



Data Requirements Definition and Data Services Options for RAPP

Brian Killough
CEOS Systems Engineering Office
(SEO)

4th GEOGLAM RAPP Workshop
Pretoria, South Africa
June 20-21, 2016





CEOS

What is CEOS?

www.ceos.org



The Committee on Earth Observation Satellites (CEOS) serves as a focal point for international coordination and data exchange to optimize societal benefit from space-based Earth observations. CEOS represents 22 countries through its 31 space agencies and 28 associate members.



CEOS agencies collaborate on a variety of international projects that address key global issues ...

Data Access ... Promoting and improving space data use

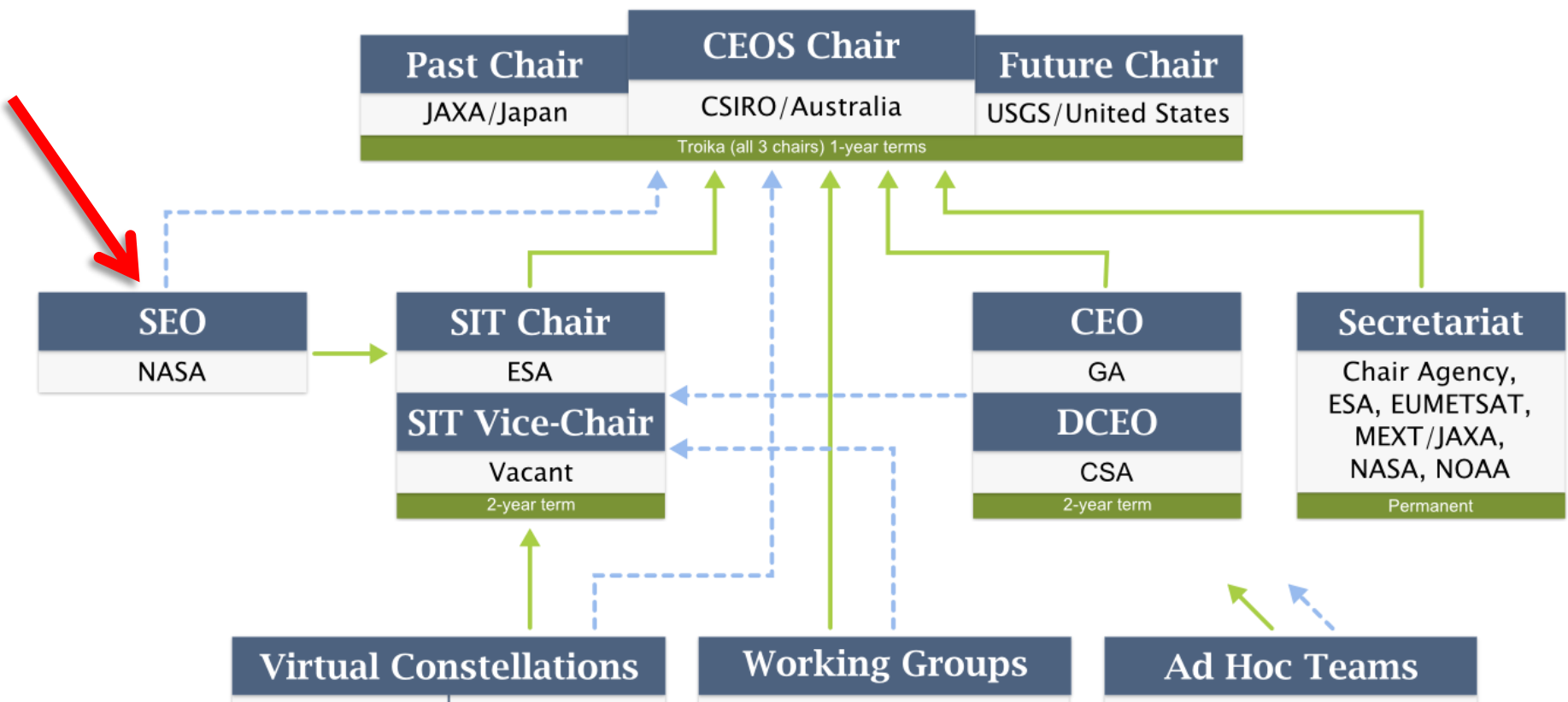
Agriculture ... Global Food Security (GEOGLAM, UN-FAO)

Deforestation ... Carbon Stocks (GFOI, UN-REDD)

Climate Change ... Global Climate Change (GCOS, UNFCCC)

Disaster Management ... Floods, Earthquakes (UNISDR)







- The observation and measurement requirements for RAPP are very similar to GEOGLAM, but there are some questions to consider ...
- We need to review the details of the following tables to be sure there is a common understanding of the requirements. Do we have the right combination of How, When and Where to obtain the needed products?
- Do we have the correct list of core and contributing missions? Do we have the personal connections and provisions in place to obtain this data?
- What are the plans for hosting and using this data (Data Services) to support the RAPP testing sites and the long-term operational plan?
- How can we utilize the CEOS Data Cube infrastructure to support RAPP?
- What user interface tools are needed to produce RAPP products? What GIS tools are used by the RAPP community (e.g. ArcGIS, QGIS, ENVI).

Requirements Table #1a



Req#	How?		When?	Where?		Rangelands and Pastures (RP)						
	Spatial Resolution	Spectral Range	Effective Observ. frequency (cloud free)*	Sample Type	Field Size	Mask	Calendar	Vegetation Cover	Variables (NDVI/LAI)	Quantity and Growth Rates	Quality and Digestibility	Environ. Variables (e.g. Soil moisture)
Coarse Resolution Sampling (>100m)												
1	>500-2000m	optical	Daily	Wall-to-Wall	All		X	X		X		
2	100-500m	optical	2 to 5 per week	Cropland Extent	All	X	X	L		X	X	
3	5-50km	microwave	Daily	Cropland Extent	All							X
Moderate Resolution Sampling (10 to 100m)												
4	10-70m	optical	Monthly (min 2 out of 3 seasons in season). Required every 1-3 years.	Cropland Extent	All	X	X					
5	10-70m	optical	~Weekly (8 days; min. 1 per 16 days)	Sample	All	X	X	X	X	X	X	X
6	10-100m	SAR Dual Polarization	~Weekly (8 days; min. 1 per 16 days)	Cropland Extent or Sample	All	X	X	X	X	X	X	X
6+	10-100m	Hyperspectral	Seasonal (3-4 per year)	Sample	All	X		X		X	X	

Requirements Table #1b



Req#	How?		When?	Where?		Rangelands and Pastures (RP)						
	Spatial Resolution	Spectral Range	Effective Observ. frequency (cloud free)*	Sample Type	Field Size	Mask	Calendar	Vegetation Cover	Variables (NDVI/LAI)	Quantity and Growth Rates	Quality and Digestibility	Environ. Variables (e.g. soil moisture)
Fine Resolution Sampling (5 to 10m)												
7	5-10m	VIS, NIR, SWIR	Monthly (min. 3 in season)	Cropland Extent	M/S							
8	5-10m	VIS, NIR, SWIR	~Weekly (8 days; min. 1 per 16 days)	Sample	All	M/S	X	X	X	X	X	X
9	5-10m	SAR Dual Polarization	Monthly	Cropland Extent or Sample	M/S	M/S		X		X		
Very Fine Resolution Sampling (<5m)												
10	<5m	VIS, NIR	3 per year (2 in season; 1 out of season); Required every 3 years	Cropland Extent	S	S		S		S	X	
11	<5m	VIS, NIR	1 to 2 per month	Refined Sample	All			S		S		

Measurements Table #2a



Req#	How?			Where?	When?		
	Core Missions (future)	Contributing Missions (future)	Spatial Resolution	Spectral Range	Sample Type	Effective Observ. frequency (cloud-free)*	Growing Season Calendar
Coarse Resolution Sampling (>100m)							
1	MODIS (1000m)	Suomi-NPP (750m) Proba-V (1000m)	>500-2000m	optical	Wall-to-Wall	Daily	all year
2	MODIS (250/500m) Sentinel-3A (500m)	Suomi-NPP (375m) Proba-V (100/333m)	100-500m	optical	Cropland Extent	2 to 5 per week	all year
3	Aqua GCOM-W1/W2	SMOS SMAP	5-50km	microwave	Cropland Extent	Daily	all year
Moderate Resolution Sampling (10 to 100m)							
4	Landsat 7/8 (30m) Sentinel-2A/2B (10-20m)	ResourceSat-2 (56m) CBERS-4 (20-40m)	10-70m	optical	Cropland Extent	Monthly (min. 2 out of season). Required every 1-3 years.	all year
5	Landsat 7/8 (30m) Sentinel-2A/2B (10-20m)	ResourceSat-2 (56m) CBERS-4 (20-40m)	10-70m	optical	Sample	~Weekly (8 days; min. 1 per 16 days)	growing season
6	Sentinel-1A/1B (C) Radarsat-2 (C), RCM (C) ALOS-2 (L)	RISAT-1/1A (C) RISAT-3 (L)	10-100m	SAR Dual Polarization	Cropland Extent or Sample	~Weekly (8 days; min. 1 per 16 days)	growing season



Req#	How?				Where?	When?	
	Core Missions (future)	Contributing Missions (future)	Spatial Resolution	Spectral Range	Sample Type	Effective observ. frequency (cloud free)*	Growing Season Calendar
Fine Resolution Sampling (5 to 10m)							
7	RapidEye	SPOT-6, SPOT-7 CBERS-4	5-10m	VIS, NIR, SWIR	Cropland Extent	Monthly (min. 3 in season)	growing season
8	RapidEye	SPOT-6, SPOT-7 CBERS-4	5-10m	VIS, NIR, SWIR	Sample	~Weekly (8 days; min. 1 per 16 days)	growing season
9	Sentinel-1A/1B (C) Radarsat-2 (C), RCM (C) ALOS-2 (L)	RISAT-1/1A (C) RISAT-3 (L)	5-10m	SAR Dual Polarization	Cropland Extent Sample	Monthly	growing season
Very Fine Resolution Sampling (<5m)							
10		Pleiades, SPOT-6, SPOT-7	<5m	VIS, NIR	Cropland Extent	3 per year (2 in season + 1 out of season); Required every 3 years	all year
11		Pleiades, SPOT-6, SPOT-7	<5m	VIS, NIR	Refined Sample	1 to 2 per month	growing season

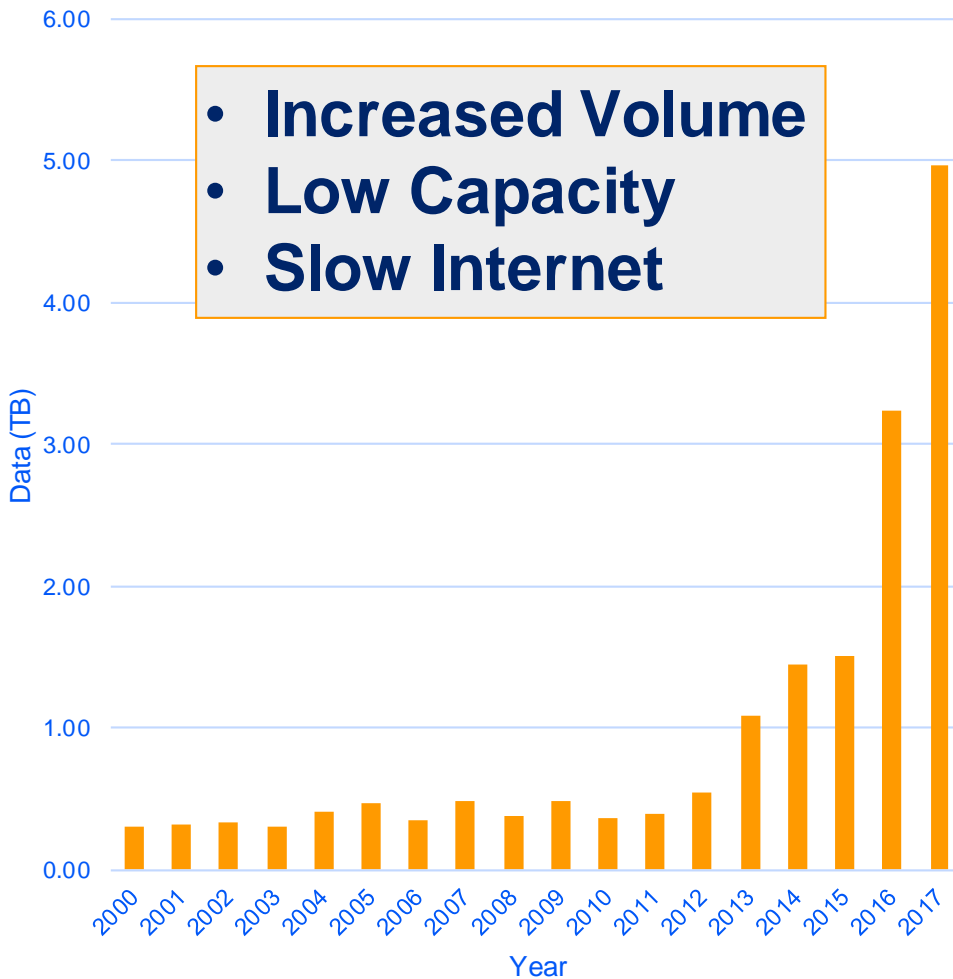


- It is assumed that most of the RAPP data supply will NOT require support from CEOS, as the data is free/open and easily obtained and understood. This includes Landsat, Sentinel, MODIS, SMAP, etc.
- If there is a desire for restricted datasets (e.g. ResourceSat-2, CBERS-4, Radarsat-2, ALOS) then CEOS will likely need to provide support.
- How about fine resolution sample data from RapidEye or Pleiades? Does this require CEOS support or is it handled locally for each test site?
- If this data is to be stored and used as “scenes”, are the proper provisions in place to accommodate this data for the test sites?

Now lets talk about Data Cubes and how they might help ...



Land Imaging Data Growth over Kenya

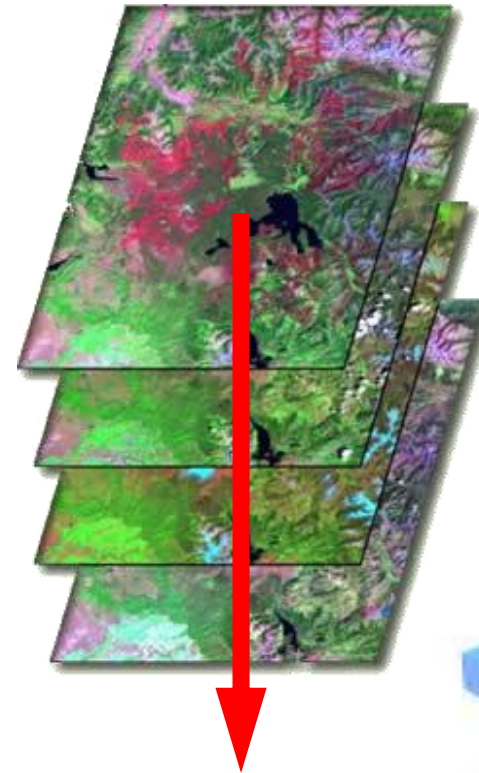


- A significant growth in land imagery data (e.g. Landsat, Sentinel) will increase data volumes by >10x in the next few years.
- Many countries lack the knowledge, infrastructure, and resources to access and use space-based data.
- Countries have requested support from CEOS for data access, storage, processing, and analysis.
- The new **CEOS Data Cube** architecture provides a solution that saves countries time and money and reduces technical complexity.



- **Data Cube** = Time-series multi-dimensional (space, time, data type) stack of spatially aligned pixels ready for analysis
- **Proven concept** by Geoscience Australia (GA) and the Australian Space Agency (CSIRO) and planned for the future USGS Landsat archive.
- **Shift in Paradigm** ... Pixels vs Scenes
- **Analysis Ready Data (ARD)** ... Dependent on processed products to reduce processing burden on users
- Supports an **infinite number of applications**, reduces data preparation time, allows time series analyses, increases interoperability of multiple datasets.
- **Open source** software approach allows free access, promotes expanded capabilities, and increases data usage.

Open Source Software
<https://github.com/data-cube>



TIME





Data



Data Cubes

Users

- Working with CEOS Space Agencies to develop requirements and plans for sustained provision of Analysis Ready Data (ARD)
- Landsat, Sentinels, MODIS, SRTM, ALOS, SPOT, and more
- Open source software, developed and sustained by CEOS
- Support for diverse datasets and grid projections
- Deployment via local computers, regional hubs (e.g. SERVIR), or computing cloud (e.g. Amazon)
- Connections to common GIS tools (ArcGIS, QGIS)
- Advanced Programming Interfaces (APIs) for users
- Prototypes in Colombia and Kenya
- Developing and testing user interfaces for custom mosaic creation, water detection and other concepts
- Investigating many capacity building options

Why not use Google Earth Engine?



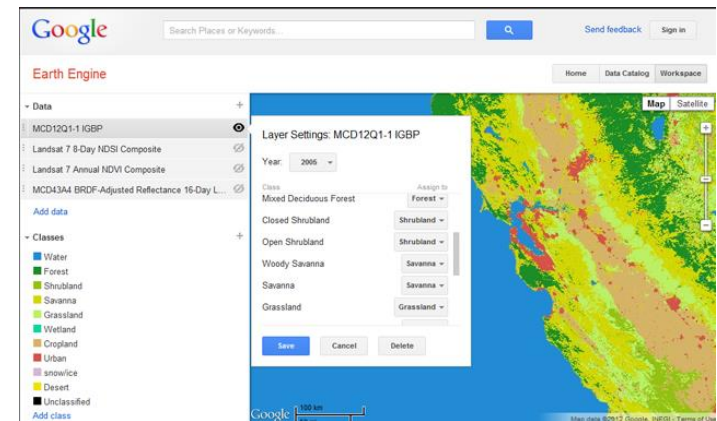
PROS

- Free, open, global datasets
- Powerful analysis tools for Javascript and Python
- Datasets are growing daily ... Adding Sentinel-1 and Sentinel-2 recently

CONS

- Commercial dependency ... Will it be sustained for the future?
- Limited time and spatial scales and limits on Python API
- Cloud-based computing only ... No options for hubs or local computing solutions
- Google “owns” all of the data. Users prefer control and ownership.
- Missing datasets

Google Earth Engine




What data is available from Google and Amazon?



Mostly Level-1 data, but some recent efforts to provide pre-processed Level-2 products by Google (Landsat and Sentinel-1).

Company	Dataset	Host	Processing/Comments
Amazon	Landsat 8, TOA	AWS S3	Level-1 Top-of-Atmosphere with orthorectification
Amazon	Sentinel-2A, TOA	AWS S3	Level-1 Top-of-Atmosphere with orthorectification
Google	Landsat 5/7/8, TOA Landsat 5/7/8, SR	Earth Engine	Level-1 Top-of-Atmosphere with orthorectification and Level-2 processed Surface Reflectance (SR) using LEDAPS
Google	Sentinel-2A, TOA	Earth Engine	Level-1 Top-of-Atmosphere with orthorectification
Google	Sentinel-1A, GRD	Earth Engine	Level-2 processed with Sentinel Toolbox. Processing includes orthorectification, removing thermal noise, radiometric calibration and terrain correction using SRTM or ASTER.



 Kenya Data Cube Portal
Custom Mosaic Tool
Home Map Tool Task Manager

Filters
Results
Output

Filters

Country: Kenya

Data Products: Landsat 8 SR

Product Type: SWIR1,NIR,RED: L7 (5,4,3), L8 (6,5,4)

Show non-clear land pixels:

Show Data Bounds:

Season: Custom Continuous

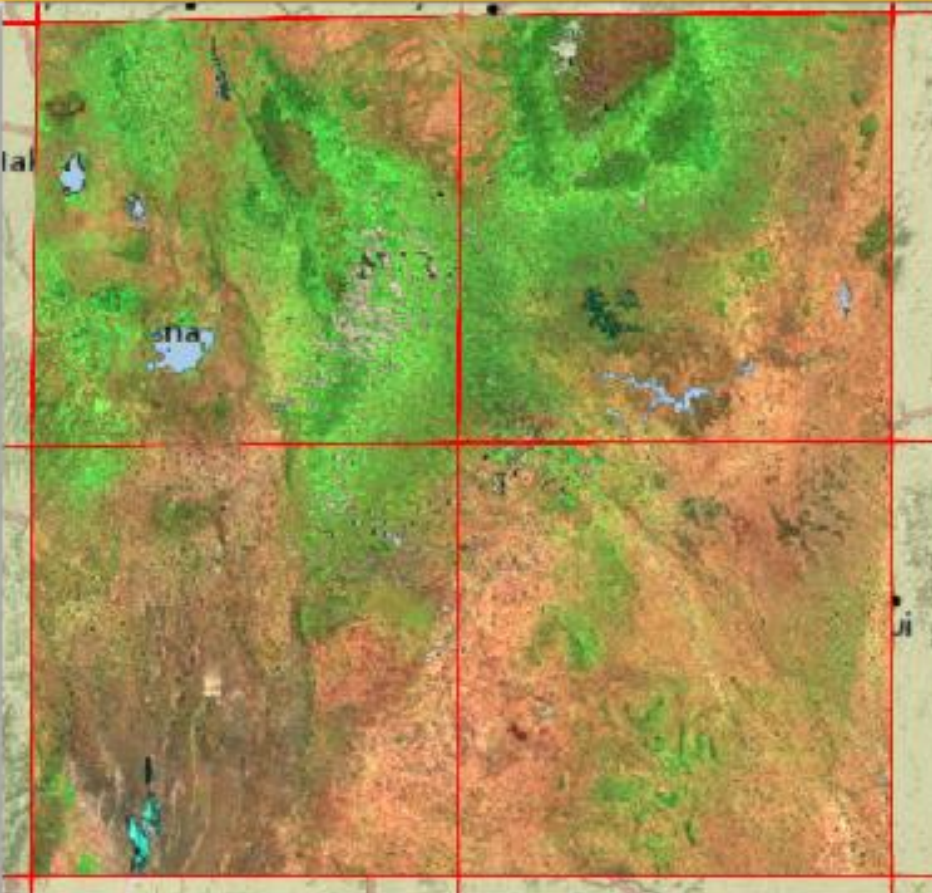
Begin Date: Year 2014 Month July

End Date: Year 2014 Month September

Shape Selector

OX

Submit Clear



Filter Selections

- Country: Kenya or Colombia
- Data Product: L7 or L8
- Product Type: 11 options
- Non-clear pixels: RED flag
- Season: Continuous or Multi
- Dates: Month-Year
- Bounding Box or Lat-Long

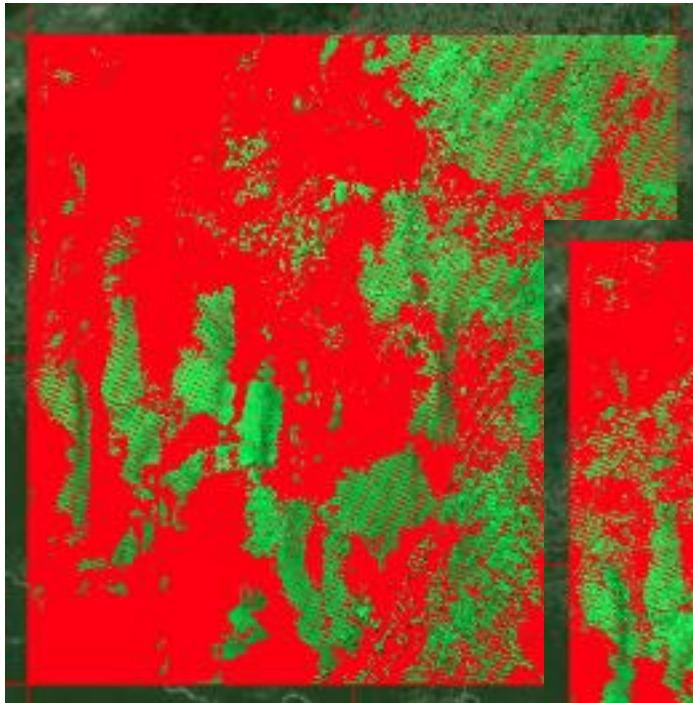


Product: **Landsat 7**

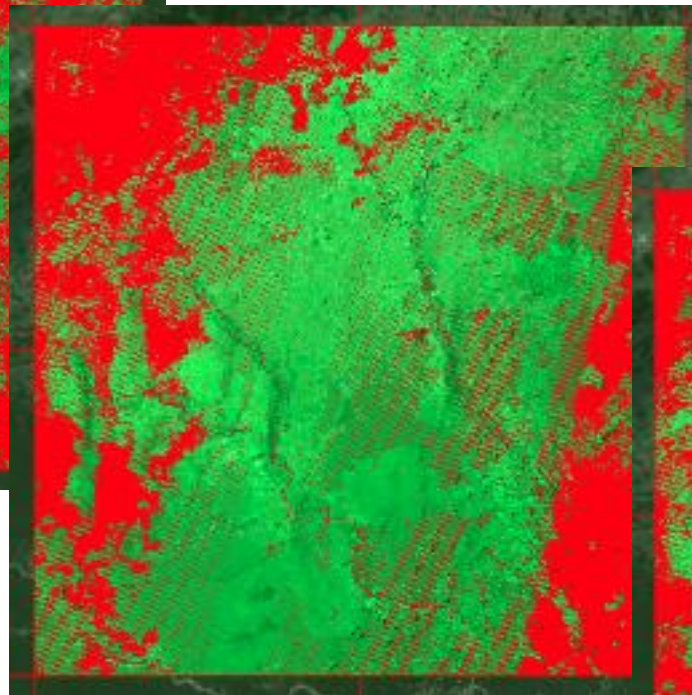
Region: Southern Colombia

Output: RGB Bands-7,4,2 (SWIR2, NIR, GREEN)

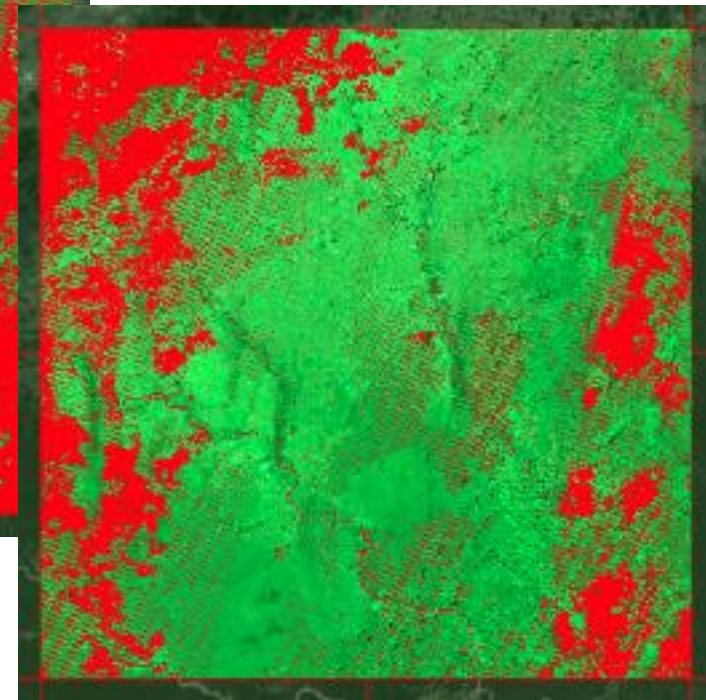
Filter (**RED**): = Cloud, Shadow, Water, No Data



January-February 2014
16 scenes
70% no data (mostly clouds)



January-April 2014
30 scenes
36% no data (mostly clouds)



January-June 2014
46 scenes
27% no data (mostly clouds)

~1 minute run time
for each mosaic with
output in GEOTIFF

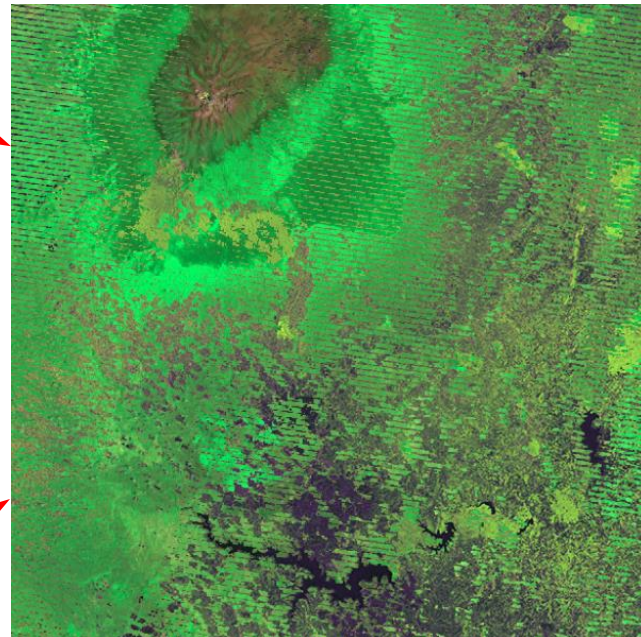


*Masked clouds, water,
and no data (banding)*

Combined Mosaic Product

An integrated mosaic was created with the Data Cube over central Kenya for Jan-Feb 2007 (dry season) using Landsat-7 (5 scenes) and the 2007 ALOS-PALSAR 25m mosaic.

Landsat-7



ALOS-PALSAR data can be used to fill areas of cloud cover and scan line "banding" within the Landsat-7 imagery for improved land classification.

ALOS-PALSAR

*Complements of JAXA
And Ake Rosenqvist*



Landsat-8

Masked clouds and water

Combined Mosaic Product

An integrated mosaic was created with the Data Cube over central Kenya for Oct-Dec 2014 using Landsat-8 and a Oct 2014 to Oct 2015 Sentinel-1A mosaic.

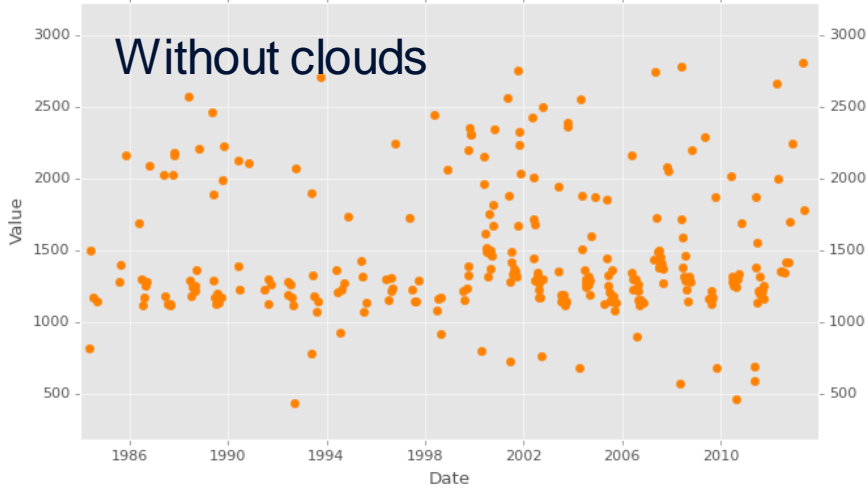
Sentinel-1A data can be used to fill areas of cloud cover within the Landsat-8 imagery for improved land classification.

Sentinel-1A

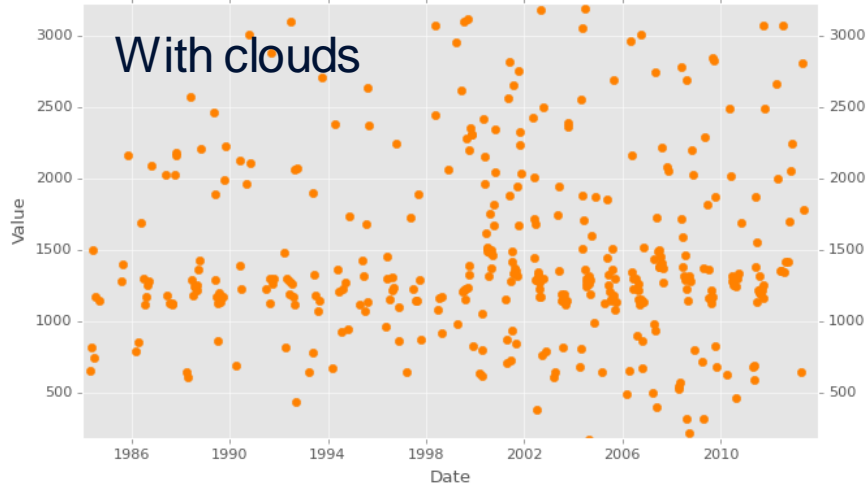
*Complements of Norut
VV-VH-NDI product*



Row/Col: Stacked TS - 48/15

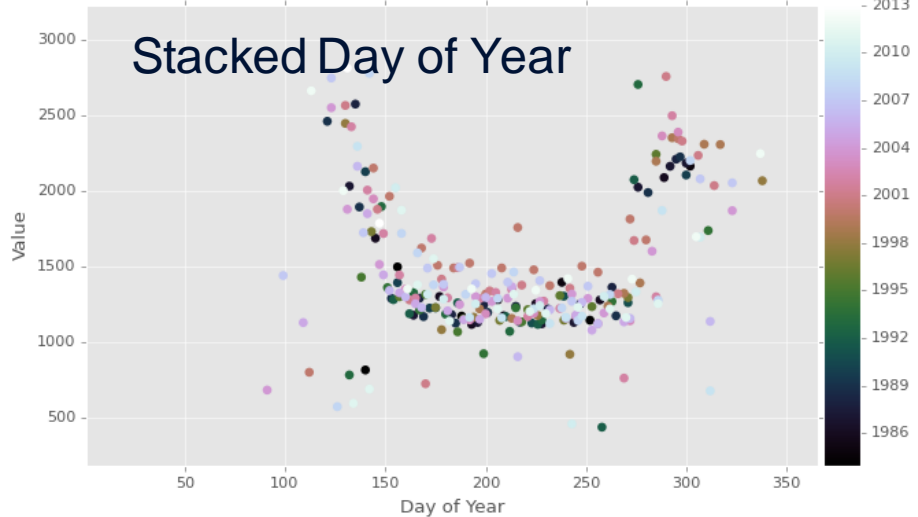


Row/Col: Stacked TS - 48/15



Sample output from QGIS using a sample Data Cube over Path-Row (035-032), USA Images on the left show the Landsat time series of SWIR (Band 5) data for one pixel. Clouds are easily removed and pixel-based change detection (e.g. CCDC) can be run.

Row/Col: Stacked TS - 48/15





- The concept of a Data Cube is not new. Many local scientists use Data Cubes to package their data for time series analyses. The “new” concept is that CEOS would like to globalize a common infrastructure.
- **USGS Land Change Monitoring Assessment and Projection (LCMAP)** initiative is planning a CONUS Data Cube. They will use a Cassandra database and a REST user interface based on their available computing infrastructure. USGS will start with a CONUS (USA only) cube and then eventually move to a global distribution of data in cube format. Their global delivery will be many years away ...
- **Australian Geoscience Data Cube (AGDC)** is focused on Australia applications and designed to utilize the National Computing Infrastructure (NCI) high performance computing facility. Their focus is only the Australia region with no plans for global expansion.



A **Data Cube Work Plan** was recently drafted to capture all of the tasks for better planning and management.

These tasks include:

- Management of the open source software GitHub
- Development of ingestors for many space datasets and ground-based climate data
- Integration with GIS tools (e.g. QGIS and ArcGIS)
- Development of user interface tools for custom mosaics, water detection, and Data Cube creation
- Definition of ARD and working with CEOS Agencies to make ARD routinely available
- External interactions with GFOI, GEOGLAM, World Bank, Amazon, Google, SERVIR, WGCapD, SilvaCarbon, GEONETCast
- Prototype testing in Kenya, Colombia, Canada and Asia-Pacific





- How can we utilize the CEOS Data Cube infrastructure to support RAPP?
- What user interface tools are needed to produce RAPP products?
- What GIS tools are used by the RAPP community (e.g. ArcGIS, QGIS, ENVI)?