



Lowering the Cost of Knowledge by Democratizing Access to Big Data

Jed Sundwall, Global Open Data Lead

Eye on Earth Symposium – 22 October 2018

“No matter who you are, most of the smartest people work for someone else.”

— Bill Joy, co-founder of Sun Microsystems

Why does AWS care about open data?

Sharing data on AWS makes it accessible to a large and growing community of researchers, entrepreneurs, and enterprises who use the AWS cloud.



Many of our public sector customers are required to make their data available to the public.



Many of our commercial sector customers rely on access to open data to develop their products.

<https://opendata.aws>

Traditional data acquisition



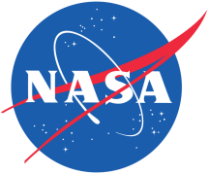
“...data must be organized, well-documented, consistently formatted, and error free. Cleaning the data is often the most taxing part of data science, and is frequently **80% of the work.**”

—*Data Driven* by DJ Patil and Hilary Mason

Undifferentiated heavy lifting

AWS Public Dataset Program

<https://registry.opendata.aws>





Earth on AWS

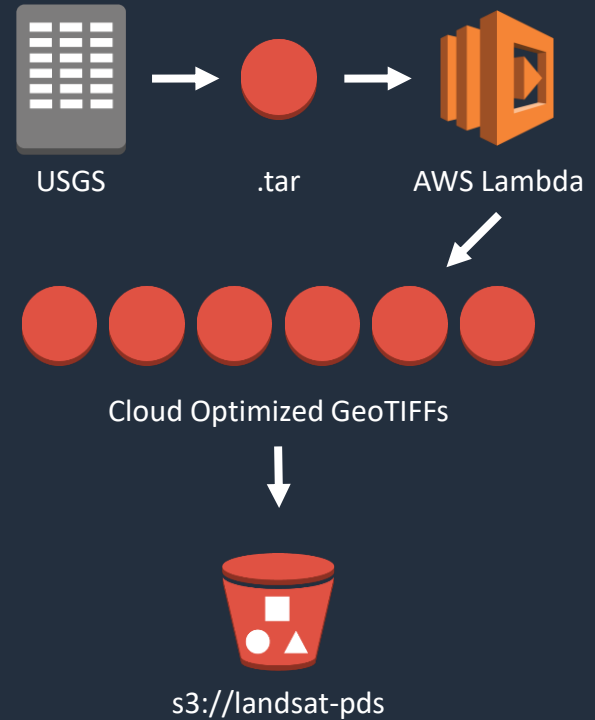
aws.amazon.com/earth

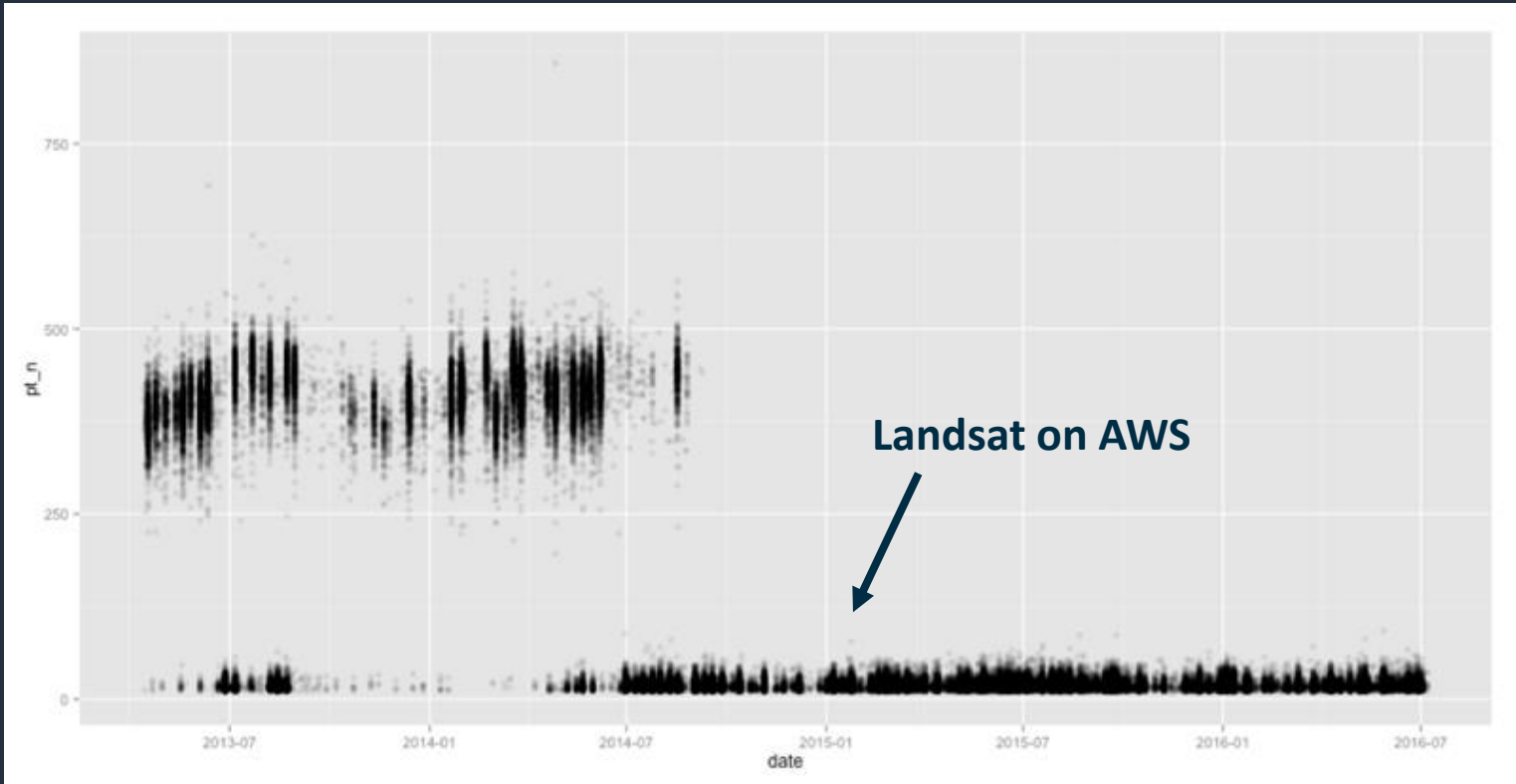
Staging data for analysis

Amazon S3 allows programmatic and precise access to data at planetary scale.

Landsat on AWS uses Cloud Optimized GeoTIFFs that allow users to get only the data they need when they need it.

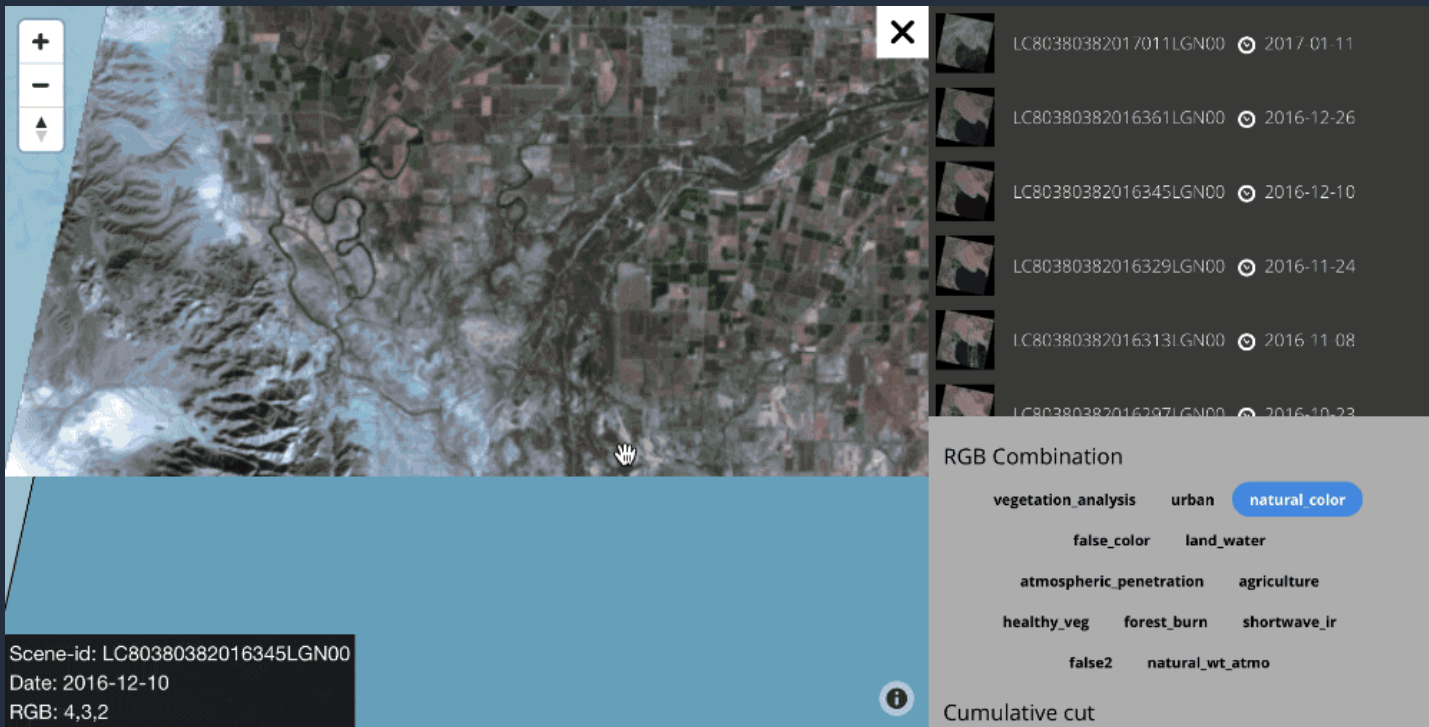
cogeo.org





Graph by Drew Bollinger (@drewbo19) at Development Seed

landsat-tiler from Mapbox



Scene-id: LC80380382016345LGN00
Date: 2016-12-10
RGB: 4,3,2

LC80380382017011LGN00 2017-01-11
LC80380382016361LGN00 2016-12-26
LC80380382016345LGN00 2016-12-10
LC80380382016329LGN00 2016-11-24
LC80380382016313LGN00 2016-11-08
LC80380382016287LGN00 2016-10-23

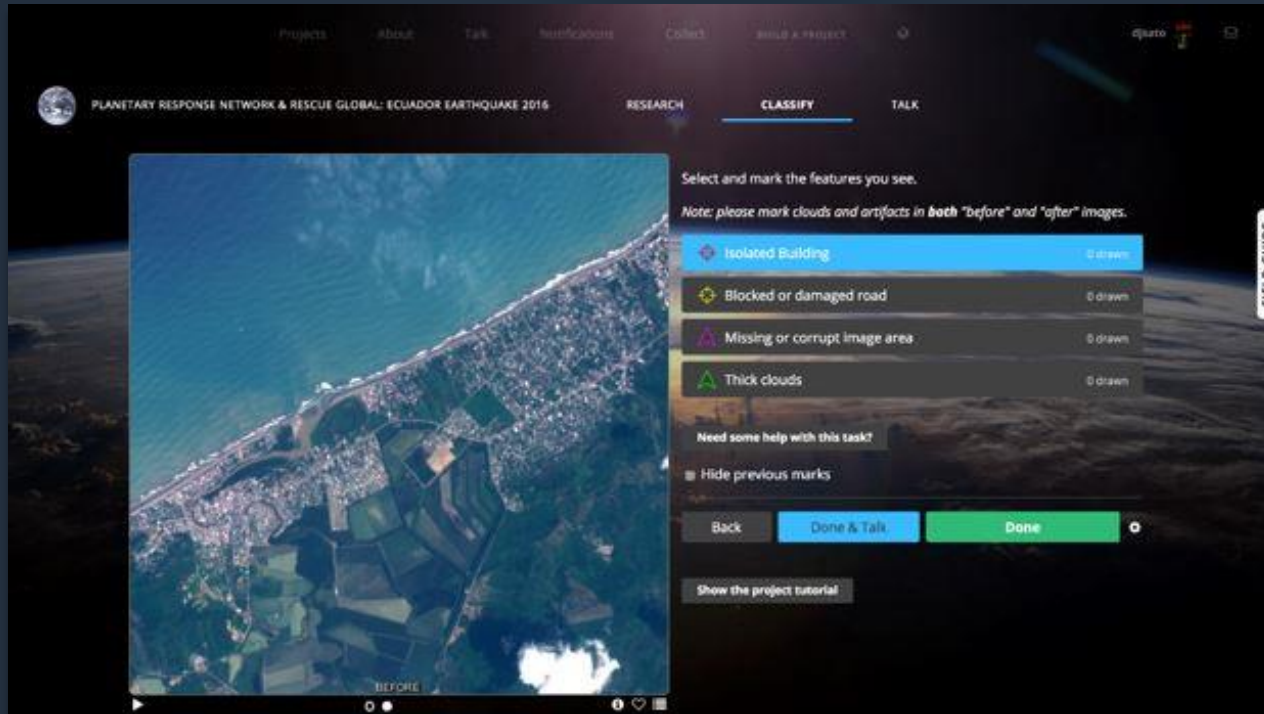
RGB Combination

vegetation_analysis urban **natural_color**
false_color land_water
atmospheric_penetration agriculture
healthy_veg forest_burn shortwave_ir
false2 natural_wt_atmo

Cumulative cut

<https://viewer.remotepixel.ca/>

Zooniverse Planetary Response Network



The screenshot displays the Zooniverse Planetary Response Network interface. At the top, there are navigation links: Projects, About, Talk, Notifications, Collect, and BUILD A PROJECT. The main header reads "PLANETARY RESPONSE NETWORK & RESCUE GLOBAL: ECUADOR EARTHQUAKE 2016". Below this, there are tabs for RESEARCH, CLASSIFY (selected), and TALK. The central part of the screen shows a satellite image of a coastal area, labeled "BEFORE" at the bottom. To the right of the image is a classification task list with the following items:

- Isolated Building (0 drawn)
- Blocked or damaged road (0 drawn)
- Missing or corrupt image area (0 drawn)
- Thick clouds (0 drawn)

Below the list, there are buttons for "Need some help with this task?", "Hide previous marks", "Back", "Done & Talk", and "Done". At the bottom, there is a button for "Show the project tutorial". A vertical "FIELD GUIDE" button is located on the right edge of the interface.



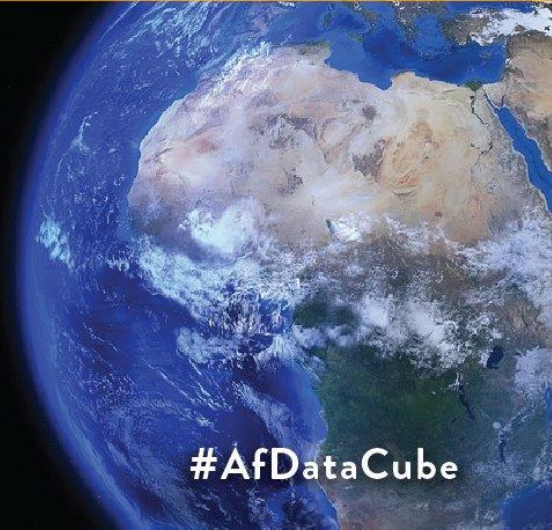
AFRICA REGIONAL DATA CUBE



The Committee on Earth Observation Satellites



WITH SUPPORT FROM



#AfDataCube

OpenStreetMap and Amazon Athena

```
SELECT * from planet
WHERE type = 'node'
      AND tags['amenity'] IN ('hospital', 'clinic', 'doctors')
      AND lon BETWEEN -15.0863 AND -7.3651
      AND lat BETWEEN 4.3531 AND 12.6762;
```

All health centers in Guinea, Sierra Leone, and Liberia

- Can be used by groups to investigate road/building quality, access to health centers, access to road networks
- Queries can be completed in seconds, even while querying over 9 years and billions of OSM elements.
- You pay only for data scanned for analysis. No server required.

Snowball Edge

- 100 TB capacity/10 Gb network
- Data encryption end-to-end
- Rugged 8.5 G impact case
- Rain and dust resistant
- Compute and storage for hybrid/edge workloads
- Rack-mountable, cluster-ready



Humanitarian Drone Data Coordination in Peru



Recently, the National Oceanic and Atmospheric Administration and Amazon Web Services (AWS) Cloud made available one of the largest datasets describing animal movement ever compiled: the Next Generation Weather Radar (NEXRAD) archive.

— Adriaan M. Dokter et al. Nature (2018)

Thank you!

Jed Sundwall

jed@amazon.com

