



THE COSMOS VIRTUAL DATA CENTER

Ralph ARCHULETA¹, Jamison STEIDL², Melinda SQUIBB³,

SUMMARY

The COSMOS Virtual Data Center (VDC) is an unrestricted web portal to strong-motion seismic data records of the United States and 14 contributing countries for use by the engineering and scientific communities. A flexible, full range of search methods, including map-based, parameter-entry, and earthquake- and station-based searches enable the web user to quickly find records of interest; and a range of display and download options allow users to view data in multiple contexts, extract and download data parameters, and download data files in convenient formats. Although the portal provides the web user a consistent set of tools for discovery and retrieval, the data files continue to be acquired, processed and managed by the data providers to ensure the currency and integrity of the data. The Consortium of Organizations for Strong-Motion Observation Systems (COSMOS) oversees the development of the VDC through a working group comprised of representatives from government agencies, engineering firms and academic institutions. New developments include a more powerful and informative interactive map interface, configurable UBC design spectra overlays on response spectra plots and enhanced download and conversion options.

INTRODUCTION

The COSMOS VDC is funded by the National Science Foundation, under the Civil and Mechanical Systems Division (CMS-0201264), and COSMOS. The core members of COSMOS, the U.S. Geological Survey, the California Geological Survey, the U.S. Army Corps of Engineers and the U.S. Bureau of Reclamation, as well as contributing members listed in Table 1, make their data available for redistribution by the VDC. COSMOS members, including representatives of the core members and members of the professional engineering community provide ongoing operations and development support through an advisory working group.

**Table 1. Agencies, Utilities, Universities and Companies
that have provided data to the COSMOS VDC.**

Advanced National Seismic System	Kandilli Observatory, Bogazici University, Turkey
Alyeska Pipeline Service Co.	Los Angeles Dept. of Water and Power

¹ Professor, Institute for Crustal Studies, University of California, Santa Barbara.

² Associate Researcher, Institute for Crustal Studies, University of California, Santa Barbara.

³ IT Specialist, Institute for Crustal Studies, University of California, Santa Barbara.

California Division of Water Resources	Los Angeles Flood Control
California Inst. of Technology	Metropolitan Water District, Los Angeles
California Strong Motion Instrument Program, California Geological Survey	Multidisciplinary Center for Earthquake Engineering Research
Centro de Investigaciones Geotecnicas	National Strong Motion Program, U.S. Geological Survey
Dept. of Civil Engineering, University of Southern California	National Survey for Seismic Protection, Armenia
Dept. of Earthquake Engineering, Indian Inst. Of Technology, Roorkee, India	Seattle Light and Power
Dept. of Geophysics and Geodesy, Santiago, Chile	Seismology Center, Central Weather Bureau, Taipei, Taiwan
EQ Hazards Team, U.S. Geological Survey	Southern California Earthquake Center
Geological Survey of Canada	Southern California Edison
Geophysical Institute, University of Alaska, Anchorage Region	Tacoma Public Utilities
Geophysics Program, University of Washington	U.S. Army Corps of Engineers
Inst. Of Geological and Nuclear Science Ltd. New Zealand	U.S. Bureau of Reclamation
Instituto de Ingenieria de la Universidad Nacional Autonoma de Mexico	Universidad CentroAmericana, San Salvador, El Salvador
IRIGM: Universite Joseph Fourier, Grenoble, France	University of California, Los Angeles
Joint Institute of the Physics of the Earth, RAS, Moscow	University of Nevada, Reno
Kiban-Kyoshin Network, Japan	Washington Dept. of Natural Resources
Kyoshin Net, Japan	Other private owners

The COSMOS Virtual Data Center began in 1996 at the Institute for Crustal Studies at the University of California, Santa Barbara, with support from the Southern California Earthquake Center (SCEC), as the Strong-Motion Data Base (SMDB). SMDB was a searchable web interface to seismic data available from SCEC and several U.S. agencies, using an Oracle database on a Sun Solaris operating system. In late 1999, the database was migrated to a Windows server running a MicrosoftSQL database. In 2004, the database has been migrated again to an Apple Xserve running a PostgreSQL database. These migrations, dictated by security, technology and cost considerations, were facilitated by the site's flexible, component-based design and reliance on cross-platform technologies for web page production. Initially, data files were available via ftp on a collocated server; but at the time of the migration to the Windows server and the incorporation of the VDC into COSMOS, the web site became a portal, a 'Virtual Data Center', with data files continuing to reside with the data providers, but passed transparently to the user through the VDC web interface.

THE VDC STRUCTURE

COSMOS maintains agreements with many data providers, including government agencies, universities, and private owners of data, to distribute their data through the VDC. These data files are pulled to the VDC to generate plots, and to extract and insert the associated metadata into the database, together with

the url from which the data files may be retrieved. These metadata and the fully relational database in which they reside enable the extensive search capabilities of the VDC. The process of metadata extraction and search is shown in Figure1.

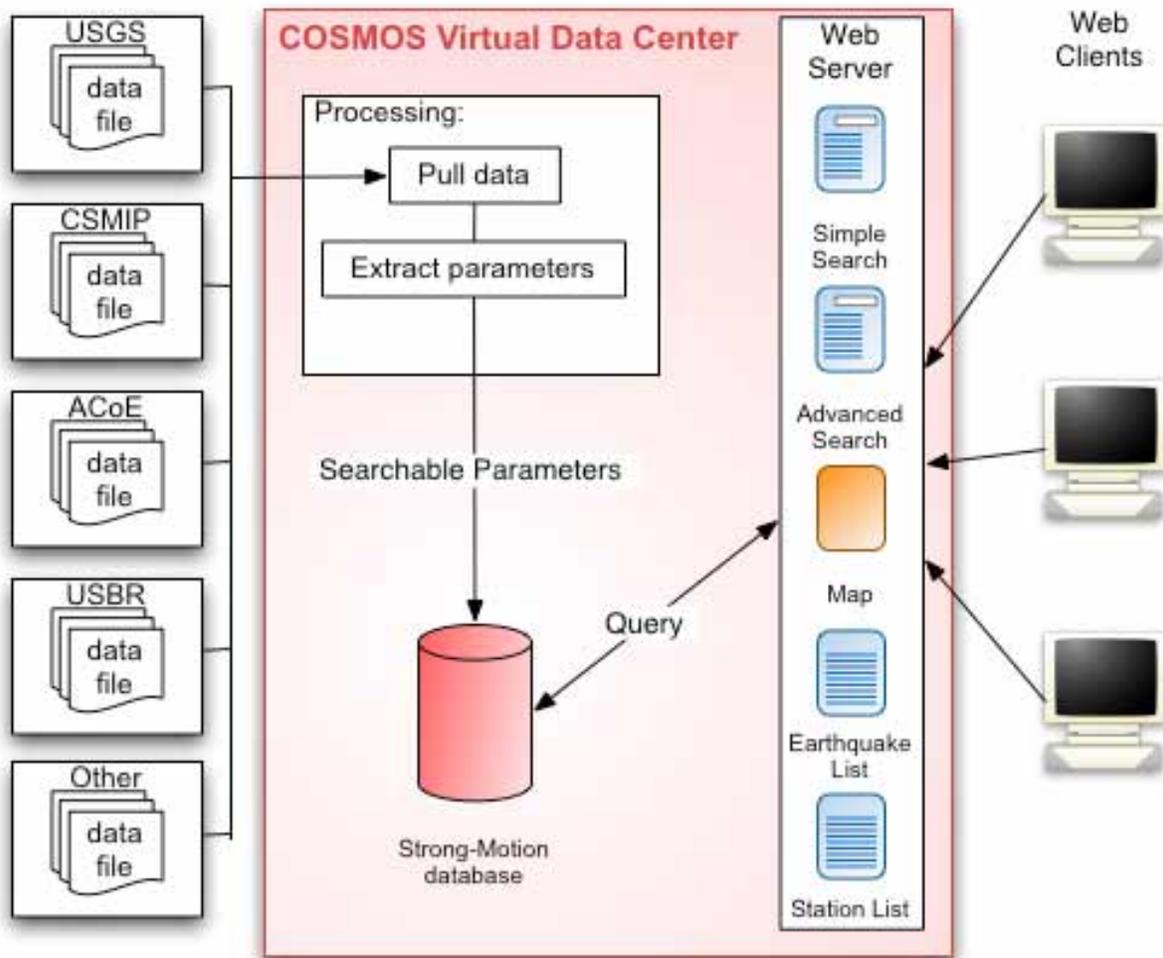


Figure 1. The COSMOS VDC Search Capabilities.

The metadata extracted from the data files is stored in a fully relational database with the structure represented in Figure 2. The VDC increased the number of tables from the original SMDB database to increase the efficiency of searches and storage and to facilitate the representation of more complex data. For example, the separation of instrument from station tables allows complex sites, such as dams or geotechnical arrays to be represented in the VDC. The addition of a format table will allow processing of standard record formats to be automated. In addition, new fields were added to the core tables to include the URLs for additional data products, such as velocity, displacement, fourier and response spectra; and to add peak ground velocity, filter values and response spectra values. These fields were then populated during the data extraction and made available through the search mechanisms represented in Figure 1.

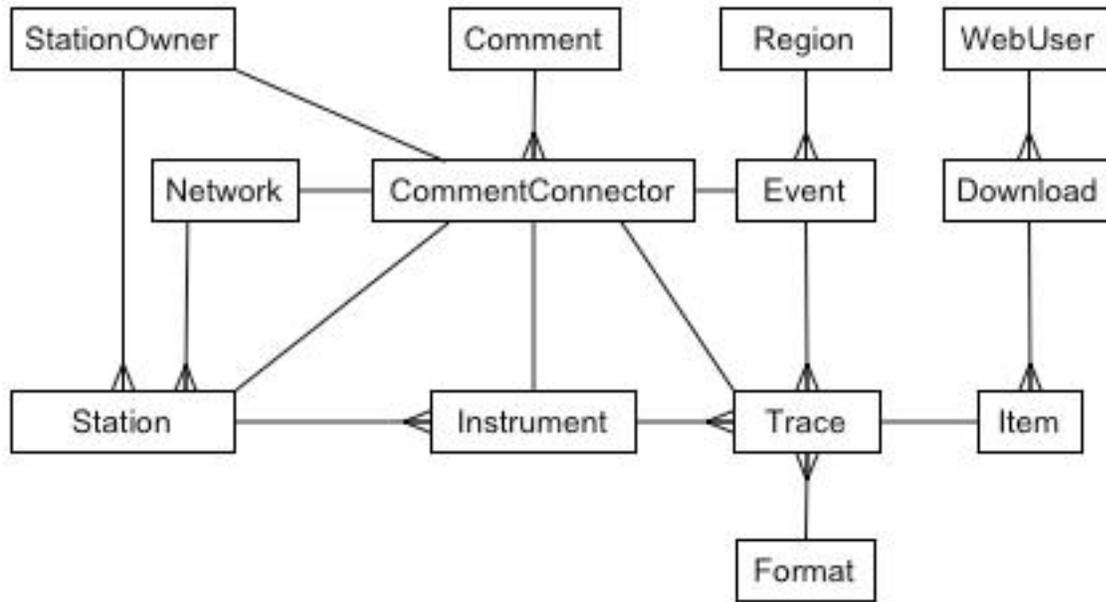


Figure 2. The Structure of the COSMOS VDC relational database.

CONTENT

As of February 2004, the VDC contains the searchable metadata for 394 earthquakes, 2,549 stations, and 18,850 traces. In the last two years substantial data sets representing earthquakes with magnitude greater than 5.0 have been added from the ChiChi, Taiwan, all New Zealand records from 1966-1999 and a array on the Pacific coast of Mexico, as well as smaller but seismically important data sets from Central Asia, Turkey and India. The VDC incorporates all data available from the USGS, CISM with magnitude greater than 5.0 in highly seismic areas and greater than 4.5 in areas of low seismicity. Recent data sets from these sources include the Denali and Nenana Alaska, San Simeon and Nisqually earthquakes. The VDC also incorporates all data from the KNet and KikNet Japanese networks with magnitude greater than 5.0, depth less than 100km and with a pga of at least 0.1g.

USAGE

In April, 2003, the VDC instituted a New Event Notice, by which users are notified by email when new earthquake events are added to the VDC. The effect of this notice, as well as the longevity and stability of the VDC website, and the ease and flexibility of data discovery is reflected in the steady increase in usage of the VDC (Figure 2).

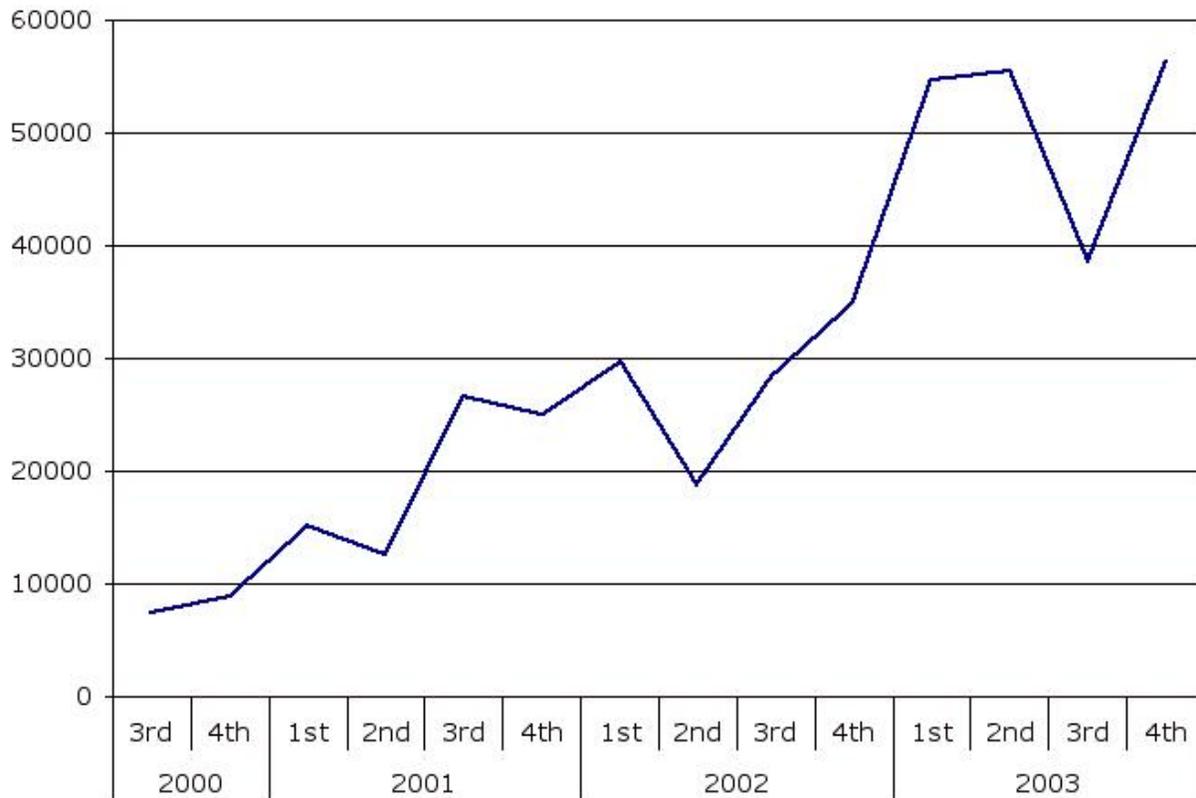


Figure 2. Number of VDC web pages accessed, 2000-2003.

SEARCH CAPABILITIES

The Map Interface

The map interface presents a world map, which displays all earthquakes and stations for which data records are available through the VDC. Users may customize the map by entering latitude and longitude ranges to view, or click on a blank area of the map to zoom in. Each earthquake symbol on this map displays name and date information in the status bar when the user moves the mouse across the symbol; and when clicked, generates an Earthquake page with data for all stations reporting that earthquake. This earthquake page displays:

- 1) Information about the earthquake itself, including location, magnitudes, mechanism, and references;
- 2) Links to another, localized map showing all of the stations which recorded this earthquake;
- 3) A table of these stations showing pga, pgv, hypocentral distance, component orientation, station name, location, owner, processor and products available for each station;
- 4) Buttons to add either all data on the page or the individual traces to a 'shopping cart' of downloadable records;
- 5) Links to acceleration and response spectra plots of the data.

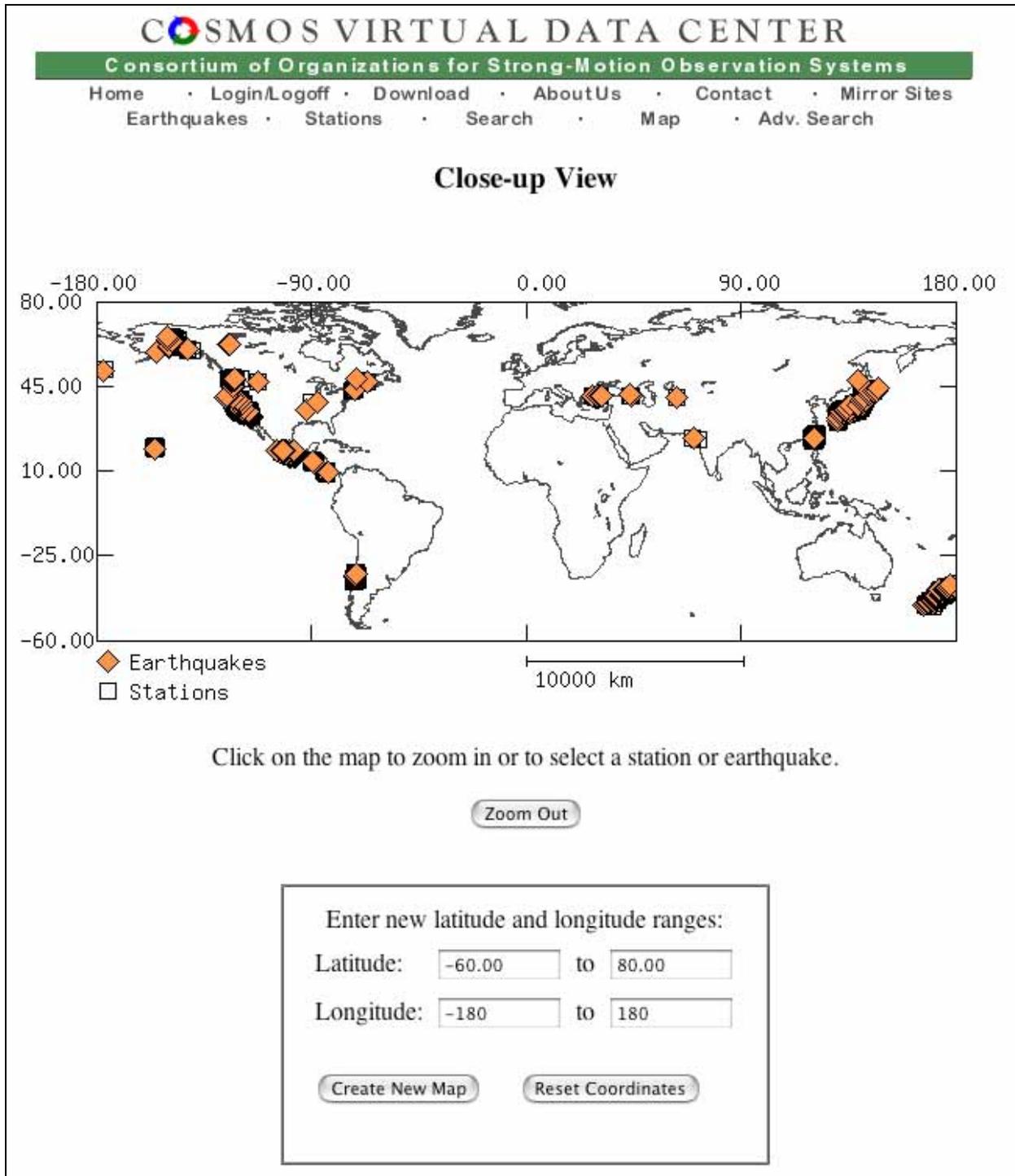


Figure 3. The COSMOS VDC Map Interface.

Each station symbol on this map displays station name in the status bar when the user moves the mouse over the symbol; and when clicked, generates a web page with data for all records associated with that station. This station page displays:

- 1) Information on the station's location, owner and geology, and the owner's web page for the station, if available;

- 2) Links to another, localized map showing all of the earthquakes recorded by this station;
- 3) A table of these earthquakes showing earthquake name and date, and pga, pgv, hypocentral distance and component orientation for this station;
- 4) Buttons to add either all data on this page or the individual traces to a 'shopping cart' of downloadable records;
- 5) Links to acceleration and response spectra plots of the data.

Earthquake and Station-Based Searches

The Earthquakes page lists all earthquakes available through the VDC by region, with a drop-down list of regions at the top of the page for quicker navigation to the region of interest. The listing for each earthquake consists of name, date, preferred magnitude, number of stations available for this earthquake and the agencies providing the data, as well as linking to the full Earthquake page for this earthquake, as described in the section on the Map interface.

The Stations page lists all stations available through the VDC by region, also with a drop-down list of regions at the top of the page. For each station the listing consists of owner and station name. When selected, the user clicks a button to display the full Station page as described in the section on the Map interface.

Parametric Searches

There are two types of parametric searches, the Simple and the Advanced. The Simple search page allows the user to enter the most common parameters:

- 1) Earthquake name
- 2) Station name or location
- 3) Earthquake mechanism
- 4) Magnitude
- 5) Peak Ground Acceleration
- 6) Hypocentral or closest distance to the fault
- 7) Station owner
- 8) Region

The user may also tailor the output to reflect station information only, earthquake information only, or all data, for the result set. The user selects enters values or ranges of values for the parameters on which he chooses to search. Output is available as 1) a list of all stations which meet the criteria with their latitude and longitude and a link to their respective Station pages; or 2) all earthquakes which meet the criteria, with their date and link to their respective Earthquake pages or, 3) if 'all data' was selected as the output type, a page containing a combination of earthquake and station page data for all records that match the criteria.

The Advanced search page allows the user to query and recover almost every field in the database. These include:

- 1) For earthquakes: name, date, longitude, latitude, depth, magnitudes and moment, mechanism, strike, rake and dip;
- 2) Region;
- 3) For station: name, owner's code, location, address, geology, and s-wave velocity for the top 30m;
- 4) For station owner: name, acronym, contact person, ftp site, web site, and address; for the network: name, acronym, owner, and website;

- 5) For the instrument: location, type, serial number, latitude and longitude;
- 6) For the trace: the url of the uncorrected, corrected acceleration, velocity, displacement, fourier spectrum and response spectrum files, epicentral, hypocentral, and closest distance to the fault, azimuth, component vertical and horizontal offset, pga, pgv, response spectrum at 5% damping at 0.1 sec, 0.3 sec, 1 sec and 3 sec.

The user chooses which fields to retrieve or search on and, on the subsequent page, enters the value or ranges of values for the fields to search on. The output is available in two forms:

- 1) a combined station-earthquake page like that returned by the simple search, or
- 2) a tab-delimited text page which contains all parameters selected on the Advanced search page for each trace returned in the result set. This may be saved as a rtf file and imported into a spreadsheet file or database.

DISPLAY AND DOWNLOAD OPTIONS

Display

Each html search result page, whether reached via a map, parameter search or list interface, presents several options for further inquiry: a link to a new map that corresponds to the stations and earthquakes in the result set, a link for each station in the result set to plots of acceleration for that station as shown in Figure 4.

Acceleration

Station: SAN LUIS OBISPO - CITY RECREATION BLDG - 864 SANTA ROSA

Station Owner: United States Geological Survey

Station Latitude & Longitude: 35.2850, -120.6610

Earthquake: SAN SIMEON, CA 2003 12 22 1915 UTC

Epicentral Distance: 61.5 km

(Use the back button on your browser to return to the previous page)

Component: 360

BASEMENT

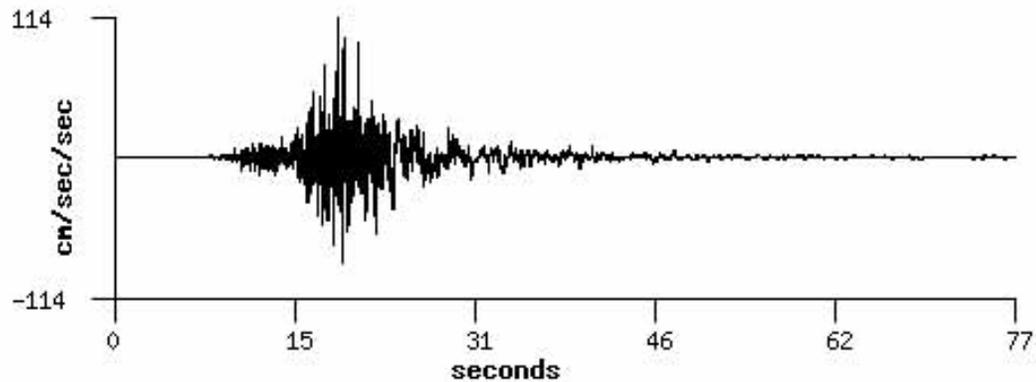


Figure 4. VDC Web page for displaying acceleration.

Each search result page also contains a link to both logarithmic and linear plots of response spectral acceleration, as shown in Figure 5, at 2% and 5% damping, when that data is available. The VDC currently provides response spectra plots only for data providers that offer that data.

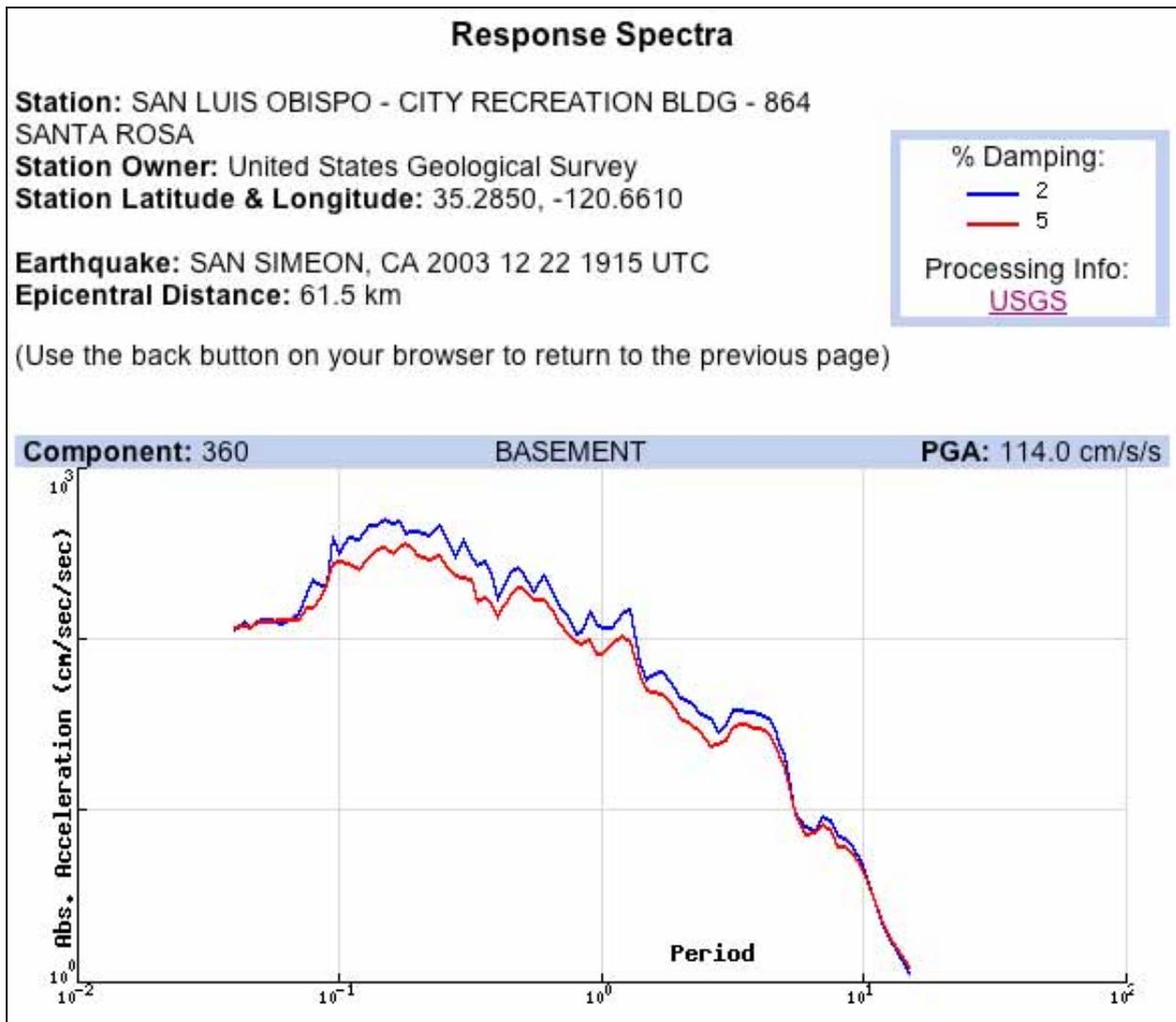


Figure 5. VDC Web page for displaying response spectra.

Downloading

After selecting the files to download from a search results page and entering his email address, the user is directed to the Download Bin. The user can see on the Download Bin page how files are configured by the arrangement of buttons. For example, in Figure 6., the single file from CSMIP contains the data for all three components, as well as combining the acceleration, velocity, displacement, fourier and response spectra in one file, while the data from the USGS station is broken into individual files for each component and data product. The user may click on one of these buttons to download the files individually, or click on the checkbox next to the button to add to a zipped set. Once up to 30 files have been selected to zip, the user clicks on the 'Zip up to 30 checked files' button to retrieve and zip the files. Data files are pulled via FTP from the data provider at the time of the user's request.

Download Bin

[About Zipping data](#)

Zip up to 30 checked files [Start over with a new bin](#)

Earthquake	Station	Instrument	Component	Download	
California Strong Motion Instrumentation Program stations:					
SAN SIMEON, CA 2003 12 22 1915 UTC	Cambria, CA Hwy 1 Caltrans Bridge Gmds		360	Acceleration-Velocity-Displacement-Fourier-Response <input type="checkbox"/>	
			UP		
			90		
United States Geological Survey stations:					
SAN SIMEON, CA 2003 12 22 1915 UTC	SAN LUIS OBISPO 864 SANTA ROSA CITY RECREATION BLDG	BASEMENT	360	Uncorrected <input type="checkbox"/>	Acceleration <input type="checkbox"/> Velocity <input type="checkbox"/> Displacement <input type="checkbox"/> Fourier <input type="checkbox"/> Response <input type="checkbox"/>
			UP	Uncorrected <input type="checkbox"/>	Acceleration <input type="checkbox"/> Velocity <input type="checkbox"/> Displacement <input type="checkbox"/> Fourier <input type="checkbox"/> Response <input type="checkbox"/>
			90	Uncorrected <input type="checkbox"/>	Acceleration <input type="checkbox"/> Velocity <input type="checkbox"/> Displacement <input type="checkbox"/> Fourier <input type="checkbox"/> Response <input type="checkbox"/>

Figure 6. The VDC Download Bin.

2004 ENHANCEMENTS TO THE VDC

Design response spectrum overlays on response spectra

In 2004, the VDC will make available configurable response spectrum overlays on the logarithmic response spectra plots (SA), expressed in g. For 1997 UBC response spectrum overlay[1], users will enter the values for C_a and C_v . For the 2000 IBC design response spectrum[2] or ground-motion response spectrum overlay, users will enter the spectral accelerations, SDS and SD1.

ArcGIS map interface for data discovery

A pilot study is currently in progress to convert the current, continental outline map to an ArcGIS map interface served by a collocated map server. This will not only shorten the time to display a map but allow the user to selectively incorporate more comprehensive data available in the form of maps, including nearby faults, geology, elevation, and nearby roads, cities and facilities.

Increase in available response spectra

The COSMOS VDC is currently developing the procedures to incorporate additional response spectra into the VDC catalog, to find and add files for older data as available, to create spectra for approved data, and to add a response spectrum module to the CosmosConverter application.

CosmosConverter: an object-oriented, distributed conversion tool

Although pulling the strong-motion data files from the data provider's sites allows for greater assurance of currency and accuracy of the data, one disadvantage for the user has been in the variety of formats in which those data are expressed. COSMOS formed a committee to develop a standard ASCII format for strong-motion data records, which is now in version 1.20 and available from the COSMOS web site, so that data providers may translate their own data into this format. Kinematics also includes on their web site a cross-platform java application to translate their instrument format (EVT) into COSMOS format v1.20 (http://www.kmi.com/eng_ftp/altussup.html).

In order to translate legacy files and files from agencies and institutions using heteromorphic formats, an object-oriented java tool, SMtoXML, was developed, with the intention of streamlining data extraction and population of the VDC database. SMtoXML converts single-channel ASCII data file formats to a

common XML format (CSMML), which was derived from the COSMOS format v.1.20. CSMML is defined by an XML schema, which describes the source file metadata by delineating the metadata names and hierarchical structure. The Castor Source Generator (<http://www.castor.org/sourcegen.html>) processes this XML Schema and generates a set of java files corresponding to the complex elements in the schema and containing methods for marshalling and unmarshalling the data from java objects to xml and vice versa. These files are inspected by the SMtoXML Mapping Framework, using reflection, to find all settable elements, their data types and restrictions, and a table of paths to the elements is defined.

To process data from a particular format, at least two files must exist for that format: a Template file, which describes the structure of the source file; and a MappingRules file which maps source file quantities to the xml path defined by the Mapping Framework. There may also exist one or more Tables for mapping information not available in the data source file (e.g. for translating codes). The Template, MappingRules and table files are xml files used as input to the SMtoXML Mapping Framework, which creates a set of java objects corresponding to the defined source data model and then maps the source set of java objects to the Castor-generated java objects, as shown in Figure 7.

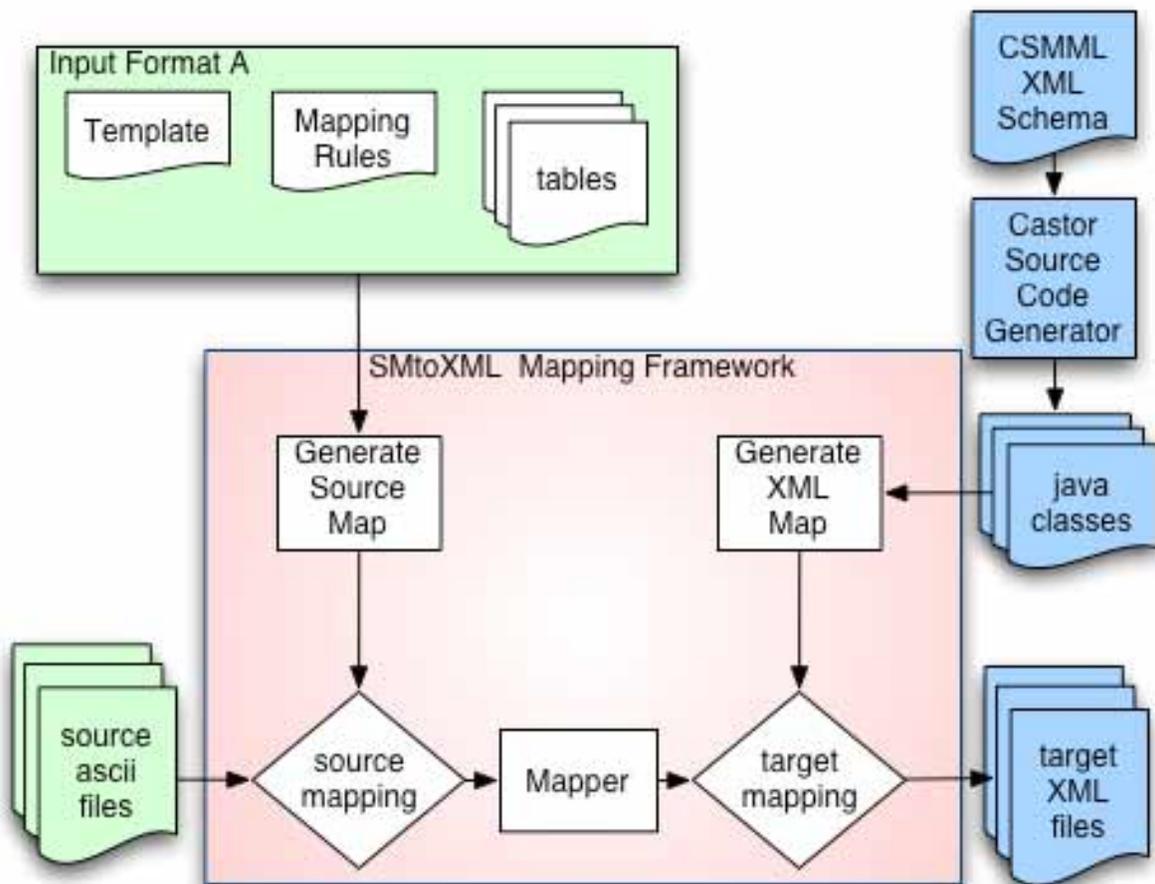


Figure 7. Strong-Motion to XML Mapping Framework.

The advantages of this process are simplicity and flexibility. Input file format changes only require the editing of a simple set of ASCII files. Output XML file format changes only require changes to the XML Schema file and recompiling the java files created by Castor. The SMtoXML program does not need to be

changed or recompiled in either case. The creation of this mapping framework also makes possible the automation of XML file creation and other data products for data providers with regularized formats. Using a single XML format to represent data internally at the VDC allows a single version of each processing scripts to be created, one each for creating acceleration plots, spectra files and COSMOS format files. The downloading, processing, conversion and extraction of the data, which can be automated and thus occur on a periodic basis.

CSMML is currently under revision in order to make the structure and format ISO compliant for date and ResponsibleParty elements[3] and GML compliant for locations[4]. ISO and GML compliance will also facilitate availability of this data through web services and use with ArcGIS mapping tools.

The CosmosConverter is an object-oriented java tool for viewing COSMOS format or CSMML format data and converting these to any of several formats. This tool incorporates the SMtoXML Framework to read Cosmos XML files or COSMOS text files and convert them to other formats, currently including html, png, xml, SAC alpha or columnar text for importation into Excel or Matlab. The tool is available via Java WebStart, a deployment technology from Sun Microsystems that manages versions of the Java Virtual Machine and jar files required for the downloadable application. WebStart may be used in on-line or off-line mode. In on-line mode, the technology connects to the originating server at start-up and downloads any jar files revised since the last time the application was updated. This tool is currently available in an alpha version at <http://db.cosmos-eq.org/converter.html>.

REFERENCES

1. 1997 Uniform Building Code, Vol. 2, International Conference of Building Officials, 1997.
2. 2000 International Building Code, International Code Council, 2000.
3. ISO 19115:2003 Geographic information – Metadata, International Organization for Standardization, 2003. ISO 8601:2000 Data elements and interchange formats – Information interchange – Representation of dates and times, International Organization for Standardization, 2000.
4. OpenGIS Geography Markup Language (GML) Implementation Specification, Version 3.00, Open GIS Consortium, OGC 02-023r4, Jan. 2003.