**NBD(NIST Big Data) Requirements WG Use Case Template Aug 11 2013**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | DOE-BER AmeriFlux and FLUXNET Networks | |
| **Vertical (area)** | | Research: Earth Science | |
| **Author/Company/Email** | | Deb Agarwal, Lawrence Berkeley Lab. daagarwal@lbl.gov | |
| **Actors/Stakeholders and their roles and responsibilities** | | AmeriFlux scientists, Data Management Team, ICOS, DOE TES, USDA, NSF, and Climate modelers. | |
| **Goals** | | AmeriFlux Network and FLUXNET measurements provide the crucial linkage between organisms, ecosystems, and process-scale studies at climate-relevant scales of landscapes, regions, and continents, which can be incorporated into biogeochemical and climate models. Results from individual flux sites provide the foundation for a growing body of synthesis and modeling analyses. | |
| **Use Case Description** | | AmeriFlux network observations enable scaling of trace gas fluxes (CO2, water vapor) across a broad spectrum of times (hours, days, seasons, years, and decades) and space. Moreover, AmeriFlux and FLUXNET datasets provide the crucial linkages among organisms, ecosystems, and process-scale studies—at climate-relevant scales of landscapes, regions, and continents—for incorporation into biogeochemical and climate models | |
| **Current**  **Solutions** | **Compute(System)** | | NERSC |
| **Storage** | | NERSC |
| **Networking** | | ESNet |
| **Software** | | EddyPro, Custom analysis software, R, python, neural networks, Matlab. |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | ~150 towers in AmeriFlux and over 500 towers distributed globally collecting flux measurements. |
| **Volume (size)** | |  |
| **Velocity**  **(e.g. real time)** | |  |
| **Variety**  **(multiple datasets, mashup)** | | The flux data is relatively uniform, however, the biological, disturbance, and other ancillary data needed to process and to interpret the data is extensive and varies widely. Merging this data with the flux data is challenging in today’s systems. |
| **Variability (rate of change)** | |  |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues) and Quality** | | Each site has unique measurement and data processing techniques. The network brings this data together and performs a common processing, gap-filling, and quality assessment. Thousands of users |
| **Visualization** | | Graphs and 3D surfaces are used to visualize the data. |
| **Data Types** | | Described in “Variety” above. |
| **Data Analytics** | | Data mining, data quality assessment, cross-correlation across datasets, data assimilation, data interpolation, statistics, quality assessment, data fusion, etc. |
| **Big Data Specific Challenges (Gaps)** | | Translation across diverse datasets that cross domains and scales. | |
| **Big Data Specific Challenges in Mobility** | | Field experiment data taking would be improved by access to existing data and automated entry of new data via mobile devices. | |
| **Security & Privacy**  **Requirements** | |  | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | |  | |
| **More Information (URLs)** | | Ameriflux.lbl.gov  www.fluxdata.org | |
| **Note:** <additional comments> | | | |