Takeaways

Earth science data challenge = DIVERSITY
Takeaways

Earth science data challenge = DIVERSITY

Better tools AND change in science practice
Takeaways

Earth science data challenge = DIVERSITY

Better tools AND change in science practice

Scientist-centered design approach and agile development
The earth system is changing

IPCC AR5 2013

Climate.gov

Atmospheric CO$_2$ at Mauna Loa Observatory

Scripps Institution of Oceanography
NOAA Earth System Research Laboratory
Understanding these changes requires complex and diverse data...

**Carbon Cycle**
- Photosynthesis
- Sensible Heat Exchange
- Transpiration
- Autotrophic Respiration
- Allocation and Vegetative Turnover
- Soil and Litter Turnover
- Ground Heat Flux
- Evaporation
- Infiltration
- Runoff

**Water Cycle**
- Atmospheric transport
- Input
- Output
- Rainfall
- Evapotranspiration
- Recycled rainfall
- Run-off

+ Energy, Nutrients, Biota ....

(Global Carbon Project 2015)
Diverse data: Monitoring ~3500 seedlings over 5 years

- Seedlings < 25 cm tall
- Trees DBH > 30 cm
Diverse data: Physical landscape

- Soil temperature
- Soil moisture
- Soil profile
- Soil surface
- Potential evaporation
- Canopy cover
- Topography / Structure
Diverse data: 3-D Structural representation of forest
Diverse and complex ecosystems
Earth science spends a lot of time translating data to information

We want to spend more time building knowledge and wisdom
A metadata reporting framework for standardization and synthesis of ecohydrological field observations

Danielle Svehla Christianson

Charuleka Varadharajan, Brad Christofferson, Matteo Detto, Boris Faybishenko, Bruno Gimenez, Kolby Jardine, Robinson I. Negron-Juarez, Gilberto Pastorello, Thomas Powell, Jeff Warren, Brett Wolfe, Nathan McDowell, Joe Wright, Niro Higuchi, Lara M. Kueppers, Jeff Chambers, and Deb Agarwal

AGU Fall Meeting | 15 Dec 2016
dschristianson@lbl.gov
NGEE Tropics: data integration into Earth System Model for carbon cycling
NGEE Tropics: data integration into Earth System Model for carbon cycling
NGEE Tropics: data integration into Earth System Model for carbon cycling

Model parameterization & benchmarking
Discovery science for model processes
Metadata are required

Example SAP-Install & Equipment Metadata

Example Tree Metadata
Metadata are required: Data interpretation, translation, & unit conversion

Example Tree Metadata

- Tree_ID
- Taxa
- Tree Height, Method, & Uncertainty
- DBH
- Height of DBH
- XYZ Coordinates
- Meas_Pos_ID
- Data Logger

Example SAP-Install & Equipment Metadata

- Sensor_ID
- Probe & Circuit Resistance
- Needle Depth
- Needle Distance
- Bark Thickness
- Sensor Height

Example Tree Metadata

- Crown Illumination Index

Data interpretation, translation, & unit conversion

Christianson et al. (in prep) 2/10
Metadata are required: Data interpretation, translation, & unit conversion. Cross-site & cross approach comparison.
Metadata are required:

- Data interpretation, translation, & unit conversion
- Cross-site & cross approach comparison
- Quality Assurance / Quality Control (QA/QC)
- Gap-filling

Example Tree Metadata

- Tree ID
- Taxa
- Tree Height, Method, & Uncertainty
- DBH
- Height of DBH
- XYZ Coordinates
- Data Logger
- Meas_Pos_ID

Example SAP-Install & Equipment Metadata

- Sensor_ID
- Probe & Circuit Resistance
- Needle Depth
- Needle Distance
- Bark Thickness
- Sensor Height
- Tree Diameter at Sensor

Crown Illumination Index
Metadata are required:

Data interpretation, translation, & unit conversion
Cross-site & cross approach comparison
Quality Assurance / Quality Control (QA/QC)
Gap-filling

Reporting metadata for Earth System Models is a CHALLENGE!
Leverage existing protocols / Iterative user-centered design process
Leverage existing protocols / **Iterative user-centered design process**

- Provider / User interviews
- Trial entry (with observation)

Feedback synthesized & incorporated
Leverage existing protocols / Iterative user-centered design process

Provider / User interviews
Trial entry (with observation)

ENSO Measurement Campaign
- Sap flow
- Soil water content (automated)
- Leaf surface temperature
- Dendrometry (automated)
- Radiation
- Leaf gas exchange
- Leaf water potential
- Non-structural carbohydrates

Feedback synthesized & incorporated
<table>
<thead>
<tr>
<th>Description</th>
<th>Initials of person entering info</th>
<th>INSTALL_DATE</th>
<th>TECHNICIANS</th>
<th>APPROACH</th>
<th>SENSOR_FIELDNAME</th>
<th>TREE_ID</th>
<th>MEAS_POS_ID</th>
<th>Measure_Location_Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exemple</td>
<td>disc</td>
<td>20160304</td>
<td>Bruno Oliva Gimenez</td>
<td>free text</td>
<td>NGEE-T_aspflow_protocol.pdf</td>
<td>G12</td>
<td>00001</td>
<td>free text</td>
</tr>
<tr>
<td>Exemple</td>
<td>disc</td>
<td>20160412</td>
<td>Bruno Oliva Gimenez</td>
<td>free text</td>
<td>NGEE-T_aspflow_protocol.pdf</td>
<td>SFM1E80T</td>
<td>00001C</td>
<td>free text</td>
</tr>
<tr>
<td>(part entering info in this row)</td>
<td>20180311</td>
<td>Kolby Jardine, Bruno Gimenez, Kleber Silva</td>
<td></td>
<td>ICT install protocol in manual rev 4.0; pg 15-29</td>
<td>ICT install protocol in manual rev 4.0; pg 15-31</td>
<td>SFM1F714</td>
<td>07626</td>
<td>sensor facing tower</td>
</tr>
<tr>
<td></td>
<td>20180311</td>
<td>Kolby Jardine, Bruno Gimenez, Kleber Silva</td>
<td></td>
<td>ICT install protocol in manual rev 4.0; pg 15-29</td>
<td>ICT install protocol in manual rev 4.0; pg 15-31</td>
<td>SFM1F70S</td>
<td>07628</td>
<td>sensor facing tower</td>
</tr>
<tr>
<td></td>
<td>20180311</td>
<td>Kolby Jardine, Bruno Gimenez, Kleber Silva</td>
<td></td>
<td>ICT install protocol in manual rev 4.0; pg 15-29</td>
<td>ICT install protocol in manual rev 4.0; pg 15-31</td>
<td>SFM1F700</td>
<td>07637</td>
<td>sensor facing tower</td>
</tr>
<tr>
<td></td>
<td>20180311</td>
<td>Kolby Jardine, Bruno Gimenez, Kleber Silva</td>
<td></td>
<td>ICT install protocol in manual rev 4.0; pg 15-29</td>
<td>ICT install protocol in manual rev 4.0; pg 15-31</td>
<td>SFM1F70W</td>
<td>07654</td>
<td>sensor facing tower</td>
</tr>
</tbody>
</table>

**Set of metadata templates (various formats)**
Data linked across spatiotemporal scales

NGEE-Tropics Multiscale Measurement Position Hierarchy
Measurement / Measurement collections should be associated with finest resolution possible.

- **Site_ID**
  - Area encompassing an identifiable study unit.
  - Location information required.

- **Plot_ID**
  - Sub-areas of site in which collections of measurement are made.

- **Location_ID**
  - Point locations: Tree, tower, crane, well, pit.

- **Meas_Pos_ID**
  - Sub-locations for permanently installed sensors.

- **Sample_ID**
  - Unique identifiers for sample-based measures.

Details submitted to Data Team; Master lists at Metadata_Central

Details described in E-Field_Log metadata template (Metadata_Central)
Data linked across spatiotemporal scales
Modular framework → flexible, scalable reporting of complex & diverse data
Modular framework → flexible, scalable reporting of complex & diverse data

Metadata Reporting Framework

- D Data files are linked to metadata files.
- M Metadata files.
- X Key metadata variables and relationships are shown.

Measurement Setting Description

- Site / Plot (on-line templates curated by Data Team)
  List of sites / plots with descriptions & contacts.
- X Site_ID / Plot_ID defined by location
Modular framework → flexible, scalable reporting of complex & diverse data
Modular framework → flexible, scalable reporting of complex & diverse data

**Metadata Reporting Framework**

- Data files are linked to Metadata files.
- Key metadata variables and relationships are shown.

**Measurement Setting Description**

**Site / Plot**
(on-line templates curated by Data Team)
List of sites / plots with descriptions & contacts.

- **Site_ID / Plot_ID** defined by location

**Tree**
List of all trees that are measured.

- **Site_ID** or **Plot_ID** ↔ **Tree_ID**
  Tree ID assigned with species and size information reported

**Locations**
List of locations for trees, towers, soil pits, etc.

- **Site_ID** or **Plot_ID** ↔ **Location_ID**
  Location ID described by xyz-coordinates

**Equipment**
List of all instruments (sensors & loggers).

- **Sensor_ID**, model, type, manual, contact associated with **Sensor_FieldName**

**Data Package Description**

**Data File Description**

**Field_Event_Log**
List of field events including installation, calibration, repairs, suspicious data, etc.

- **Tree_ID / Location_ID** & install details define **Meas_Pos_ID**

- **Sensor_FieldName** ↔ **Meas_Pos_ID**

**< Measurement specific > - Install**
Description of installation details.

- **Tree_ID / Location_ID & install details** define **Meas_Pos_ID**

**< Measurement specific > - < Details >**
Description of additional measurement details, including sample-based campaigns.

8/10
Modular framework → flexible, scalable reporting of complex & diverse data

**Metadata Reporting Framework**
- Template File or On-line template
- Data files are linked to Metadata files.
- Key metadata variables and relationships are shown.

**Measurement Setting Description**
- Site / Plot (on-line templates provided by Data Teams)
  - List of sites / plots with descriptions & contacts.
- Site_ID / Plot_ID defined by location
- Tree_ID assigned with species and size information reported
- Location_ID described by xyz-coordinates
- Tree_ID / Location_ID & install details define Meas_Pos_ID
- Sensor_FieldsName associated with Sensor_FieldName

**Data Package**
- **Data Package Description** (entered with data submission)
  - Site_IDs, Data description, Authors, Privacy

**Data File Description**
- **Data File(s)**
  - Sample-based data includes Sample_ID
  - Field_Event_Log
    - List of field events including installation, calibration, repairs, suspicious data, etc.
  - <Measurement specific> - Install
    - Description of installation details.
  - <Measurement specific> - <Details>
    - Description of additional measurement details, including sample-based campaigns.
Modular framework → flexible, scalable reporting of complex & diverse data

**Metadata Reporting Framework**

- Template File or On-line template
- Data files are linked to Metadata files
- Key metadata variables and relationships are shown

**Measurement Setting Description**

- **Site / Plot**
  - Site / Plot (on-line templates curated by Data Team)
  - List of sites / plots with descriptions & contacts.

- **Tree**
  - List of all trees that are measured.
  - Tree_ID or Plot_ID defined by location

- **Locations**
  - List of locations for trees, towers, soil pits, etc.
  - Site_ID or Plot_ID linked to Location_ID
  - Location_ID described by xyz-coordinates

- **Equipment**
  - List of all instruments (sensors & loggers).
  - Sensor_ID, model, type, manual, contact associated with Sensor_FieldName

**Data Package Description**

- **Data Package Description**
  - (entered with data submission)
  - Site_IDs, Data description, Authors, Privacy

- **Raw or processed data files**
  - Data file(s)
  - Sample-based data includes Sample_ID

**Data File Description**

- **File Submission Metadata**
  - Data FIle <filename> associated with:

- **E-Field Log** <filename>

- **Site_ID / Tree_ID / Meas_Pos_ID**

- **Data Column Description** <filename>
  - (for automatically-generated data files; sample-based data may have column description embedded in data file)
Modular framework → flexible, scalable reporting of complex & diverse data
Modular framework → flexible, scalable reporting of complex & diverse data

**Metadata Reporting Framework**
- Data files are linked to Metadata files.
- Key metadata variables and relationships are shown.

**Measurement Setting Description**
- Site / Plot (on-line templates curated by Data Team)
- List of sites / plots with descriptions & contacts.
- Site_ID / Plot_ID defined by location

**Tree**
- List of all trees that are measured.
  - Site_ID or Plot_ID ↔ Tree_ID
  - Tree_ID assigned with species and size information reported

**Locations**
- List of locations for trees, towers, soil pits, etc.
  - Site_ID or Plot_ID ↔ Location_ID
  - Location_ID described by xyz-coordinates

**Equipment**
- List of all instruments (sensors & loggers).
  - Sensor_ID, model, type, manual, contact associated with Sensor_FieldName

**Field_Event_Log**
- List of field events including installation, calibration, repairs, suspicious data, etc.

**E-Field_Log**
- Description of installation details.
  - Tree_ID / Location_ID & install details define Meas_Pos_ID
  - Sensor_FieldName ↔ Meas_Pos_ID

**< Measurement specific > - Install**
- Description of additional measurement details, including sample-based campaigns.

**Data Package**
- **Data Package Description** (entered with data submission)
  - Site_IDs, Data description, Authors, Privacy

**Raw or processed data files**
- **Data File(s)**
  - Data_file <filename> associated with:
  - E-Field_Log <filename>
  - Site_ID / Tree_ID / Meas_Pos_ID
  - Data_Column_Description <filename>
  - Data_Column_Description (for automatically-generated data files, sample-based data may have column description embedded in data file)

- **Data File Description**
  - Meas_Pos_ID assigned to data columns (for files with more than 1 sensor)
Usability = Benefits to NGEE Tropics

Data submissions for 6 measurement types at 6 sites
Usability = Benefits to NGEE Tropics

Data submissions for 6 measurement types at 6 sites

NGEE Tropics repository autofill for Site / Plot info
Usability = Benefits to NGEE Tropics

Data submissions for 6 measurement types at 6 sites
NGEE Tropics repository autofill for Site / Plot info
Formal communication between data providers & users
Usability = Benefits to NGEE Tropics

Data submissions for 6 measurement types at 6 sites
NGEE Tropics repository autofill for Site / Plot info
Formal communication between data providers & users
Organization of field data
Usability = Benefits to NGEE Tropics

Data submissions for 6 measurement types at 6 sites
NGEE Tropics repository autofill for Site / Plot info
Formal communication between data providers & users
Organization of field data
Sap flow synthesis across 9 sites
Usability = Benefits to NGEE Tropics

Data submissions for 6 measurement types at 6 sites
NGEE Tropics repository autofill for Site / Plot info
Formal communication between data providers & users
Organization of field data
Sap flow synthesis across 9 sites

Areas of development

Metadata reporting platforms
Quantify costs of flexible reporting
1) Developed a metadata reporting framework for model-data integration
   - Links data across spatiotemporal scales
   - Enables flexible, scalable reporting of complex and diverse data
   - Is usable by data providers
1) Developed a metadata reporting framework for model-data integration
   - Links data across spatiotemporal scales
   - Enables flexible, scalable reporting of complex and diverse data
   - Is usable by data providers

2) Iterative user-centered design with providers and users critical
A metadata reporting framework for standardization and synthesis of ecohydrological field observations

1) Developed a metadata reporting framework for model-data integration
   → Links data across spatiotemporal scales
   → Enables flexible, scalable reporting of complex and diverse data
   → Is usable by data providers

2) Iterative user-centered design with providers and users critical

Thank you!

Danielle Svehla Christianson
dschristianson@lbl.gov
AGU Fall Meeting | 15 Dec 2016
AmeriFlux Data Processing

Integrating automated and manual data management across software technologies and an international network to generate timely data products

Danielle Svehla Christianson, Norm Beekwilder, Stephen Chan, You-Wei Cheah, Housen Chu, Sigrid Dengel, Fianna O’Brien, Gilberto Pastorello, Megha Sandesh, Margaret S. Torn, Deb Agarwal

AGU 2017 Fall Meeting
 Flux-Met data enables understanding of carbon, water, energy cycling

Sensitive and involved measurements
Diverse observations
AmeriFlux is a voluntary network of independent scientists at 300 sites

Dedicated to provide shared, standardized, high quality data products
Many challenges to managing flux-met data across 300 independent sites

Flux-met data is complex

→ Automated / organized diagnostics, issue tracking at hierarchical levels
Many challenges to managing flux-met data across 300 independent sites

Flux-met data is complex
   → Automated / organized diagnostics, issue tracking at hierarchical levels

300 sites = 300 approaches!
   → Standardization, automated Format QA/QC with auto-repair
Many challenges to managing flux-met data across 300 independent sites

Flux-met data is complex
  ➔ Automated / organized diagnostics, issue tracking at hierarchical levels

300 sites = 300 approaches!
  ➔ Standardization, automated Format QA/QC with auto-repair

Lots of moving pieces yet no single off-the-shelf software solution
  ➔ Integrated processing and communication tracking system
Many challenges to managing flux-met data across 300 independent sites

Flux-met data is complex
  ➔ Automated / organized diagnostics, issue tracking at hierarchical levels

300 sites = 300 approaches!
  ➔ Standardization, automated Format QA/QC with auto-repair

Lots of moving pieces yet no single off-the-shelf software solution
  ➔ Integrated processing and communication tracking system

Goal: Timely, quality data processing with traceable provenance ...
  ... with minimal manual effort.
Many challenges to managing flux-met data across 300 independent sites

Flux-met data is complex
   → Automated / organized diagnostics, issue tracking at hierarchical levels

300 sites = 300 approaches!
   → Standardization, automated Format QA/QC with auto-repair

Lots of moving pieces yet no single off-the-shelf software solution
   → Integrated processing and communication tracking system

Goal: Timely, quality data processing with traceable provenance ... ... with minimal manual effort.

1 Agile development with user feedback
2 Interdisciplinary and diverse team
AmeriFlux BASE Data Processing Pipeline

Does file format comply with standard?

Format QA/QC

Does data pass quality checks?

Data QA/QC

BASE Data Product available to network

BASE Publish

BASE = the standard AmeriFlux data product consisting of half-hourly / hourly time series of 114 variable types
AmeriFlux data requires hierarchical tracking and communications
AmeriFlux data requires hierarchical tracking and communications.
AmeriFlux data requires hierarchical tracking and communication.

- How many sites have current BASE?
- Which sites need Data Team action?
- What is the average processing time?
- Is the BASE product current?
- What data issues are present?
- What is the next step?
- What files/uploads comprise BASE?

What is the status of my upload?
What action needs to happen?
Does my file meet format standard?

Data QA/QC, generate BASE

Format QA/QC
AmeriFlux data requires hierarchical tracking and communication.

- **Network**
  - Is the BASE product current?
  - What data issues are present?
  - What is the next step?
  - What files/uploads comprise BASE?

- **Data Processing Dashboard**
  - How many sites have current BASE?
  - Which sites need Data Team action?
  - What is the average processing time?

- **Site**
  - Is the BASE product current?
  - What data issues are present?
  - What is the next step?
  - What files/uploads comprise BASE?

- **Data Upload**

- **Full data record**

- **Status Table**
  - What is the status of my upload?
  - What action needs to happen?
  - Does my file meet format standard?

- **Data QA/QC, generate BASE**

- **QA/QC Report**

- **Format QA/QC** 5/15
AmeriFlux data requires hierarchical tracking and communication.

**Network**
- Data Upload
- Site
- Full data record

**Data Processing Dashboard**
- Is the BASE product current?
- What data issues are present?
- What is the next step?
- What files / uploads comprise BASE?

**Email communication (JIRA)**
- Email communication
- Data QA/QC, generate BASE
- Format QA/QC

**Status Table**
- What is the status of my upload?
- What action needs to happen?
- Does my file meet format standard?

**QA/QC Report**
Processing Modules in Python

SQL Database used to track data processing

WCF and WebAPI Web Services
Format QA/QC is automated

- Uploaded
- Format Checker
- File Ready for BASE
- AutoRepaired File
- Format Fixer
- File NOT AutoRepaired
- File NOT Considered for BASE
- Original File
- File AutoRepaired
Format QA/QC is automated

FP-In Standard specifies:
- Filename
- Variable names
- Timestamps
- Temporal Resolution
- Missing value format
- Missing data
- Gap-filled
Format QA/QC is automated

- Uploaded
- Format Checker
- File Ready for BASE
- File NOT AutoRepaired
- File NOT Considered for BASE
- QA/QC Report
- Email communication (JIRA)

Pass = Warning, OK
FAIL = Error, Critical
Format QA/QC Report and JIRA for communications about uploaded files

<table>
<thead>
<tr>
<th>Filename</th>
<th>Site ID</th>
<th>Start Year</th>
<th>End Year</th>
<th>Current Status</th>
<th>Date of Status Update</th>
<th>Uploader</th>
<th>Upload Date</th>
<th>Site Contact</th>
</tr>
</thead>
</table>
Format QA/QC Report and JIRA for communications about uploaded files
Format QA/QC Report and JIRA for communications about uploaded files

Norm Beekwilder commented on [QAQC-614]

Re: Format Results - Review requested | US-Sne data uploaded on Aug 24, 2017

Dear Kyle S Hemes,

Thank you for uploading data for US-Sne on Aug 24, 2017 (see complete file list below).

In the context of the new processing for the AmeriFlux database, we are applying a new file Format QA/QC scheme. Format QA/QC is conducted to assess the compliance of your data submission with AmeriFlux FP-In format. This step is critical to ensure that your data is formatted properly and will be interpreted correctly by us and other data users. Details about the format requirements can be found at [http://ameriflux.lbl.gov/half-hourly-hourly-data-upload-format/](http://ameriflux.lbl.gov/half-hourly-hourly-data-upload-format/). Data that passes the Format QA/QC checks will be automatically queued for Data QA/QC, the next step in the AmeriFlux data processing pipeline.

We have processed your data through our Format QA/QC scheme. The results are listed below.

**REVIEW REQUESTED**

We will proceed to Data QA/QC; however, we found the following potential issues. Some issues may impact Data QA/QC results if left unaddressed. If you would like to address these Format QA/QC issues, please re-upload your data at [https://ameriflux.lbl.gov/data/upload-data/](https://ameriflux.lbl.gov/data/upload-data/) and/or reply to this email. See [http://ameriflux.lbl.gov/half-hourly-hourly-data-upload-format/](http://ameriflux.lbl.gov/half-hourly-hourly-data-upload-format/) for AmeriFlux FP-In formatting instructions. To proceed with Data QA/QC, we fixed issues as detailed below. Data QA/QC results will be sent in a separate email.

These automatic fixes were attempted to address issues encountered in the following files US-Sne_HH_201605240030_201706010000.csv:

- Filename component fixed: ts-start (start time)

Please correct these issues in subsequent data submissions.

View the status of your uploaded files at [https://ameriflux.lbl.gov/qaqc-reports-data-team/](https://ameriflux.lbl.gov/qaqc-reports-data-team/). Links to view the Format QA/QC report for each file are at the end of this email.

We appreciate your help with standardizing the data submission format. We hope that fixing any identified issues will not take too much time from your work, but it is necessary to enable timely data processing. Please reply to this email with any questions. You can track communications on this Format QA/QC report at QAQC-614 using your AmeriFlux account ID and password to login.

Sincerely,

AMP Data Team

**List of uploaded file(s) and corresponding Format QA/QC Report link:**

Data QA/QC is semi-automated.
Data QA/QC is semi-automated

- File Ready for BASE
- Combiner
- Data QAQC Checker
- Curation
- Manual assessment
- BASE Attempt Failed
- Published BASE–BADM

Diagnostics generated for:
- Outliers
- Diurnal & Seasonal Cycles
- Standard Units
- Multivariate comparison
- Radiation shift + shadow errors
- USTAR
Data QA/QC is semi-automated

File Ready for BASE

Combiner

Data QAQC Checker

Curation

Manual assessment

BASE Attempt Failed

Proposed BASE

QA/QC Report

Published BASE–BADM

BASE–BADM Generation

Email communication (JIRA)
# QA/QC Report (Beta): Data

This page is under active development. Please bear with us as we make the QA/QC reports more user friendly. For more information, please contact ameriflux-support@lbl.gov.

This report details the results of the AmeriFlux QA/QC data processing pipeline. For more information, see

- How to Read This Report
- Result Code Definitions
- FAQ

### Base Generation Report: US-ADR

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Status</th>
<th>Report ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERROR</td>
<td>PHYSICAL LIMITS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OK</td>
<td>PAIRWISE VARIABLE COMPARISON</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WARNING</td>
<td>FILE COMBINER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WARNING</td>
<td>DIURNAL SEASONAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OK</td>
<td>FORMAT QAQC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERROR</td>
<td>RADIATION TIMESHIFT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Click to View Details
Data QA/QC Report and JIRA for site level analysis
Data QA/QC Report and JIRA for site level analysis

### QA/QC Report (Beta)

This page is under active development.

This report details the results of the QA/QC process for the Base Generation Report: US-ADR.

#### How to Read This Report

- **Result Code Definitions**
- **FAQ**

#### Base Generation Report: US-ADR

<table>
<thead>
<tr>
<th>Error</th>
<th>Base Generation: US-ADR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERROR</td>
<td>BASE Generation: US-ADR</td>
</tr>
<tr>
<td>ERROR</td>
<td>PHYSICAL LIMITS</td>
</tr>
<tr>
<td>OK</td>
<td>PAIRWISE VARIABLES</td>
</tr>
<tr>
<td>WARNING</td>
<td>FILE COMBINER</td>
</tr>
<tr>
<td>WARNING</td>
<td>DIURNAL SEASONALITY</td>
</tr>
<tr>
<td>OK</td>
<td>FORMAT QA/QC</td>
</tr>
<tr>
<td>ERROR</td>
<td>RADIATION TIMES</td>
</tr>
</tbody>
</table>

#### Data QAQC Results

**Type:** Data QAQC Results  
**Priority:** Medium  
**Affects Version/s:** None  
**Component/s:** None  
**Labels:** Results_Sent  
**Process ID(s):** 11086  
**Site ID:** US-ADR  
**Temporal Resolution:** HH  
**FTP Link:** ftp://ftp01.fluxdata.org/ameriflux_downloads/data/US-ADR_0403460\*11086\*output  
**sandbox:** [DATA QAQC]\*  

#### Description

QAQC completed with the following results: critical(0), error(12), warning(6), ok(63)

#### Attachments

- Drop files to attach, or browse.

#### Issue Links

  - REPLACE WITH...  
  - FIXED  
  - FIXED  
  - ATTEMPT DAT...  

#### SLAs

- **-15:03** Time to resolution within 24h
- **6:28** Time to first response within 8h

#### People

- **Assignee:** Unassigned  
  - Assign to me
- **Reporter:** Michael T. Moreo
- **Request participants:** andraski@usgs.gov
- **Organizations:** US-ADR
- **Votes:** Vote for this issue  
  - Start watching this issue
- **Watchers:**

#### Service Desk request

- **Request type:** Data QAQC  
- **Customer status:** Publishable  
- **Channel:** JIRA  
- **View customer request id**

#### Dates

- **Created:** 07/Dec/17 10:27 AM  
- **Updated:** 07/Dec/17 11:56 AM
Data QA/QC is semi-automated
## QA/QC for the Data Team

Use the table below to determine which sites have data that require action.

### Links:
- Report Summary Table
- Creator UI

### Table:

<table>
<thead>
<tr>
<th>Site</th>
<th>Process Type</th>
<th>Priority</th>
<th>Published Version</th>
<th>Progress</th>
<th>Status</th>
<th>Reports</th>
<th>Issue</th>
<th>Last Activity</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-ADR</td>
<td>Tower</td>
<td>Not Set</td>
<td>V</td>
<td></td>
<td>Data ready to be published</td>
<td>11080</td>
<td>QMOC-1143</td>
<td>2017-12-07</td>
<td>Michael Moreo</td>
</tr>
<tr>
<td>US-Srr</td>
<td>Tower</td>
<td>Not Set</td>
<td>V</td>
<td></td>
<td>Data ready to be published</td>
<td>11077</td>
<td>QMOC-1143</td>
<td>2017-12-01</td>
<td>Brian Bergamaschi</td>
</tr>
<tr>
<td>US-Tw1</td>
<td>Tower</td>
<td>Not Set</td>
<td>V</td>
<td></td>
<td>Data ready to be published</td>
<td>11038</td>
<td>QMOC-1130</td>
<td>2017-11-21</td>
<td>Dennis Baldocchi</td>
</tr>
<tr>
<td>US-S02</td>
<td>Tower</td>
<td>Not Set</td>
<td>V</td>
<td></td>
<td>Ready for Data QA/QC</td>
<td>11004</td>
<td>QMOC-1104</td>
<td>2017-11-16</td>
<td>Quinn Thomas</td>
</tr>
<tr>
<td>US-Tw4</td>
<td>Tower</td>
<td>Not Set</td>
<td>V</td>
<td></td>
<td>Data ready to be published</td>
<td>11011</td>
<td>QMOC-1095</td>
<td>2017-11-13</td>
<td>Dennis Baldocchi</td>
</tr>
<tr>
<td>US-Tw4</td>
<td>Tower</td>
<td>Not Set</td>
<td>V</td>
<td></td>
<td>Data ready to be published</td>
<td>11012</td>
<td>QMOC-1092</td>
<td>2017-11-13</td>
<td>Dennis Baldocchi</td>
</tr>
</tbody>
</table>

### Statuses:
- ⬤ ⬤ ⬤ ⬤ ⬤: Format issues: Waiting for AmeriFlux Data Team
- ⬤ ⬤ ⬤ ⬤ ⬤: Format issues: Waiting for Tower Team
- ⬤ ⬤ ⬤ ⬤: Data issues: Waiting for Tower Team
- ⬤ ⬤ ⬤ ⬤ ⬤: Data issues: Waiting for AmeriFlux Data Team
- ⬤ ⬤ ⬤ ⬤ ⬤: Data ready to be published
- ⬤ ⬤ ⬤ ⬤ ⬤ ⬤: Data Products Up-to-Date

### Key:
- ▲ - Data Team action required
- ✔ - Complete
Tracking sites across the network

### Data Processing Status

Use the table below to determine which sites have data that require action.

#### Key:
- ▼ = Site Team action required
- ✓ = Complete

<table>
<thead>
<tr>
<th>Site</th>
<th>Published Version</th>
<th>Progress</th>
<th>Status</th>
<th>Issue</th>
<th>Last Activity</th>
<th>Uploader</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-Tw1</td>
<td>2-1</td>
<td></td>
<td>Data ready to be published</td>
<td>QAQC-1130</td>
<td>2017-11-21</td>
<td>Sara Knox</td>
</tr>
<tr>
<td>US-Tw4</td>
<td>4-1</td>
<td></td>
<td>Data ready to be published</td>
<td>QAQC-1095</td>
<td>2017-11-13</td>
<td>Elke Eichelmann</td>
</tr>
<tr>
<td>US-Myb</td>
<td>3-1</td>
<td>▼</td>
<td>Data issues: Waiting for Tower Team</td>
<td>QAQC-1094</td>
<td>2017-11-13</td>
<td>Samuel Chamberlain</td>
</tr>
<tr>
<td>US-Tw4</td>
<td>3-1</td>
<td></td>
<td>Data ready to be published</td>
<td>QAQC-1092</td>
<td>2017-11-13</td>
<td>Elke Eichelmann</td>
</tr>
<tr>
<td>US-Sne</td>
<td>1-4</td>
<td>✓</td>
<td>Published as version 1-4</td>
<td>QAQC-809</td>
<td>2017-09-12</td>
<td>Kyle S Hemes</td>
</tr>
<tr>
<td>US-Snd</td>
<td>2-1</td>
<td></td>
<td>Ready for Data QA/QC</td>
<td>QAQC-666</td>
<td>2017-08-30</td>
<td>AmeriFlux Data Team</td>
</tr>
<tr>
<td>US-Tw2</td>
<td>1-1</td>
<td></td>
<td>Ready for Data QA/QC</td>
<td>QAQC-649</td>
<td>2017-08-30</td>
<td>AmeriFlux Data Team</td>
</tr>
<tr>
<td>US-Ton</td>
<td>9-1</td>
<td>▼</td>
<td>Format issues: Waiting for Tower Team</td>
<td>QAQC-553</td>
<td>2017-08-17</td>
<td>Siyan Ma</td>
</tr>
<tr>
<td>US-Var</td>
<td>9-1</td>
<td>▼</td>
<td>Format issues: Waiting for Tower Team</td>
<td>QAQC-409</td>
<td>2017-07-31</td>
<td>Siyan Ma</td>
</tr>
</tbody>
</table>
Success?

Goal: Timely, quality data processing with traceable provenance ...  
... with minimal manual effort (< 1 hours / site).

Prior system
→ Data team time 1-2 weeks
→ First response ~ 6 months

Format QA/QC:
→ Data team time ~15 min
→ First response < 1 day

Data QA/QC:
→ Data team time ~2 hours
→ First response ~ 1-2 weeks

Published: 67 site years
Key Features:

- Automated Format QA/QC
- Automated generation of draft emails
- Data QA/QC diagnostics with highlights
- Organization of info for manual assessment
- JIRA issue tracking
Key Features: Automated
  → Format QA/QC
  → Automated generation of draft emails
  → Data QA/QC diagnostics with highlights
  Organization of info for manual assessment
  JIRA issue tracking

Challenges: Simplifying communications of complex process
  Organization of info for manual assessment
  Fully integrating / maintaining software tools
Agile development with user feedback

Interdisciplinary and diverse team

Key Features:
- Automated
- Format QA/QC
- Automated generation of draft emails
- Data QA/QC diagnostics with highlights
- Organization of info for manual assessment
- JIRA issue tracking

Challenges:
- Simplifying communications of complex process
- Organization of info for manual assessment
- Fully integrating / maintaining software tools
AmeriFlux Data Processing

Integrating automated and manual data management across software technologies and an international network to generate timely data products

Danielle Svehla Christianson, Norm Beekwilder, Stephen Chan, You-Wei Cheah, Housen Chu, Sigrid Dengel, Fianna O’Brien, Gilberto Pastorello, Megha Sandesh, Margaret S. Torn, Deb Agarwal

AGU 2017 Fall Meeting

Thank you!
dschristianson@lbl.gov | Contact us at ameriflux-support@lbl.gov.