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

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | DOE-BER Subsurface Biogeochemistry Scientific Focus Area | |
| **Vertical (area)** | | Research: Earth Science | |
| **Author/Company/Email** | | Deb Agarwal, Lawrence Berkeley Lab. daagarwal@lbl.gov | |
| **Actors/Stakeholders and their roles and responsibilities** | | LBNL Sustainable Systems SFA 2.0, Subsurface Scientists, Hydrologists, Geophysicists, Genomics Experts, JGI, Climate scientists, and DOE SBR. | |
| **Goals** | | The Sustainable Systems Scientific Focus Area 2.0 Science Plan (“SFA 2.0”) has been developed to advance predictive understanding of complex and multiscale terrestrial environments relevant to the DOE mission through specifically considering the scientific gaps defined above. | |
| **Use Case Description** | | Development of a **G**enome-**E**nabled **Wa**tershed **S**imulation **C**apability (**GEWaSC)** that will provide a predictive framework for understanding how genomic information stored in a subsurface microbiome affects biogeochemical watershed functioning, how watershed-scale processes affect microbial functioning, and how these interactions co-evolve. While modeling capabilities developed by our team and others in the community have represented processes occurring over an impressive range of scales (ranging from a single bacterial cell to that of a contaminant plume), to date little effort has been devoted to developing a framework for systematically connecting scales, as is needed to identify key controls and to simulate important feedbacks. A simulation framework that formally scales from genomes to watersheds is the primary focus of this GEWaSC deliverable. | |
| **Current**  **Solutions** | **Compute(System)** | | NERSC |
| **Storage** | | NERSC |
| **Networking** | | ESNet |
| **Software** | | PFLOWTran, postgres, HDF5, Akuna, NEWT, etc |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Terabase-scale sequencing data from JGI, subsurface and surface hydrological and biogeochemical data from a variety of sensors (including dense geophysical datasets) experimental data from field and lab analysis |
| **Volume (size)** | |  |
| **Velocity**  **(e.g. real time)** | |  |
| **Variety**  **(multiple datasets, mashup)** | | Data crosses all scales from genomics of the microbes in the soil to watershed hydro-biogeochemistry. The SFA requires the synthesis of diverse and disparate field, laboratory, and simulation datasets across different semantic, spatial, and temporal scales through GEWaSC. Such datasets will be generated by the different research areas and include simulation data, field data (hydrological, geochemical, geophysical), ‘omics data, and data from laboratory experiments. |
| **Variability (rate of change)** | | Simulations and experiments |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues) and Quality** | | Each of the sources samples different properties with different footprints – extremely heterogeneous. Each of the soruces has different levels of uncertainty and precision associated with it. In addition, the translation across scales and domains introduces uncertainty as does the data mining. Data quality is critical. |
| **Visualization** | | Visualization is crucial to understanding the data. |
| **Data Types** | | Described in “Variety” above. |
| **Data Analytics** | | Data mining, data quality assessment, cross-correlation across datasets, reduced model development, statistics, quality assessment, data fusion, etc. |
| **Big Data Specific Challenges (Gaps)** | | Translation across diverse and large 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)** | | A wide array of programs in the earth sciences are working on challenges that cross the same domains as this project. | |
| **More Information (URLs)** | | Under development | |
| **Note:** <additional comments> | | | |