What is “good” data quality?

- For time series?
- For metadata?
MUSTANG: Modular Utility for STAistical kNowledge Gathering

- New IRIS automated data quality assurance system
- Assesses data in the IRIS SEED archive
- Recalculates measurements when data or algorithms change (in development)
- **Web service** design - query MUSTANG using URLs
- Easy to add new metrics
- Can include measurements from other institutions
**MUSTANG now serves 45 metrics**

<table>
<thead>
<tr>
<th>State of health flags</th>
<th>Data continuity</th>
<th>Time series amplitude statistics</th>
<th>Noise analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>amplifier_saturation</td>
<td>max_gap</td>
<td>sample_max</td>
<td>noise-psd</td>
</tr>
<tr>
<td>calibration_signal</td>
<td>max_overlap</td>
<td>sample_mean</td>
<td>noise-pdf</td>
</tr>
<tr>
<td>clock_locked</td>
<td>num_gaps</td>
<td>sample_median</td>
<td>noise-mode-timeseries</td>
</tr>
<tr>
<td>digital_filter_charging</td>
<td>num_overlaps</td>
<td>sample_min</td>
<td>pct_above_nhnm</td>
</tr>
<tr>
<td>event_begin</td>
<td></td>
<td>sample_rms</td>
<td>pct_below_nhnm</td>
</tr>
<tr>
<td>event_end</td>
<td></td>
<td>sample_snr</td>
<td></td>
</tr>
<tr>
<td>event_in_progress</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glitches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>missing_padded_data</td>
<td>percent_availability</td>
<td>cross_talk</td>
<td>m2_tides*</td>
</tr>
<tr>
<td>spikes</td>
<td>channel_uptime</td>
<td>dc_offset</td>
<td>orientation_check</td>
</tr>
<tr>
<td>suspect_time_tag</td>
<td>station_completeness</td>
<td>dead_channel_exp</td>
<td>polarity_check</td>
</tr>
<tr>
<td>telemetry_sync_error</td>
<td>data_latency</td>
<td>max_stalta</td>
<td>timing_drift</td>
</tr>
<tr>
<td>timing_correction</td>
<td>feed_latency</td>
<td>num_spikes</td>
<td>transfer_function</td>
</tr>
<tr>
<td>timing_quality</td>
<td>total_latency</td>
<td>asl_coherence</td>
<td></td>
</tr>
</tbody>
</table>

* in development
The MUSTANG system
Where do I find MUSTANG?

http://service.iris.edu/mustang/

**IRIS DMC Web Services**

**Services implementation: MUSTANG**

**Request tools**

<table>
<thead>
<tr>
<th>Service interface</th>
<th>Version</th>
<th>Summary</th>
<th>Return options</th>
</tr>
</thead>
<tbody>
<tr>
<td>measurements</td>
<td>v.1</td>
<td>The main MUSTANG web service returning measurements for metrics relating to station data quality.</td>
<td>XML (default), text, CSV, JSON, JSONP</td>
</tr>
<tr>
<td>noise-psd</td>
<td>v.1</td>
<td>Returns Power Spectral Density estimates of seismic data and can generate aggregate plots.</td>
<td>Text - CSV, XML, Plot (PNG)</td>
</tr>
<tr>
<td>noise-pdf</td>
<td>v.1</td>
<td>Returns Probability Density Functions in frequency &quot;bins&quot; and can generate aggregate plots.</td>
<td>Text - CSV, XML, Plot (PNG)</td>
</tr>
<tr>
<td>noise-mode-timeseries</td>
<td>v.1</td>
<td>Returns PDF Mode Timelines at select frequencies and can generate plots.</td>
<td>Text - CSV, XML, Plot (PNG)</td>
</tr>
<tr>
<td>metrics</td>
<td>v.1</td>
<td>The metrics web service returns a description of available metrics in a variety of formats</td>
<td>XML, HTML, XSD, JSON, JSONP</td>
</tr>
<tr>
<td>targets</td>
<td>v.1</td>
<td>The targets web service returns a list of stations and channels for a given metric.</td>
<td>Text</td>
</tr>
</tbody>
</table>
Channel Orientations

URL Builder: measurements v.1

- **Service interface**
- **URL builder**
- **Help**
- **Revisions**

Use this form to build a URL to the measurements web service. Notice that as you edit the form, the link is automatically updated.

**Metric:** Orientation Check Metric

### Targets

- **SNCLQ filter or Target?**
  - **Filter**
  - **Target**
- **Network:** C
- **Station:** GO08
- **Location:**
- **Channel:**

### Temporal Constraints

- **Parameter:**
  - **Time window**
  - **Start/end**
- **Start time:** 2013-01-01T00:00:00
- **End Time:** 2015-01-01T00:00:00

### Value Constraints

- **Parameter:**
- **Condition:** None
- **Value:**

### Sorting

Both fields must be defined for sorting to be applied.

- **Parameter:** start
- **Order by:** Ascending

More URL parameters described here

- **URL Builder (client) example**
- Text boxes provide parameters for the URL at the bottom
- This URL returns a CSV file that Excel can read
Channel Orientations

- **orientation_check** finds observed channel orientations for shallow \( M \geq 7 \) events by
  - Calculating the Hilbert transform of the Z component (\( H\{Z\} \)) for Rayleigh waves
  - Cross-correlating \( H\{Z\} \) with trial radial components calculated at varying azimuths until the correlation coefficient is maximized
  - The observed channel orientation is difference between the calculated event back azimuth and observed radial azimuth

Channel Orientation Analysis

orientation_check measurements from 2013 and 2014 for C.GO03 having correlation coefficients > 0.4.

A discrepancy with the C.GO08 metadata orientation was found and reported using this metric.

This median observed Y azimuth differed from the metadata by -1.33 degrees.

This value was omitted from the median because it exceeded two standard deviations.
Monitoring Mass Position

Latest Assessment of Seismic Station Observations (LASSO)

- LASSO (client) example
- http://lasso.iris.edu
- Extends MUSTANG by
  - converting mass position counts into volts using metadata
  - Displaying “virtual networks”
GPS Clock Locking

**MUSTANG databrowser**

- **Plot Type and Metric**
  - Plot Type: Network Boxplots
  - Metric: Daily Flag Count: Clock locked
- **Plot Options for Boxplots**
  - Checkboxes: Outliers, Box Plot Explained

**Data Source and Timespan**

- **Network - Station - Location - Channel**
  - IU, GRFO.BHZ
  - IU, MACI.BHZ
  - IU, KBL.BHZ
- **Time Span**
  - Start: 2015-05-01
  - End: 2015-05-15

Access Speed (sec): data load=3.27, plot=0.03, R total=3.30
data URL

---

**Databrowser (client) example**

- [http://ds.iris.edu/mustang/databrowser/](http://ds.iris.edu/mustang/databrowser/)

**Extends MUSTANG by**

- Providing a variety of plot types
- Plotting related channels (or metrics) as a group
Dead channels

- **Scripting (client) example**
  - Quickly retrieves one month of measurements for the IU network
  - Summarizes potential data problems based on metrics thresholds

<table>
<thead>
<tr>
<th>dead</th>
<th>dead_channel_exp &lt; 0.3 &amp;&amp; pct_below_nlnm &gt; 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>dead</td>
<td></td>
</tr>
<tr>
<td>dead</td>
<td>IU,TRQA.10,BH1,M</td>
</tr>
<tr>
<td>dead</td>
<td>IU,TRQA.10,BH2,M</td>
</tr>
<tr>
<td>dead</td>
<td>IU,TRQA.10,BH2,M</td>
</tr>
<tr>
<td>dead</td>
<td>IU,TRQA.10,BHZ,M</td>
</tr>
<tr>
<td>dead</td>
<td>IU,WCI.00,BHZ,M</td>
</tr>
<tr>
<td>dead</td>
<td>IU,WCI.00,BHZ,M</td>
</tr>
</tbody>
</table>

- Analysts can focus on verifying potential problems – what else can we learn about IU.WCI?
Dead channels

- **MUSTANG pdf-noise service**
- **IU.WCI.00.BHZ** isn’t completely dead – it still records some energy

![Graph showing microseisms and Nepal Earthquake](image-url)
Dead channels

- MUSTANG noise-mode-timeseries service
- The problem started on 2014/08/27.
Would you like to learn more about:

- Data quality assessment?
- How you can use MUSTANG and it’s clients?
- Data quality of stations archived at IRIS?

If so, I invite you to talk to me during the week and/or attend the Thursday afternoon breakout session.
Thank you –
I look forward to talking with you.