



Project Matsu:

Developing open source methods for cloud-based processing and analysis of earth science data and data products

Maria Patterson, PhD

Open Science Data Cloud

Center for Data Intensive Science (CDIS)

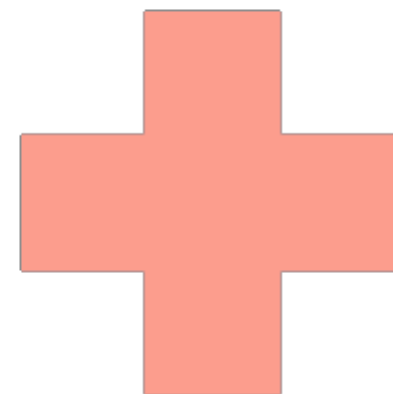
University of Chicago

OSDC PIRE Workshop, 17 June, 2014





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.

Turning earth science observations into actionable information



Processing and serving data

- Generating (EO-1) satellite L1 and L2 data
- Web Coverage Processing Service
- Hadoop-based Matsu Wheel

Aggregating and displaying data products

- Namibia flood dashboard
- Matsu Wheel analytic reports
- Preliminary Geoserver Web Map Service to work with Open Geosocial API

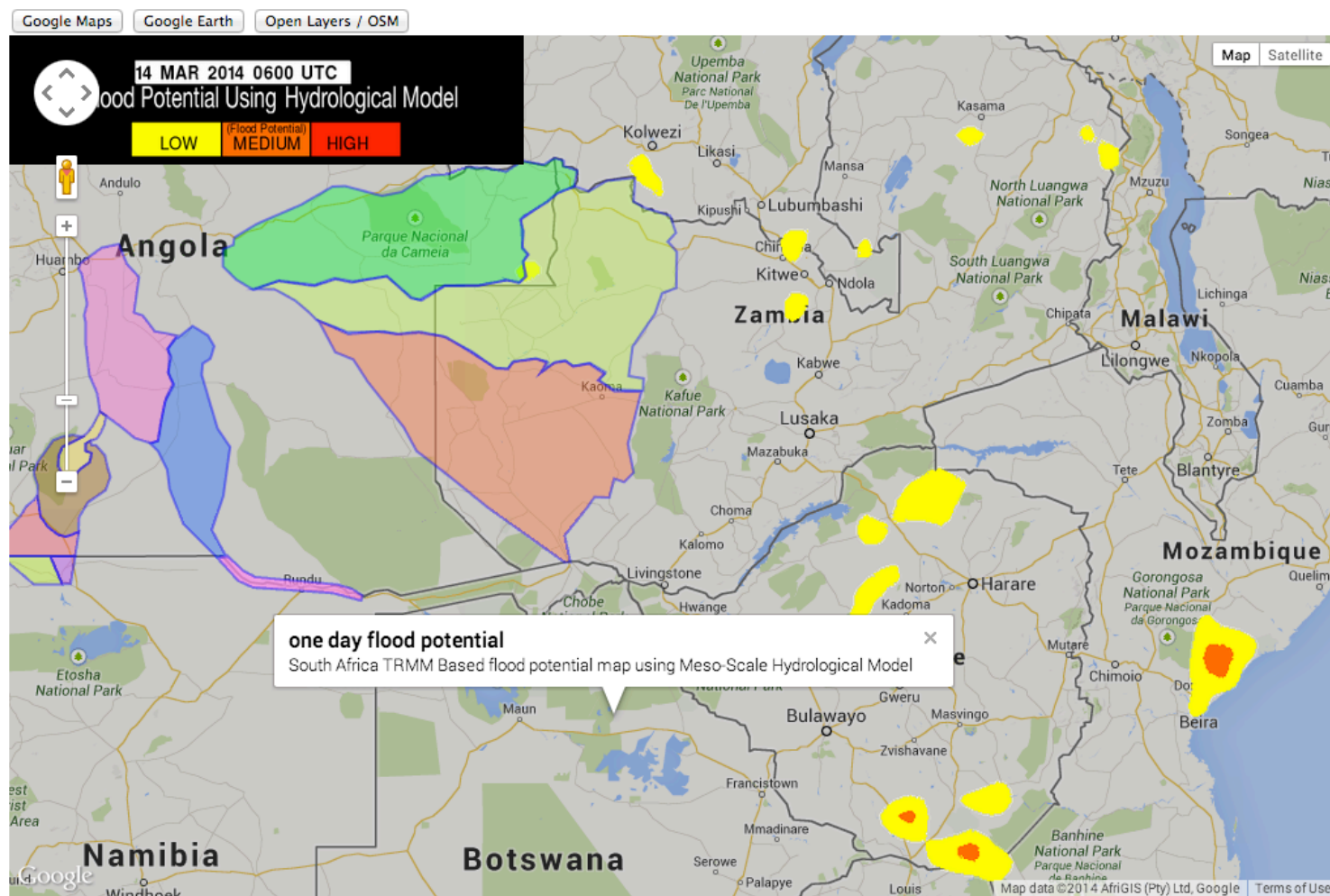


Namibia Flood Dashboard

SensorWeb enabled for early flood warning

<http://matsu.opencloudconsortium.org/namibiaflood>
<http://sensorweb.nasa.gov/>

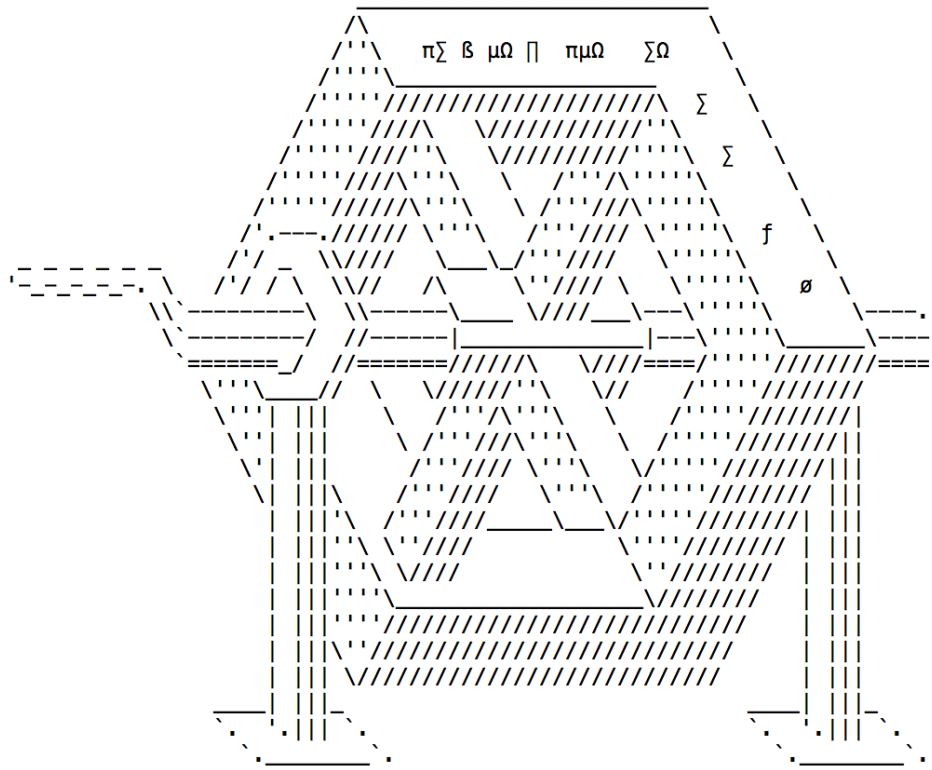
- ▼ River Stations
- ▼ SensorWeb Layers
 - Catchments
 - River Gauges
 - GDACS
 - Current/Past Floods
- ▼ Water Lines and Areas
- ▼ Satellite Overlays
- ▼ Ground Pics
- ▼ Kavango Radarsat Data
- ▼ Cuvelai Radarsat Data
- ▼ TRMM Rainfall Accumulation and Flood Forecast
 - 1 Day Forecast
 - 24 Hour GFS Forecast Accumulation
 - 48 Hour GFS Forecast Accumulation
 - 3 Hour Accumulation
 - 24 Hour Accumulation
 - 72 Hour Accumulation
- ▼ Global Scene Counts



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

The Matsu Wheel

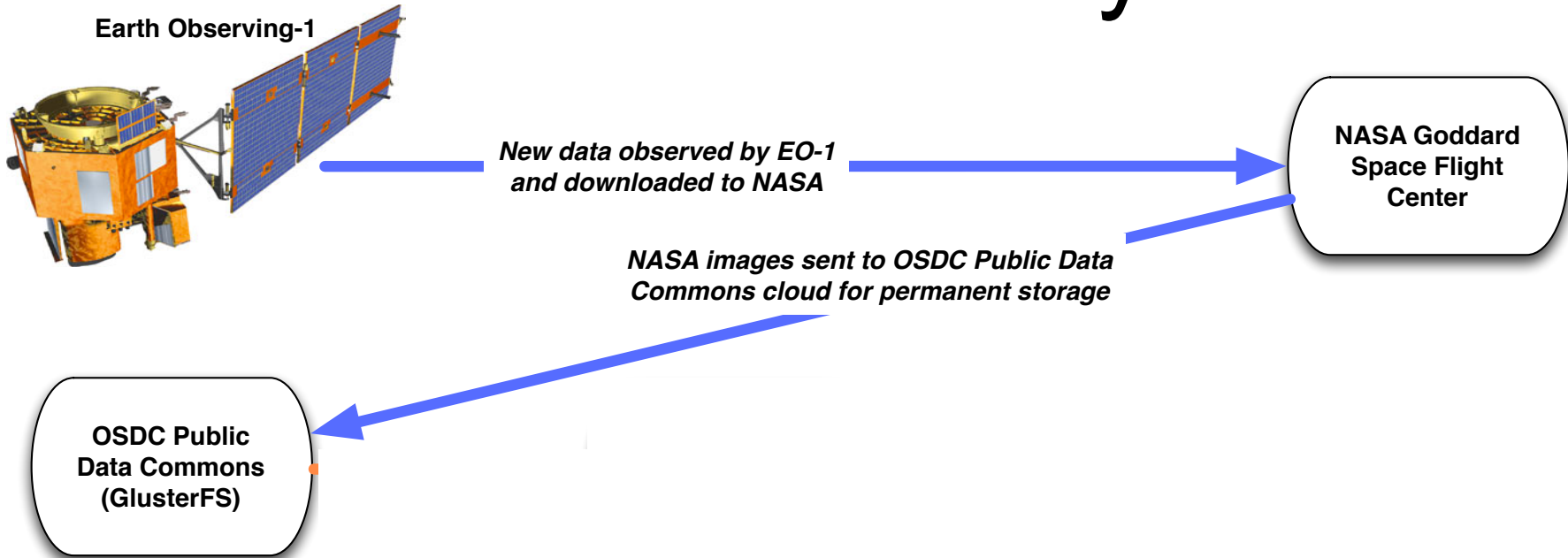


0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, ...

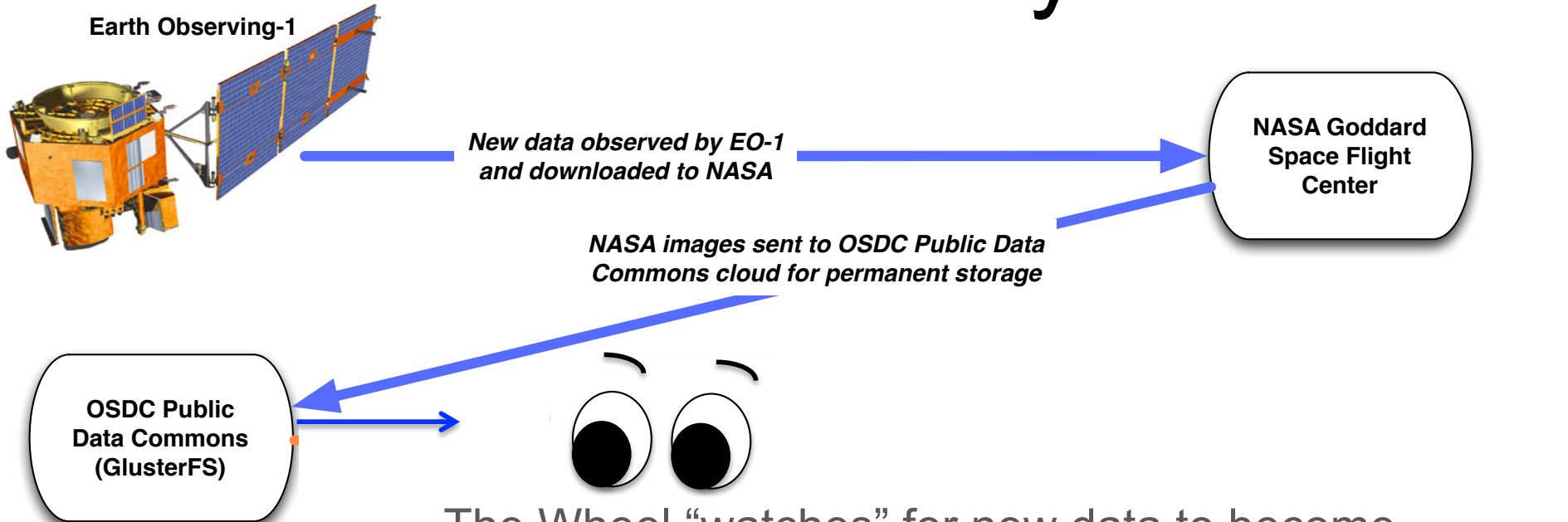
processing all the data
as it comes in

- Hadoop-based over Skidmore over 25 nodes with 800 compute cores and 784 GB of compute RAM, 261 TB raw storage
- EO-1 Level 0 images are received daily from NASA, transformed into various Level 1 products, converted (SequenceFiles), uploaded (HDFS), and MapReduced (analytic) once a day to build the Wheel analytic reports.

Matsu Analytic Wheel

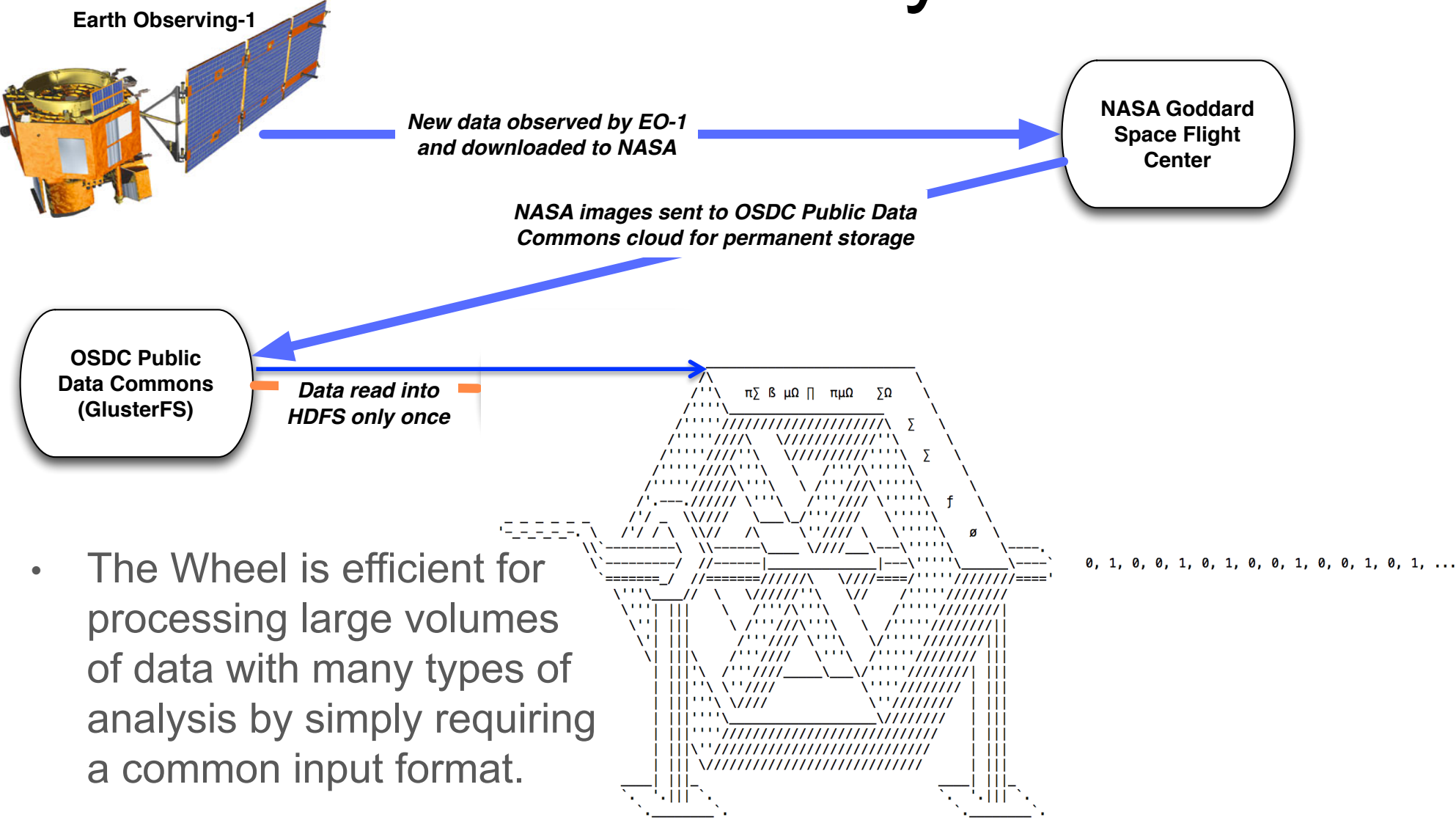


Matsu Analytic Wheel



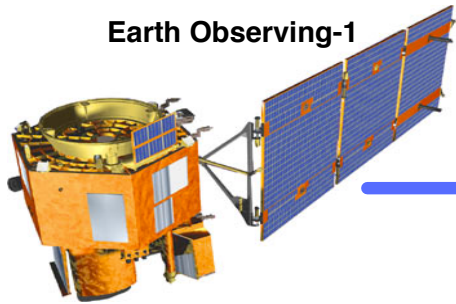
- The Wheel “watches” for new data to become available, using Apache Storm.
- When new data are detected, loaded into Hadoop’s distributed file system for analysis using MapReduce.
- The Wheel analytics run each night, daily reports available the morning after data are received.

Matsu Analytic Wheel



- The Wheel is efficient for processing large volumes of data with many types of analysis by simply requiring a common input format.

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

Contour ID	Cluster Score	Contour Score	lat,lng	Area (Pixels)	Area (Meters)	color	Spectral Signature
C2-97081-0P1	0	0.1898	139.706143433,-30.9385603559	13.8046	11062.8979	green	wavelength
C2-97081-0P1	0	0.0763	139.79377387,-30.8504859316	54.0231	43236.9527	green	wavelength
C2-97081-0P1	0	0.0917	139.799505582,-30.8611007022	281.1866	225174.5817	green	wavelength
C2-97081-0P1	0	0.2390	139.711392036,-31.1875949205	27.1649	21026.4541	green	wavelength
C2-97081-0P1	0	0.1585	139.848832889,-30.4135174894	221.0056	176181.6474	purple	wavelength

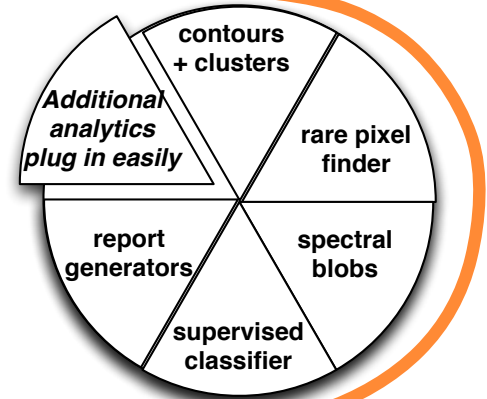
Analytic reports generated by Wheel are accessible via web browser

Metadata stored

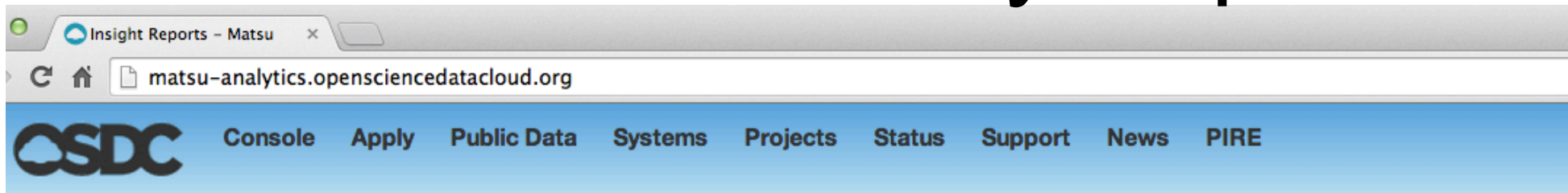
NoSql Database (Accumulo)

Analytic results stored

Secondary analysis can be done from analytic database



Matsu Wheel Daily Reports



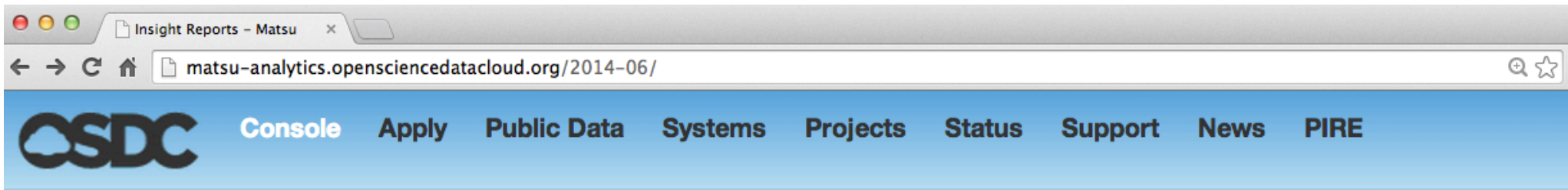
EO-1 Level 0 images are received daily from NASA and are transformed into various Level 1 products. This transformation does not run each day. Level 1G images are converted (SequenceFiles), uploaded (HDFS), and MapReduced (analytic) once a day to build the Insight Reports. If the Level 1G is not available at that time, it is processed the next day, but not all days have images. Daily processing began in July 2013. Previous to that, a few select days from 2010 and 2012 were processed.

The analytics are run, and the reports are generated, on Skidmore, an Open Science Data Cloud resource intended for computational projects. The main service provide by Skidmore is Hadoop over 25 nodes with 800 compute cores and 784 GB of compute RAM. The raw storage available is 261 TB.

Name	Last modified	Size
2014-06/	04-Jun-2014 10:18	-
2014-05/	01-Jun-2014 10:23	-
2014-04/	20-Apr-2014 12:23	-
2014-03/	20-Apr-2014 11:47	-
2014-02/	20-Apr-2014 11:06	-
2014-01/	16-Apr-2014 16:13	-
2013-12/	25-Mar-2014 17:18	-
2013-11/	25-Mar-2014 17:18	-
2013-10/	25-Mar-2014 17:18	-
2013-09/	25-Mar-2014 17:18	-
2013-08/	25-Mar-2014 17:18	-
2013-07/	25-Mar-2014 17:18	-
2012/	25-Mar-2014 17:18	-
2010/	25-Mar-2014 17:18	-

matsu-analytics.opensciencedatacloud.org

Matsu Wheel Daily Reports



EO-1 Level 0 images are received daily from NASA and are transformed into various Level 1 products. This transformation does not run each day. Level 1G images are converted (SequenceFiles), uploaded (HDFS), and MapReduced (analytic) once a day to build the Insight Reports. If the Level 1G is not available at that time, it is processed the next day, but not all days have images. Daily processing began in July 2013. Previous to that, a few select days from 2010 and 2012 were processed.

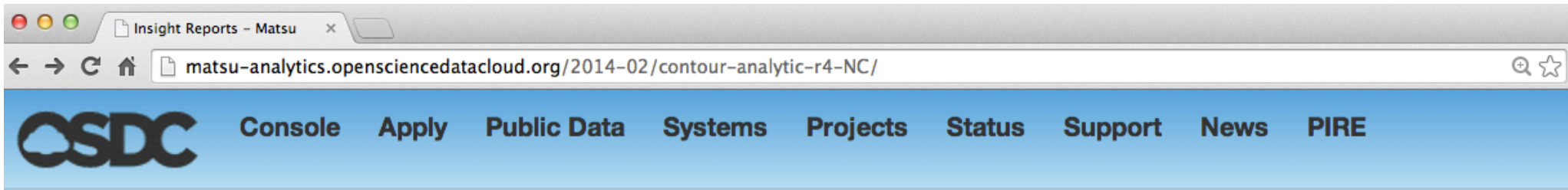
The analytics are run, and the reports are generated, on Skidmore, an Open Science Data Cloud resource intended for computational projects. The main service provide by Skidmore is Hadoop over 25 nodes with 800 compute cores and 784 GB of compute RAM. The raw storage available is 261 TB.

Name	Last modified	Size
contour-analytic-r4/	04-Jun-2014 10:55	-
contour-analytic-r4-NC/	04-Jun-2014 10:38	-

 [Insight Reports Home Page](#)

matsu-analytics.opensciencedatacloud.org

Matsu Wheel Daily Reports



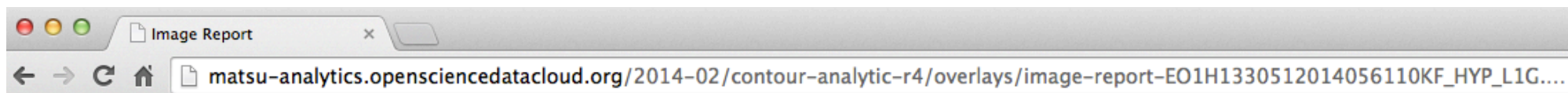
EO-1 Level 0 images are received daily from NASA and are transformed into various Level 1 products. This transformation does not run each day. Level 1G images are converted (SequenceFiles), uploaded (HDFS), and MapReduced (analytic) once a day to build the Insight Reports. If the Level 1G is not available at that time, it is processed the next day, but not all days have images. Daily processing began in July 2013. Previous to that, a few select days from 2010 and 2012 were processed.

The analytics are run, and the reports are generated, on Skidmore, an Open Science Data Cloud resource intended for computational projects. The main service provide by Skidmore is Hadoop over 25 nodes with 800 compute cores and 784 GB of compute RAM. The raw storage available is 261 TB.

Name	Last modified	Size
overlays/	17-Apr-2014 12:34	-
summary-report-2014-059.html	17-Apr-2014 03:12	7.3K
summary-report-2014-058.html	17-Apr-2014 02:43	7.5K
summary-report-2014-057.html	17-Apr-2014 02:14	7.7K
summary-report-2014-056.html	17-Apr-2014 01:45	8.1K
summary-report-2014-054.html	17-Apr-2014 01:10	7.7K

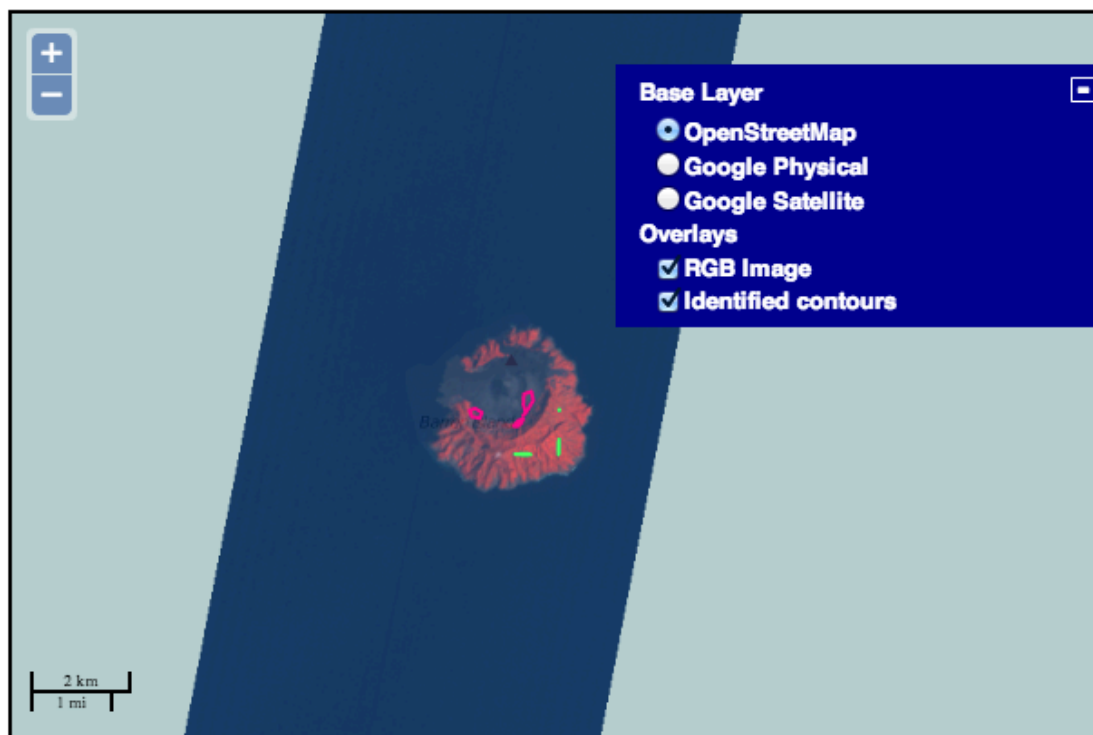
matsu-analytics.opensciencedatacloud.org

Matsu Wheel Daily Reports



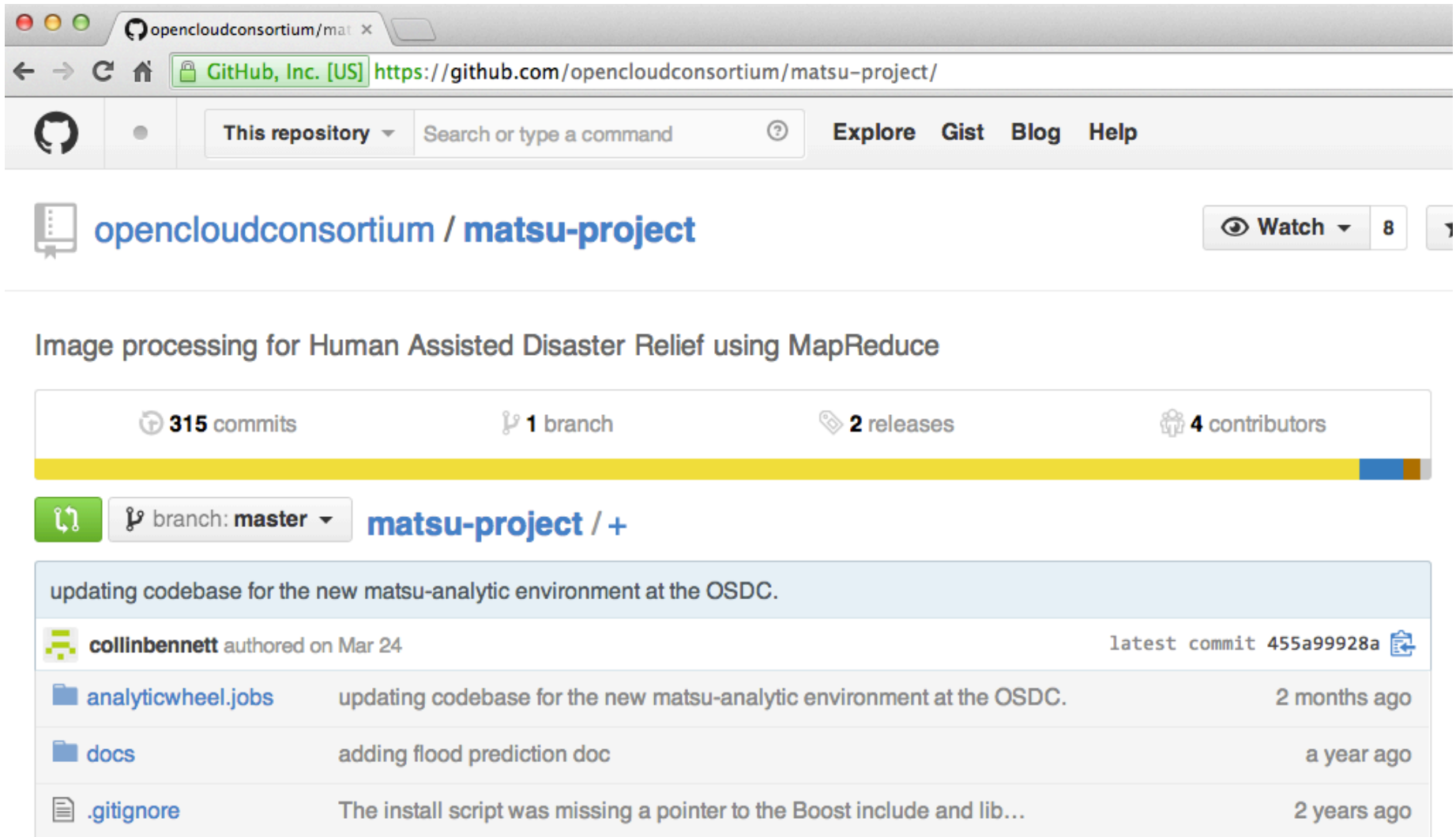
Matsu Analytic Image Report

Collection Date	2014-02-25 (day 056)
Analysis Date	Thu Apr 17 01:59:44 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	EO1H1330512014056110KF_HYP_L1G
Number of Bands	242



Contour ID	Cluster Score	Contour Score	lat, lng	Area (Pixels)	Area (Meters)	color	Spectral Signature
C2-33051-	414	1.3330	93.8671531666, 12.2691714755	6.1221	387.8394	COLOR	wavelengths

Matsu Wheel is open source



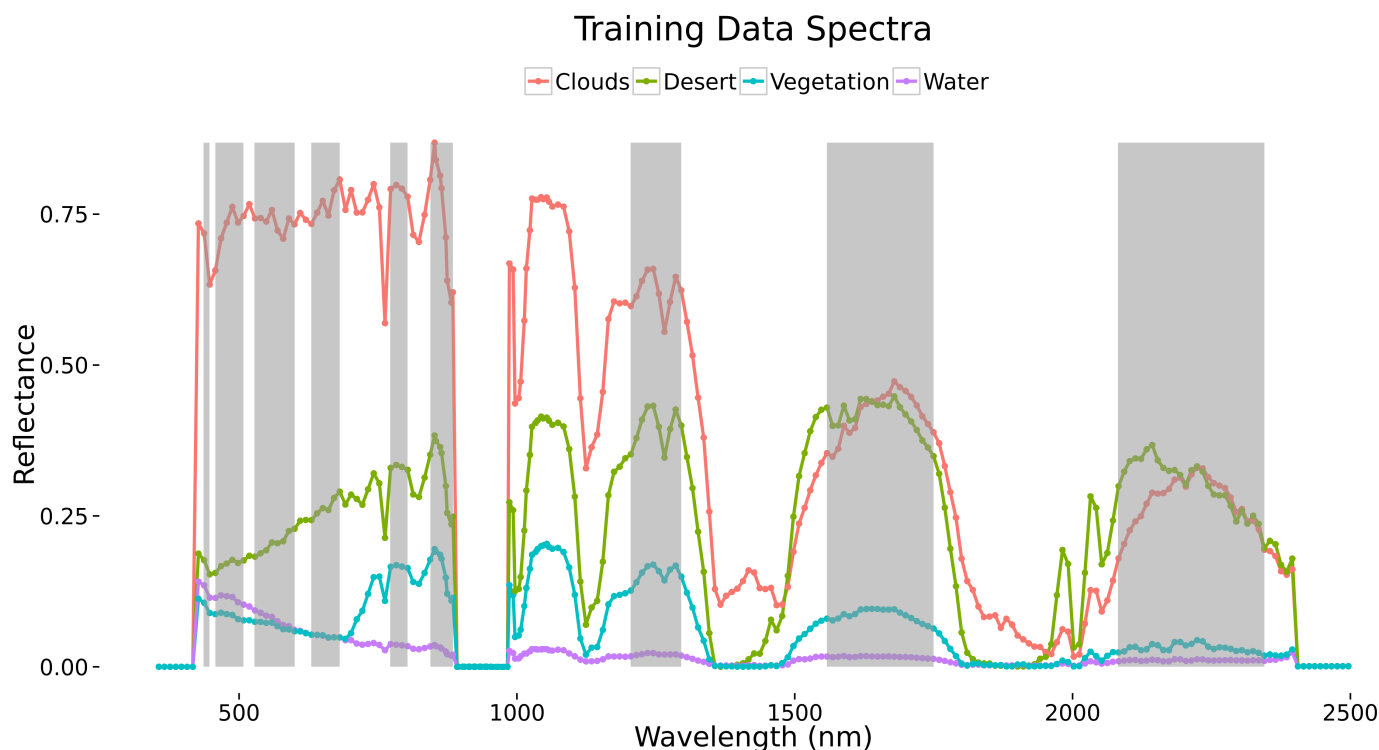
The screenshot shows the GitHub repository page for `opencloudconsortium/matsu-project`. The repository is titled "Image processing for Human Assisted Disaster Relief using MapReduce". It has 315 commits, 1 branch, 2 releases, and 4 contributors. The latest commit is by `collinbennett` on Mar 24, with the message "updating codebase for the new matsu-analytic environment at the OSDC." The commit hash is `455a99928a`. The repository contains the following files and folders:

File/Folder	Description	Last Commit
<code>analyticwheel.jobs</code>	updating codebase for the new matsu-analytic environment at the OSDC.	2 months ago
<code>docs</code>	adding flood prediction doc	a year ago
<code>.gitignore</code>	The install script was missing a pointer to the Boost include and lib...	2 years ago

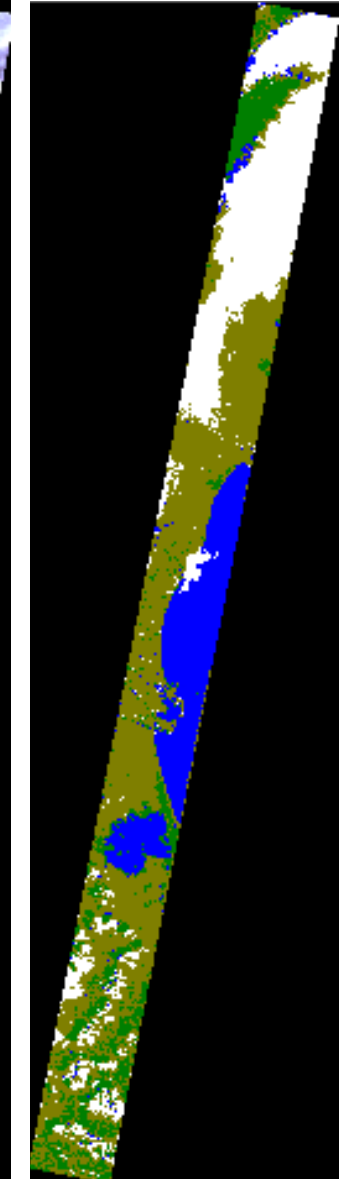
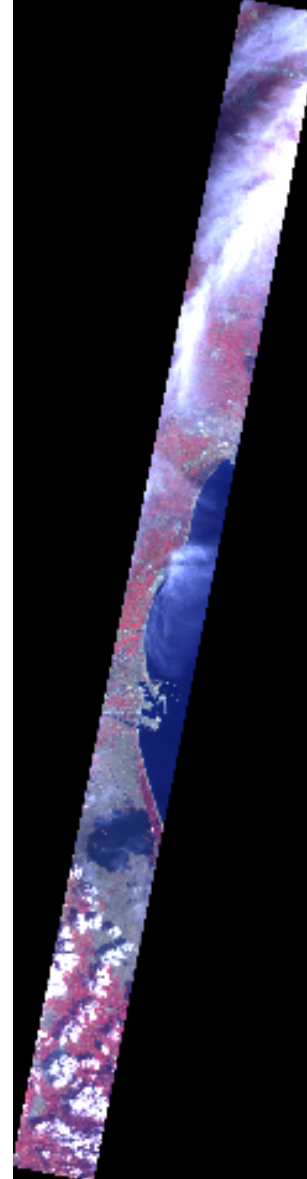
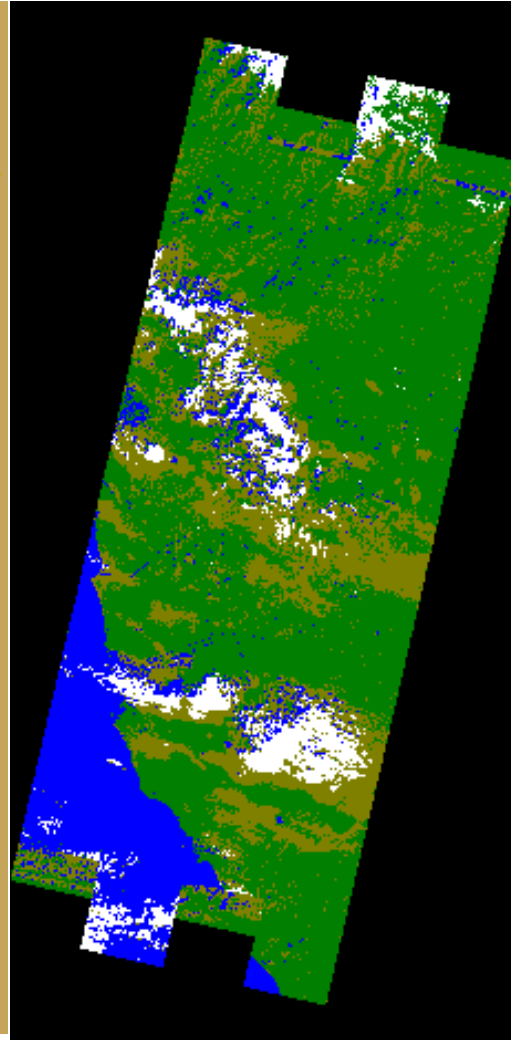
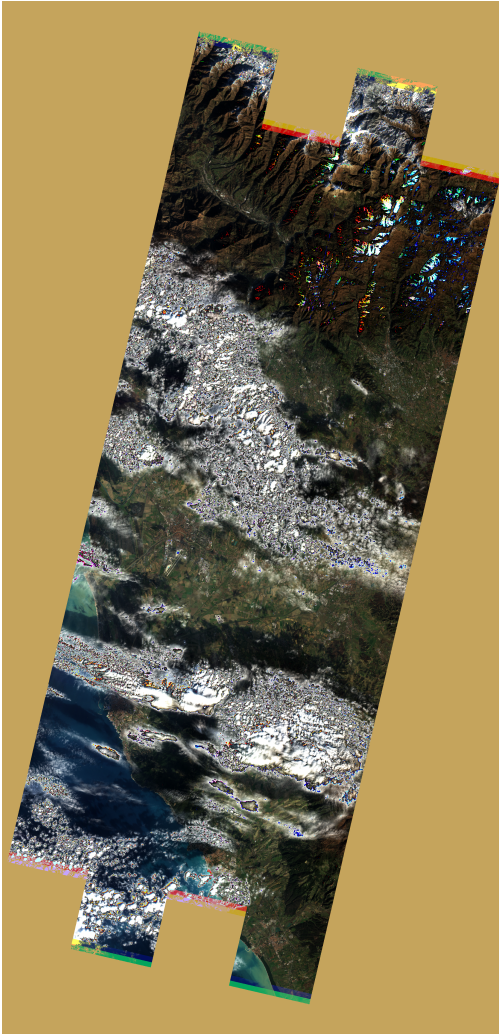
github.com/opencloudconsortium/matsu-project

New wheel analytic (beta): Support Vector Machine (SVM) classifier

- A supervised machine learning classification algorithm
- Train the classifier by hand classifying areas in a set of training images
- Beta classifier has 4 classes: clouds, dry land, vegetation, water



New wheel analytic (beta): Support Vector Machine (SVM) classifier



Continuing work

- SVM classifier adapt regionally to geographic area (classes and training set depend on geography)
- Incorporate SVM classifier into Matsu Wheel
- Additional wheel analytics, data
- Web Map Service and tiling using Open Geospatial Consortium compliant Geoserver
- Interoperating with Open Geosocial API
 - Deliver products so end users can easily display from geoJson, topoJson formats in github, MapBox and redistribute maps over Facebook and other social media
 - See Pat Cappelaere's work
<http://www.slideshare.net/cappelaere/open-geosocial-api>