Cloud-based analysis of NASA satellite data on OSDC

Project Matsu:

analysis of earth science data and data products

OSDC PIRE Workshop, 8 June, 2015

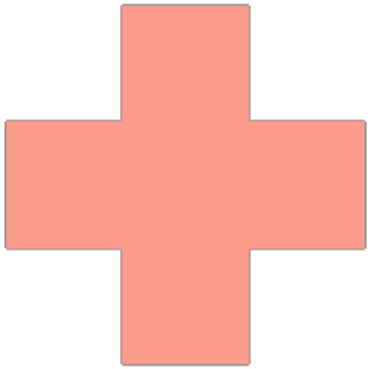
THE UNIVERSITY OF CHICAGO

OSDC OPEN SCIENCE DATA CLOUD

OCC OPEN CLOUD CONSORTIUM

NASA

Project Matsu



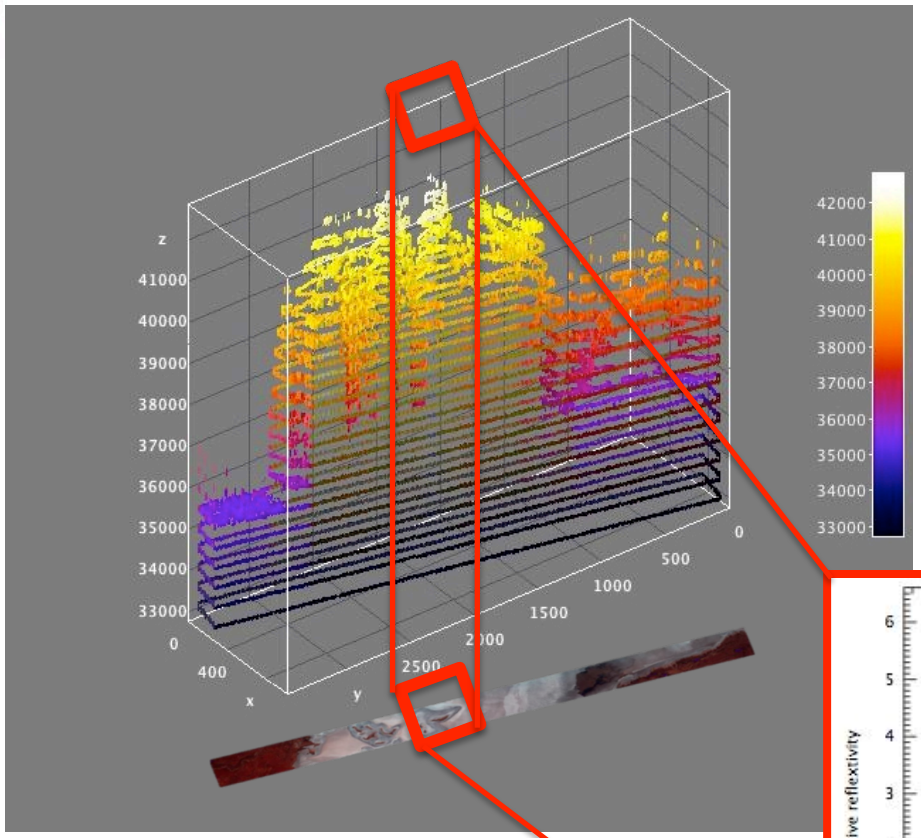
- Matsu, or Mazu, is a goddess of the sea said to protect fishermen and sailors.
- Initially formed in response to the 2010 Haiti earthquake, the name was chosen in the spirit of aiding those in need.
- A collaboration between members of the Open Cloud Consortium, NASA (lead, Dan Mandl at NASA GSFC), and others like the Namibian Department of Hydrology, involved with NASA's SensorWeb.
- Turning earth science observations into knowledge and information.

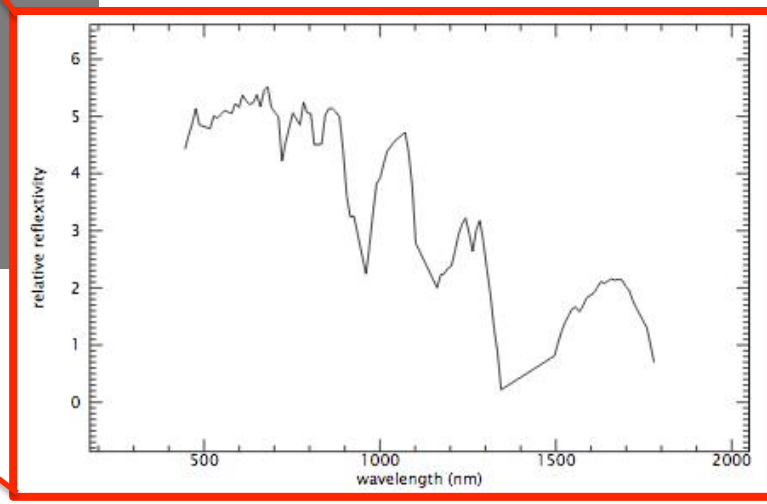
Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Earth Observing-1 (EO-1)

- Earth Observing-1 launched in Nov 2000 as a one year mission.
- The OSDC is used by NASA to process Earth Observing 1 (EO-1) satellite imagery from
  - Advanced Land Imager (ALI)
    - 9 simultaneous wavelength bands from 0.48–2.35 µm with 30-meter resolution plus a panchromatic band with higher 10-meter spatial resolution
    - 37 km x 42 km
    - Compare to Landsat 7
  - Hyperion imaging spectrometer
    - 242 wavelength bands 0.357–2.576 µm with 10-nm bandwidth
    - 7.7 km x 42 km

Hyperspectral satellite image data - Hyperion





Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Turning earth science observations into actionable information

Processing and serving data

- Generating (EO-1) satellite L1 and L2 data
- Web Coverage Processing Service for individual scenes
- Hadoop-based Matsu "Wheel" of analytics for processing all daily scenes

Aggregating and displaying data products

- Namibia flood dashboard
- Matsu "Wheel" analytic reports
- Web Map Service to publish data to an Open Geosocial API

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Turning earth science observations into actionable information

Processing and serving data

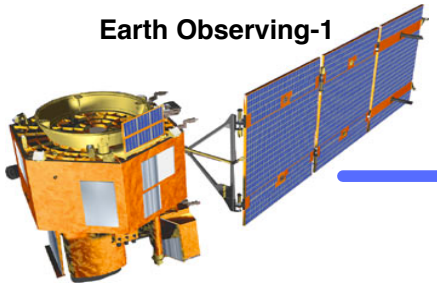
- Generating (EO-1) satellite L1 and L2 data
- Web Coverage Processing Service for individual scenes
- Hadoop-based Matsu "Wheel" of analytics for processing all daily scenes

Aggregating and displaying data products

- Namibia flood dashboard
- Matsu "Wheel" analytic reports
- Web Map Service to publish data to an Open Geosocial API

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Daily EO-1 Processing



New data observed by EO-1 and downloaded to NASA

NASA Goddard Space Flight Center

NASA images sent to OSDC Public Data Commons cloud for permanent storage

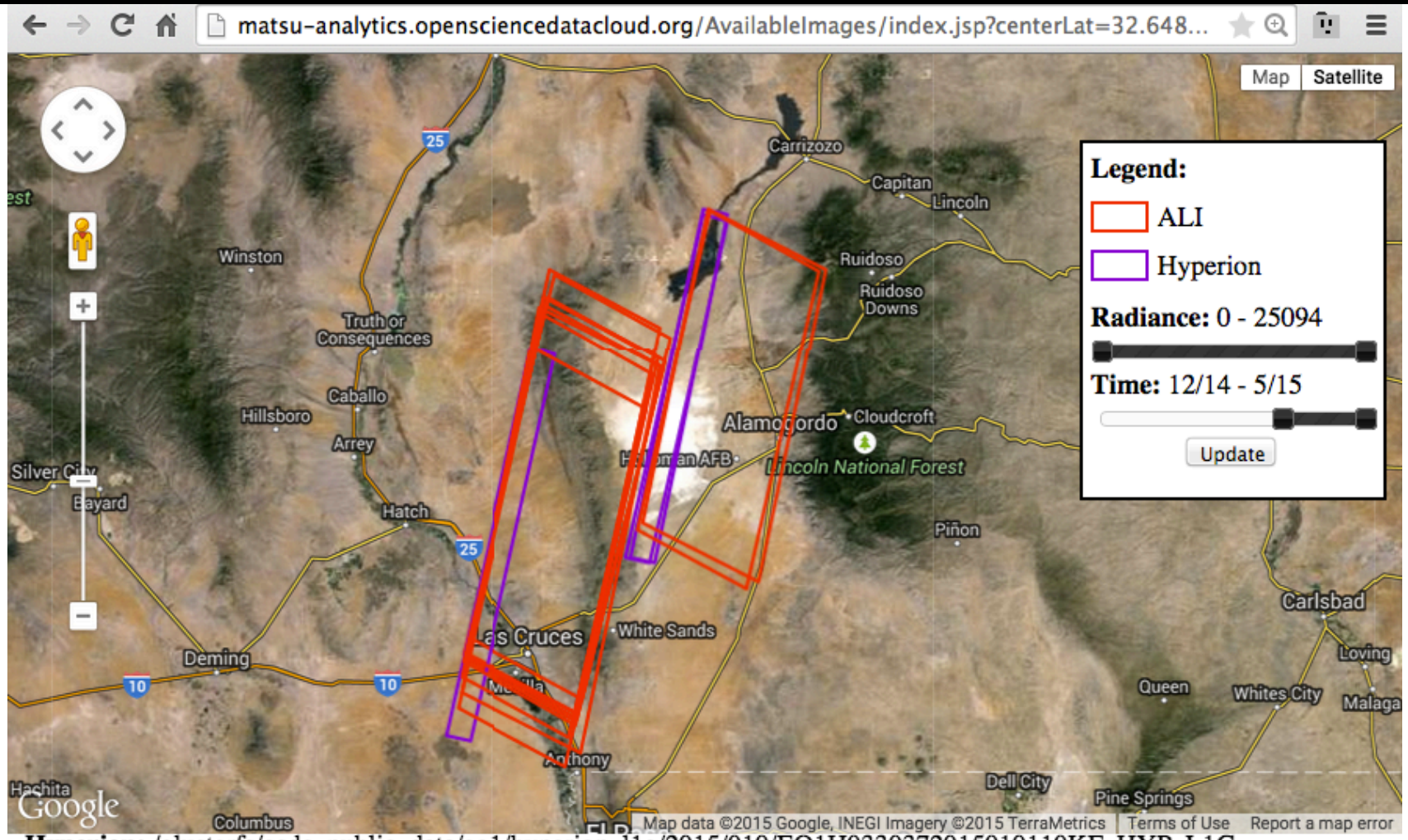
OSDC Public Data Commons (GlusterFS)

NASA researchers process EO-1 on OSDC virtual machines

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Available images for Level 1 EO-1 scenes

matsu-analytics.opensciencedatacloud.org/AvailableImages/



Hyperion: /glusterfs/osdc_public_data/eo1/hyperion_l1g/2015/010/EO1H0330372015010110KF_HYP_L1G
/glusterfs/osdc_public_data/eo1/hyperion_l1g/2015/021/EO1H0330372015021110P3_HYP_L1G
/glusterfs/osdc_public_data/eo1/hyperion_l1g/2015/056/EO1H0330372015056110KE_HYP_L1G

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Turning earth science observations into actionable information

Processing and serving data

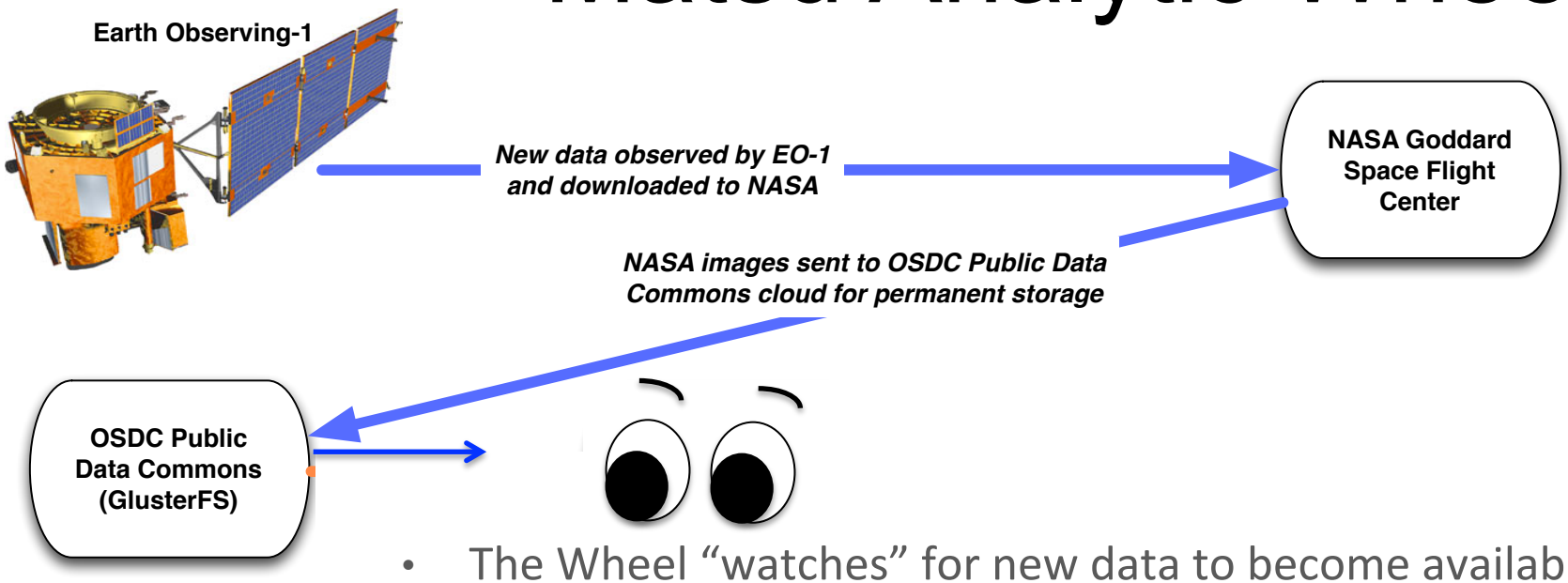
- Generating (EO-1) satellite L1 and L2 data
- Web Coverage Processing Service for individual scenes
- Hadoop-based Matsu "Wheel" of analytics for processing all daily scenes

Aggregating and displaying data products

- Namibia flood dashboard
- Matsu "Wheel" analytic reports
- Web Map Service to publish data to an Open Geosocial API

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Matsu Analytic Wheel

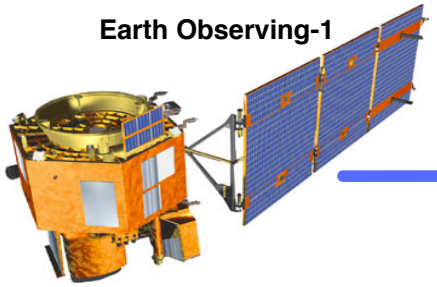


- The Wheel "watches" for new data to become available, using Apache Storm.
- When new data are detected, loaded into Hadoop's distributed file system for analysis using MapReduce.
- The Wheel analytics run each night, daily reports available the morning after data are received.

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Matsu Analytic "Wheel"

Earth Observing-1



New data observed by EO-1 and downloaded to NASA

NASA Goddard Space Flight Center

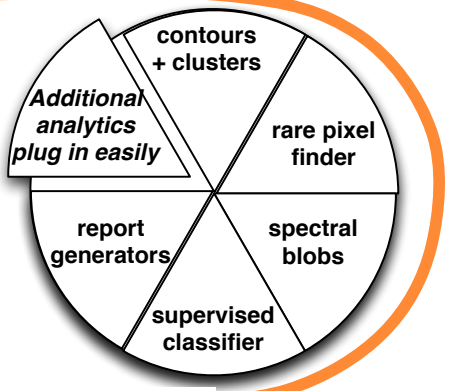
NASA images sent to OSDC Public Data Commons cloud for permanent storage

OSDC Public Data Commons (GlusterFS)

Data read into HDFS only once

HDFS

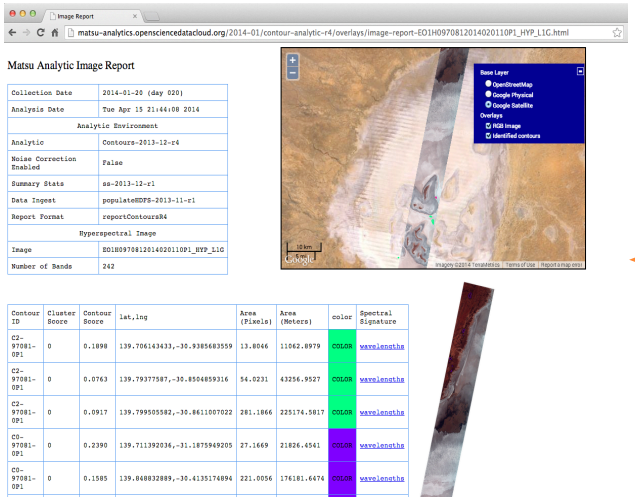
Wheel analytics run over data using MapReduce



Metadata stored

NoSql Database (Accumulo)

Analytic results stored



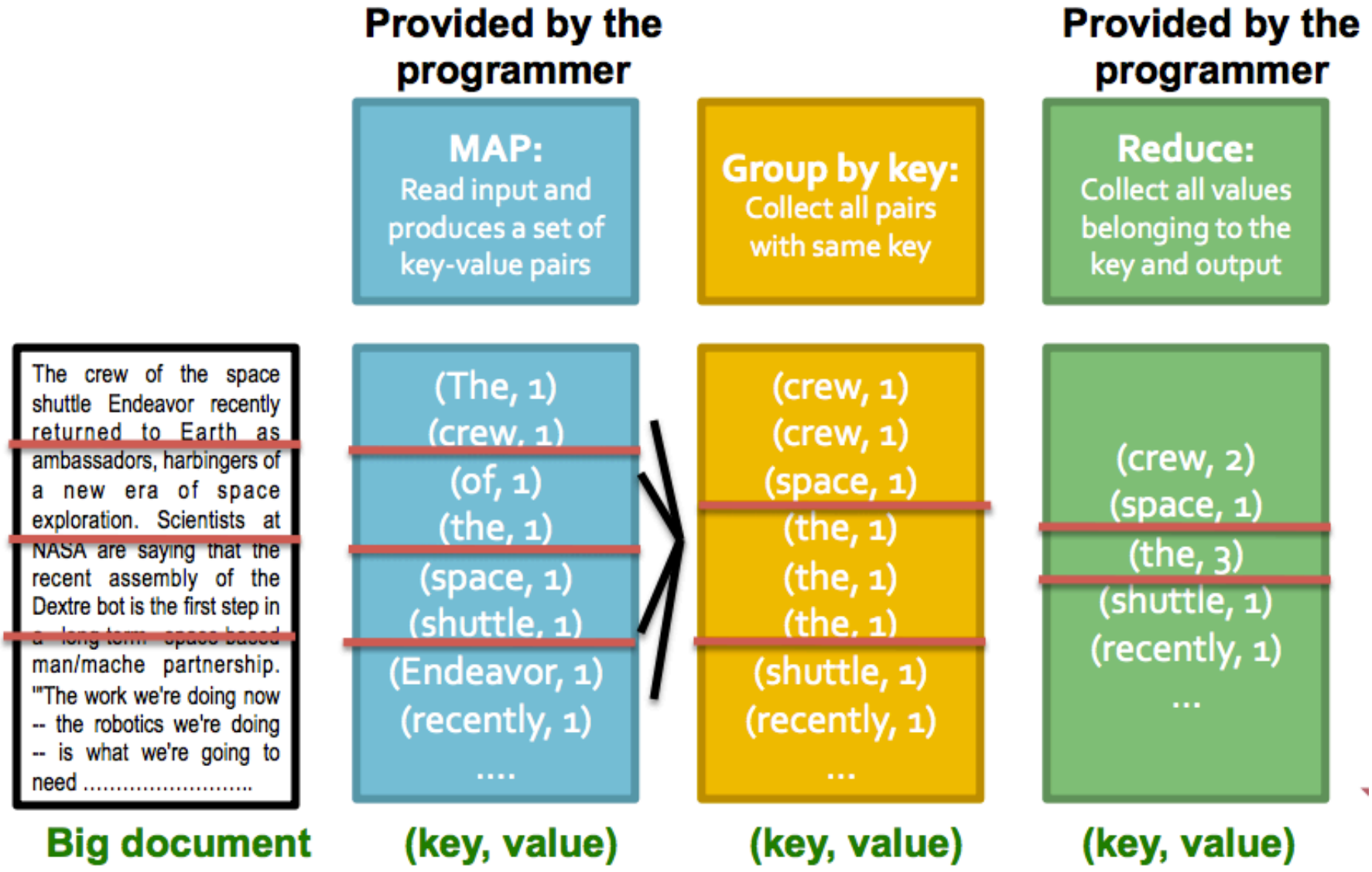
Analytic reports generated by Wheel are accessible via web browser

Secondary analysis can be done from analytic database

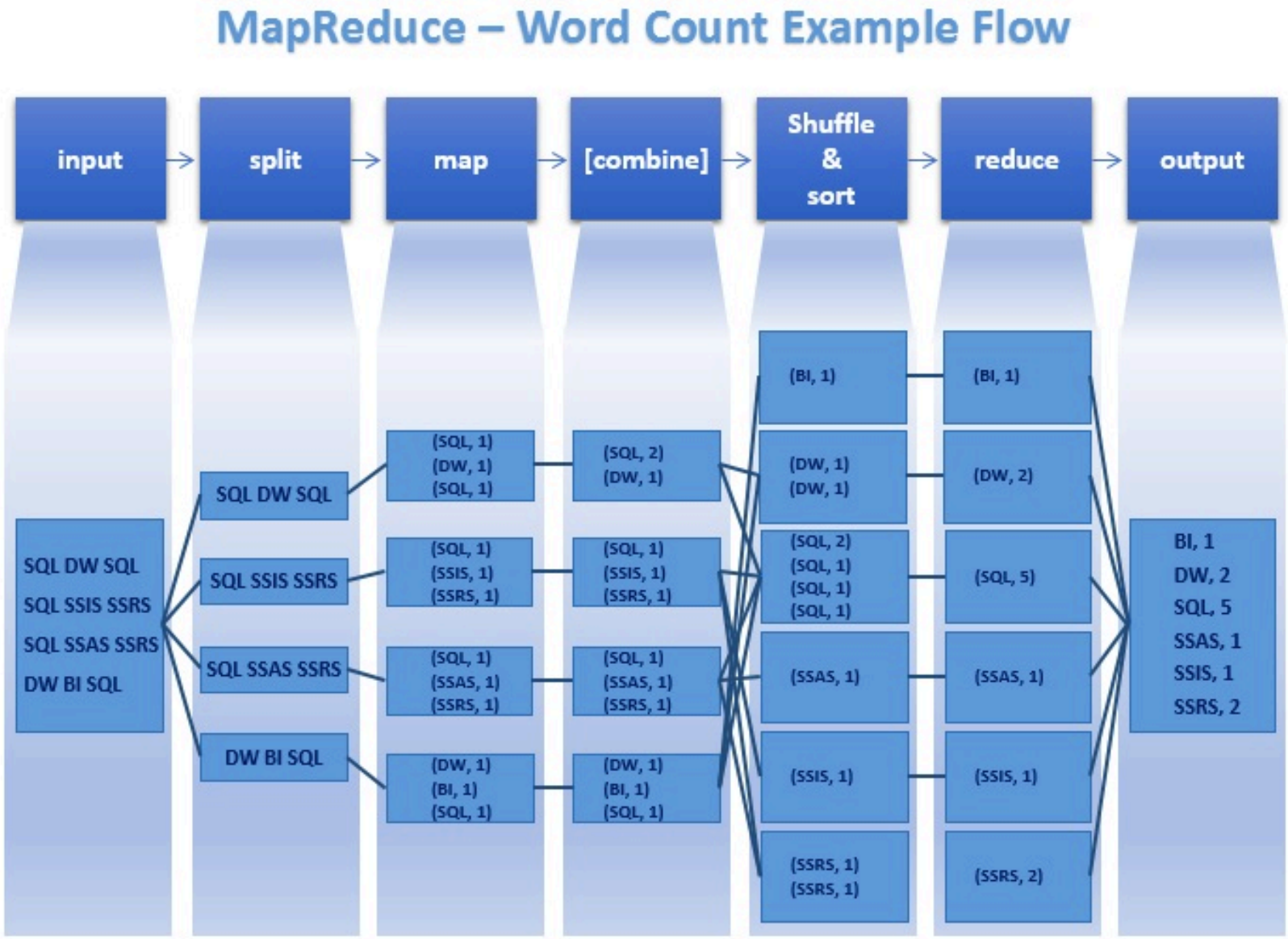
Maria Patterson (mtpatter@uchicago.edu)

Center for Data Intensive Science, University of Chicago

Hadoop-based analytics



Hadoop-based analytics



Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

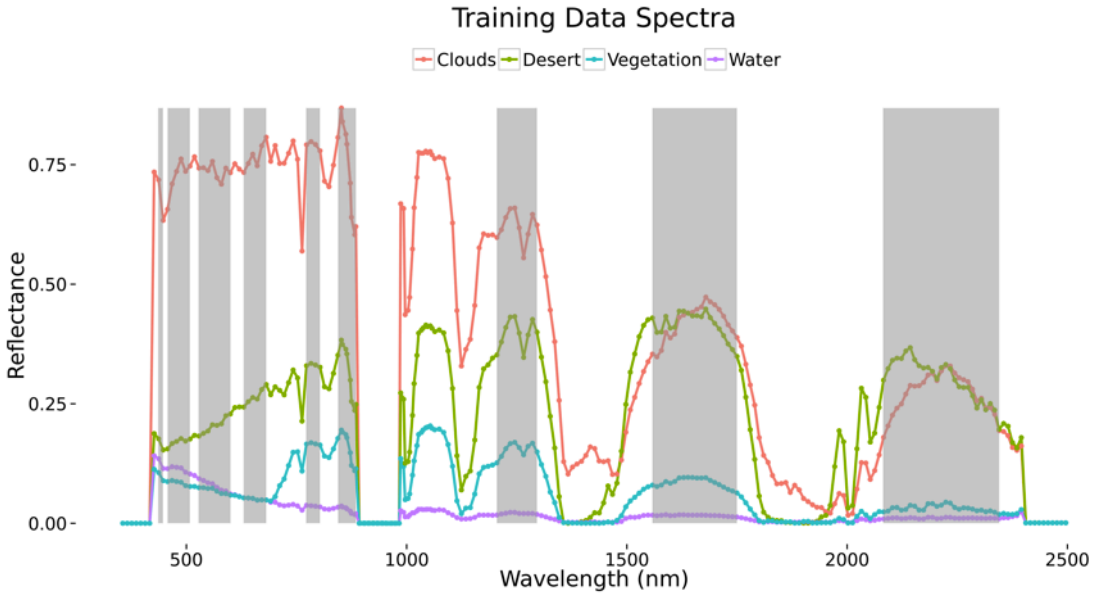
Hadoop-based analytics

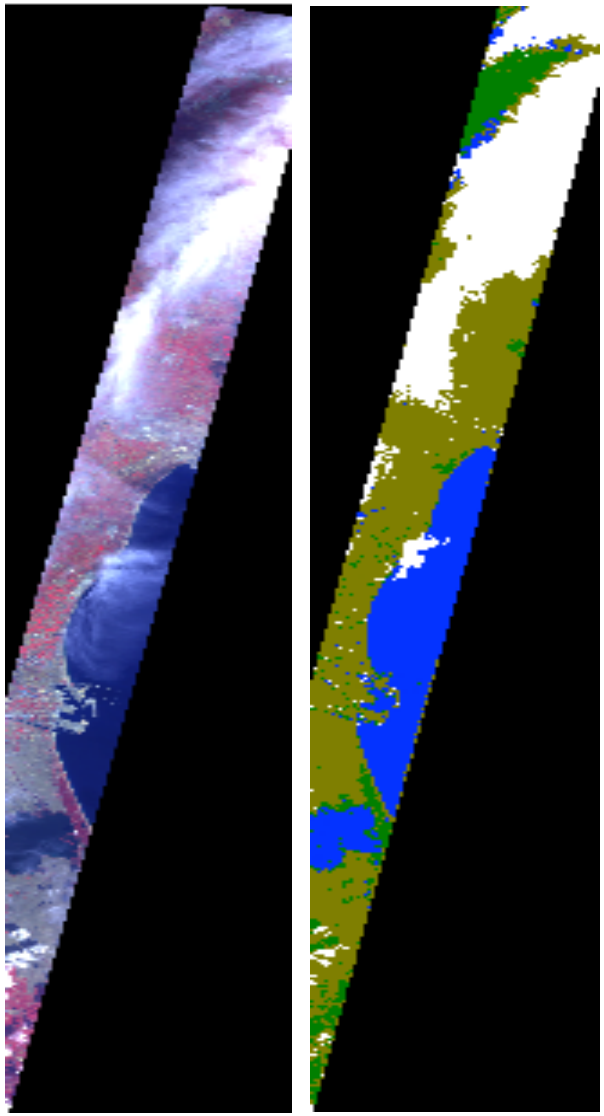
```
map(key, value):
// key: document name; value: text of the document
   for each word w in value:
      emit(w, 1)
```

```
reduce(key, values):
// key: a word; value: an iterator over counts
      result = 0
      for each count v in values:
            result += v
      emit(key, result)
```

SVM-based land cover classifier

Support vector machine model to predict new data based on training set data



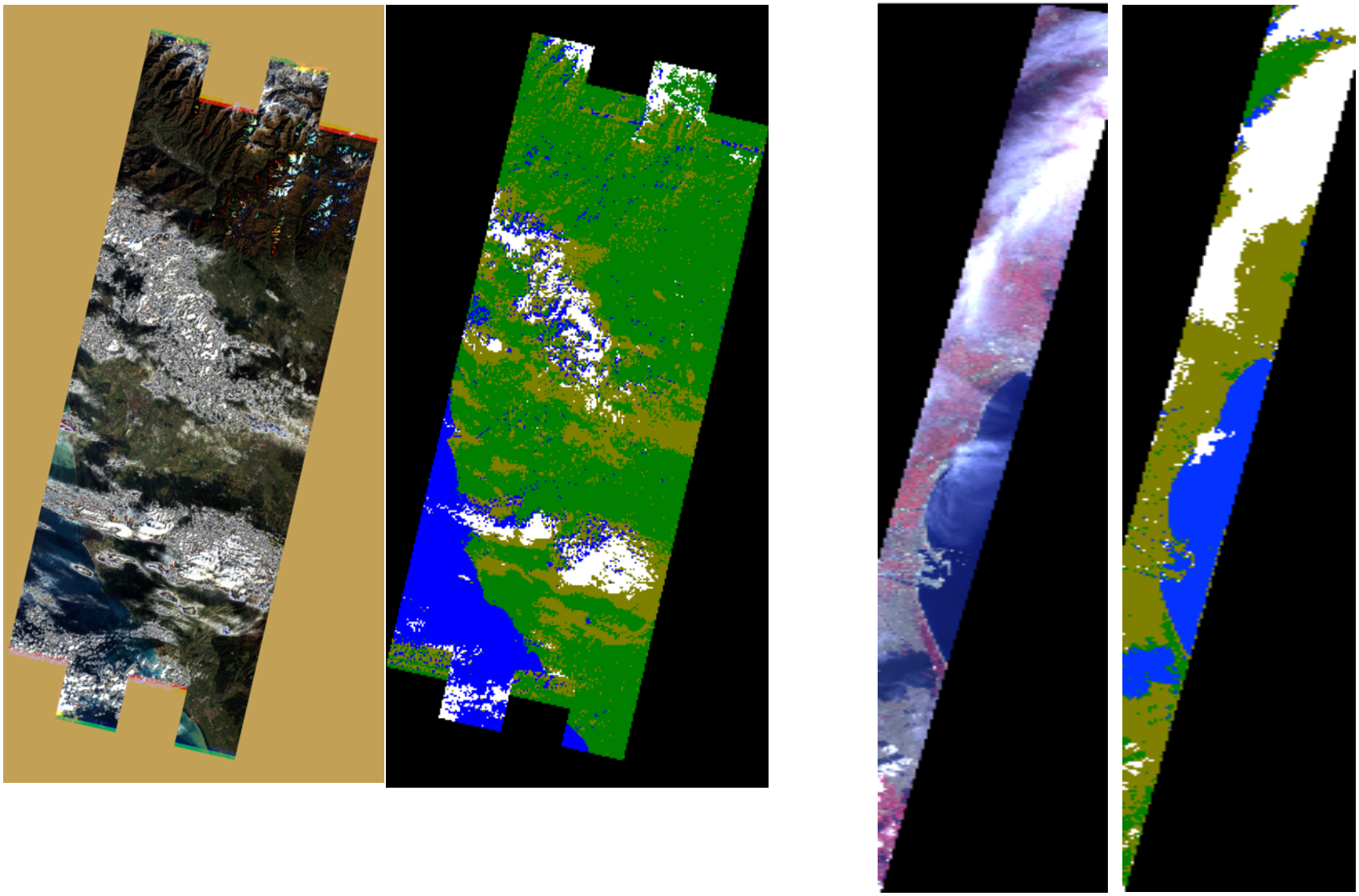


- Classifier Mapper

```
// Key: Scene; Value: pixels in scene
for each pixel p in value:
    landclass = svm.predict(p)
    emit(p, landclass)
```

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Wheel analytic:
SVM-based land cover classifier



Maria Patterson (mtpatter@uchicago.edu)

Center for Data Intensive Science, University of Chicago

Matsu "Wheel" Spectral Anomaly Detector

- "Contours and Clusters" – looks for physical contours around spectral clusters
  - PCA analysis applied to the set of reflectivity values (spectra) for every pixel, and the top 5 components are extracted for further analysis.
  - Pixels are clustered in the transformed 5-D spectral space using a k-means clustering algorithm.
  - For each image, k = 50 spectral clusters are formed and ranked from most to least extreme using the Mahalanobis distance of the cluster from the spectral center.
  - For each spectral cluster, adjacent pixels are grouped together into contiguous objects.

→ returns geographic regions of spectral anomalies that are scored again as anomalous (0 least , 1000 most) compared to a set of "normal" spectra, constructed for comparison over a baseline of time

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Turning earth science observations into actionable information

Processing and serving data

- Generating (EO-1) satellite L1 and L2 data
- Web Coverage Processing Service for individual scenes
- Hadoop-based Matsu "Wheel" of analytics for processing all daily scenes

Aggregating and displaying data products

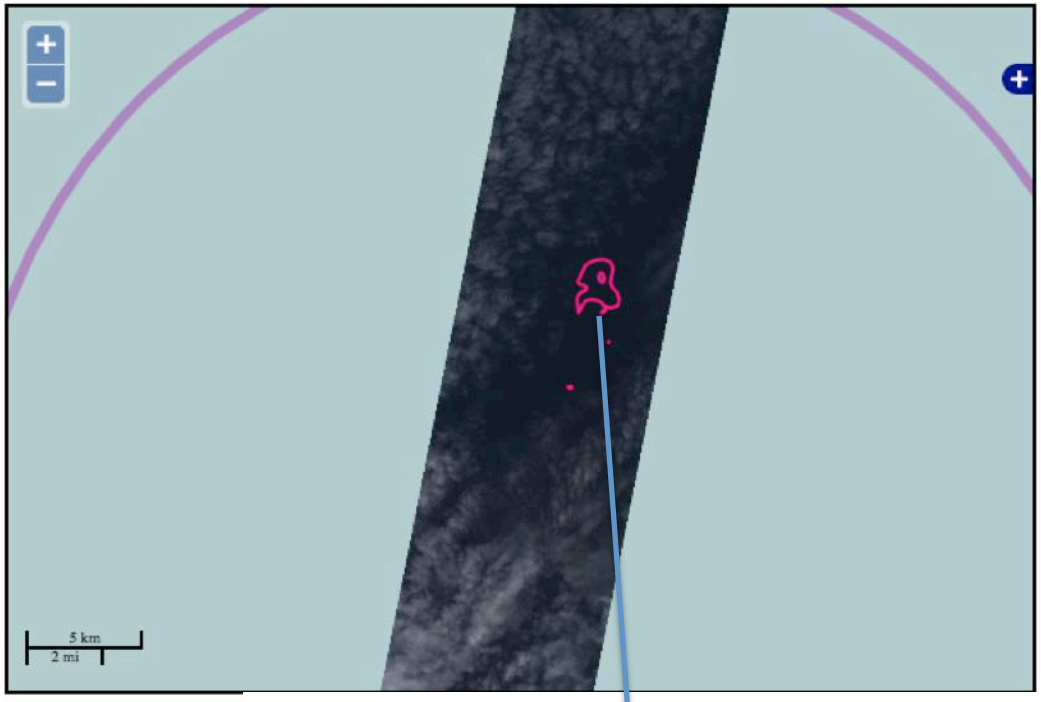
- Namibia flood dashboard
- Matsu "Wheel" analytic reports
- Web Map Service to publish data to an Open Geosocial API

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

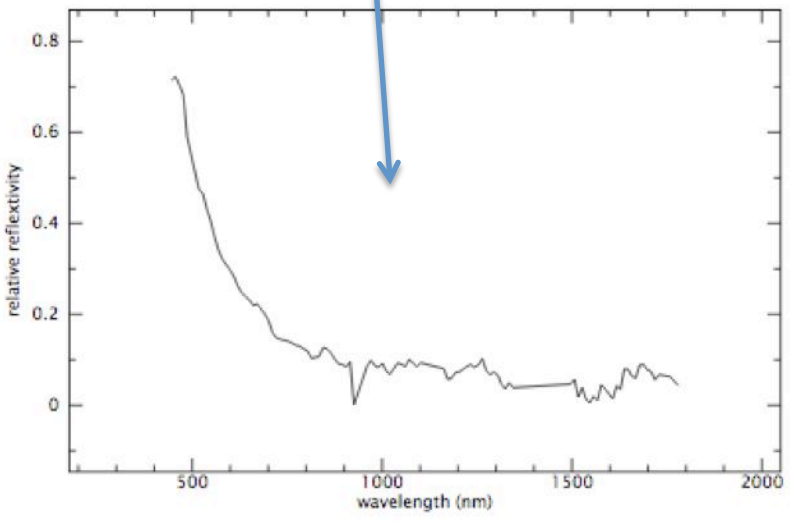
Spectral anomaly detected: Nishinoshima active volcano, Dec, 2014

Matsu Analytic Image Report

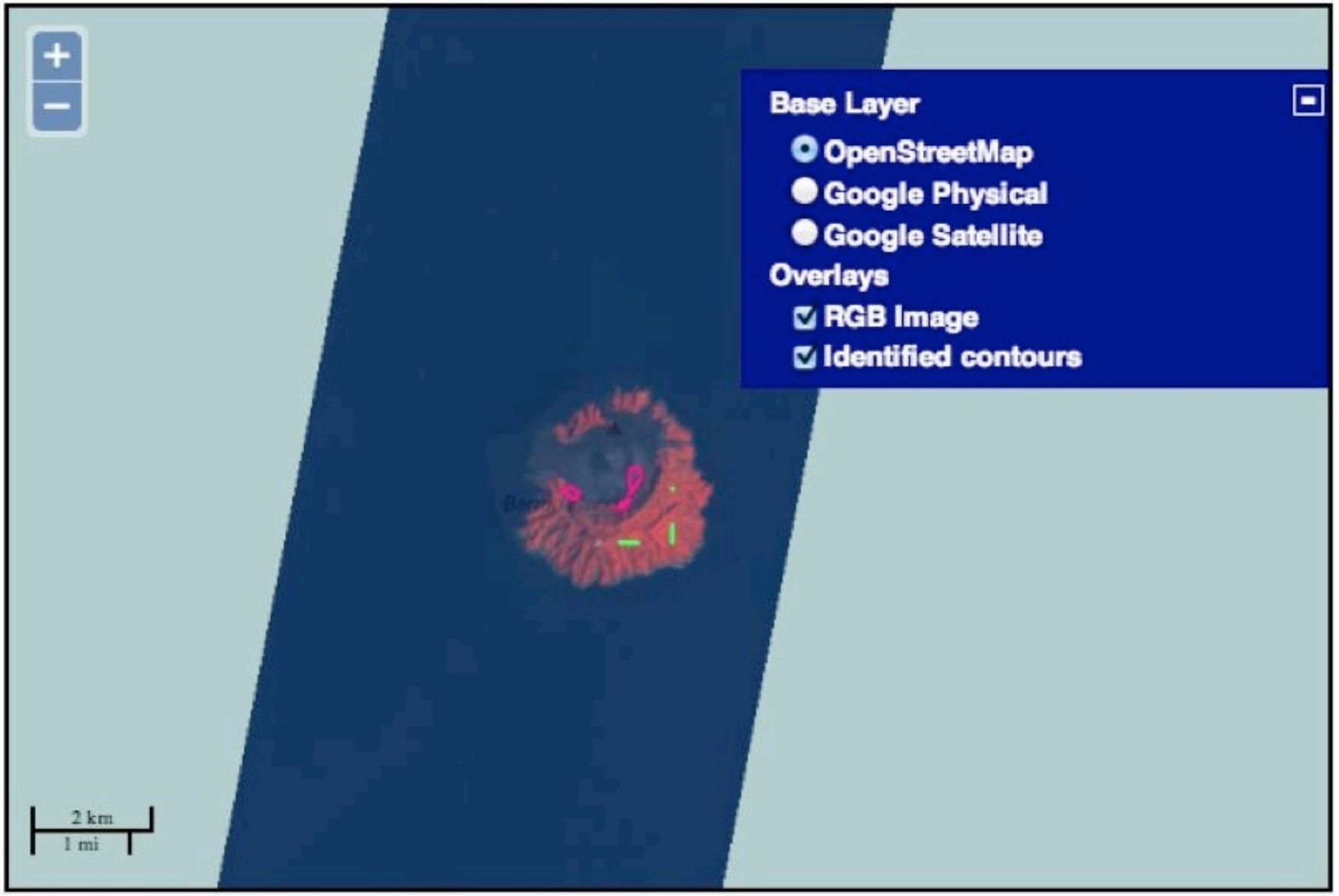
Collection Date	2014-12-02 (day 336)
Analysis Date	Wed Dec 17 12:27:25 2014
Analytic Environment	
Analytic	Contours-2013-12-r4
Noise Correction Enabled	False
Summary Stats	ss-2013-12-r1
Data Ingest	populateHDFS-2013-11-r1
Report Format	reportContoursR4
Hyperspectral Image	
Image	EO1H1050412014336110KF_HYP_L1G
Number of Bands	242



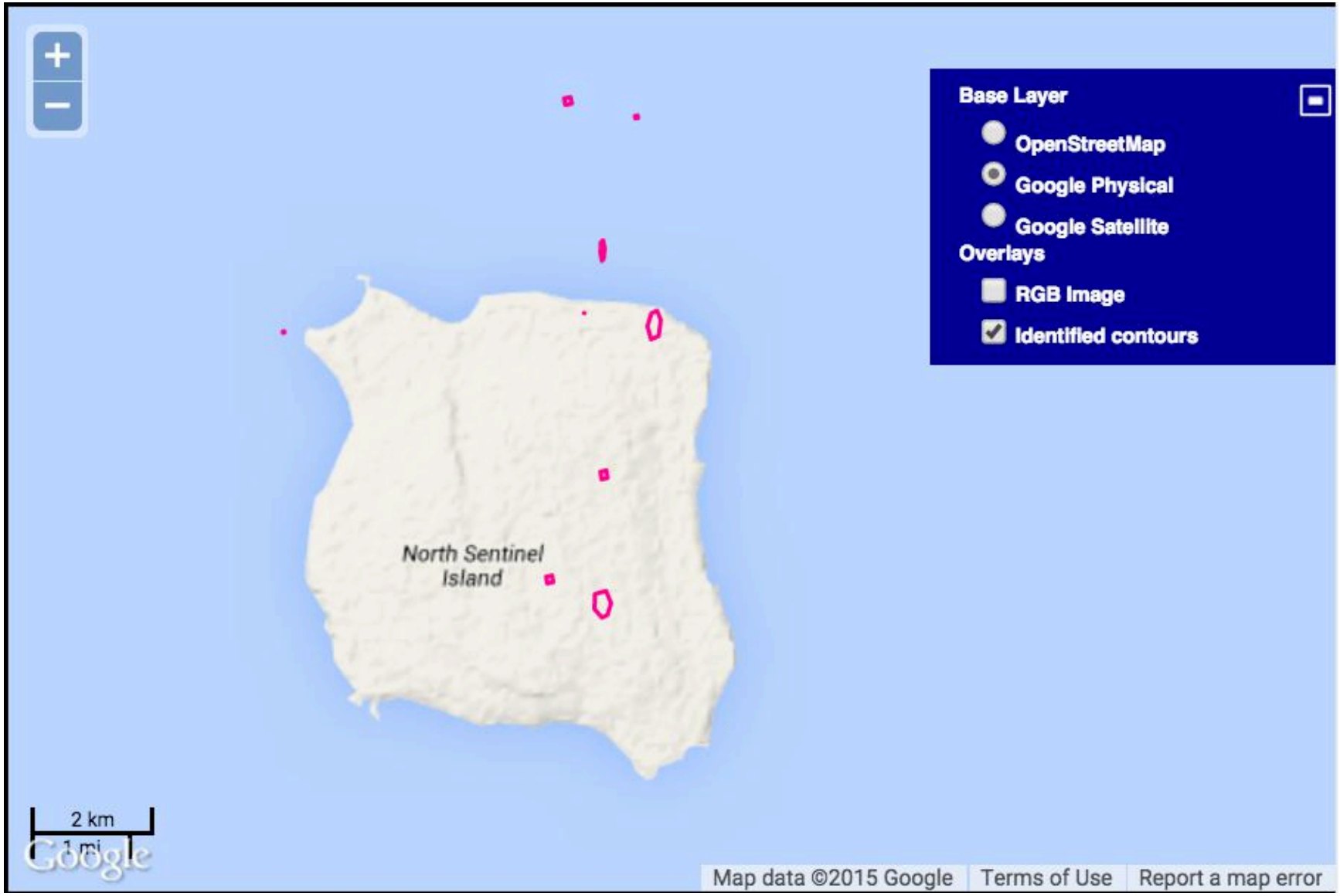
Contour ID	Cluster Score	Contour Score	lat,lng	Area (Pixels)	Area (Meters)	color
C1-05041-0KF	351	0.9719	140.886733625,27.2918559268	7.9589	6259.0137	COLOR
C1-05041-0KF	351	1.0807	140.897972808,27.3285963336	2447.4154	1925311.5337	COLOR
C1-05041-0KF	351	1.1266	140.899385769,27.3310296144	66.3332	52183.5335	COLOR
C1-05041-0KF	351	1.4893	140.900233529,27.3190516554	8.5744	6744.6581	COLOR
C1-05041-0KF	351	0.9264	140.902293378,27.3081518463	0.6165	484.8863	COLOR



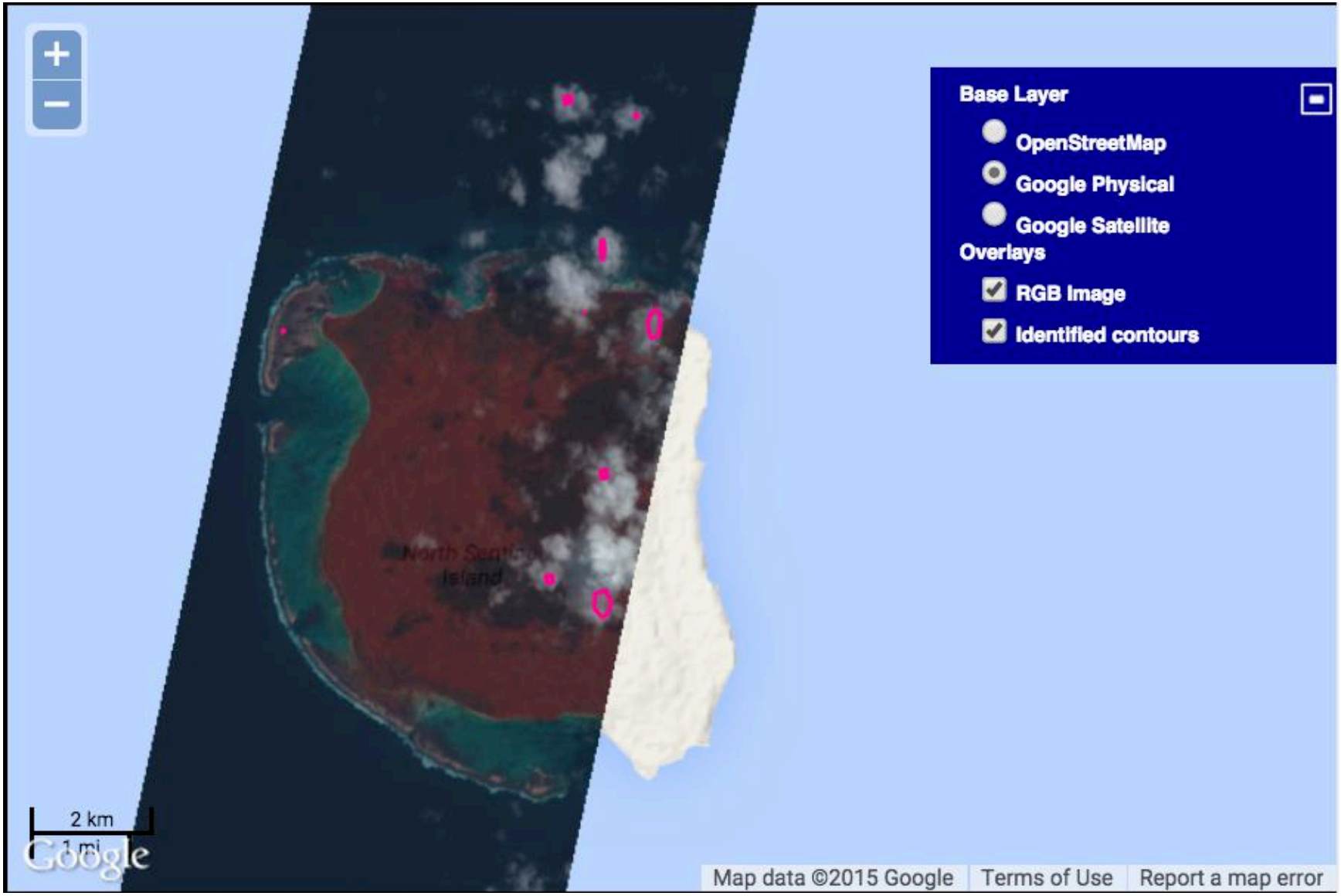
Spectral anomaly detected:
Barren Island active volcano, Feb, 2014



Spectral anomaly detected:
North Sentinel Island fires, May, 2014



Spectral anomaly detected:
North Sentinel Island fires, May, 2014



Turning earth science observations into actionable information

Processing and serving data

- Generating (EO-1) satellite L1 and L2 data
- Web Coverage Processing Service for individual scenes
- Hadoop-based Matsu "Wheel" of analytics for processing all daily scenes

Aggregating and displaying data products

- Namibia flood dashboard
- Matsu "Wheel" analytic reports
- Web Map Service to publish data to an Open Geosocial API

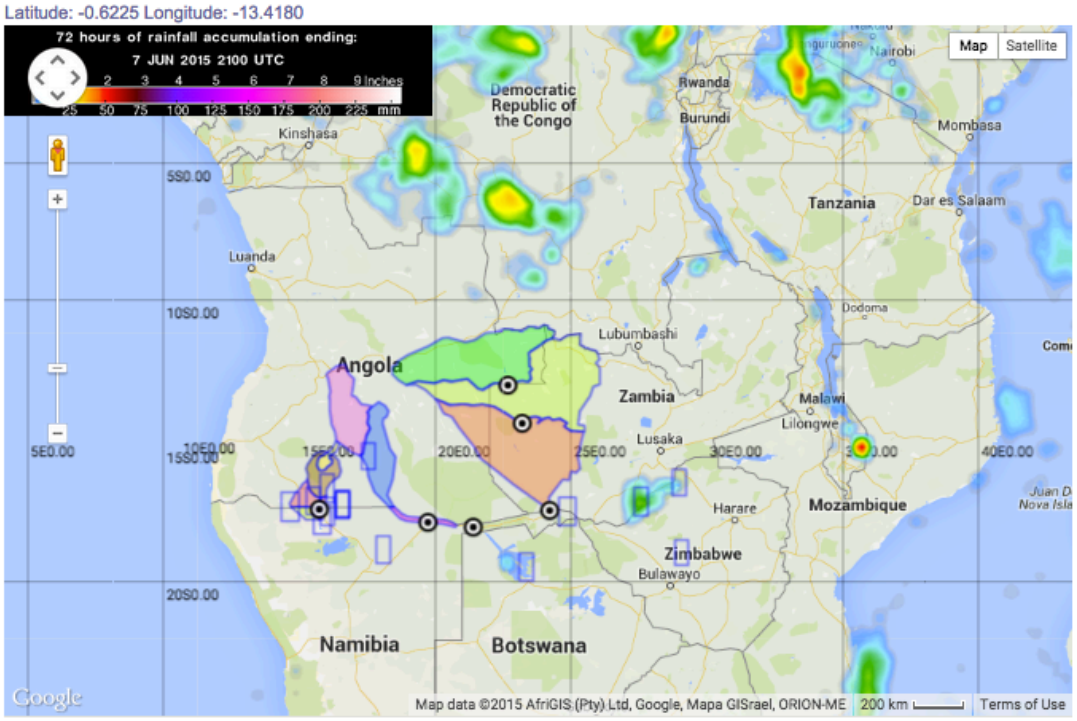
Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Namibia Flood Dashboard

Flood and waterborne disease risk management in Southern Africa

matsu-namibiaflood.opensciencedatacloud.org

- SensorWeb Layers
- Water Lines and Areas
- Satellite Overlays
- Ground Pics
- Kavango Radarsat Data
- Cuvelai Radarsat Data
- TRMM Rainfall Accumulation and Flood Forecast
  - 1 Day Forecast
  - 24 Hour GFS Forecast Accumulation
  - 48 Hour GFS Forecast Accumulation
  - 3 Hour Accumulation
  - 24 Hour Accumulation
  - 72 Hour Accumulation
- Global Scene Counts



Legend:

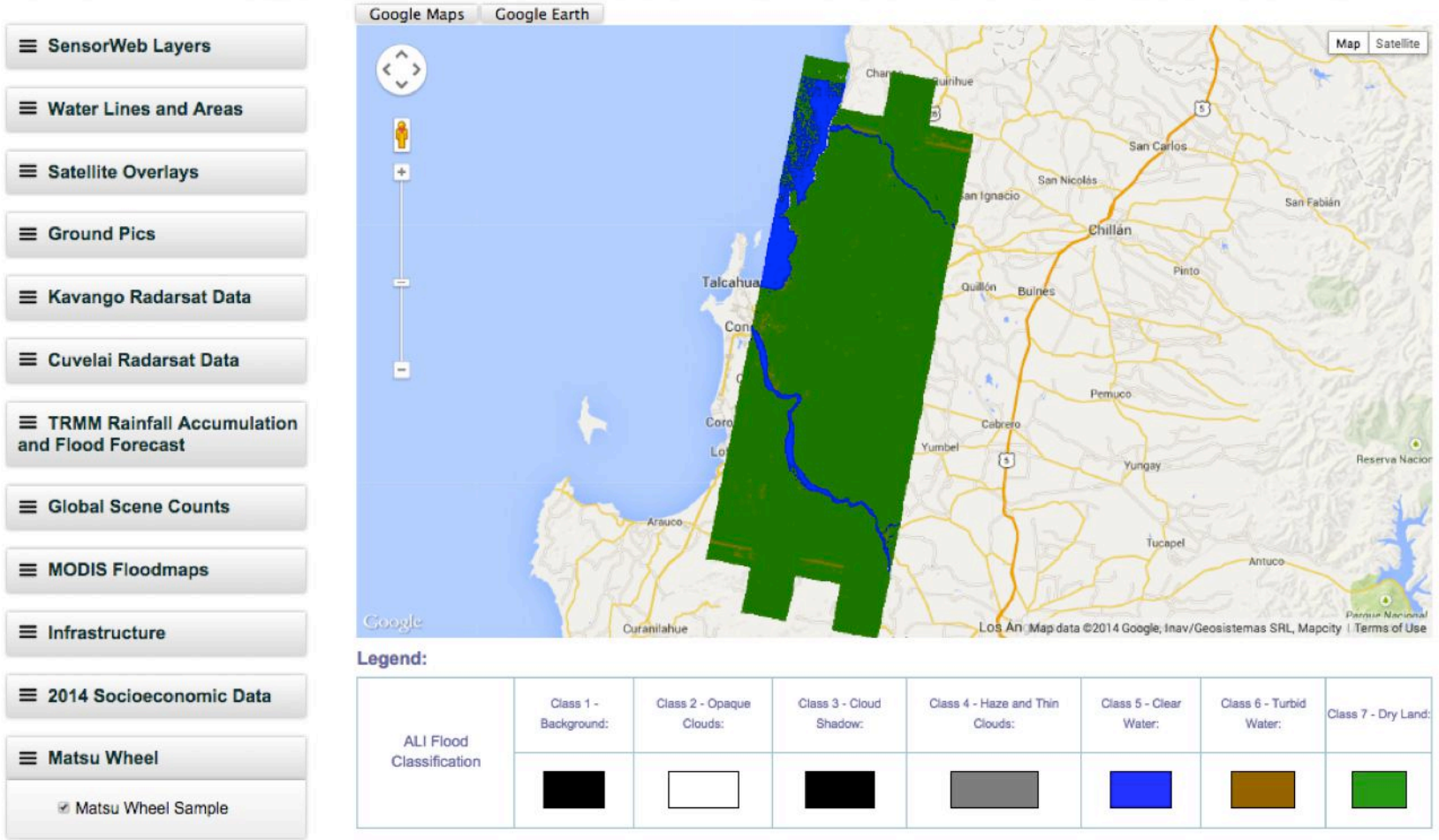
ALI Flood Classification	Class 1 - Background:	Class 2 - Opaque Clouds:	Class 3 - Cloud Shadow:	Class 4 - Haze and Thin Clouds:	Class 5 - Clear Water:	Class 6 - Turbid Water:	Class 7 - Dry Land:

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Namibia Flood Dashboard

Flood and waterborne disease risk management in Southern Africa - landcover classifier wheel analytic

matsu-namibiaflood.opensciencedatacloud.org



Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*

Sample code on OSDC

https://www.opensciencedatacloud.org/support/quickstart.html#osdc-eo-1-quick-start-tutorial

OSDC EO-1 Quick Start Tutorial

Now we'll take you step by step through a demo using NASA's Earth Observing-1 dataset. In this tutorial, we will show you how to use OSDC to visualize and perform a simple example analysis of NASA satellite imagery data. You will perform many tasks common to using the OSDC during this demo like launching an instance, ssh'ing, in addition to those specific to analysis.

Here we will show you how to use Python to

- create png false-color images from GeoTiff data,
- use a machine algorithm to classify each pixel of a scene as desert, water, cloud, or vegetation,
- view GeoTiffs and save the results of your classification as an image.

Launch the OSDC EO-1 Instance

In the console, under 'Images and Snapshots', scroll down to find the section labeled 'All Snapshots'. Here's you'll want to find and launch the snapshot called 'OSDC_DatasetExplorer_EO1'. We recommend using a medium instance.

When you ssh in to both the login node and the instance, make sure and add both the "A" and the "X" flags. The A is for key forwarding, the X is for X11 forwarding. IE: `ssh -AX <username>@sullivan.opensciencedatacloud.org` and then `ssh -AX ubuntu@<INSTANCE.IP>`. If you're doing a lot of GUI work like looking at plots and images, you'll want to use this X flag often.

Once you're in the instance, cd and run all commands from

About the Data

NASA's Earth Observing-1 satellite (EO-1) was launched in 2000 for the purpose of studying new technologies in remote earth imaging. On the OSDC, we host data from EO-1's two primary scientific instruments, the Hyperion imaging spectrometer and the Advanced Land Image (ALI). In this tutorial we will be working with ALI data.

The ALI instrument acquires data in 9 different wavelength bands from 0.48 – 2.35 micron with 30-meter resolution plus a panchromatic band with higher 10-meter spatial resolution. The standard 'scene' (image) size projected on the Earth's surface equates to 37 km x 42 km (width x length). Hyperion has similar spatial resolution but higher spectral resolution, observing in 242 band channels from 0.357 – 2.576 micron with 10-nm bandwidth. Hyperion scenes have a smaller standard footprint width of 7.7 km.