Scientific Data

Sponsor: National Center for Atmospheric Research (NCAR)

Team Members

- Seongmin Choi - Deployment Lead
- Bobby Crimi - Research Lead
- Connor Guerrieri - Source Control Lead
- Bo Han - Test Lead
- Hannah Keller - Team Lead
- Hannah Thomas - Documentation Lead

Project Description

Design a web-based application that allows users to choose subsets of web-based climate data by geographical area and time range and perform a sequence of analysis operations to the chosen dataset.

The output of the application is then:

- A copy of the subset data used
- Results of analysis
- Plots if applicable
- A log of reproducible steps for the entire analysis process

Project Description

Big picture requirements:

- Easily reproducible analysis steps, including data used
- Automated expertise
- Consistent workflow

A major issue in the scientific community today is the ability for scientists to easily reproduce each other's results. The main focus of this project is to create an application that allows users to easily share and reproduce the outcome of a given workflow.

Data

- NetCDF data format - self documenting data format for storing scientific data
- NARCCAP data - North American Regional Climate Change Assessment Program
  - Climate change data for only North America
  - 50 km resolution
  - 30 year simulations
  - Data collected every 3 hours
  - 5 years of data ~2GB
- Variety of options to pull/handle data
  - OPeNDAP
  - OpenClimateGIS
  - NCO

Project Concerns

Large concerns:

- Determining subsets of data to load into workspace
  - NetCDF files are very large (~2Gb)
- Making the results reproducible and shareable
- Extensibility

Small concerns:

- Creating a fluid working environment
- Getting results back as quickly as possible
- Incorporation of many different tools

Process Model

Agile - Adaptive

- No concrete requirements
- Open-ended implementation
- Research required
- Want an extensible application
- Fits with sponsor process model