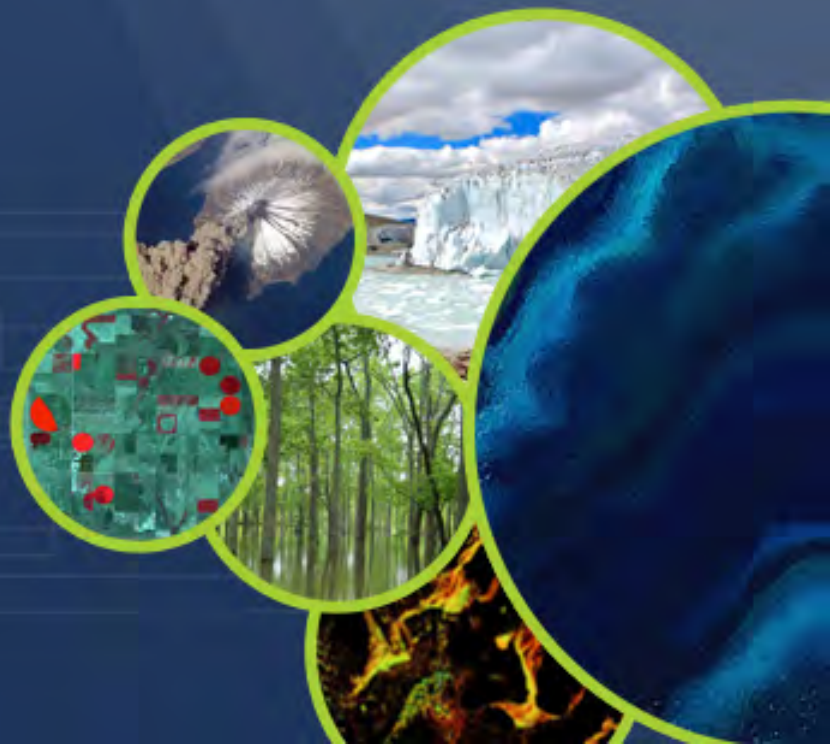
The Africa Regional Data Cube (ARDC)

A summary of the implementation plans

Africa Regional Data Cube
Training Workshop
Nairobi, Kenya
May 9, 2018 (Day #1)

Dr. Brian Killough
CEOS Systems Engineering Office
NASA Langley Research Center



Africa Regional Data Cube (ARDC)

Ghana
Kenya
Senegal
Sierra Leone
Tanzania

- Managed by Strathmore University (Nairobi, Kenya)
- Supported by Amazon Web Services (2-year grant)

Key ARDC Stakeholders

- GPSDD ... Primary lead and initiator of the ARDC
- Strathmore University ... Responsible for core ARDC operations and data management
- NASA (CEOS Systems Engineering Office) and Analytical Mechanics Associates (AMA) ... Technical lead for initial deployment, training and implementation
- Amazon Web Services (AWS) ... Provider of 2-year cloud computing grant
- GEO ... Promotion of the ARDC among African users and relevant global groups

CEOS GEO GROUP ON EARTH OBSERVATIONS Global Partnership for Sustainable Development Data sbs Strathmore Business School aws

Potential ARDC Stakeholders

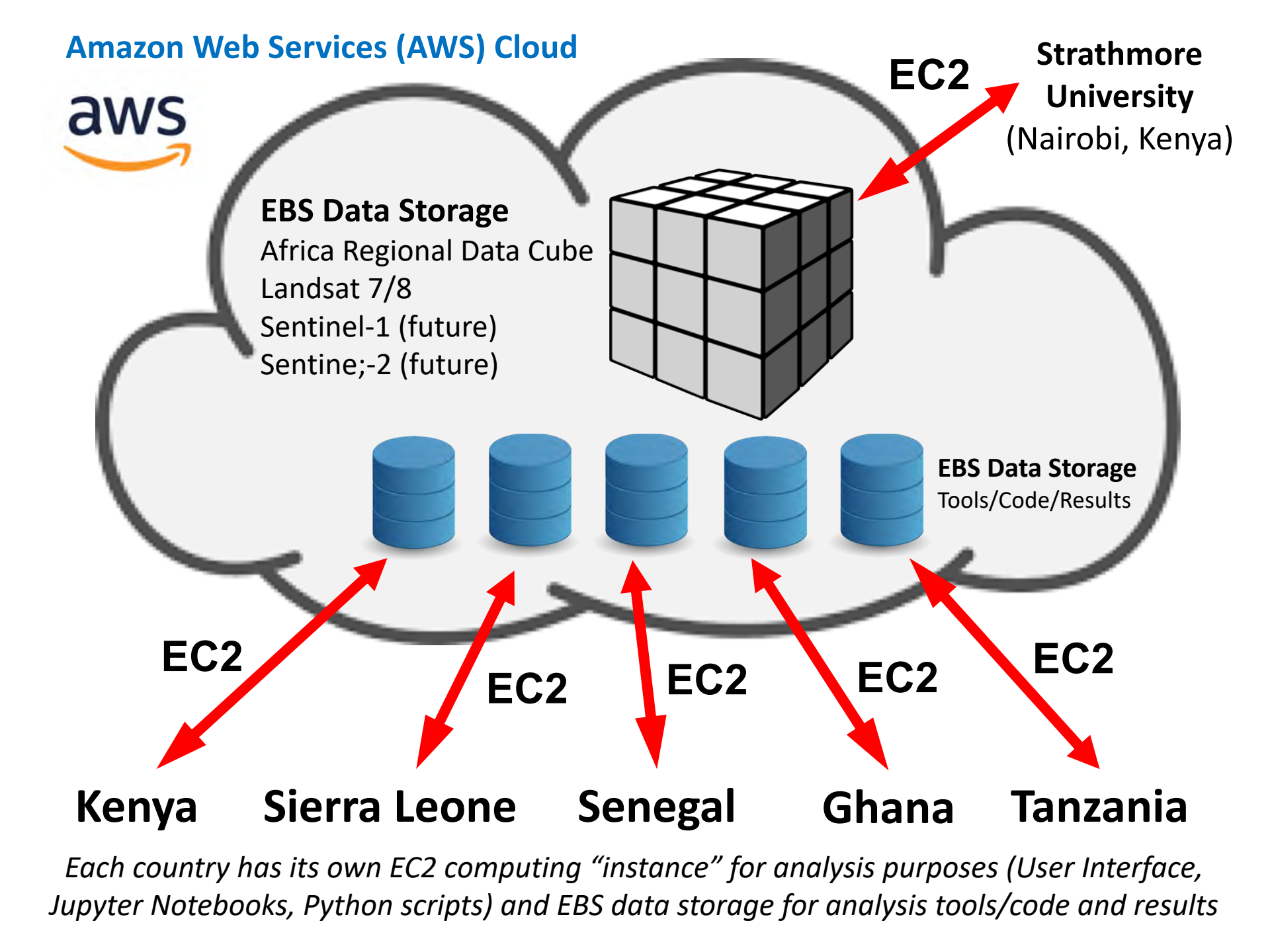
- UK-Catapult ... Potential to support Sentinel-2 data acquition, processing (via ARCSI), and ingestion in the cube + capacity building for Strathmore
- UK-Rhea ... Potential to merge their Uganda Data Cube with the ARDC and add application algorithms and capacity building for users
- Radiant Earth ... Interested in using the cube to develop and test algorithms for the region with a focus on agriculture
- UNEP ... Chief Scientist of UNEP has Headquarters in Nairobi)
- AfriGEOSS
- AfriGAM

Strathmore Roles and Responsibilities

- ARDC core data, algorithms, and user interface management
- Managing the use of Amazon credits
- Adding new satellite data to the ARDC (moving toward automation)
- Branding and advertisement of the ARDC throughout Africa
- Training for local users
- Managing ARDC accessibility for researchers and students
- Building an educational degree centered on the use of the ARDC
- Evaluation of sustainable operation options including cloud-based deployment (e.g. Amazon), local deployment (e.g. Strathmore computing systems), and network enhancement (e.g. Liquid Telecom).

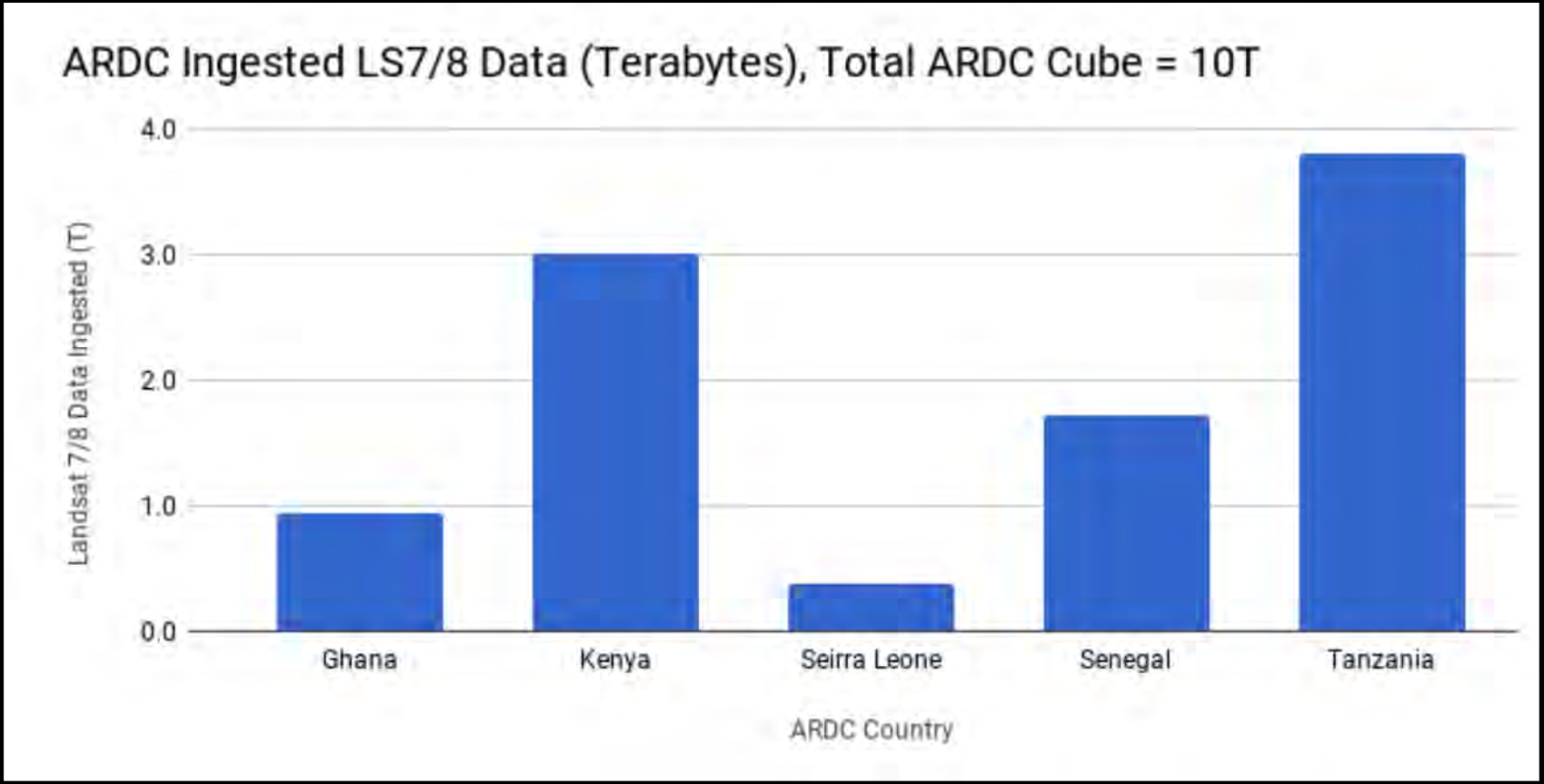
How does Cloud Computing Work?

- A computing “cloud” is a global network of connected servers that allow you to store and access data as well as use computing power for analyses. You will never know where the computers are located but you will notice the performance is outstanding! The ARDC will be using the Amazon Web Services (AWS) cloud.
- Data Storage ... All of the data (data cubes, algorithms, software) is stored in the “cloud”. The ARDC will use the AWS Elastic Block Storage (EBS) method for storing data, as it works very well for rapid computing.
- Data Analysis ... Access to data for analyses is done through “computing instances”. The ARDC will use the AWS Elastic Cloud Compute (EC2) method for accessing and analysing data.



Each country has its own EC2 computing "instance" for analysis purposes (User Interface, Jupyter Notebooks, Python scripts) and EBS data storage for analysis tools/code and results

Data Cube Storage



The ARDC cloud implementation

- The ARDC will be using the Amazon Web Services (AWS) cloud.

- Data Storage ... One large EBS unit is used by Strathmore to store the core cube and software and each country is given smaller EBS units to store data analysis results or algorithms.
- Data Analysis ... One large EC2 is used by Strathmore to manage the core cube and software and each country is given their own EC2 instances for performing analyses and operating their user interface.
- EC2 computing instances are always “ON” and do not incur more costs with increased number of users. The performance will slow down with many users running processes at the same time.
- Countries (governments) are responsible for access control. They can decide to limit access or open access to anyone. Performance and data storage will be the issues with a large number of users.
- If given access, anyone can purchase an EC2 “instance” and connect to the ARDC. For example, the NASA SEO will use our own EC2 resources to connect to the ARDC for testing.

Enjoy your new Data Cube!