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

|  |  |  |
| --- | --- | --- |
| **Use Case Title** | UAVSAR Data Processing, Data Product Delivery, and Data Services | |
| **Vertical (area)** | Scientific Research: Earth Science | |
| **Author/Company/Email** | Andrea Donnellan, NASA JPL, [andrea.donnellan@jpl.nasa.gov](mailto:andrea.donnellan@jpl.nasa.gov); Jay Parker, NASA JPL, [jay.w.parker@jpl.nasa.gov](mailto:jay.w.parker@jpl.nasa.gov) | |
| **Actors/Stakeholders and their roles and responsibilities** | NASA UAVSAR team, NASA QuakeSim team, ASF (NASA SAR DAAC), USGS, CA Geological Survey | |
| **Goals** | Use of Synthetic Aperture Radar (SAR) to identify landscape changes caused by seismic activity, landslides, deforestation, vegetation changes, flooding, etc; increase its usability and accessibility by scientists. | |
| **Use Case Description** | A scientist who wants to study the after effects of an earthquake examines multiple standard SAR products made available by NASA. The scientist may find it useful to interact with services provided by intermediate projects that add value to the official data product archive. | |
| **Current**  **Solutions** | **Compute(System)** | Raw data processing at NASA AMES Pleiades, Endeavour. Commercial clouds for storage and service front ends have been explored. |
|  | **Storage** | File based. |
|  | **Networking** | Data require one time transfers between instrument and JPL, JPL and other NASA computing centers (AMES), and JPL and ASF.  Individual data files are not too large for individual users to download, but entire data set is unwieldy to transfer. This is a problem to downstream groups like QuakeSim who want to reformat and add value to data sets. |
|  | **Software** | ROI\_PAC, GeoServer, GDAL, GeoTIFF-suporting tools. |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | Data initially acquired by unmanned aircraft. Initially processed at NASA JPL. Archive is centralized at ASF (NASA DAAC). QuakeSim team maintains separate downstream products (GeoTIFF conversions). |
|  | **Volume (size)** | Repeat Pass Interferometry (RPI) Data: ~ 3 TB. Increasing about 1-2 TB/year.  Polarimetric Data: ~40 TB (processed)  Raw Data: 110 TB  Proposed satellite missions (Earth Radar Mission, formerly DESDynI) could dramatically increase data volumes (TBs per day). |
|  | **Velocity**  **(e.g. real time)** | RPI Data: 1-2 TB/year. Polarimetric data is faster. |
|  | **Variety**  **(multiple datasets, mashup)** | Two main types: Polarimetric and RPI. Each RPI product is a collection of files (annotation file, unwrapped, etc). Polarimetric products also consist of several files each. |
|  | **Variability (rate of change)** | Data products change slowly. Data occasionally get reprocessed: new processing methods or parameters. There may be additional quality assurance and quality control issues. |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues)** | Provenance issues need to be considered. This provenance has not been transparent to downstream consumers in the past. Versioning used now; versions described in the UAVSAR web page in notes. |
|  | **Visualization** | Uses Geospatial Information System tools, services, standards. |
|  | **Data Quality** | Many frames and collections are found to be unusable due to unforseen flight conditions. |
|  | **Data Types** | GeoTIFF and related imagery data |
|  | **Data Analytics** | Done by downstream consumers (such as edge detections): research issues. |
| **Big Data Specific Challenges (Gaps)** | Data processing pipeline requires human inspection and intervention. Limited downstream data pipelines for custom users.  Cloud architectures for distributing entire data product collections to downstream consumers should be investigated, adopted. | |
| **Big Data Specific Challenges in Mobility** | Some users examine data in the field on mobile devices, requiring interactive reduction of large data sets to understandable images or statistics. | |
| **Security & Privacy**  **Requirements** | Data is made immediately public after processing (no embargo period). | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | Data is geolocated, and may be angularly specified. Categories: GIS; standard instrument data processing pipeline to produce standard data products. | |
| **More Information (URLs)** | <http://uavsar.jpl.nasa.gov/>, <http://www.asf.alaska.edu/program/sdc>, <http://quakesim.org> | |
| **Note:** <additional comments> |  | |

**Note: No proprietary or confidential information should be included**