Deriving Earth Science Data Analytics Requirements

**Goal oriented Earth Science Data Analytics (ESDA)**
reveal requirements for needed data analytics tools/techniques

**Earth Science Data Analytics: Definition**
The process of examining, preparing, reducing, and analyzing large amounts of spatial (multi-dimensional), temporal, or spectral data using a variety of data types to uncover patterns, correlations and other information, to better understand our Earth.

**Earth Science Data Analytics: Goals**
- To glean knowledge
- To forecast/predict/model
- To derive new analytics tools

**Earth Science Data Analytics: Initial Requirements**
- Ingest from various sources; Homogenize data; Visualization; Sampling; Gridding
- Access large datasets; High speed processing; Subsetting, mining, machine learning
- Homogenize data; Intercomparison statistics; Pattern recognition
- Seek heterogeneous data relationships; Ingest from various sources; Image processing
- Looking for Community input
- Data exploration; Neural networks; Math/Stat modeling: Near Real Time data
- Access very large datasets; homogenize data; visualization

**Types of Analytics**
- Data Preparation
- Data Reduction
- Data Analysis

**Tools**
- R, SAS, Python, Java, C++
- SPSS, MATLAB, Minuitab
- CPLEX, GAMS, Gauss
- Tableau, Spotfire
- VBA, Excel, MySQL
- Javascript, Perl, PHP
- Open Source Databases
- PIO, NCL, Parallel NetCDF
- AWS, Cloud Solutions, Hadoop
- MPI, GIS, ROI-PAC, GDAL

**Techniques**
- Statistics functions
- Machine Learning
- Data Mining
- Natural Language Processing
- Linear/Non-linear Regression
- Logical Regression
- Time Series Models
- Clustering
- Decision Tree
- Factor Analysis
- Principal Component Analysis
- Neural Networks
- Bayesian Techniques
- Text Analytics
- Graph Analytics
- Visual Analytics
- Map Reduce

**Integrated Systems**
- EarthServer (http://www.earthserver.eu)
- NASA Earth Exchange (https://nex.nasa.gov/nex/)
- EDEN (http://cda.ornl.gov/projects/eden/#)
- EARTH_DATA (http://earthdata.nasa.gov)
- Giovanni (http://giovanni.gsfc.nasa.gov/giovanni/)

**Earth Science Data Analytics: Exemplary Tools, Techniques, Integrated Systems**

**Research Data Sharing without barriers**
Federaion of Earth Science Information Partners
Fostering connections to make data matter

**The good news… Earth Science Data Analytics: Preparing for the Future**
Central England NERC Training Alliance

**Big data analysis to fuel environmental research at Reading University**

2nd Annual Graduate Workshop on Environmental Data Analytics
July 27-31, 2015

**... offering degrees in Data Science**
**... summer school on Big Data Analytics**
**... online master’s degree in data analytics**

**Earth Science Data Analytics: Looking Ahead**
- Complete Gap Analysis between ESDA requirements and current tools/technologies
- Continue to evolve tools/techniques to address growing scope of the ‘Internet of Things’

**Motivation**
How can we maximize the usability of large heterogeneous datasets to glean knowledge out of the data?

**Methodology**
Categorize/Analyze ESDA use cases; derive data analytics requirements; associate tools/techniques; perform gap analysis

**Compiled from:**
- http://practicalanalytics.co/predictive-analytics-101/

**Thanks to the work of the Earth Science Information Partners (ESIP) Federation, Earth Science Data Analytics (ESDA) Cluster**