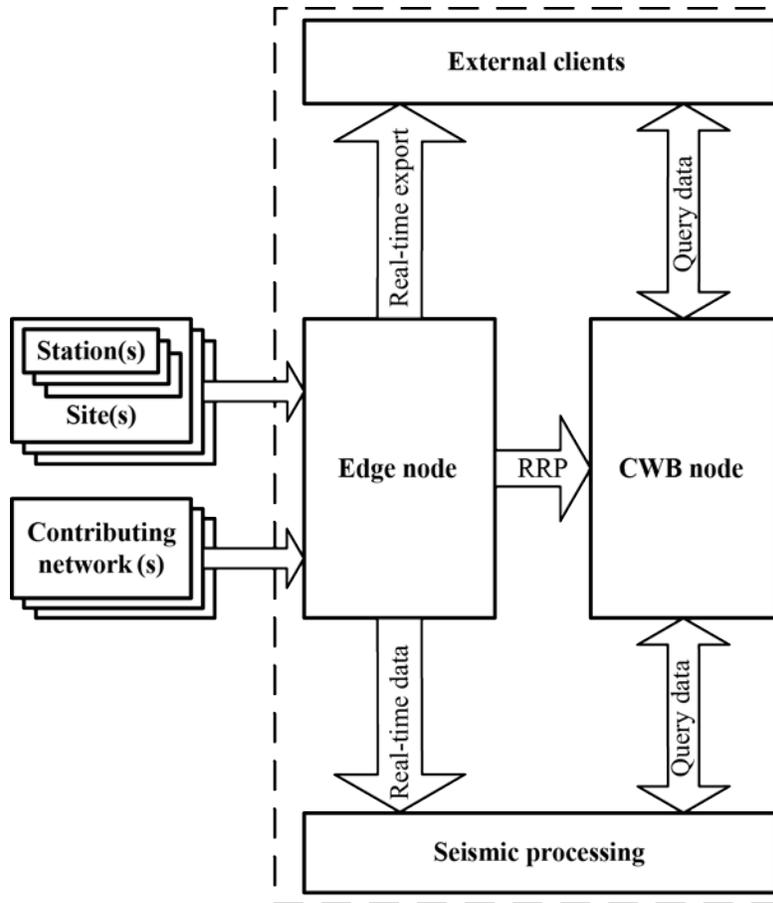
Figure 4. Diagram showing the typical (National Earthquake Information Center) use case for the Edge/Continuous Waveform Buffer (CWB) software. One or more Edge nodes handle acquisition from stations and contributing networks while feeding the waveform data to external organizations, CWB node(s) as a local archive, and various seismic processing applications. (RRP, ring replication protocol)

Also shown in figure 4, the CWB provides long-term data archive and query/retrieval services to both the external clients and the seismic processing system.

Continuous Waveform Buffer Archiving Use Case

Figure 5 shows that the Edge software (CWB configuration) can be easily integrated into an Earthworm, ANSS Earthquake Monitoring System (AQMS), or Antelope installation in place of a Classic or Winston Waveserver. Simply install the software onto a suitable server, configure the CWB to use an Earthworm-style UDP Tracewire or Earthworm Import or Export as the input source, and enable Earthworm and Winston Waveserver queries in QueryMom.
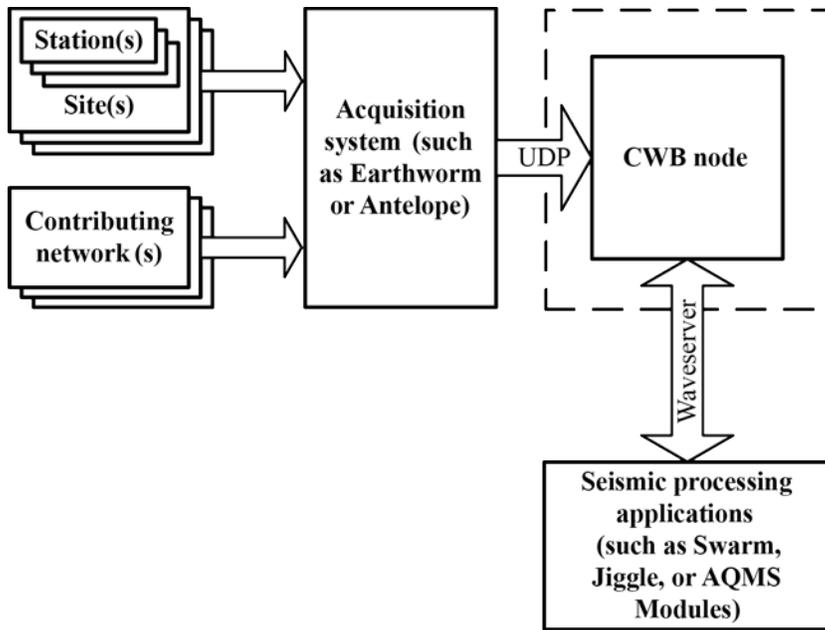
Figure 5. Diagram showing an alternate use case for Continuous Waveform Buffer (CWB) archiving with the Edge/CWB software. Here acquisition is handled by a different system, and only the CWB archiving functionality is used to support seismic processing. (AQMS, Advanced National Seismic System Earthquake Monitoring System; UDP, user datagram protocol)

Additional Development

Expanding the capabilities of the Edge/CWB software would benefit the user community. Software development to support additional DASs, a variety of external data file formats, and more installations of the Edge/CWB software are among foreseeable needs. We expect that the data acquisition, processing, and distribution needs of the USGS National Geomagnetism Program (NGP) as well as those of international organizations will increase, leading to support for continued development of the Edge/CWB software. Loading a CWB with historical data could provide an archival and retrieval system, of particular use to international organizations.

How to Download

The Edge/CWB software is freely available for download at the following link: ftp://hazards.cr.usgs.gov/CWBQuery/EdgeMomBundle.tar.gz.
Included at that link is a detailed document that covers Edge/CWB installation and configuration.
Efforts are underway transfer the Edge/CWB software from the file transfer protocol (FTP) site to the USGS section of GitHub. Currently, the Edge/CWB documentation wiki is available at https://github.com/usgs/edgecwb/wiki.
The requirements to use the software are as follows:
- Server with at least two processors, each with at least two cores
- At least 8 GB of system random access memory
- Linux operating system (Redhat, or Scientific Linux) with X11 support

9

- Oracle Java JDK 6.0 or higher
- MySQL 5.0 or Postgres 9.1 or higher (no-db option available)
- For an Edge node configuration, a data partition of at least 100 GB; for a CWB configuration, the size of data partitions vary on the basis of local needs

Summary

The Edge/Continuous Waveform Buffer (CWB) project is an effort to encompass all of the software for data acquisition written by the U.S. Geological Survey National Earthquake Information Center in a single software package.

The Edge and CWB software system handles the acquisition, storage, and distribution of seismic waveform within the National Earthquake Information Center system infrastructure.

The Edge/CWB software capabilities include direct station acquisition, data import, short- and long-term archiving, data distribution, query services, and playback.

The software design and architecture supports acquisition (Edge) and archiving (CWB) use cases.

The Edge/CWB software is continually developed to support the data acquisition, processing, and distribution needs of the National Earthquake Information Center, U.S. Geological Survey National Geomagnetism Program, and other organizations.

The Edge/CWB software is available for download at the following link: ftp://hazards.cr.usgs.gov/CWBQuery/EdgeMomBundle.tar.gz.