Cyberinfrastructure Tools for Managing GAMUT Data and Infrastructure, Part 2

12/12/2013

Amber Spackman Jones

Logan River Sites

Red Butte Sites

Provo River Sites

MDF LoggerNet Server

Base Station

LoggerNet Backup

ODM Streaming Data Loader

.dat Datalogger Text Files

Backup Script

Backup Script

MDF Database Server

Automated QC Checks

Logan ODM Raw Data

QC Data

Automated QC Checks

Red Butte ODM Raw Data

QC Data

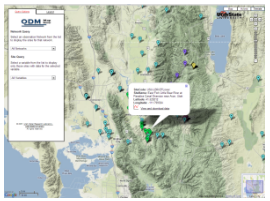
Automated QC Checks

Provo ODM Raw Data

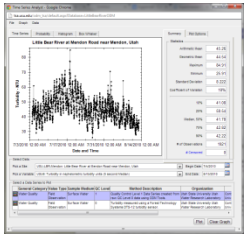
QC Data

Staging of Current Conditions

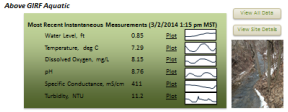
MDF Web Server



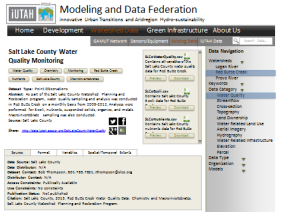
Map Based Display



Time Series Display (and other plot types)



Current Conditions



File-based Archival and Publication

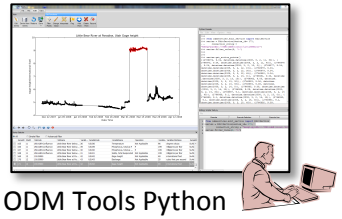
Archival of original .dat files

Backup of ODM Databases

UofU Data Store

Virtual Machine Backups

Visualization and QC of Sensor Data Stream



ODM Tools Python

Datalogger Program Management

- Need to adhere to workflow/protocol.
- What is the status of datalogger program updates?
- Can use separate tables for monitoring efforts external to the baseline programs.
- Can use offsets and constants in a variables/ constants table.

Planned Update

for addition/deletion of long term variable, adding tables for adamptive sampling, etc.

Data for 1-2 time steps may be lost in this process

Technician modifies datalogger program. Saves it as new baseline version or sub-version

Technician documents changes and rationale in changelog file.

Uploaded to iUTAH cloud share

Technician communicates with other technicians, ensures that all sites are running same baseline program.

Technician notifies Jeff and Amber via email.

Jeff or Amber performs the following on the LoggerNet Server:

- -Downloads latest data from the site.
- -Runs the Streaming Data Loader.
- -Pauses the Streaming Data Loader.
- -Moves the old datalogger file into a backups folder.
- -Sends the new datalogger program to the site
- -Modifies the Streaming Data Loader to account for any changes to the datalogger file.
- -Restarts the Streaming Data Loader.

Jeff or Amber notifies technician that process is complete.

Urgent Update

for diagnostic or troubleshooting purposes

Some data may be lost under this scenario!

iUTAH GAMUT Technicians

iUTAH CI Team

Technician modifies datalogger program. Saves it as a diagnostic version.

Technician documents changes and rationale in changelog file.

Uploaded to iUTAH cloud share

Technician notifies Jeff and Amber via email.

If available, Jeff or Amber will:
-Download data to Loggernet Server just prior to diagnostic program being sent.
-Pause Streaming Data Loader.

If there is no central data download prior to sending new program, all data not yet loaded to database will be lost.

Technician sends diagnostic program to site.

Troubleshooting/ Diagnostic Period

Site is taken offline. It is assumed that any data collected during this period will not be stored in the database.

Technician downloads data resulting from diagnostic program if it is to be saved.

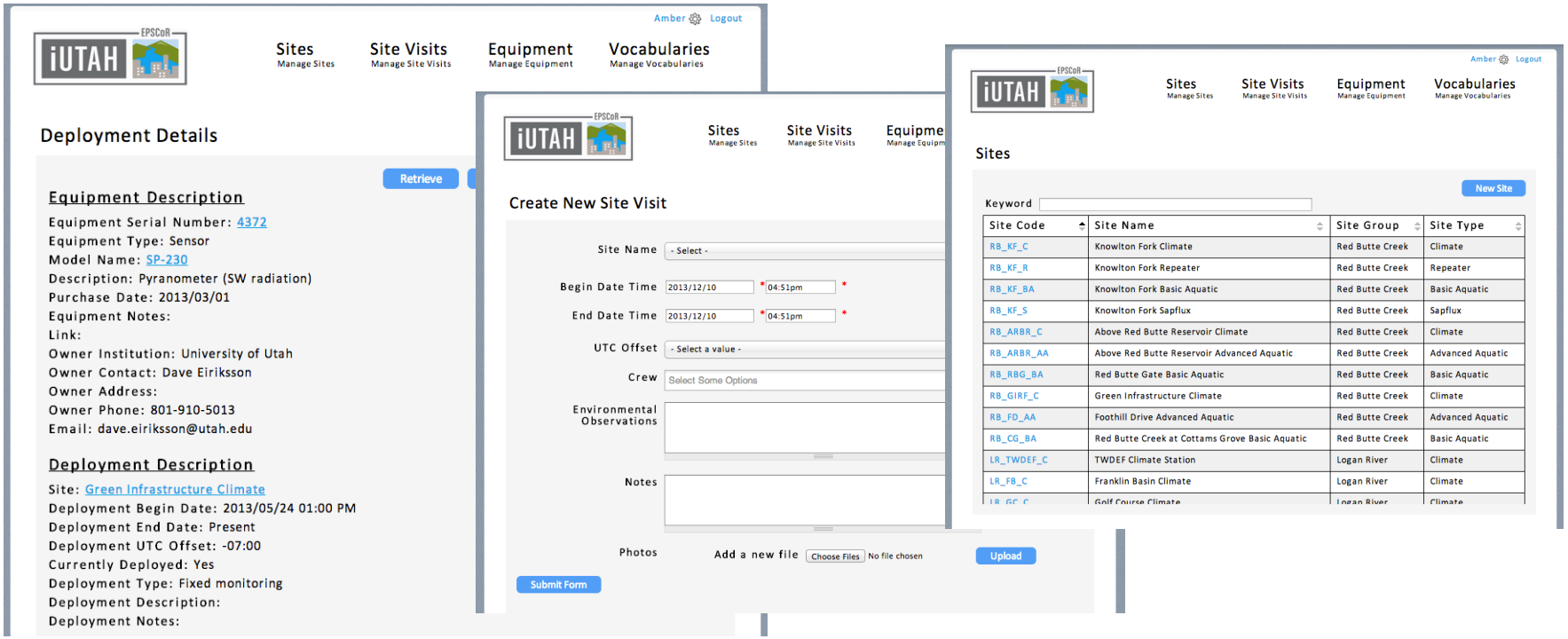
Technician sends regular program to the site.

Technician notifies Jeff and Amber that site is back to normal operation.

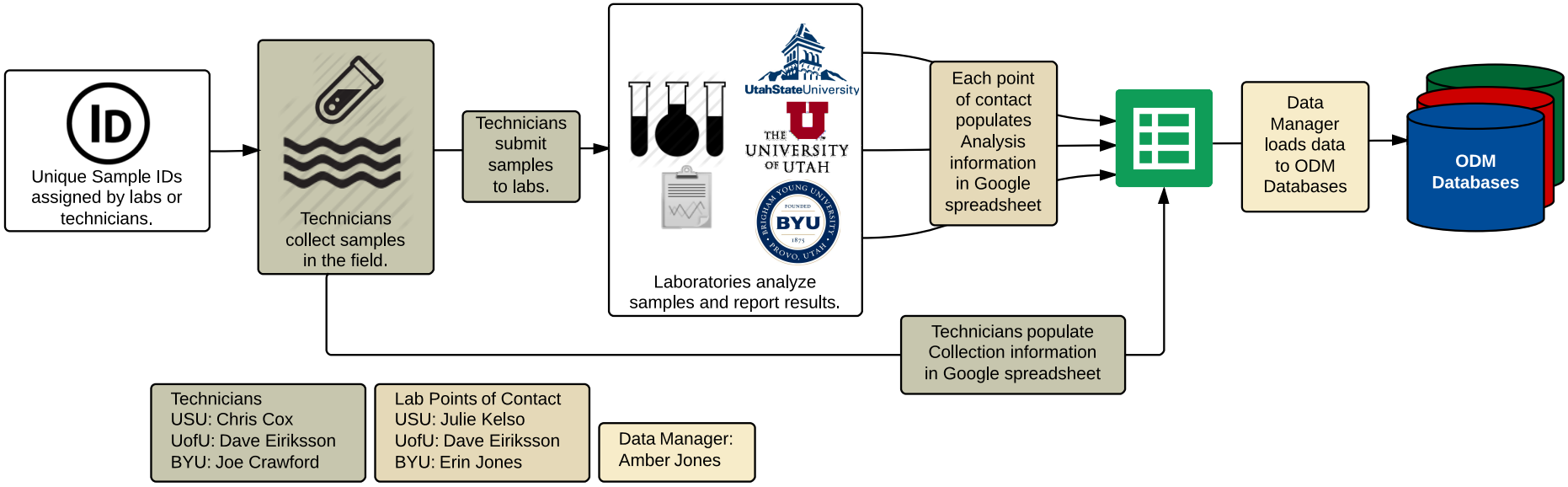
Jeff or Amber restart the Streaming Data Loader for the site.

Equipment Management

- http://data.iutahepscor.org/gamutmanagement/
- Feedback? Questions?
- Compile list of issues/questions/feature requests

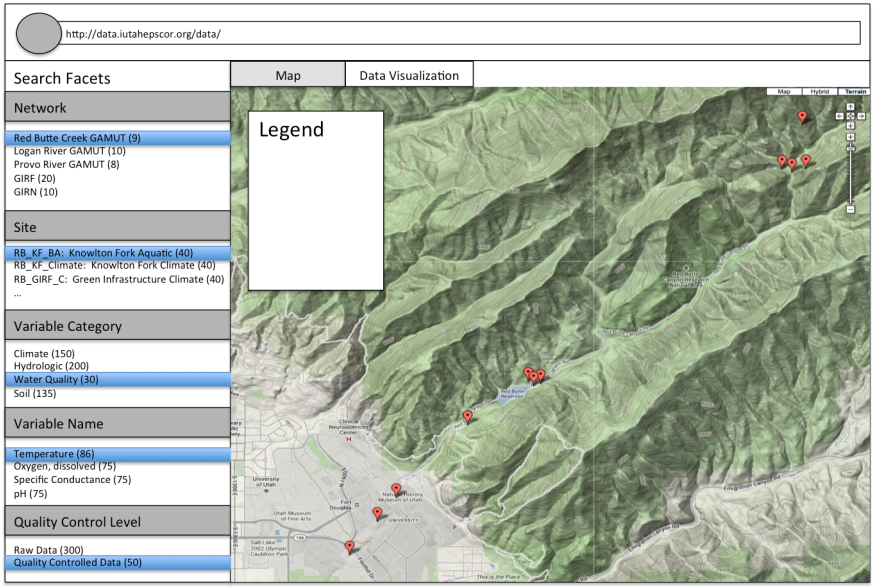


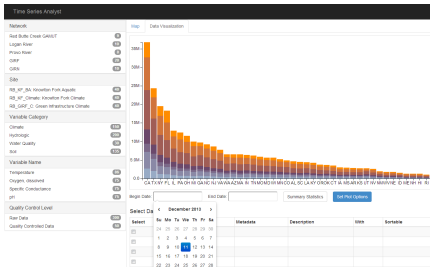
Samples Data Management

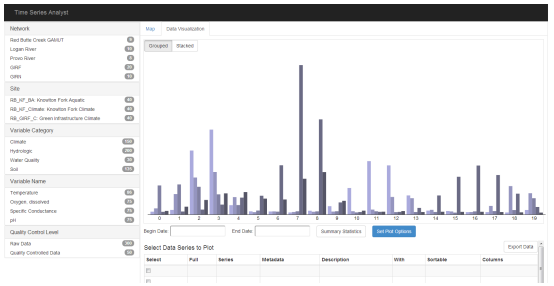


- Feedback on the workflow.
- Need to format such that lab POCs can enter results.
- Need to complete samples table.
- Should update regularly.
- Should we include sampling events in the Equipment Management?
- -9999 represents “No Data”

Time Series Analyst Overhaul







- Faceted searching.
- Integrated visualization and map search interface.
- Various plotting options and summary statistics will be available.
- Data series export will be available.

Sensor Quality Assurance/ Quality Control

- Quality Assurance: "set of processes or steps taken to ensure that the sensor network and protocols are developed and adhered to in a way that minimizes inaccuracies in the data produced. The purpose of QA is to produce high-quality data while minimizing the need for corrective measures to improve data quality."
- Quality Control: "occurs after the data are generated and tests whether they meet the necessary requirements for quality outlined by the end users."

Campbell, J. L., Rustad, L. E., Porter, J. H., Taylor, J. R., Ethan, W., Shanley, J. B., … Dereszynski, E. W. (2013). Quantity is Nothing without Quality. *BioScience*, *63*(7), 574–585. doi:10.1525/bio.2013.63.7.10

Quality Assurance

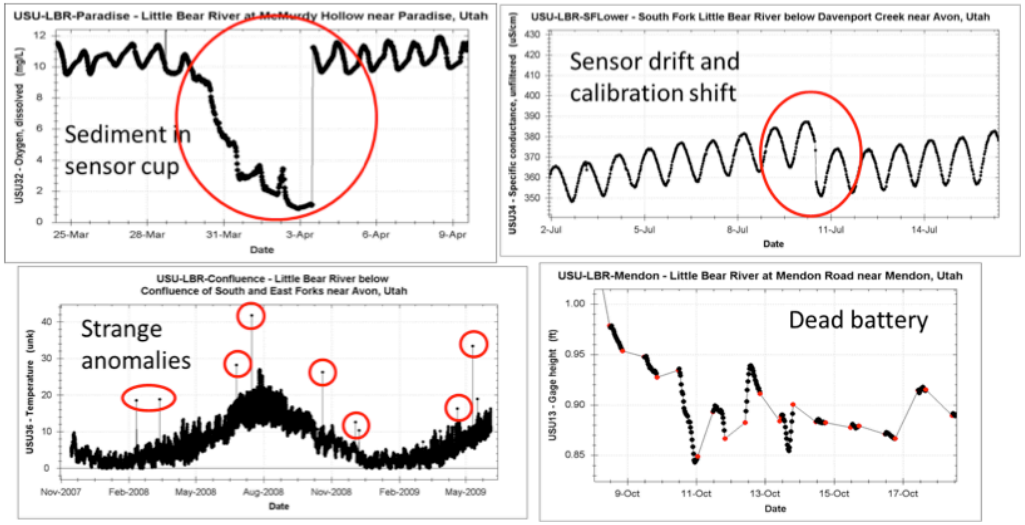
Developing a plan:

- Replicate sensors
- Factory maintenance schedule
- Field maintenance schedule and procedures
- Sensor calibration schedule and procedures
- Data checking: visually, automated alerts
- Recording and tracking events: "Field technicians are often aware of sensor-related inaccuracies resulting from routine maintenance, repairs, or other interruptions of service. Tracking these events is crucial for identifying and understanding the origin of inaccurate data."
- Equipment management database/interface is a quality assurance tool to tracking equipment, events, deployments, servicing, etc.

Quality Control

Types of QC Tests that can be implemented:

1. **Range:** tests can be based on extreme values or statistical measures (e.g., two standard deviations from the mean), can be specific to temporal variability (daily rate of change, different range for different seasons).
2. **Persistence:** check for constant value over some period of time.
3. **Change in Slope:** check that the rate of change is realistic.
4. **Internal consistency:** evaluate differences between related sensors.
5. **Spatial Consistency:** make comparisons between sites.

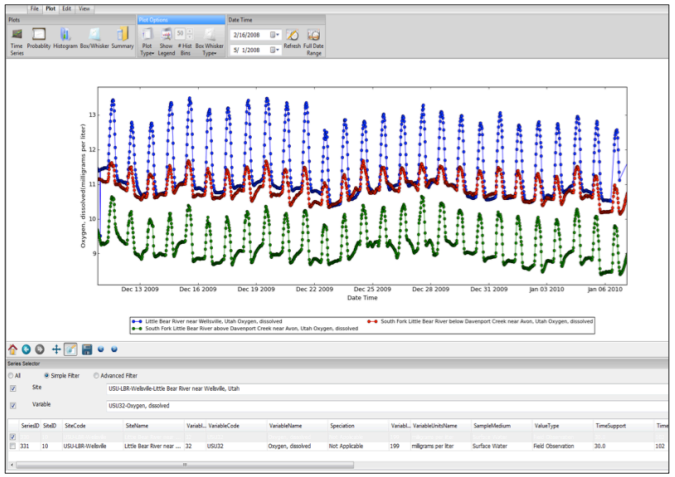


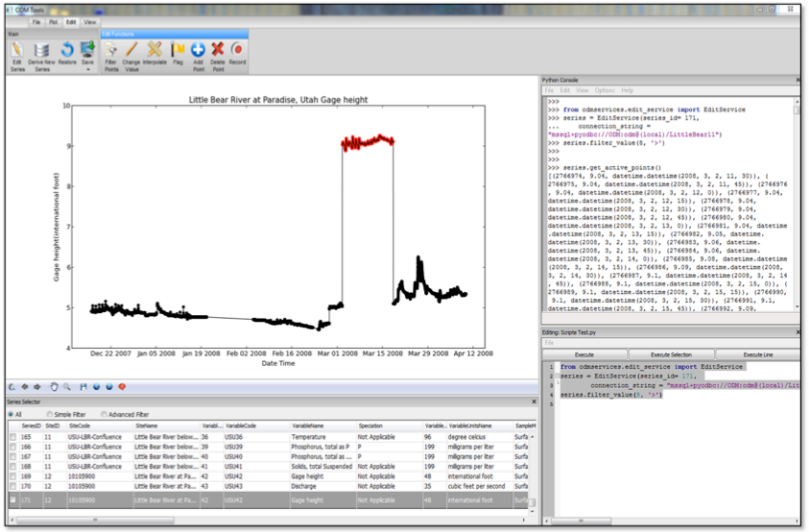
Automated Checks, Alerts, Rules

- Straddles QA (sending alerts) and QC (running tests and flagging for data quality).
- Working on developing processes for regularly (daily?) checking through the data to generate email alerts.
- Currently developing as procedures in SQL Server databases: can write SQL code to define checks.
- Could use Campbell Scientific software (RTMC) for alerts. Eventually plan to develop tool with functionality to define alerts and trigger flags.
- Questions:
  - What do we actually want to check?
  - What rules to generate alerts?
  - What format of alerts?
  - What rules to generate flags?
  - What formats of flags?

Post Processing

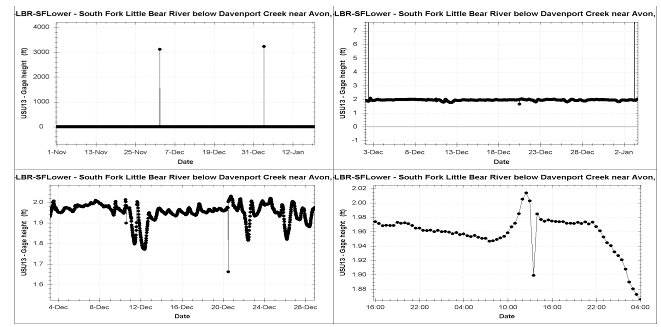
- Can't do everything automatically!
- ODM Tools Python: new, improved program
  - Improved plotting interface
  - Advanced series selection
  - Scripting for data editing: track provenance, reproducible
- What specific tools to build into ODM Tools Python?
  - Delete
  - Interpolate
  - Add Value
  - Smoothing
- Alternatively, could write scripts in Python, Matlab, R, etc.
- Need a plan for post-processing QC
  - Which variables?
  - What steps for each variable?



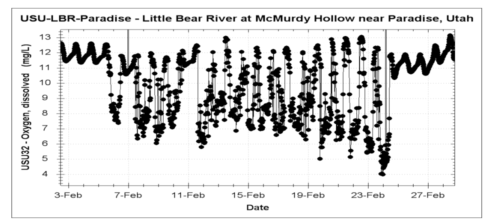


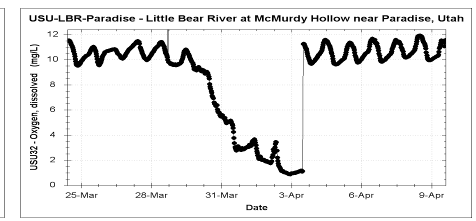
Post Processing

- Need a plan for post-processing QC
  - Which variables?
  - What steps for each variable?

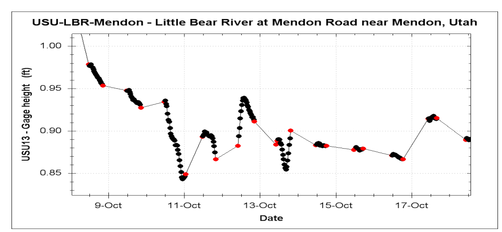


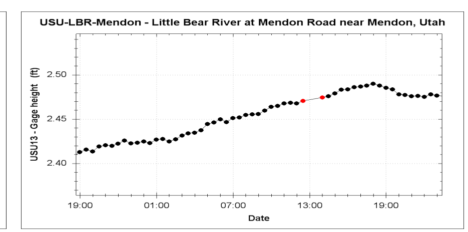
- Correcting anomalous data points
  - Deletion
  - Linear Interpolation
  - Other methods?



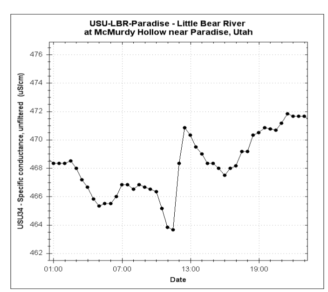


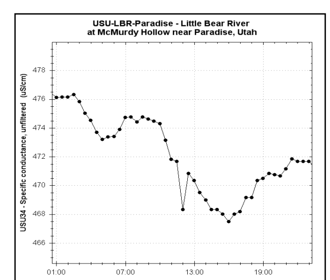
- Correcting data gaps
  - Linear Interpolation
  - Other methods?

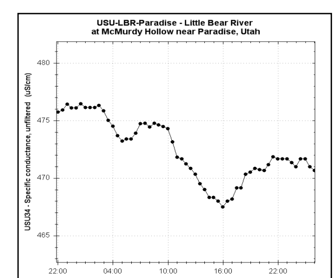




- Adjusting for Drift
  - Linear Drift Correction
  - Other methods?







Data QAQC Workflow/Timeline

1. Schedules established/followed for maintenance and calibration in GAMUT Sensors/Fieldwork SOP
2. Datalogger programs/files recorded and curated
3. Offsets and other constants stored in tables on each datalogger to reduce post-processing
4. Data streaming to ODM databases
5. Technicians maintain visual check on data
6. Rules established for automated alerts- run on database or on Loggernet
7. Events recorded using equipment management interface
8. Rules established for automated flagging
9. Steps/rules established for post-processing QC
10. Post-processing QC conducted by technicians
11. Approved datasets released within one year of collection