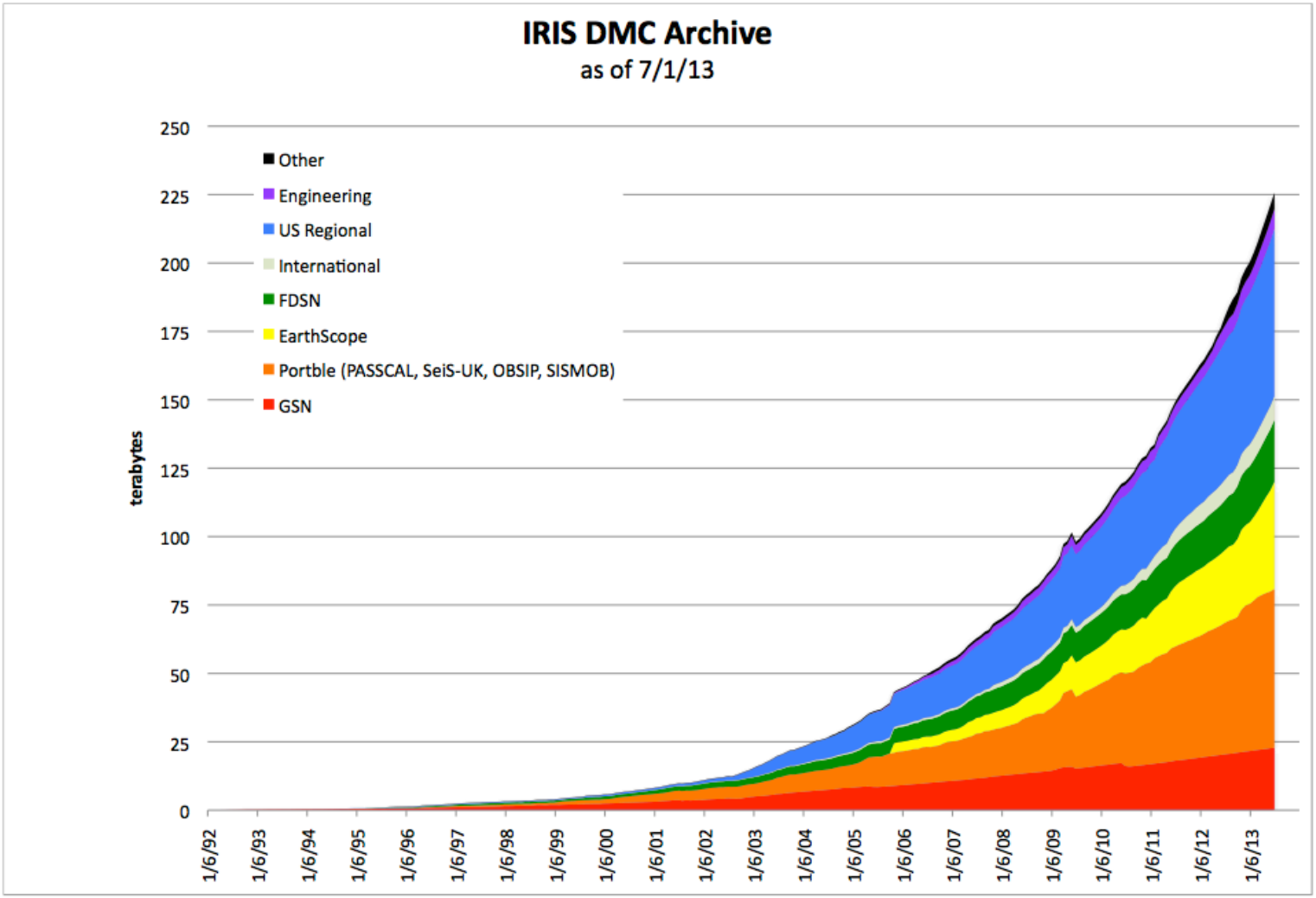
IRIS Services, Products, Quality Assurance Efforts, and Potential Links to High Performance Computing in the Era of BIG DATA

By T. Ahern, M. Bahavar, R.Casey, C. Trabant, A. Clark, A. Hutko, R. Karstens, Y. Suleiman, B. Weertman

Primary TOPICS

- Data Access Services – a new paradigm
  - Improved internal and external ease of use
- Products – stepping stones to further research
- Improved Quality Assurance
- Developing connections to HPC environments

IRIS' Crown Jewel



IRIS Data Services Challenge

- The data holdings are large!
- How do we develop simple methods to discover, access, and utilize the data?
- How can we assist researchers in early stages of their research?
- How can we support tools that are commonly used in the community?
- How can IRIS improve the quality of global seismological data?

IRIS Services – service.iris.edu

- FDSN Web services
  - dataselect
  - station
  - event
  - Documentation
- IRIS web services
  - timeseries
  - rotation
  - sacpz
  - resp
  - evalresp
  - virtualnetwork
  - traveltime
  - Flinnengdahl
  - distaz
  - products

Programmatic support is widespread

Modern computer languages that include support for basic web services include:

- Java
- Perl
- Python
- PHP
- MatLab
- JavaScript
- R (e.g. Rcurl)
- C#
- C/C++ (multiple libraries)

Perl Fetch scripts: command line access

http://service.iris.edu/clients/

FetchData

FetchEvent

FetchMetadata

FetchData options

FetchData retrieves miniSEED, simple metadata, SEED RESP and/or SAC Poles and Zeros using the following selection criteria:

- Network, Station, Location and Channel
  - all optional, can contain '*' and '?' wildcards, virtual networks supported
- Start and end time range
- Geographic box or circular region

Selections: command line, selection list file or BREQ_FAST file

FetchData example

Request 1 hour of GSN/ANMO long-period vertical (LHZ) data and simple metadata for 2010-2-27 M8.8 Chilean earthquake:

$ FetchData

- -N IU –S 'ANMO' –L 00 –C 'LHZ'
- -s 2010-02-27,06:34:00 -e 2010-02-27,07:34:00
- -o /data/Chile-GSN-LHZ.mseed
- -m /data/Chile-GSN-LHZ.metadata

Convert the miniSEED to SAC with metadata

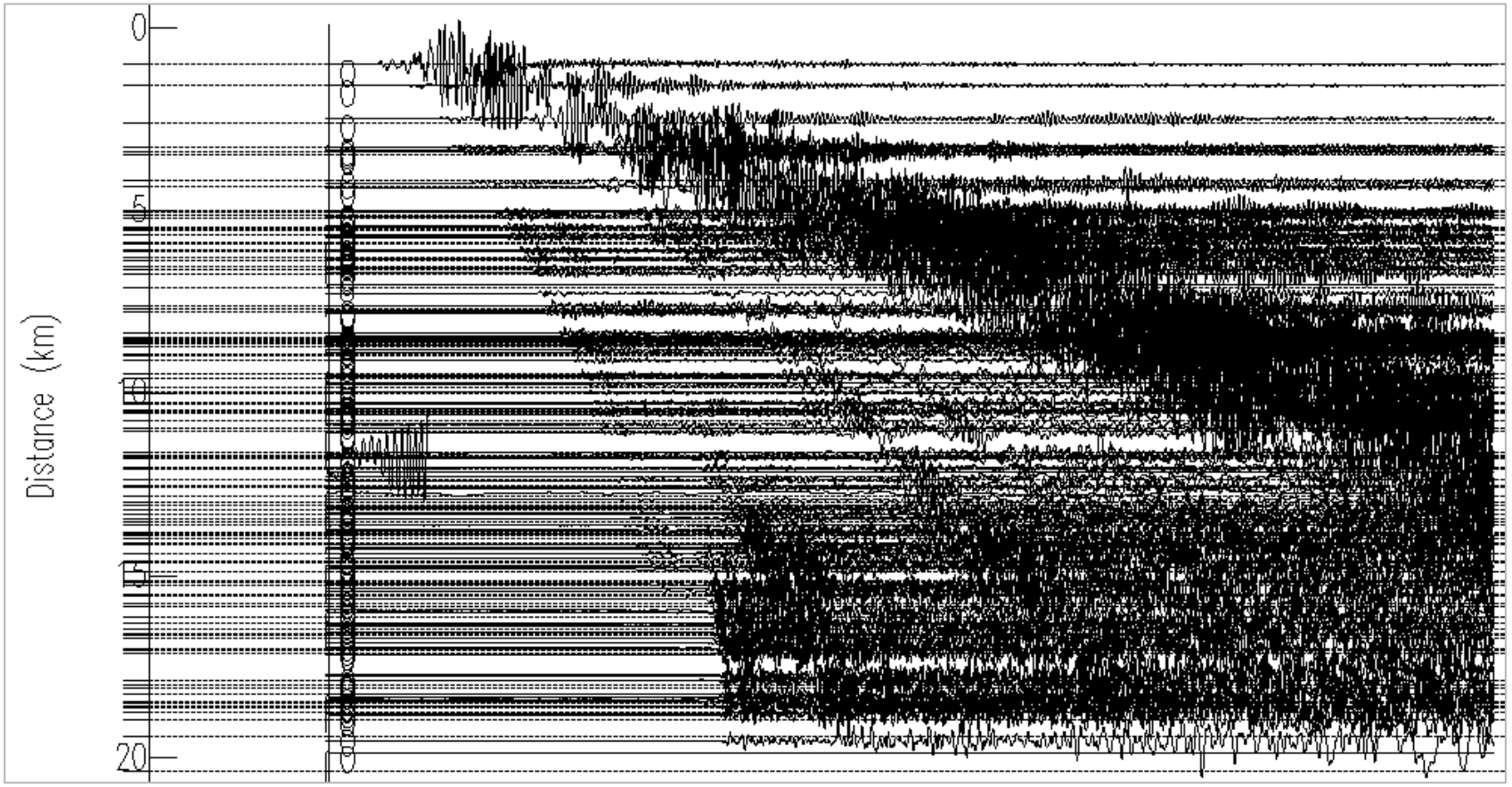
$ mseed2sac Chile-GSN-LHZ.mseed –m Chile-GSN-LHZ.metadata

- -E '2010,058,06:34:11/-36.122/-72.898/22.9'

FetchData example results

2 minutes later…

121 SAC files and a quick-n-dirty record section:



Performance

- WS-dataselect has been shown to be able to deliver 1 terabyte of data per day to a single remote user

FetchEvent options

FetchEvent retrieves event information from **ws-event** and prints simple ASCII output. Events can be selected using these criteria:

- Start and end time range
- Geographic box or circular region
- Depth range
- Magnitude range and type
- Catalog and contributor
- IRIS event ID

Other options:

- Include secondary origins (default is primary only)
- Order results by magnitude or time
- Limit to origins updated after a specific date

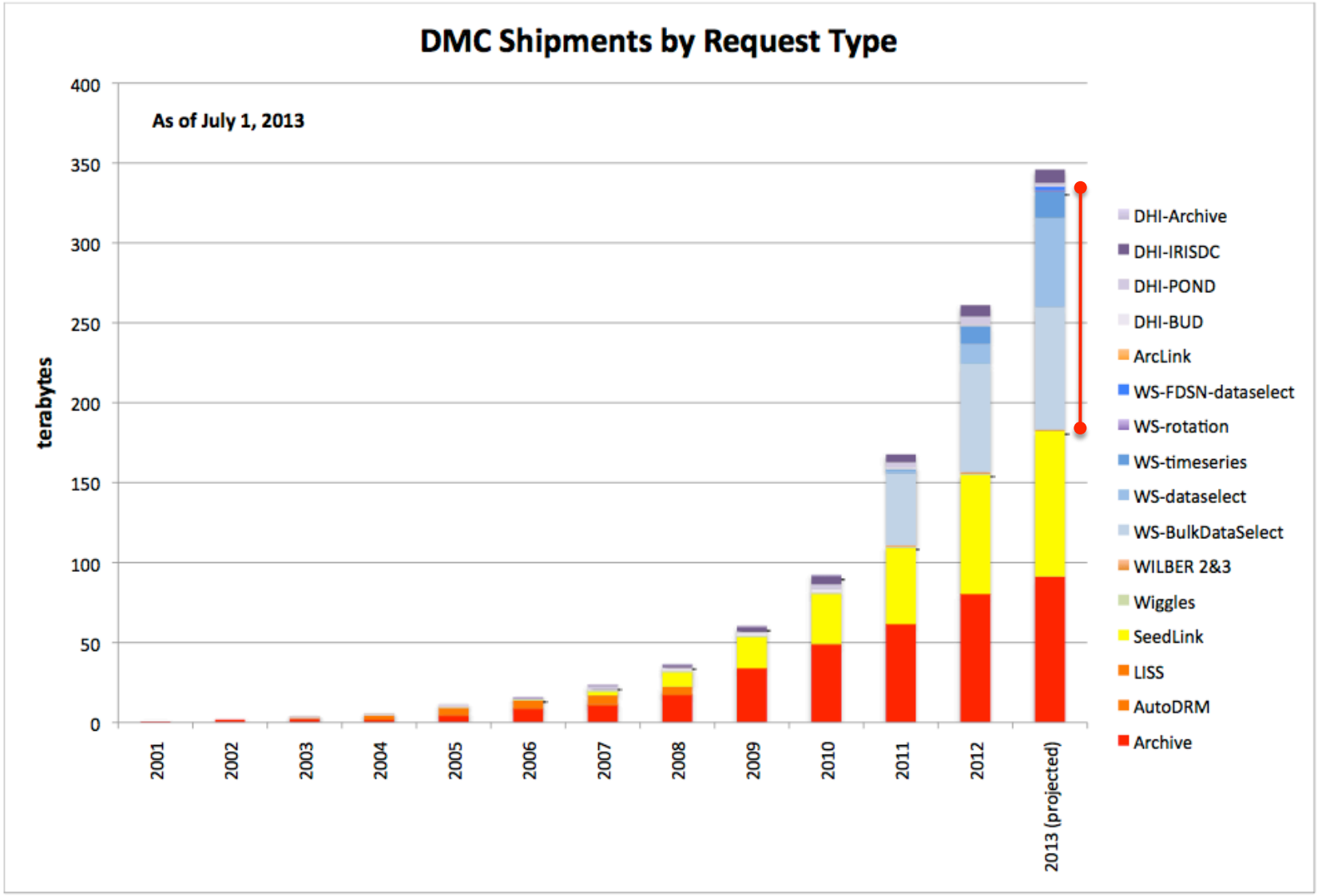
FetchEvent example

Request events for a 20 minute period including secondary origins:

$ FetchEvent -s 2010-2-27,6:30 -e 2010-2-27,6:50 -secondary

```
2844994 |2010/02/27 06:59:31.9200|-33.975    | -72.119  |   21.3|NEIC|NEIC PDE|NEIC PDE-M|MB,5.3,NEIC|OFF COAST OF CENTRAL CHILE
         2010/02/27 06:59:33.9400|-33.971    | -72.088  |   35.0|PDE-W|ANF|ANF,67832
2844993 |2010/02/27 06:58:28.0800|-31.72     | -72.156  |   35.0|NEIC|NEIC PDE|NEIC PDE-M|MB,5.2,NEIC|OFF COAST OF CENTRAL CHILE
         2010/02/27 06:58:28.0800|-31.72     | -72.156  |   35.0|PDE-W|ANF|ANF,67831
2844992 |2010/02/27 06:56:26.1900|-34.35     | -72.197  |   34.8|NEIC|NEIC PDE|NEIC PDE-M|MB,5.6,NEIC|NEAR COAST OF CENTRAL CHILE
2844991 |2010/02/27 06:56:03.1200| 36.073    |-117.878  |    2.5|PAS|NEIC PDE|NEIC PDE-M|ML,2.9,PAS|CALIFORNIA-NEVADA BORDER REGION
2844990 |2010/02/27 06:52:34.0200|-34.867    | -72.614  |   35.0|NEIC|NEIC PDE|NEIC PDE-M|MB,6.2,NEIC|NEAR COAST OF CENTRAL CHILE
         2010/02/27 06:52:34.0200|-34.867    | -72.614  |   35.0|PDE-W|ANF|ANF,67830
2844989 |2010/02/27 06:51:17.6500|-31.663    | -69.141  |   39.8|NEIC|NEIC PDE|NEIC PDE-M|MB,6.0,NEIC|SAN JUAN PROVINCE, ARGENTINA
         2010/02/27 06:51:17.6500|-31.663    | -69.141  |   39.8|PDE-W|ANF|ANF,67829
2844988 |2010/02/27 06:47:23.5900|-33.655    | -72.033  |   35.0|NEIC|NEIC PDE|NEIC PDE-M|MB,5.6,NEIC|OFF COAST OF CENTRAL CHILE
         2010/02/27 06:47:23.5900|-33.655    | -72.033  |   35.0|PDE-W|ANF|ANF,67828
2844986 |2010/02/27 06:34:11.5300|-36.122    | -72.898  |   22.9|NEIC|NEIC PDE|NEIC PDE-M|MS,8.5,NEIC|MW,8.8,UCMT|MB,7.2,NEIC|MW,8.8
         2010/02/27 06:35:14.5000|-35.98     | -73.15   |   23.2|GCMT|GCMT|GCMT,C201002270634A
         2010/02/27 06:34:11.5300|-36.122    | -72.898  |   22.9|PDE-W|ANF|ANF,67827
2844987 |2010/02/27 06:34:16.2900|-54.027    |-133.64   |   10.0|NEIC|NEIC PDE|NEIC PDE-M|MB,5.6,NEIC|PACIFIC-ANTARCTIC RIDGE
```

The success of web services

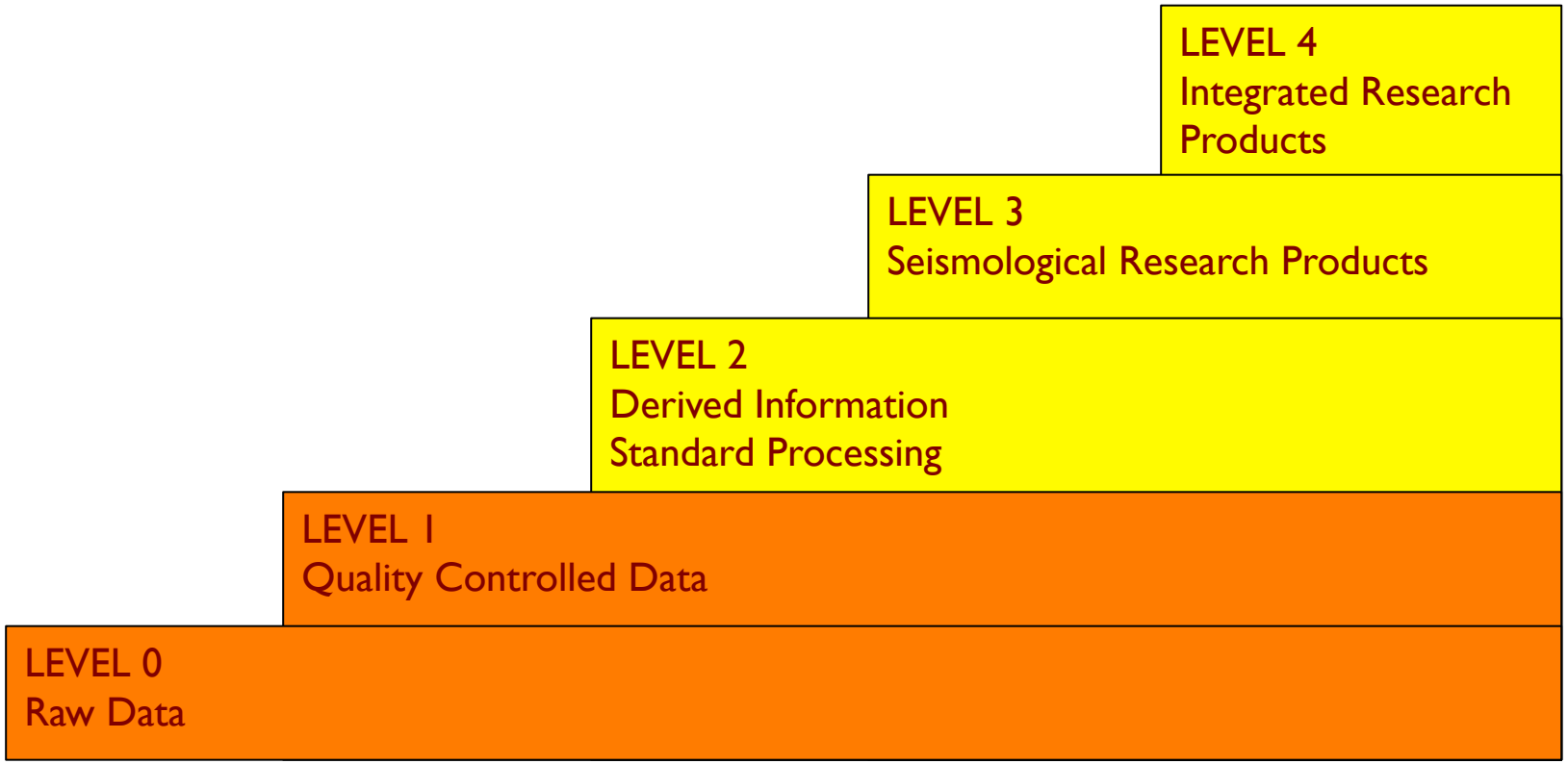


International Coordination

- FDSN web services are well coordinated between Europe and the US
  - Intend to promote them elsewhere
    - Canada, Japan, China, SE Asia
- Many developers producing ws aware clients
  - ObsPy
  - SOD
  - jWeed
  - WILBER 3

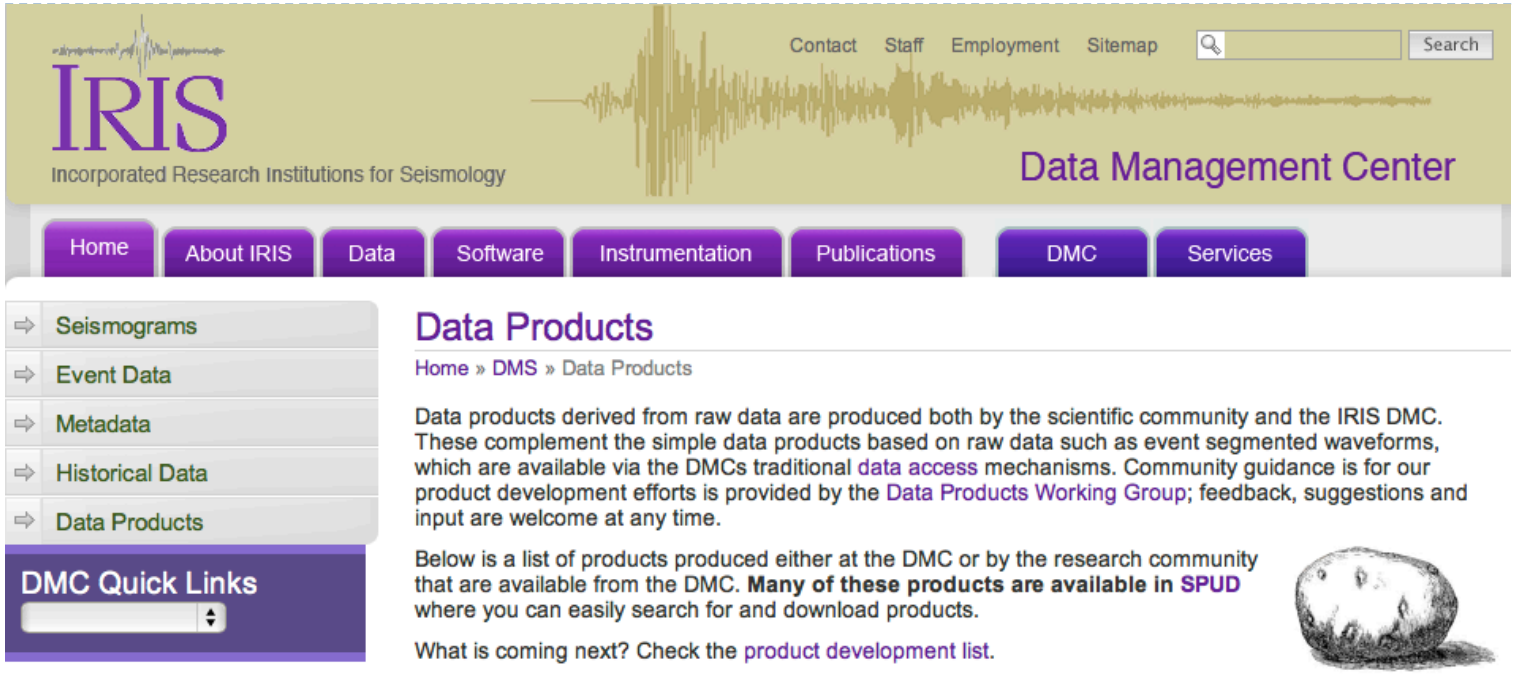
EFFORTS in Higher Level Products

Adapted from National Research Council Committee on Data Management and Computation (CODMAC)



Products from IRIS

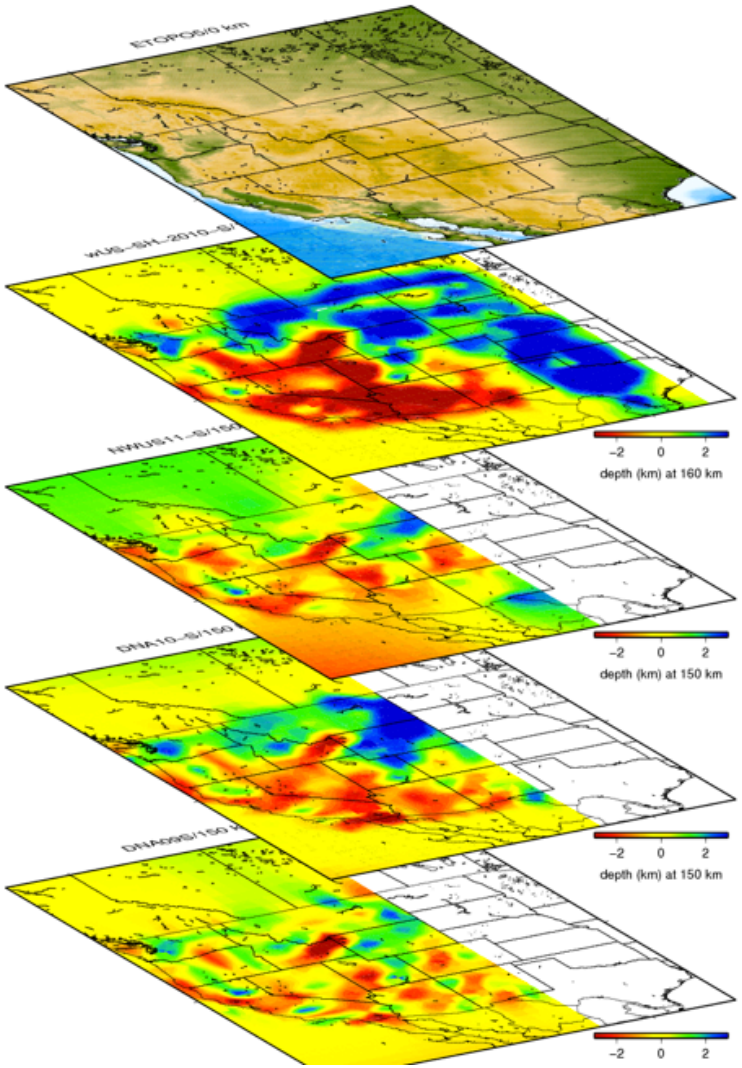
http://www.iris.edu/dms/products/



PRODUCT

IRIS Earth Model Collaboration

Home » DMS » Data Products » EMC



View

Horizontal Slice | **Slice Stack** | Generalized X-section | Cross-section Stack | Velocity Trend | Depth Profile

Earth models

default settings are loaded for each new model as long as they **are not** modified

☑ add topography as the top layer Note: adding topography will significantly increase the plot time!

extreme colors: ○ [■ < min. ■ > max. □ no value] ◉ saturate min./max. colors

☐ **lock settings** (to avoid loading default settings when changing the model)

slice 1 (top) wUS-SH-2010 (% dVs), Schmandt & Humphreys, 2010

section color **depth**

palette seis ☐ invert colors 160 km

dVs -3 % 3 % annotation interval 2 label depth

☐ **lock settings** (to avoid loading default settings when changing the model)

slice 2 NWUS11-S (% dVs), James, Fouch, Carlson & Roth, 2011

section color **depth**

palette seis ☐ invert colors 150 km

dVs -3 % 3 % annotation interval 2 label depth

☐ **lock settings** (to avoid loading default settings when changing the model)

slice 3 DNA10-S (% dVs), Obrebski, Allen, Pollitz & Hung, 2011

section color **depth**

palette seis ☐ invert colors 150 km

dVs -3 % 3 % annotation interval 2 label depth

☐ **lock settings** (to avoid loading default settings when changing the model)

slice 4 DNA09 (% dVs), Obrebski, Allen, Xue & Hung, 2010

section color **depth**

palette seis ☐ invert colors 150 km

dVs -3 % 3 % annotation interval 2 label depth

Area

Asia | Africa | Americas | Australasia | Europe | Japan | N. America | S. America | US | World |

Products Help

SPUD: The Searchable Product Depository

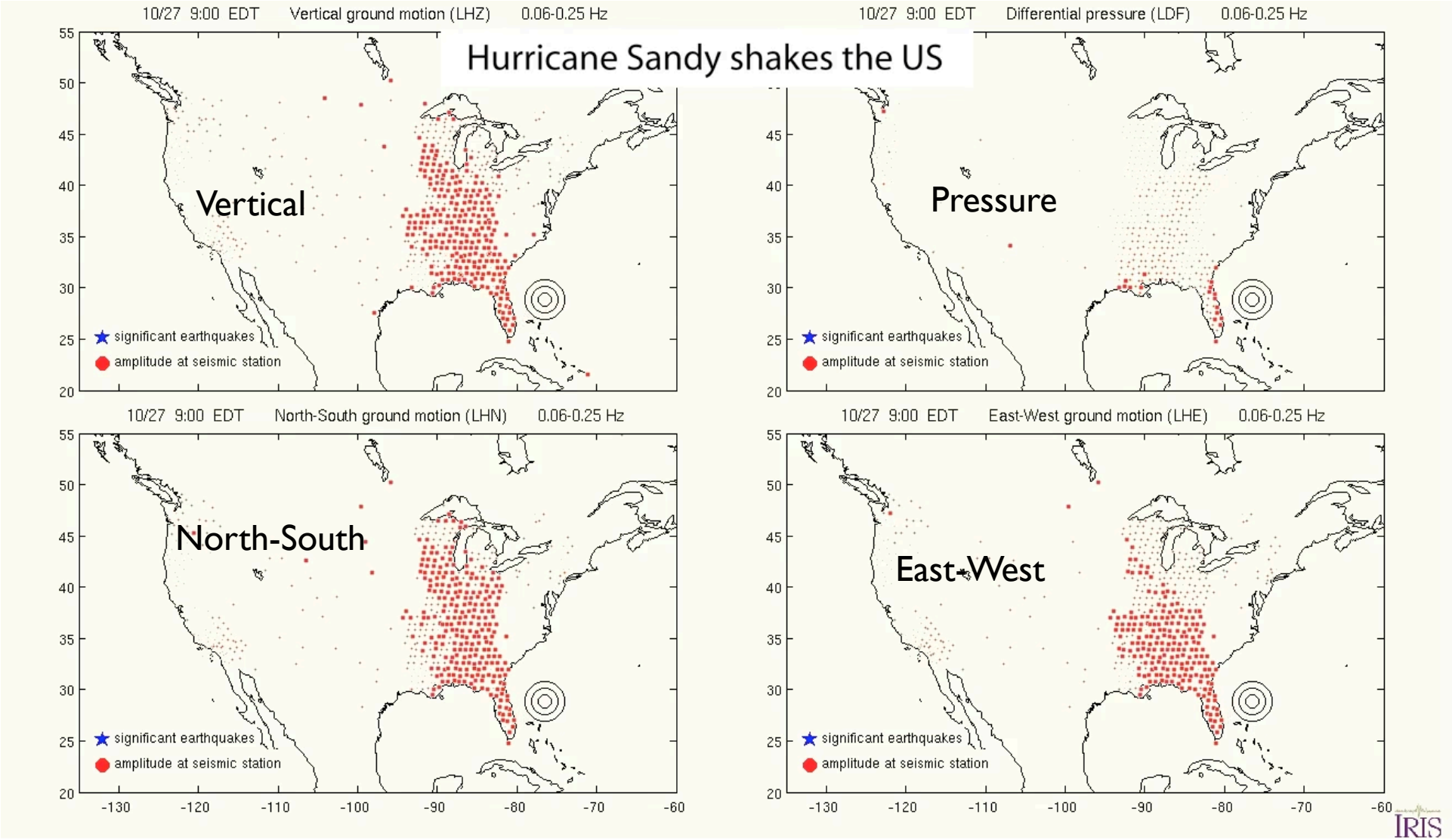
IRIS

Searchable **P**rod**U**ct ***D***epository (event products all products)

Special Event Products

Hurricane Sandy: very bad for New York City, $75B in damage overall

Hurricane Sandy: very interesting seismic noise source



Special Event Archive

"Special events" are events considered to be of interest to the research community. These pages serve to collect and share preliminary research results.

We welcome any additional contributions that might be of interest to the research community. Please add **comments** at the base of each special event page with links and attribution. **Important Note:** Comments are moderated and new comments will be automatically disabled after a period of time.

View special events by year: [2013 | 2011].

You can also search for special event pages based on how they have been tagged.

Chelyabinsk, Russia bolide (meteor)

Posted on Feb. 19, 2013

The meteor strike in Siberia was reported as 9:20am local, about UTC 03:22 Feb 15, 2013. Approx Lat-Lon: 55.15 N 61.41 E

Read more »

IRIS

North Korea nuclear explosion

Posted on Feb. 12, 2013

On Feb 12, 2013, a magnitude 5.1 event was recorded in North Korea. Preliminary results suggests that a nuclear test took place.

Read more »

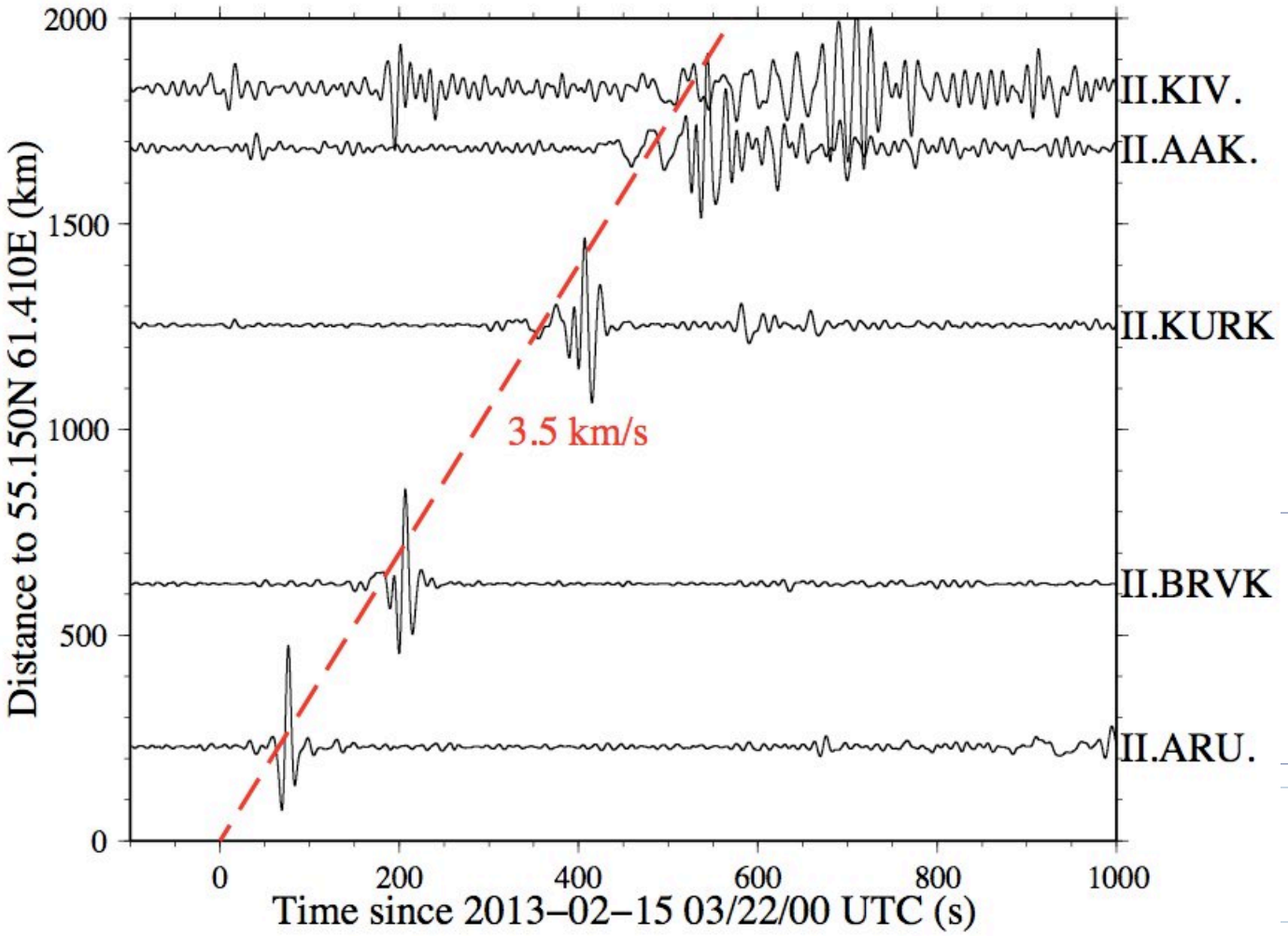
Tohoku, Japan earthquake

Posted on March 11, 2011

A magnitude 9.0 earthquake has occurred near the east coast of Honshu, Japan, as a result of thrust faulting on or near the subduction zone interface plate boundary between the Pacific and North American plates. This page collects many disparate data visualizations from across the websphere.

Read more »

Russian bolide seen by Global Seismic Network stations
(atmospheric to ground coupling generated long period surface waves)



but…..

Special Event: North Korea nuclear explosion

Posted: 2013-02-12 Last updated: 2013-02-21

We welcome any additional contributions that might be of interest to the research community. Please add to the **moderated** comments below.

On Feb 12, 2013, a magnitude 5.1 event was recorded in North Korea. Preliminary results suggests that a nuclear test took place.

Event parameters (from USGS)

These are preliminary results and are subject to change without notice. Please check the USGS page for the latest official information.

Magnitude	5.1
UTC Time	Tuesday, February 12, 2013 at 02:57:51 UTC
Local Time	Tuesday, February 12, 2013 at 11:57:51 (UTC+9) at epicenter
Location	41.307°N, 129.076°E
Depth	0.0km (0.0 miles)
Region	ENE of Sungjibaegam, North Korea
Distances	24km (15mi) ENE of Sungjibaegam, North Korea 24km (15mi) ENE of Sungjibaegam, North Korea 35km (22mi) WNW of Hau-ri, North Korea 43km (27mi) NNW of Kilju, North Korea 379km (235mi) NE of Pyongyang, North Korea
Details	USGS

2009 & 2013 test had *very* similar locations!

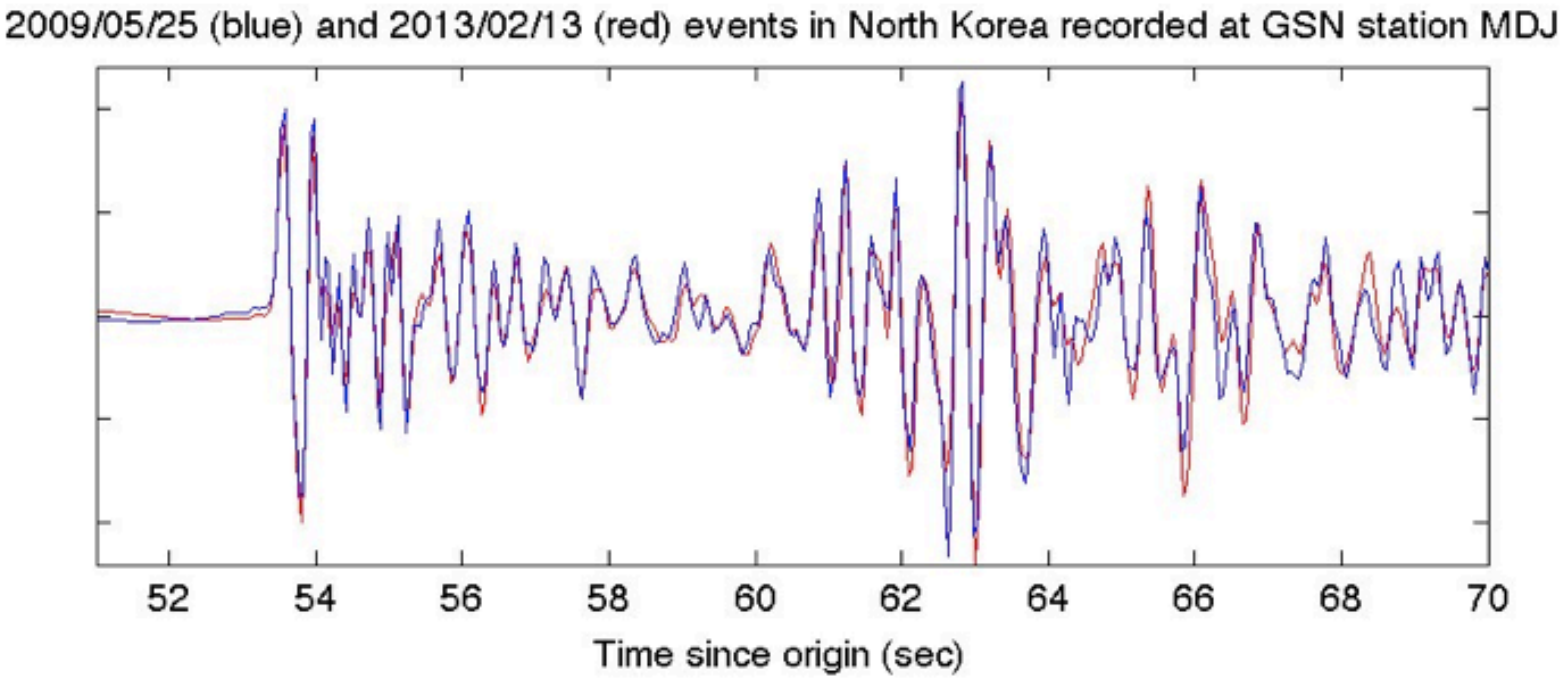


Figure 3: *Strikingly similar vertical displacement records from 2009 and 2013 North Korean nuclear tests, with amplitudes normalized. (IRISDMC)*

On-demand synthetic seismograms

We are computing a complete GF database for:

* High resolution 2D axisymmetric SEM (maybe 0.5 Hz?)
* All source depths/distances
* Seven 1D reference models (PREM, AK135, PREMoceanic…
* Available on demand/command line to anyone through IRIS
* Returns synthetic seismograms: filtered, GCMT or any moment tensor convolved

ETH: Tarje Nissen-Meyers, Martin viel Driel, Niloufar Abolfathian
IRIS: Alex Hutko & Chad Trabant

Quality Assurance Using MUSTANG

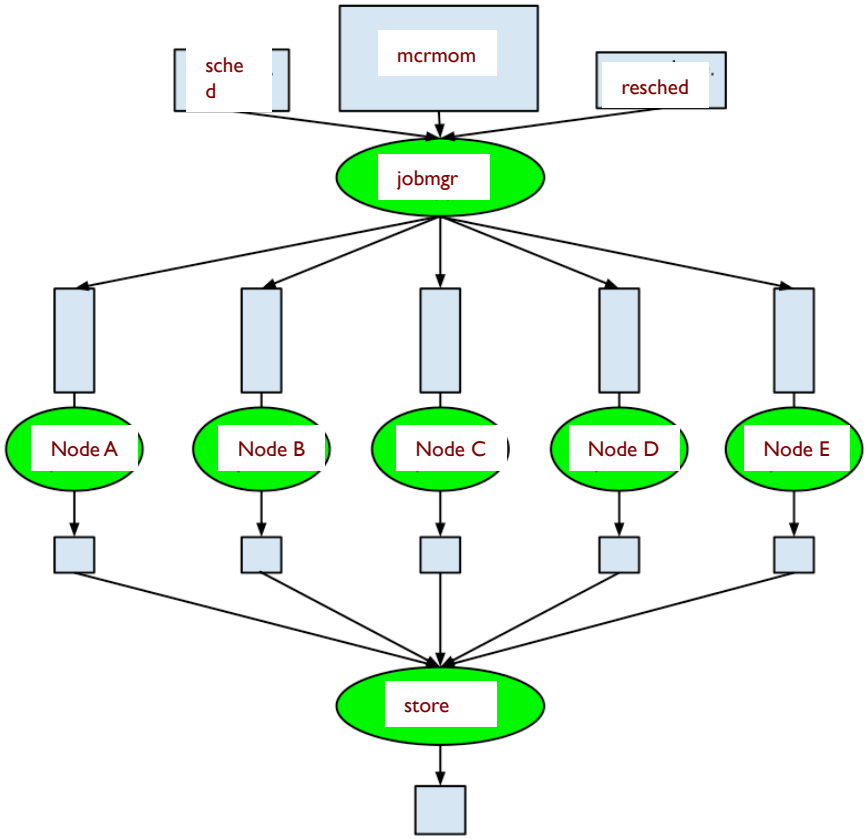
Modular Utility for Statistical Knowledge Gathering

- What is MUSTANG
  - A system initially providing ~two-dozen QA metrics
  - Web service architecture and accessible
  - Crawls through all data in the archive
  - Changes in data and metadata trigger recalculation
  - Integration with IRIS Web Services suite
  - Can be part of a larger network of QA systems

How is MUSTANG designed?

- Consists of 3 major components
  - A Master Scheduler (MCR)
  - A central storage system (BSS)
  - A metrics compute cluster

STALTA_IU@2012-04-13 00:00	2012-06-27 00:08:53	2012-06-27 04:08:10	completed	100%
STALTA_IU@2012-04-12 00:00	2012-06-27 00:00:52	2012-06-27 03:55:07	completed	100%
STALTA_IU@2012-04-10 00:00	2012-06-26 23:20:46	2012-06-27 03:04:45	completed	100%
STALTA_IU@2012-04-04 00:00	2012-06-26 23:20:45	2012-06-27 03:52:07	completed	100%
STALTA_IU@2012-04-09 00:00	2012-06-26 19:31:52	2012-06-26 23:59:16	completed	100%
STALTA_IU@2012-04-08 00:00	2012-06-26 19:29:51	2012-06-26 23:19:38	completed	100%
STALTA_IU@2012-04-07 00:00	2012-06-26 19:15:50	2012-06-26 23:19:54	aborted	95%
STALTA_IU@2012-04-06 00:00	2012-06-26 19:07:50	2012-06-27 00:06:56	completed	100%
STALTA_IU@2012-04-05 00:00	2012-06-26 15:02:43	2012-06-26 19:28:39	completed	100%
STALTA_IU@2012-04-03 00:00	2012-06-26 15:00:42	2012-06-26 19:07:03	completed	100%
STALTA_IU@2012-04-02 00:00	2012-06-26 14:59:52	2012-06-26 19:29:58	completed	100%
STALTA_IU@2012-04-01 00:00	2012-06-26 14:59:22	2012-06-26 19:15:30	completed	100%



Metrics Project Status

- Simple metrics development
  - Includes development of data acquisition, messaging, metadata processing, and other foundational details
  - Gaps, STA/LTA, Overlaps, Availability, Max/Min/Mean/Median values, RMS
  - SNR – event based using tau-p
  - Data Latency adapted from existing QUACK code
  - Polarity reversal will follow SNR
  - Linearity is challenging
  - State of health metrics

Metrics Project Status (2)

- Multiple time series metrics
  - Station percent completeness
  - Multiple station min/max/mean/median
  - Other metrics being worked on
- Complex processing – in pipeline
  - PSD algorithm just completed
    - Processing just beginning
    - Calculations do not have instrument corrections applied
  - PDF plots will be generated dynamically to support aggregation and spectral differencing

More Metrics in Development

- Coherence of two separate time series
- Cross-correlation of two separate channels
- Differencing in PDFs, Aggregate PDFs
- Percent difference above HNM
- Check channel orientation – finding max coherence
- Compare cross-spectrum of two co-located channels
- Compare data to synthetic tide

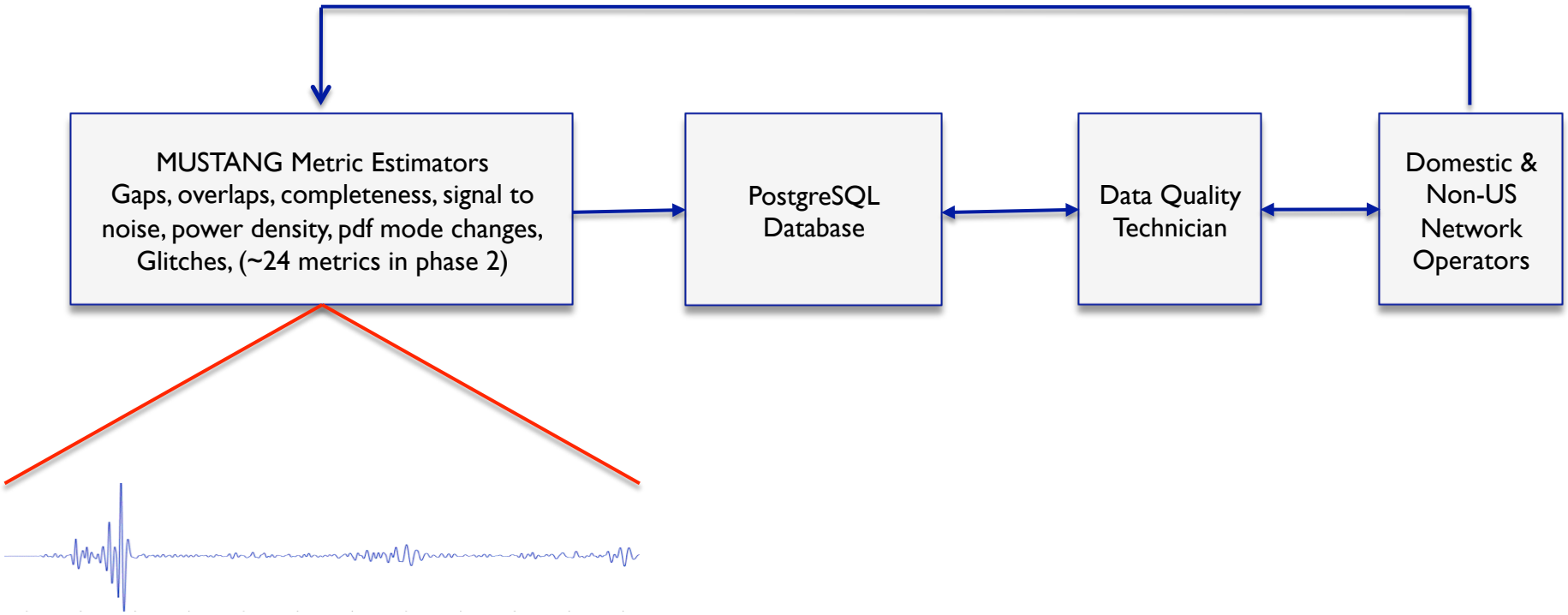
Later Phase

- Additional metrics to be produced
  - Look for spectral trends through mode differencing
  - Timing integrity check by comparing to TauP
  - Correlation of data to atmospheric data
  - Ping or glitch detection
  - Histogram of DC offsets
  - Dead channel detector

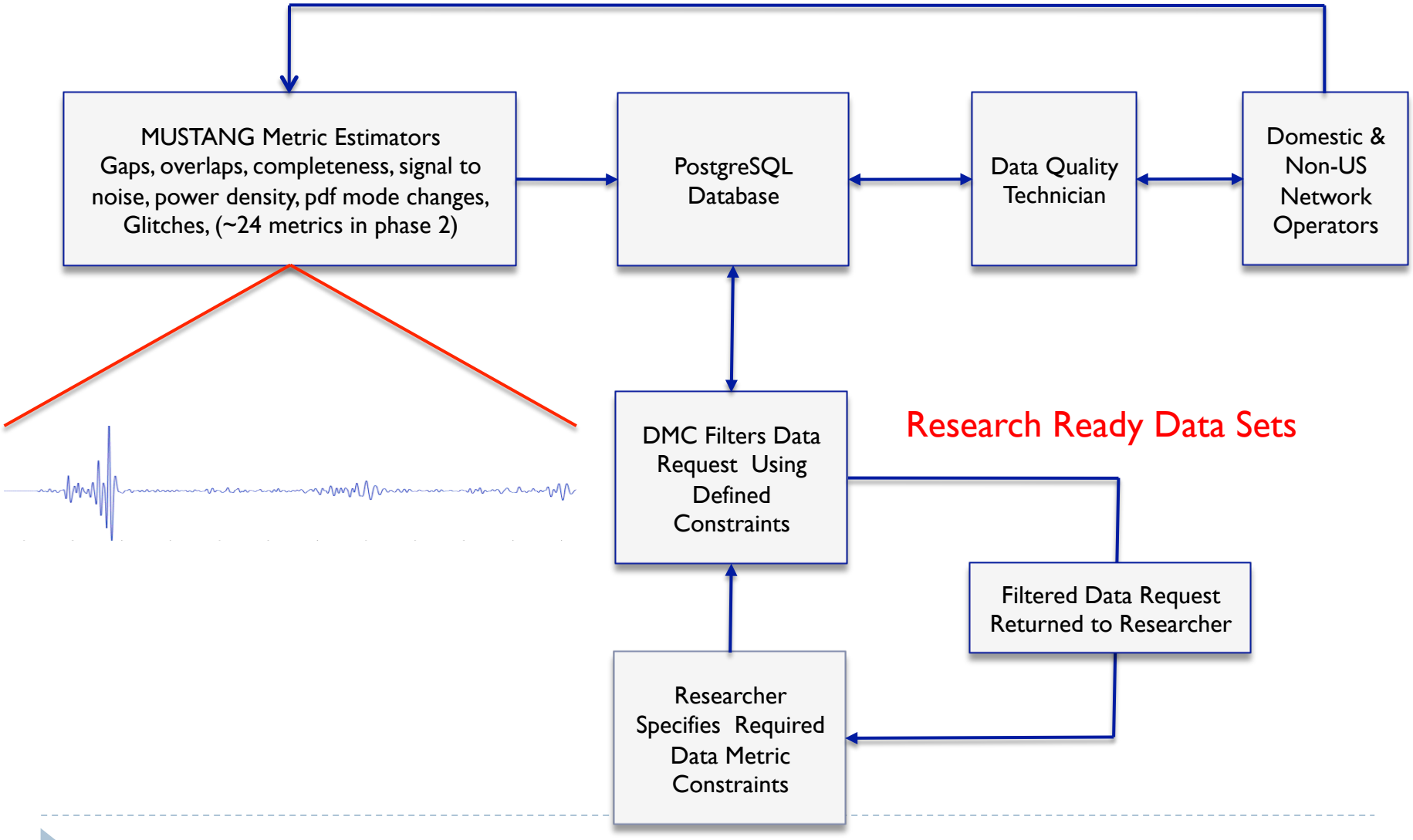
Visualization Client -LASSO

- Flagship visualization client
  - Provide ability to easily browse metrics data
  - Provide ability to generate plots of indicated metrics
  - Provide ability to organize results in web page
- Intended audiences
  - Network operators
  - Scientific users

IRIS DMC:
Enhanced Quality Assurance



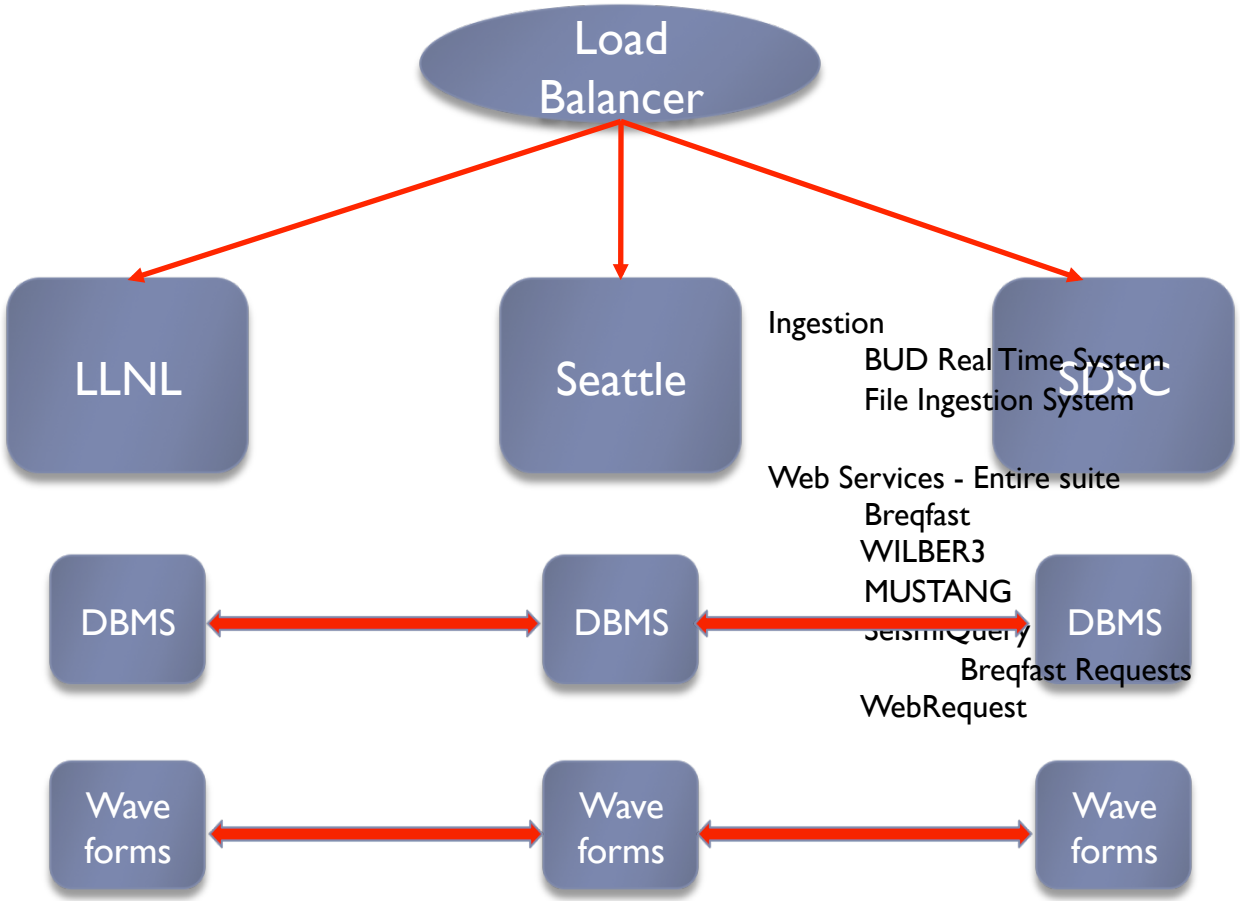
IRIS DMC:
Research Ready Data Sets



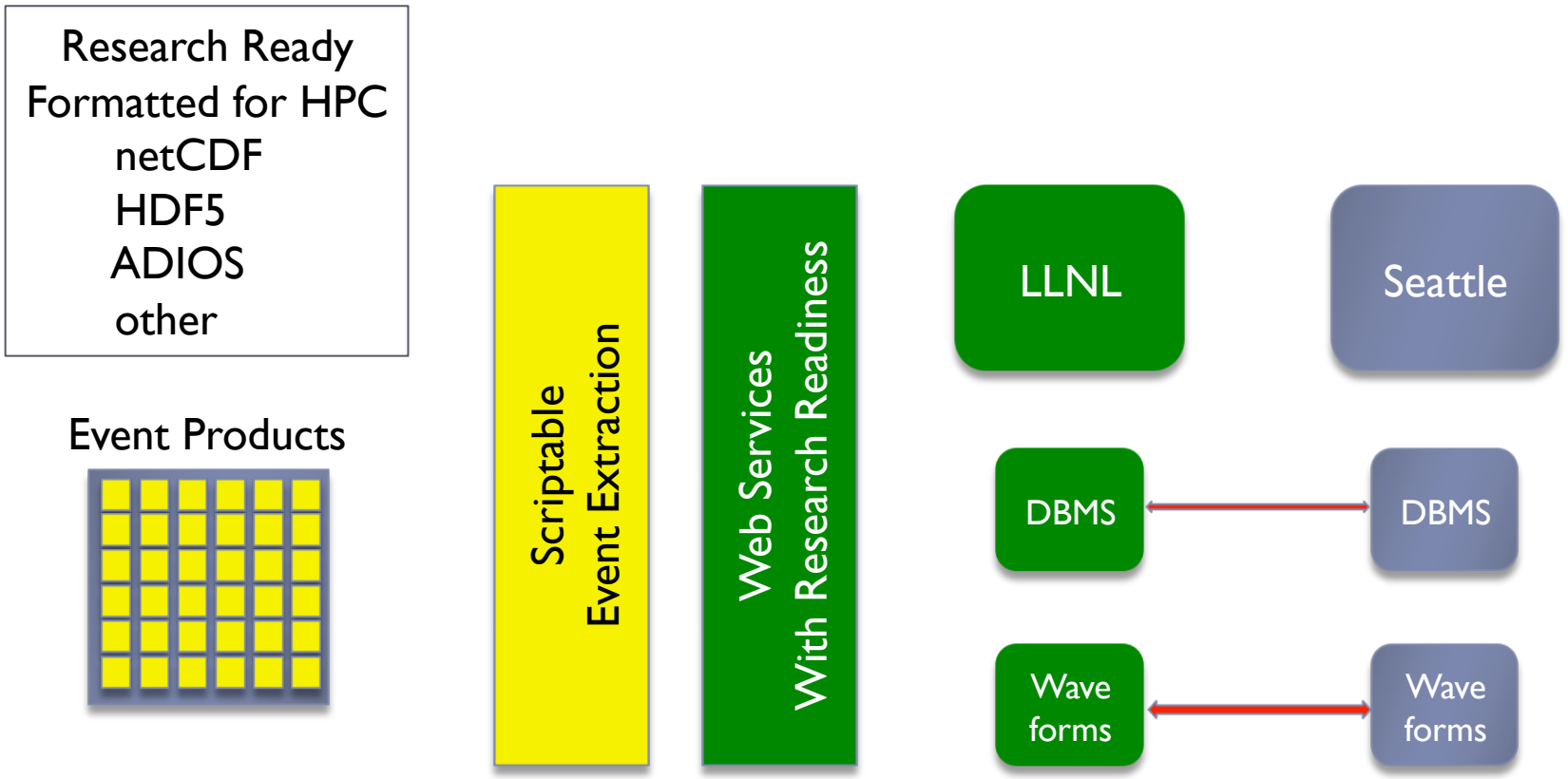
Auxiliary Data Center

- IRIS currently operates an Active Backup System in Boulder, CO at UNAVCO
  - Replication of time series and DBMS
  - And other key items such as software source, etc.
- We wish to move toward a fully functional auxiliary data center model
  - LLNL
  - SDSC
  - Argonne
- This can provide "cycles close to data"

Multiple Fully Functioning DMCs



Links with High Performance Computing



Coordination with University Researchers

- Builds on IRIS DMC Strengths
  - Provide access to hi-graded event products
  - Plumbing between the archive and HPC environment streamlined
- Builds on LLNL strengths
  - Data Mining
  - Algorithmic processing on an HPC environment
- Fosters Collaboration

Some short live demonstrations

- Fetch data
- Conversion to SAC

- The entire GSN in 2 minutes per event

THANK YOU FOR YOUR ATTENTION