GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

Building Stronger Data Ecosystems to Address SDGs Data Needs

Data Science in Africa Workshop
@ the African University of Science and Technology, Abuja
Nov. 15, 2018

The Global Sustainable Development Goals

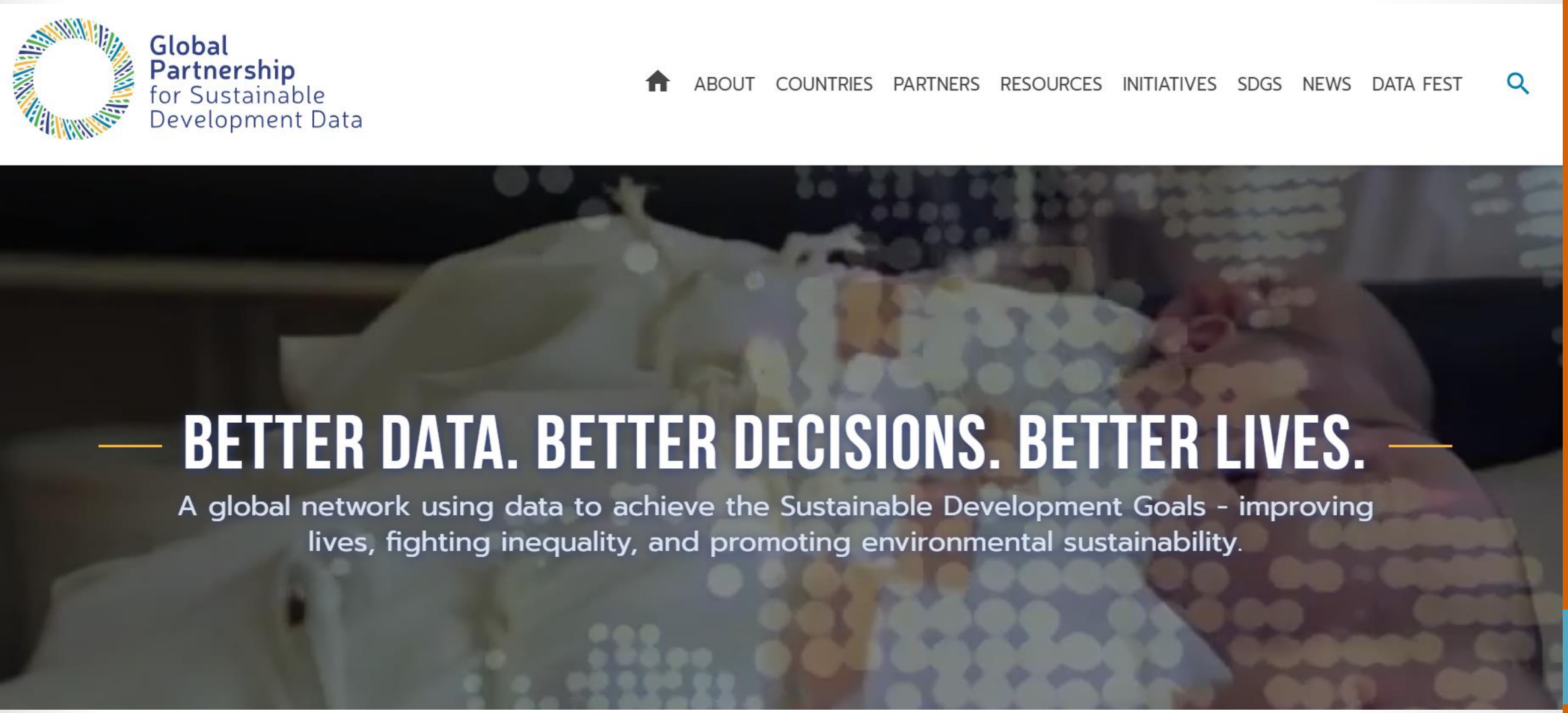
MDGs (2000-2015)		SDGs (2015-2030)
Developing country focused	→	Universal
Social	→	Social, Economic, and Environmental
Foreign Aid	→	Domestic Investment, Private Flows, and Aid
Official Statistics and Administrative Data	→	Big Data, Citizen Generated Data, Geospatial and Earth Observation Data, Open Data, and more

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

- 17 Goals, 169 Targets, 230 Indicators = Huge Data Needs

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

About the GPSDD



GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

Our Approach

- We **focus on data systems:** a system that integrates people, institutions and technology, organized and incentivized to promote data production, share and use.

- We **work with people and institutions who want to be in the forefront of change.**

- We are **demand-driven.**

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

THE CHALLENGES:

Data are not available, dynamic, disaggregated, high quality, useable, accessible, open, or used effectively.

- Data on entire groups and key issues are unavailable.
- Data are not dynamic or disaggregated.
- Data quality is poor and major gaps remain.
- Data that exist are often not useable.
- Data that are useable are not accessible or open.
- Data that are accessible are often not used effectively.

DATA CHALLENGES LEAVE TOO MANY BEHIND

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

Problems of Focus

- Huge data gaps, poor quality, illiteracy and inaccessibility.
- Lack of capacity to produce and ineffective use of data for decision making, empowerment, accountability, innovation.
- Local innovations and new technologies are neither shared broadly enough or scaled for maximum impact.
- The public, social and development sectors lag far behind in harnessing the great potential of the data revolution for sustainable development.
- There is a worrying siloed stakeholder landscape, both across sectors and data communities.

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

Key issues and challenges?

Enabling Environment

Political

Technical

Coordination+
Collaboration

Capacity

Evidence & learning

DATA FOR WHAT?

Improved Decision-Making and Policy

Increased Citizen Empowerment

Power Innovation and Entrepreneurship

To Achieve and Monitor Sustainable Development

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

What Are We Doing to Help?

1. Political Advocacy and creating/strengthening institutional mechanisms.

2. Brokering partnerships and connections to global, regional and local resources.

3. Enabling Thematic and Sectoral Data Gaps Assessment.

4. Establishing Data Ecosystem to address sectoral data and service challenges.

5. Provision of Data Infrastructure and Dataskill4SDGs.

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

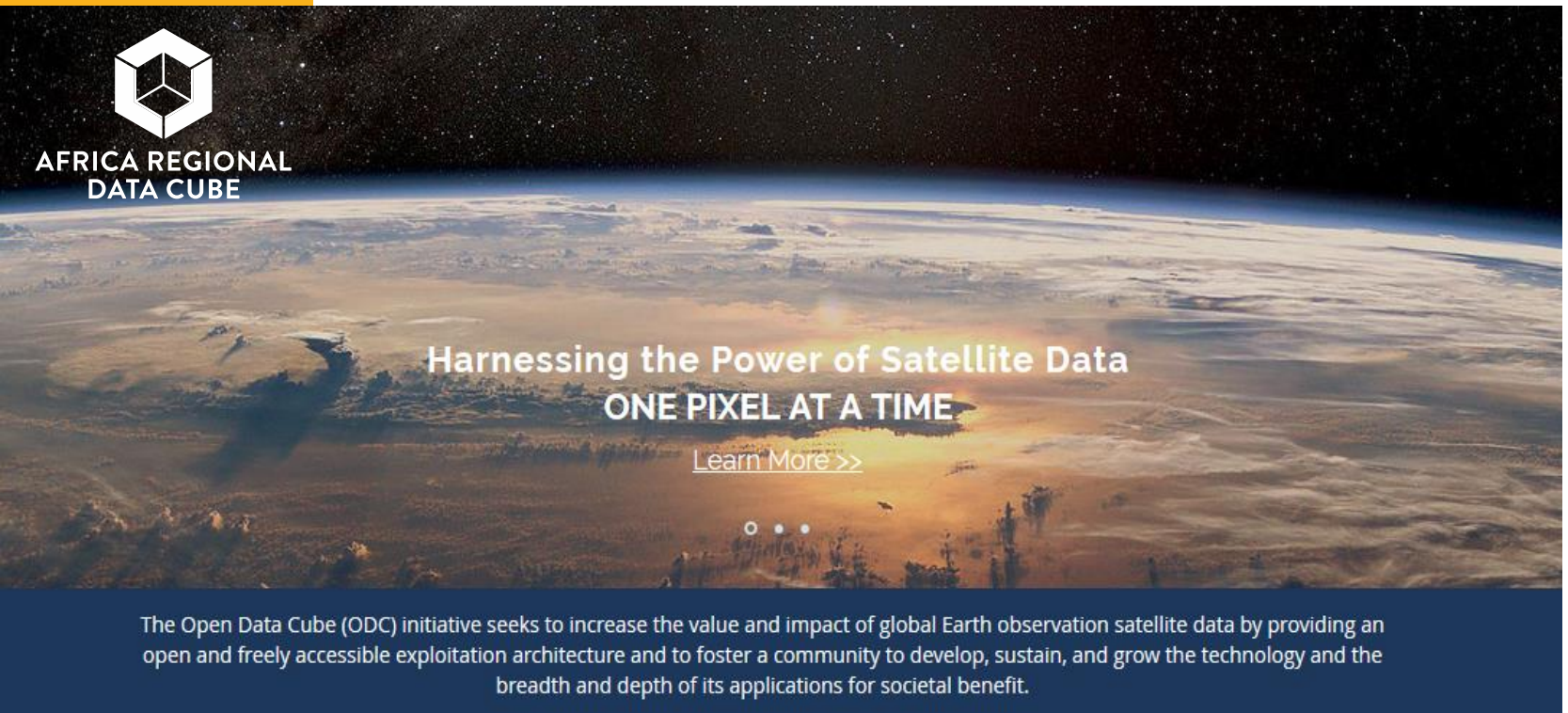
Earth Observation Data

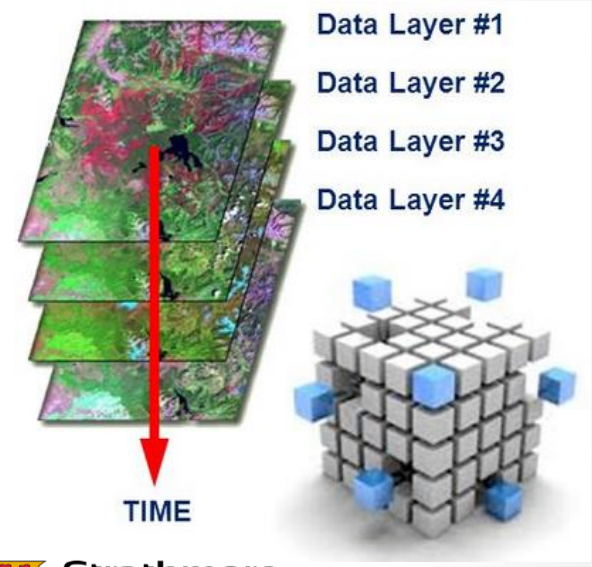
Global Partnership for Sustainable Development Data

Africa Regional Data Cube

A data cube provides analytically-ready data across decades allowing for easily accessible geospatial analysis on key issues. The initial focus for the data cube will be on algorithms to address priorities identified by GPSDD partners across **5 countries**:

Ghana | Kenya | Senegal I Sierra Leone | Tanzania





Strathmore UNIVERSITY

NASA

CEOS Committee on Earth Observation Satellites

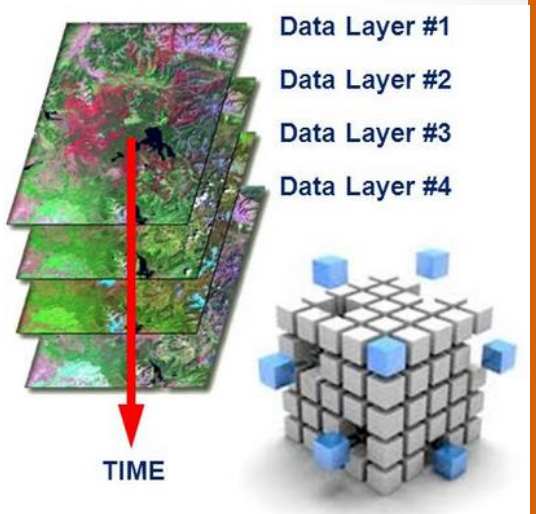
esa

GEO GROUP ON EARTH OBSERVATIONS

Global Partnership for Sustainable Development Data

Africa Regional Data Cube

- **Data Cube** = Time-series stack of spatially aligned pixels ready for analysis
- **Analysis Ready Data (ARD) ...** Data which has been pre-processed and ready for analyses, thereby reducing the burden on users
- **Open source** software approach allows free access, promotes expanded capabilities, and increases data usage.
- **Unique features:** exploits time series, increases data interoperability, and supports many new applications.



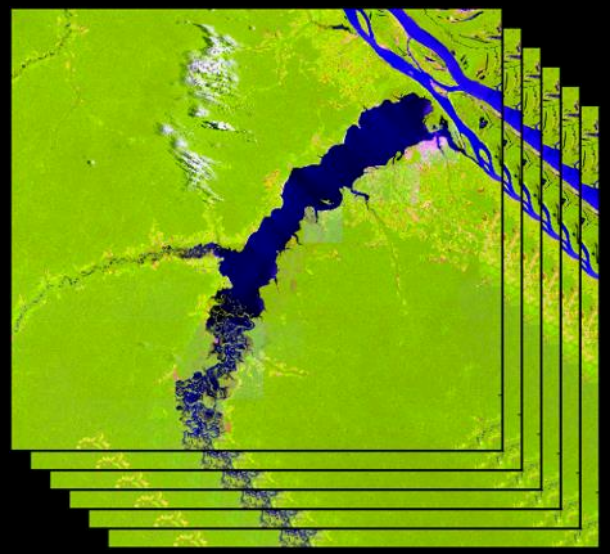
NASA

CEOS Committee on Earth Observation Satellites

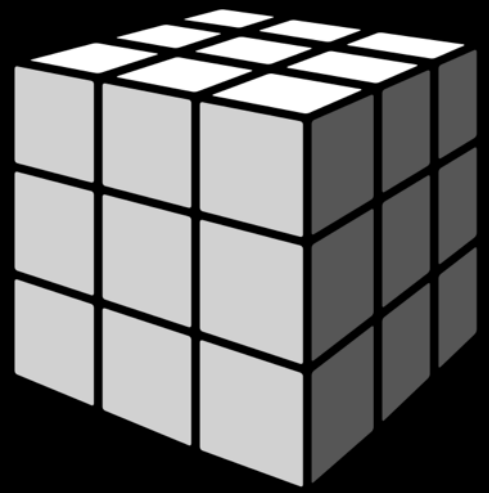
esa

GEO GROUP ON EARTH OBSERVATIONS

What is a Data Cube?



8000 Landsat images over 17 years



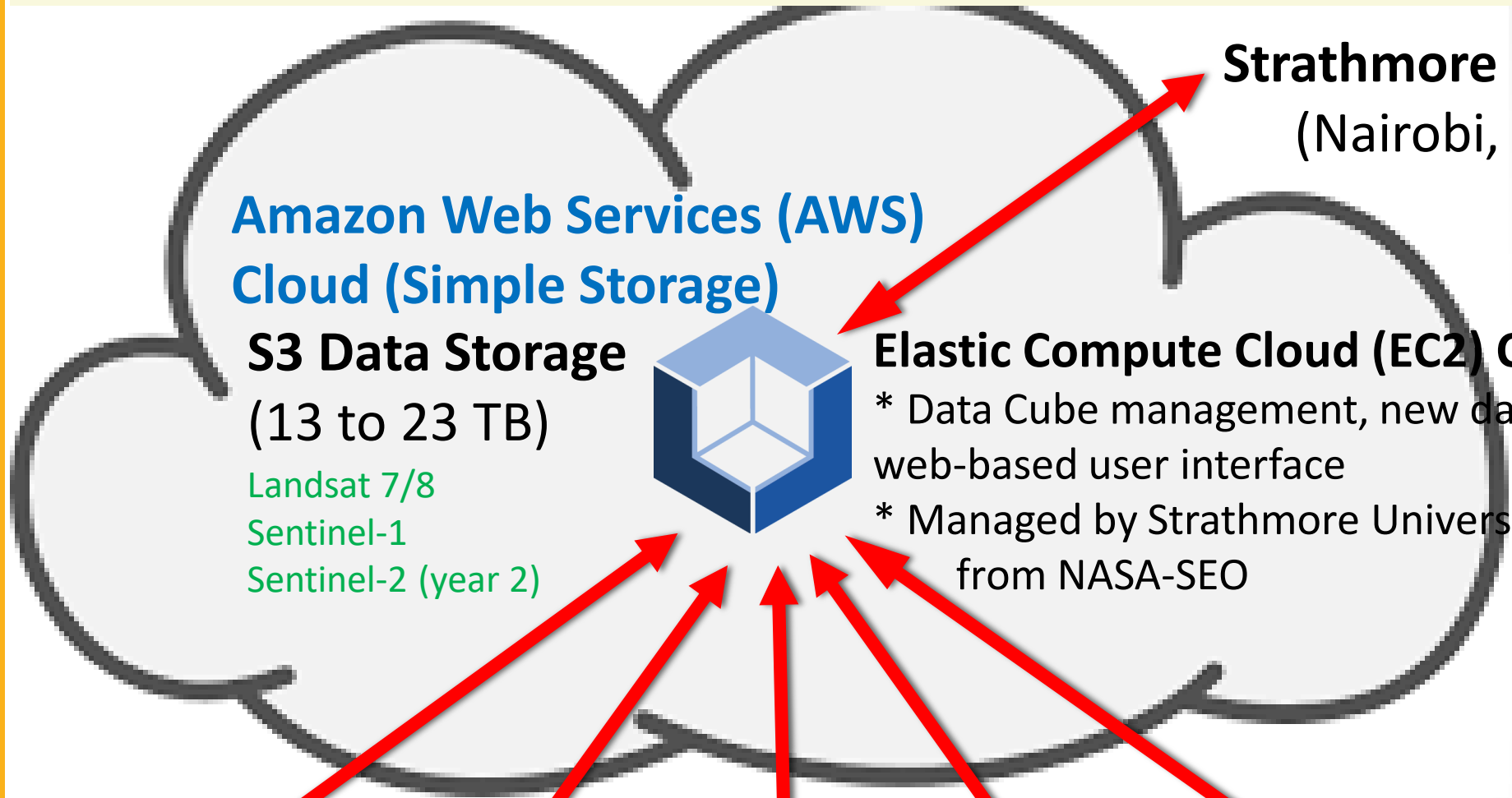
1000 Data Cube storage units
(1° x 1° x 1 year)
Processed and ready!

8x data compression

Why do we need Data Cubes?

- Makes it easier to use satellite data by reducing preparation time
- Makes it easier to query data in time and space for powerful application products
- Provides free and open source software and algorithms for the cloud or a local computer
- Allows community development, shared capacity building and organized algorithms
- Enables efficient time series analyses and data interoperability
- Because it works … just ask Australia!

African Regional Data Cube (ARDC) Operational Model



Each country has its own EC2 Computing "instance" for analysis purposes (User Interface and Jupyter Notebooks), but S3 data storage is shared among countries in the AWS cloud

Some Priority Use Cases in Africa

1. **illegal mining:** locating mining sites to produce Mining Map and understand land cover changes and water extent/quality around those locations.

2. **Land Cover/land degradation**: detection of land cover changes to highlight land degradation which could be due to mining, logging, etc.

3. **Landslides/Flooding**: detection of potential and actual land shifts and areas prone to flooding and landslides.

4. **Urbanization:** understanding urbanization rate around Freetown and rate of land consumption.

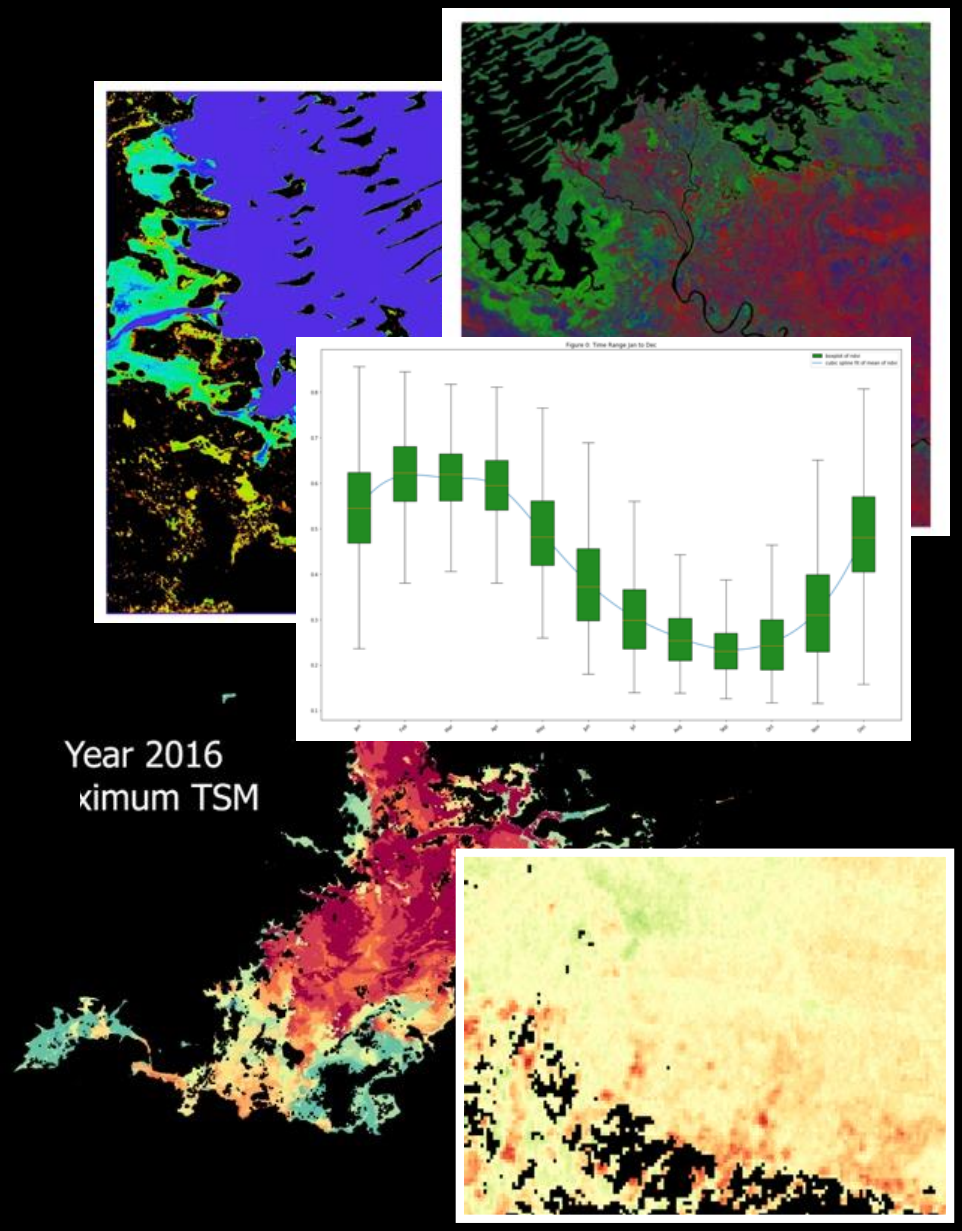
5. **Water mgt**: Water extent and quality (for drinking and agriculture)

6. Rangeland quality (no fertilizers are used)

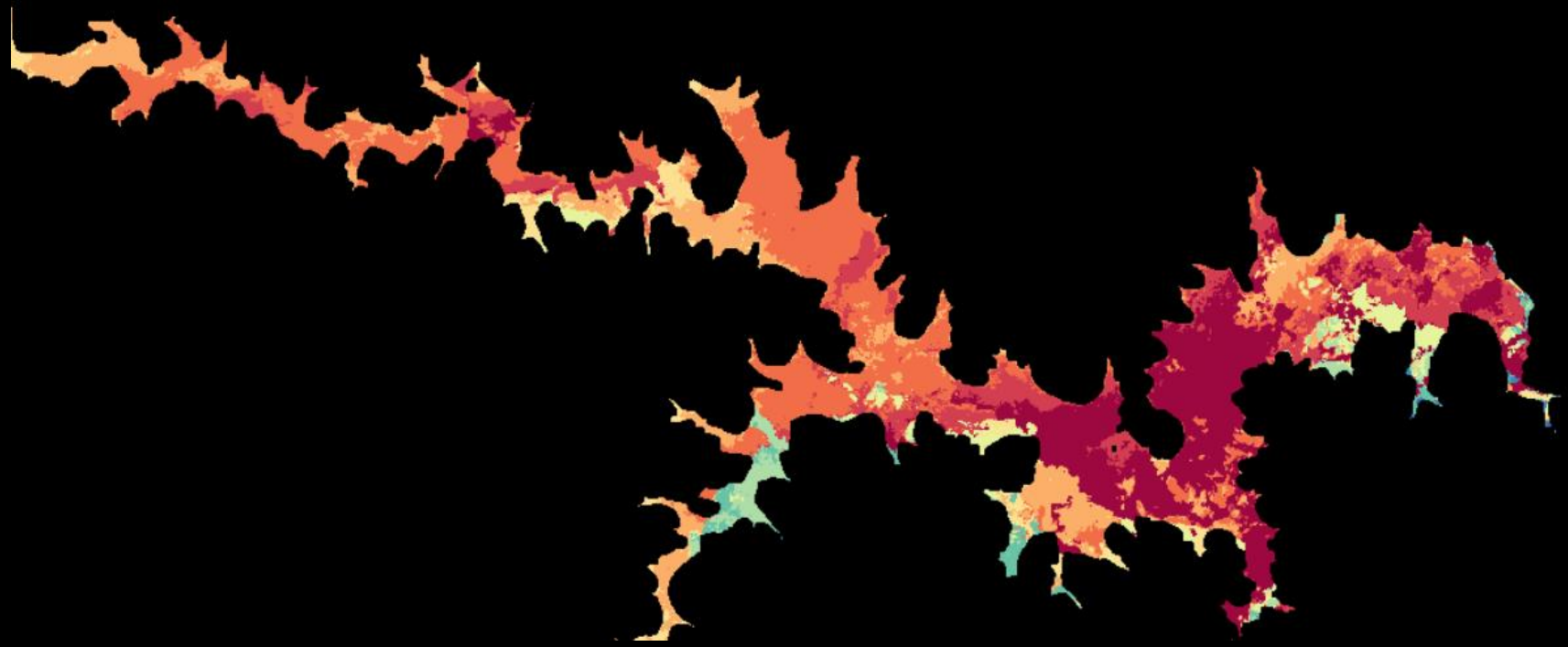
GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

Application Algorithms

- Cloud-free Mosaics: Recent Pixel, Median, Geomedian, Max-NDVI
- Spectral Indices: NDVI, EVI, NDBI, NDSI, NDWI, Fractional Cover, Tasseled Cap
- Land Classification:
  K-Means, Random Forest
- Water: WOFS (Australia), Sentinel-1 WASARD,
  Total Suspended Matter
- Land Change: PyCCD (USGS),
  PCA (Colombia), NDVI Anomaly, Sentinel-1 Radar (Deutscher), Coastal Change, Landslide Risk

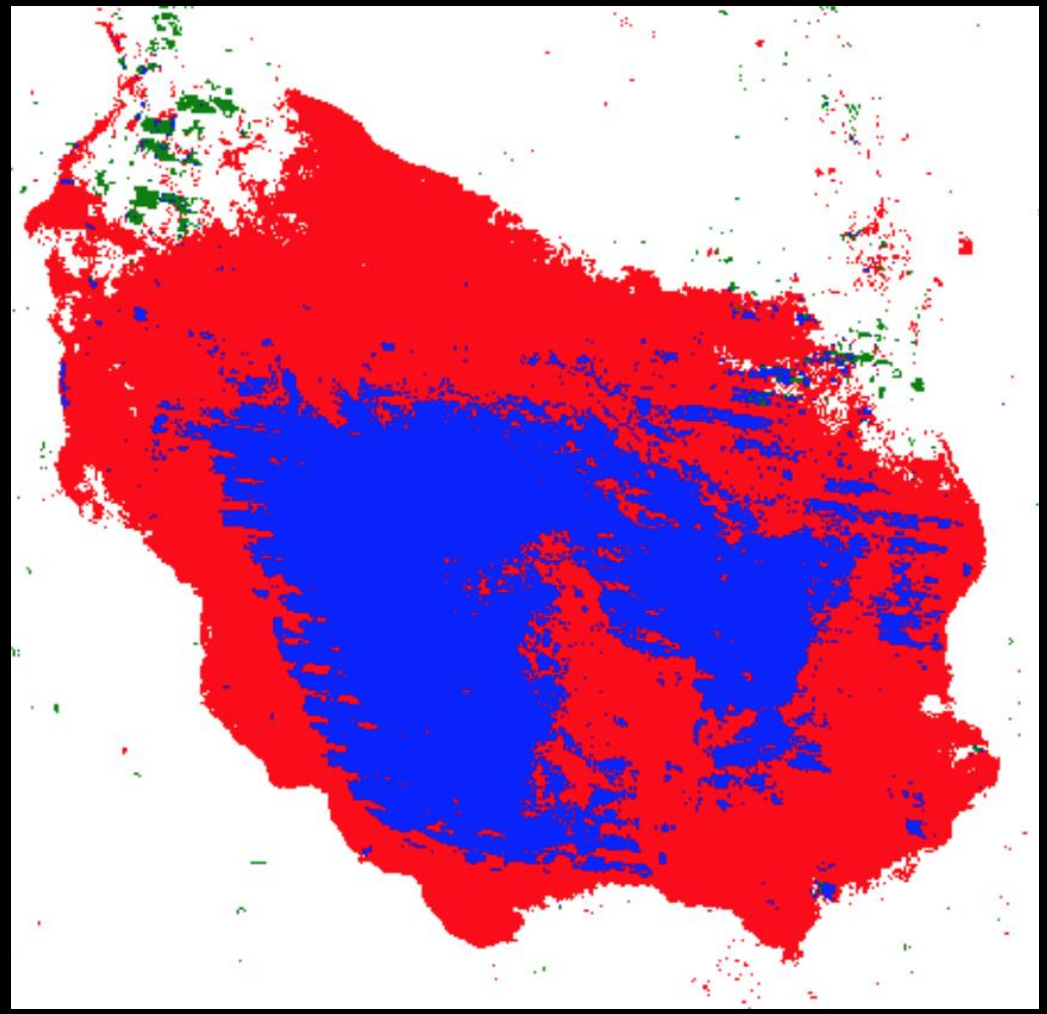


SDG 6.3.2 - Water Quality



Water Quality (total suspended sediment) variability (2013 to 2017) along the Masinga Dam (near Nairobi, Kenya. These Data Cube results can be used to modify water sampling locations and timing with satellite data.

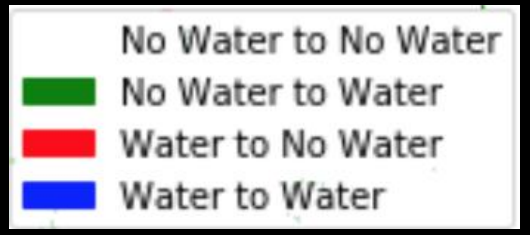
SDG 6.6.1 - Water Extent



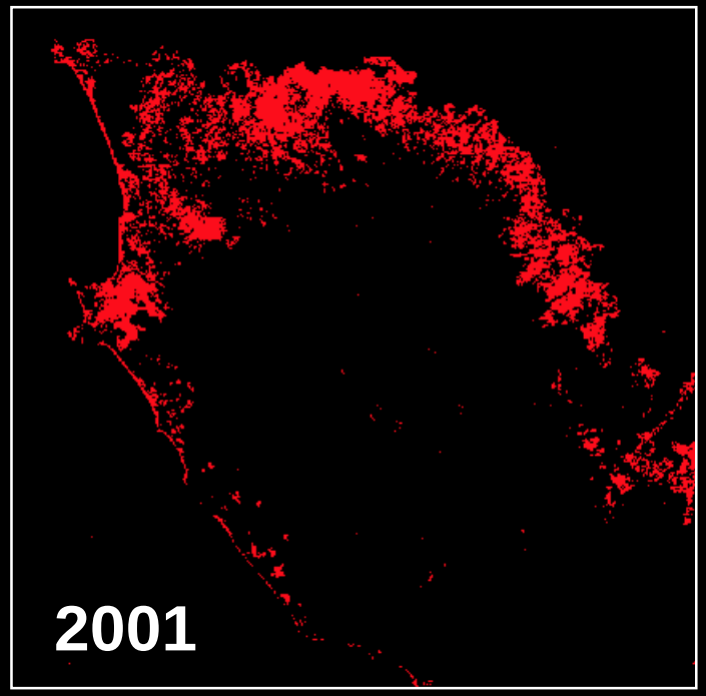
Lake Sulunga in Tanzania

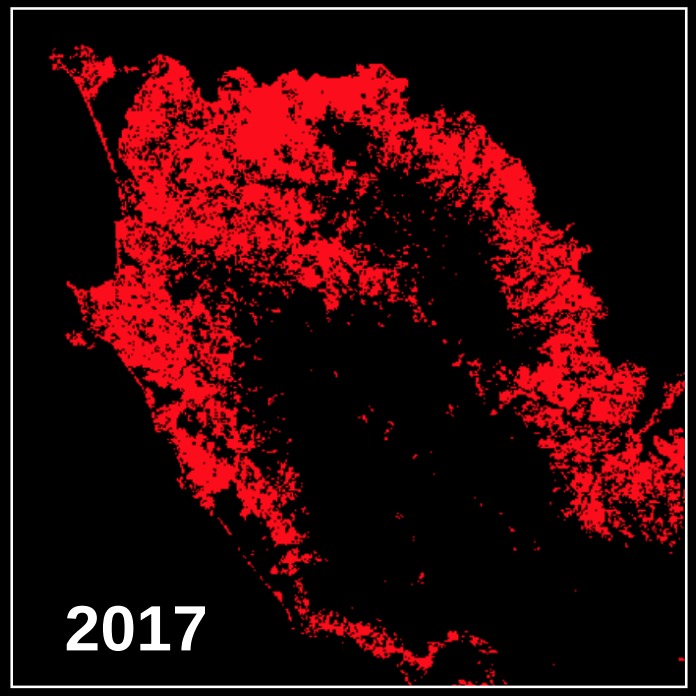
Highly variable water extent over 17 years.

From 2001 to 2017 the lake lost 62% of its water area



SDG 11.3.1 - Urbanization

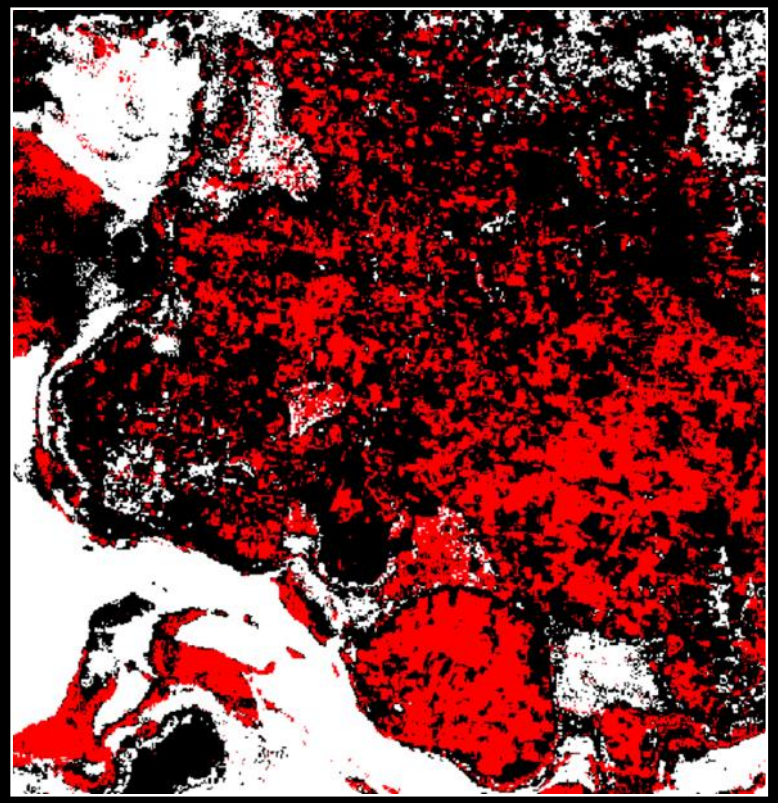




Urbanization in Freetown, Sierra Leone
The change in urban area from over 17 years was ~150% compared to a population growth of ~80%.

What does this tell us? **People are moving out of the cities and taking over more land.**

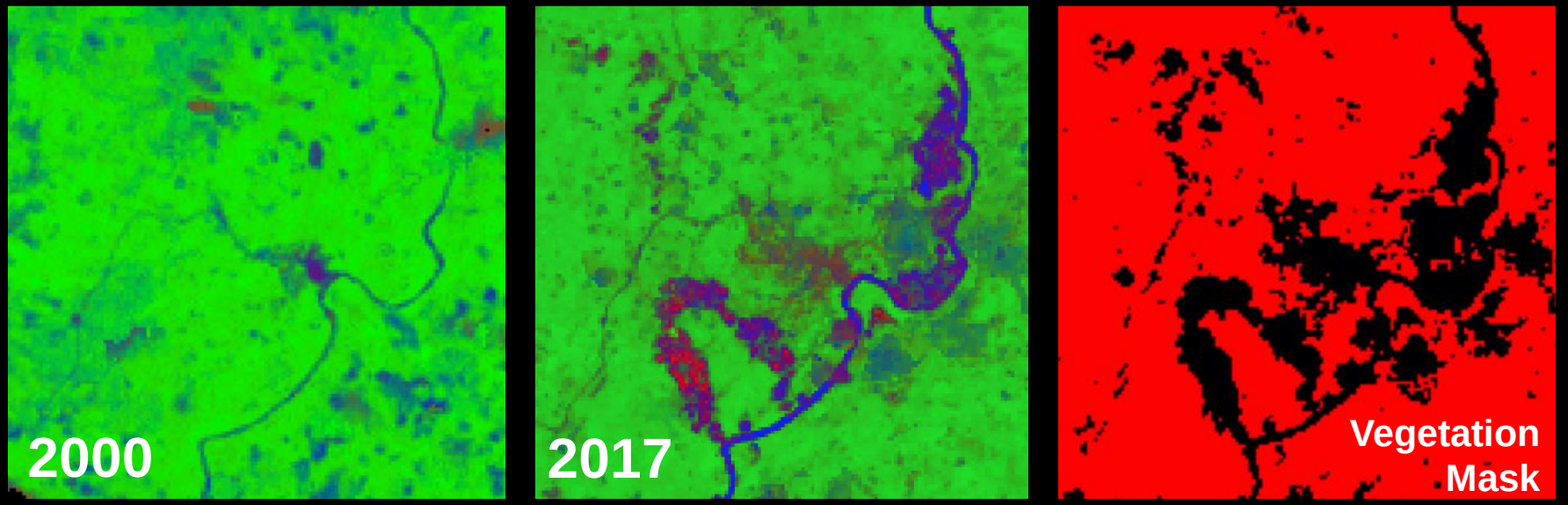
SDG 15.1.1 – Forest Area



Usinge Forest in western Tanzania
The results of the Data Cube (left) can be compared with the Global Forest Watch product (right).

Using an NDVI threshold approach, there is a loss of 10.6% of the land to deforestation from 2004 to 2014.

SDG 15.3.1 – Land Degradation



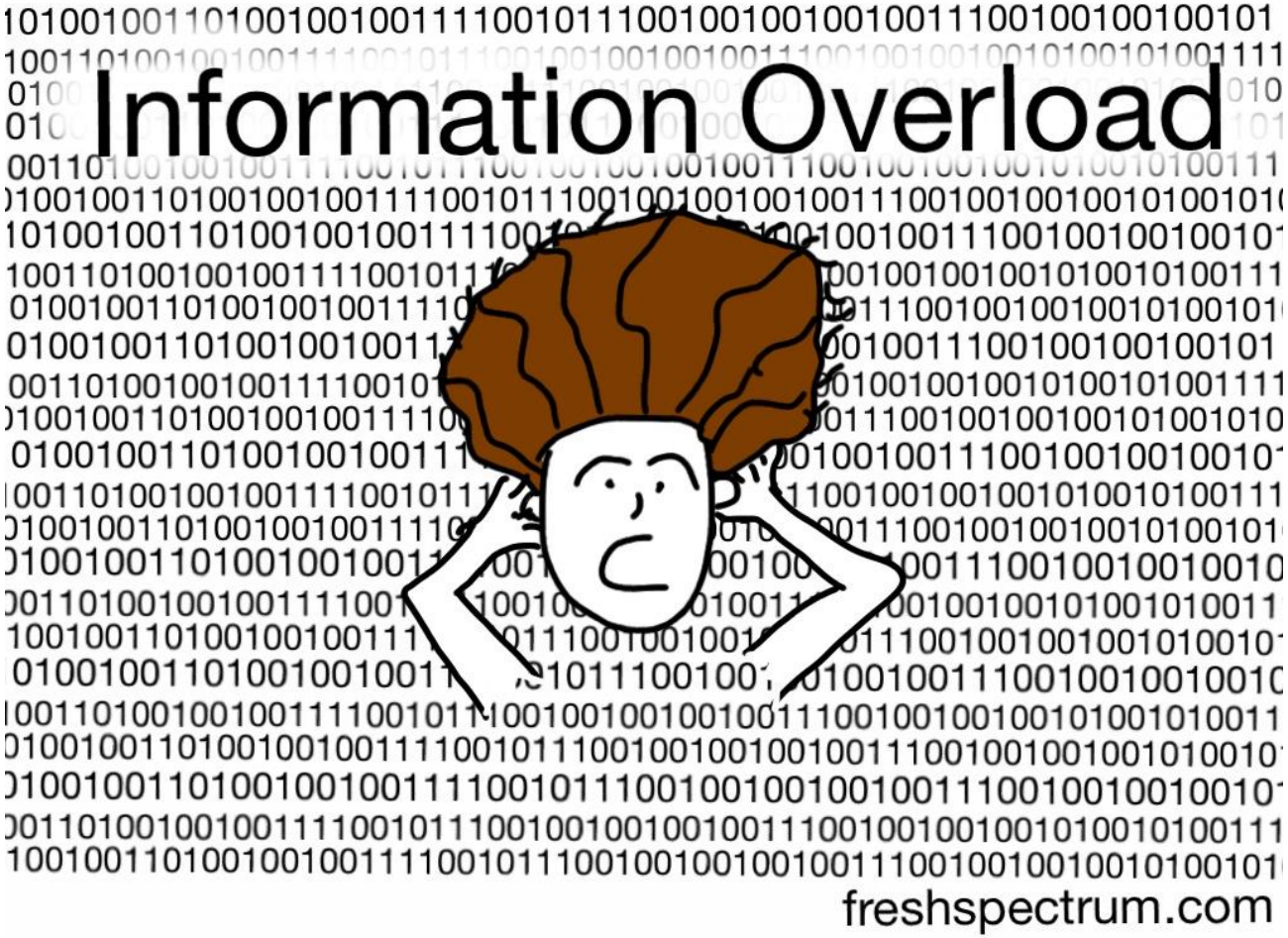
Mining along the Ankobra River in Ghana

There is a 13% loss in dense vegetation from 2000 to 2017. These illegal mines have a significant impact to land and water resources.

Introduction to API Highways

Problem

- Data often not available in a form ideal for application/visualization development
- APIs, documentation, examples are scattered and/or hidden



Global Partnership for Sustainable Development Data

Solution

Developer-friendly, application-ready

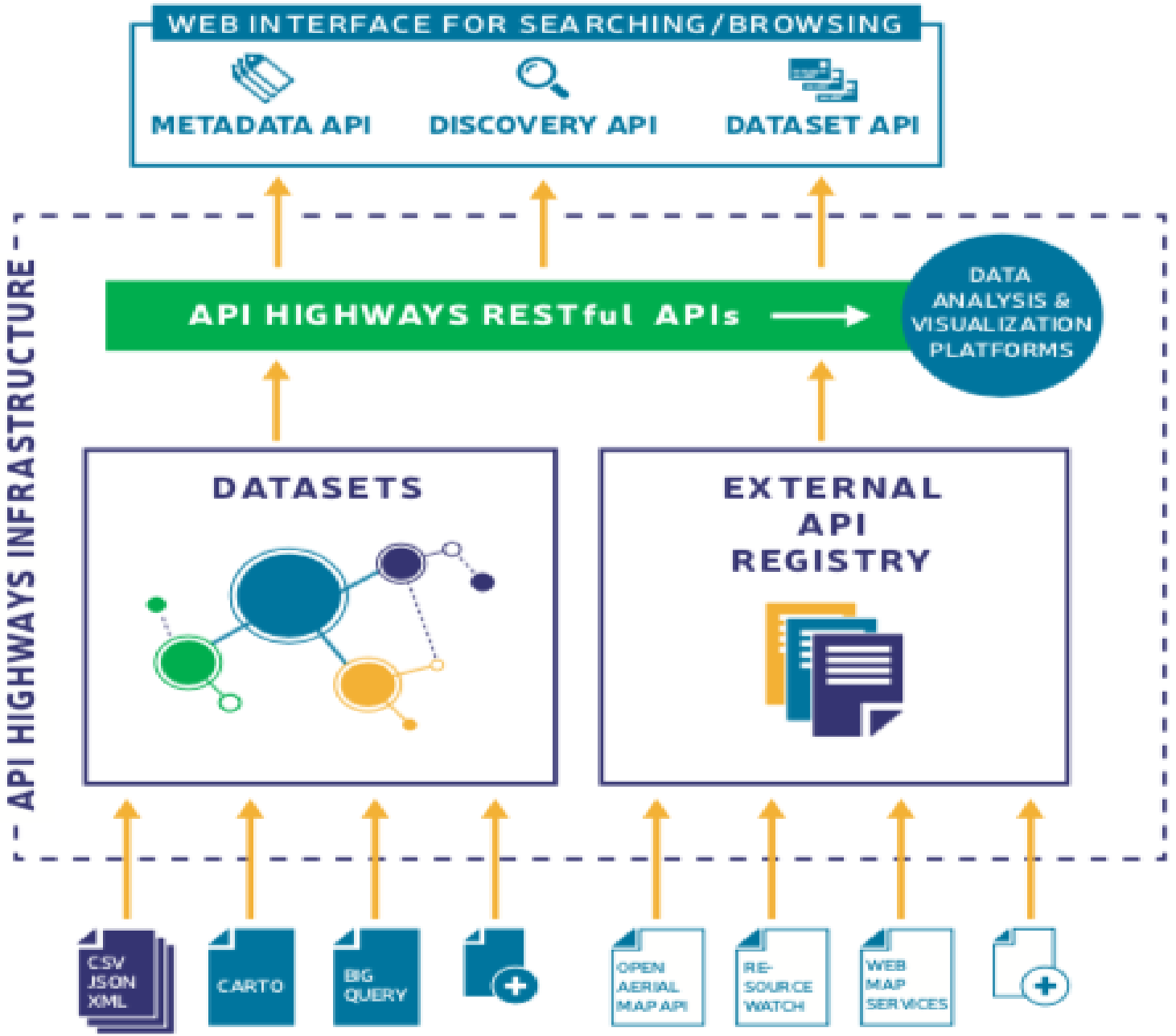
An open source developer community

Data discoverability

Use case development

Global Partnership for Sustainable Development Data

Infrastructure Overview



Global Partnership for Sustainable Development Data

Data Resource : GPSDD API Highway Infrastructure

- https://apihighways.org/

  - ✓ https://apihighways.org/data-sets

  We have curated 3000 datasets unto the infrastructure

Global Partnership for Sustainable Development Data

Opportunities from Mobile Big Data

Data from the mobile sector constitutes a growing new source of data: Examples: tourism statistics; population statistics (population density and mapping); urban population and inter-city migration; commuting statistics; traffic flow Stat and planning; tracking unemployment etc.

Mobile Big Data could assist in filling data gaps in thematic areas: to monitor SDG goals where data is scarce: E.g. Mapping socioeconomic status by analyzing airtime credit and mobile phone datasets etc.

Open algorithms: A new paradigm for using private data for social good

By Thomas Roca, Emmanuel Letouzé | 18 July 2016

devex
Do Good. Do It Well.™

AFD

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA

SUSTAINABLE DEVELOPMENT
SOLUTIONS NETWORK
A GLOBAL INITIATIVE FOR THE UNITED NATIONS

DANE
Para tomar decisiones

DNP Departamento Nacional de Planeación

ANSD
Agence Nationale de la Statistique et de la Démographie

Telefónica

sonatel

Pilot

Interest

Tariq Khokhar @tkb · 2h

Shoutout to the OPAL project - "bring the algorithm to the data" - more at: opalproject.org #UNDataForum

Open Algorithms (OPAL)

OPAL

bring the code to the data

OPAL is a techno-political platform and an environment that aims to leverage data for development by unlocking the potential of private data for public good in a privacy conscientious, inclusive scalable, socially and economically sustainable manner.

DATA POP ALLIANCE

Imperial College London

mit media lab

orange™

WORLD ECONOMIC FORUM

Global Partnership for Sustainable Development Data

Some Use Cases: Mobile Data for Sustainable Dev

- Use of mobile data in the Health sector: disease tracking and data on health.
- Mobile Data can be harnessed to provide supplementary real time data on the population's access to basic social services/improvement of services: e.g. people accessing Health, Water and Sanitation, Education, Agricultural extension services etc. across the country.
- Mobile data can provide the NSO supplementary real time, up-to-date demographic data and proxy poverty index data of the population.
- ICT surveys can help among others, to provide a country with regular ICT relevant SDG indicators which can be got from Mobile Network Operators.

Global Partnership for Sustainable Development Data

Mobile Big Data Access Challenges

Challenge	Description
Access: Legal and Regulatory	• Huge Legal and Regulatory constraints (lack of clarity on responsible data use; data protection issues) • Lack of data sharing protocols or policies
Technical	• Most NSOs lack the infrastructure to store and tools to analyze such data • Lack of standards (collection method, validation etc.)
Capacity/Institutional	• Poor engagement of the private sector by Govt/NSOs • Lack of capacity in processing and analyzing such large unstructured data sets (reliance on 3rd party) • Location of processing and methodology

Global Partnership for Sustainable Development Data

What Do We Expect from this Community

- Establish a sustainable Community of Practice on AI and Data Science for the SDGs that will proactively partner with the Govt. (NBS and key MDAs).
- Collaborate and partner with the National Statistical Office, to explore how the power and capabilities of AI and advanced Data science can better support the National Statistical System.
- Help create greater awareness on Data and its effective and responsible use for policy and development planning at all levels of governance.
- Actively promote data story-telling skills and data journalism using various innovative and compelling approaches.

Thank You!

- Contact Information:

Victor Ohuruogu

Program Manager,

Global Partnership for Sustainable Development Data

vohuruogu@data4sdgs.org

http://www.data4sdgs.org/

GLOBAL PARTNERSHIP
FOR SUSTAINABLE DEVELOPMENT DATA