

# DEMYSTIFYING GEOSPATIAL BIG DATA, GEOSPATIAL DATA SCIENCE, GEOAI & GEO-ANALYTICS

27 JUNE 2023

StatsTalk-Africa Series: Demystifying Big Data and Official Statistics in Africa

Aster Denekew Yilma  
Geospatial Information Officer  
Geospatial Information Management Section,  
African Centre for Statistics,  
United Nations Economic Commission for  
Africa

# OUTLINE

- What is Geospatial Big Data?
  - Characteristics of Geospatial Big Data
  - Sources of Geospatial Big Data
  - How big is Geospatial Big data (volume)?
  - Geospatial Big Data Platforms
- What is Geospatial Data Science?
  - GeoAI and Big data analytics
  - GeoAI and Geospatial Data Science Tools (AI / ML)
  - Applications of Geospatial Big Data and GeoAI
- Role of ECA in promoting Big Data and AI

# WHAT IS GEOSPATIAL BIG DATA?

---

Geospatial Big Data: subset of Big Data, having spatial component.

Big data incorporating a spatial component: “GEOSPATIAL BIG DATA” becomes topic for research and development community: academia, industry, government and international organizations including UN

---

A significant portion of big data is geospatial data, and the size of such data is growing rapidly

---

Managing and processing geospatial big data to help decision-making becomes an important scientific and societal issue

# CHARACTERISTICS OF GEOSPATIAL BIG DATA



Geospatial data has always been **big data**.



*Geospatial Big data* typically refers to spatial data sets exceeding capacity of standard computing systems.



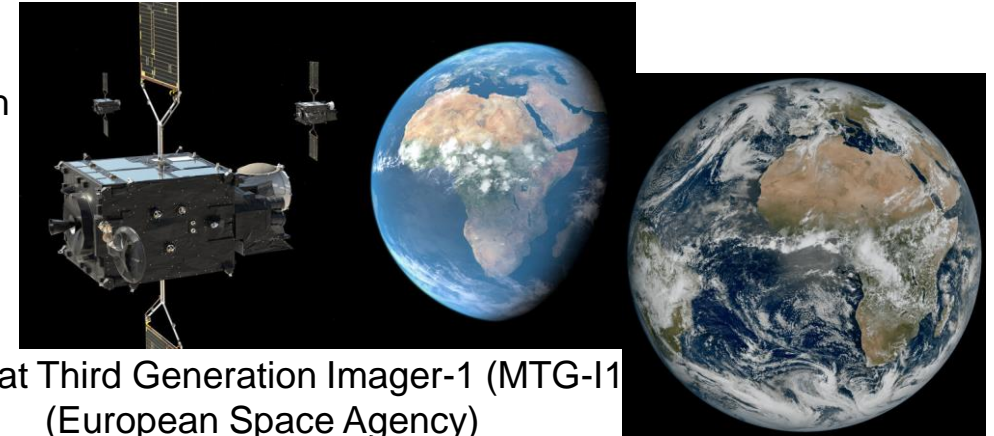
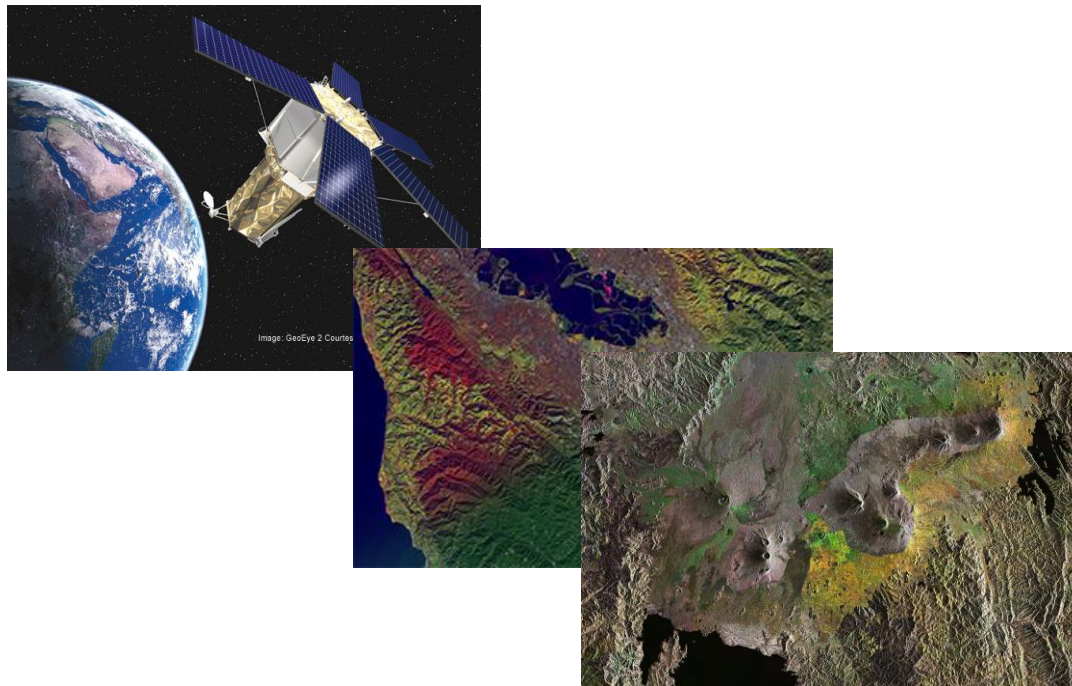
Geospatial Big Data is characterized by the **5V**

Volume  
Velocity  
Variety  
Veracity  
Value

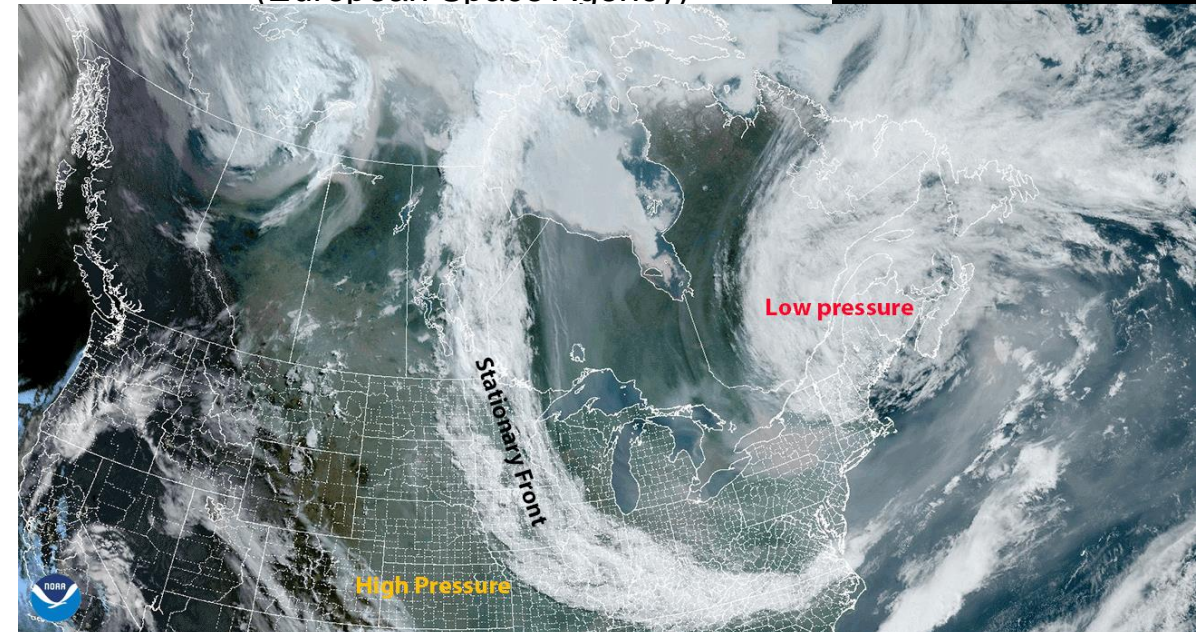


# SOURCES OF GEOSPATIAL BIG DATA

- **Earth Observation Satellites:** Satellites are the most extensively used Earth observation technology for gathering geospatial information.
- Satellite imagery are sources of geospatial big data
- Weather satellites collect large amount of weather and climate data



Meteosat Third Generation Imager-1 (MTG-I1)  
(European Space Agency)



NOAA satellites are monitoring Wild fires across Canada, June 2023

# SOURCES OF GEOSPATIAL BIG DATA

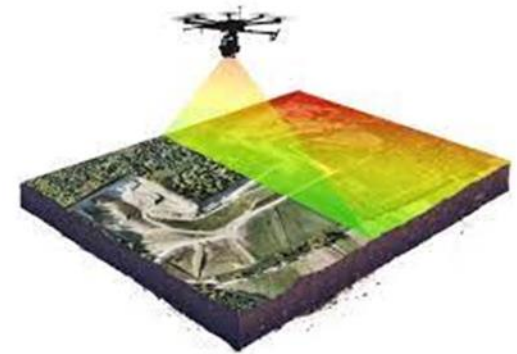
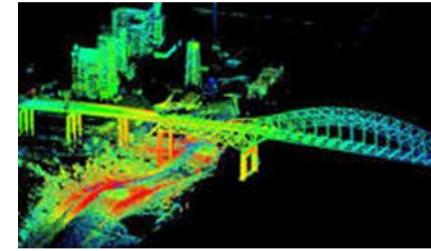
- **GPS (Global Positioning System): provide positioning and navigation information**
  - One of the Global Navigation Satellite Systems (GNSS): Galileo, GLONASS, BeiDou,
  - GPS technology is widely utilized for applications such as mapping, surveying, tracking, precision agriculture, etc.
  - Embedded in mobile devices, vehicles (navigation)



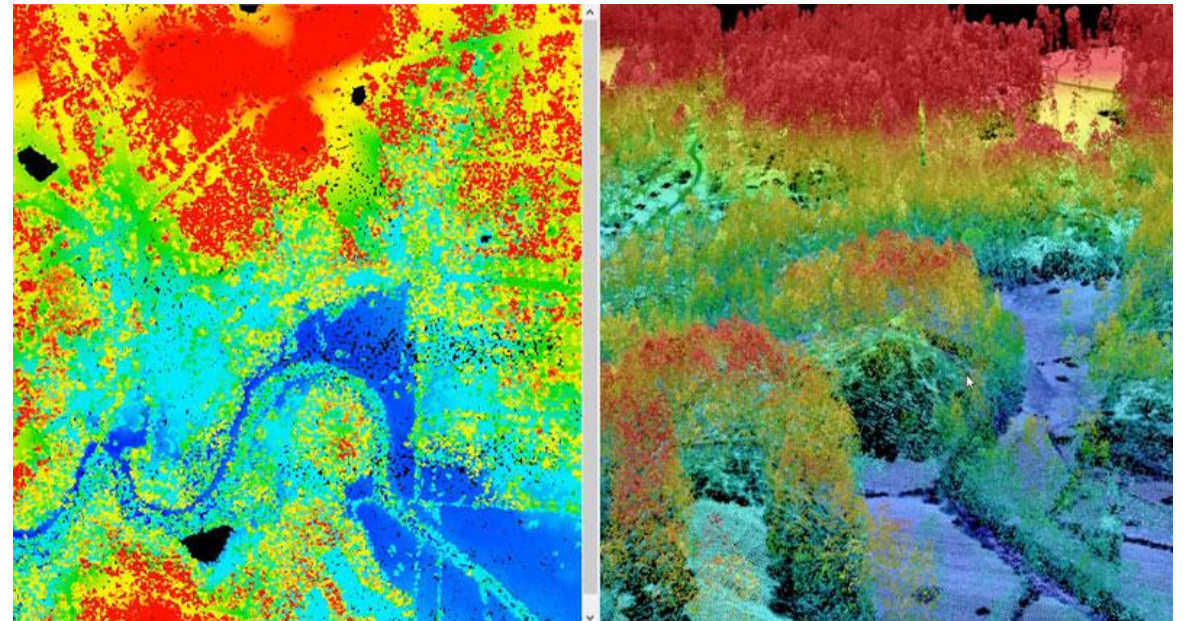


# SOURCES OF GEOSPATIAL BIG DATA

- **LiDAR (Light Detection and Ranging):** is a remote sensing method used to map the surface of the Earth using laser light to measure ranges (variable distances) to the Earth.
  - Point cloud data are acquired from LiDAR
  - Used to generate elevation/terrain models, 3D City model, mapping infrastructure (e.g. power lines)



LiDAR (Light Detection and Ranging)



Point cloud datasets

# SOURCES OF GEOSPATIAL BIG DATA

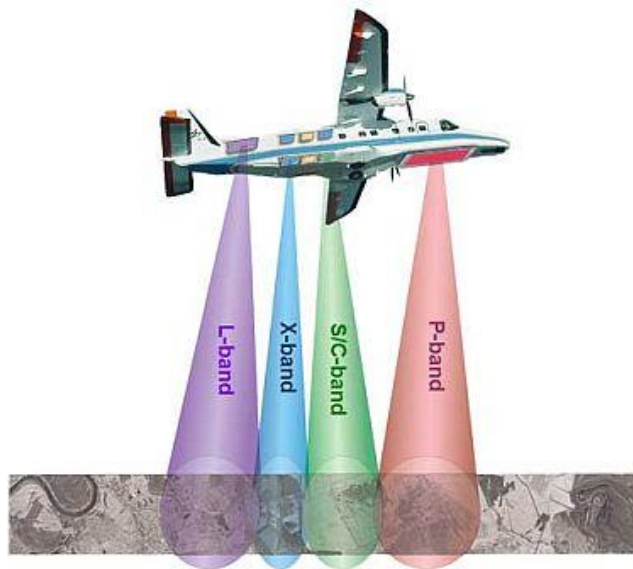
- **UAVs (Unmanned Aerial Vehicles):** UAVs, better known as drones, are utilized for capturing high-resolution aerial images of small areas, such as buildings or farm fields.
- Full motion videos or high-resolution photogrammetric images
- Object detection, change detection and real-time tracking at finer scales
- Used for applications such as agriculture, forestry, and mining, environmental, urban, and hazards assessment applications ...





# SOURCES OF GEOSPATIAL BIG DATA

- Radar system penetrate clouds; Clouds obscure satellite imagery
- Weather radars used for weather detection and early warning
- Airborne radar systems used for mapping large area



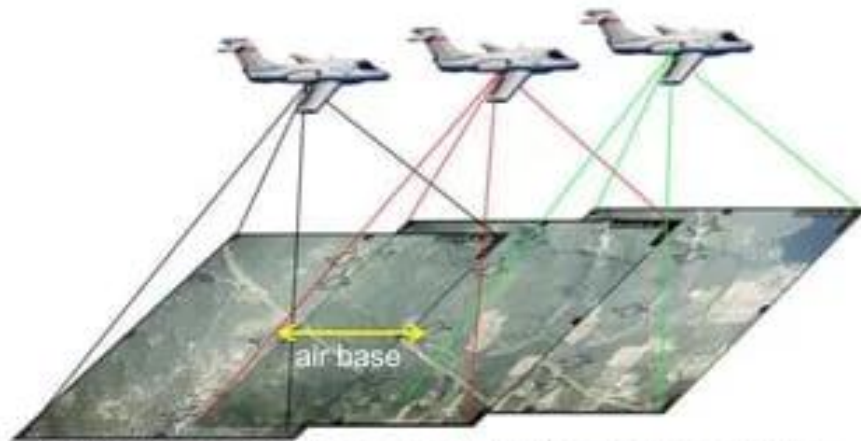
Airborne SAR (Synthetic Aperture Radar)



High Resolution Airborne SAR Image

# SOURCES OF GEOSPATIAL BIG DATA

**Aerial photography:** Aerial photography captures photographs of the Earth's surface using high-resolution cameras mounted on a plane or helicopter.



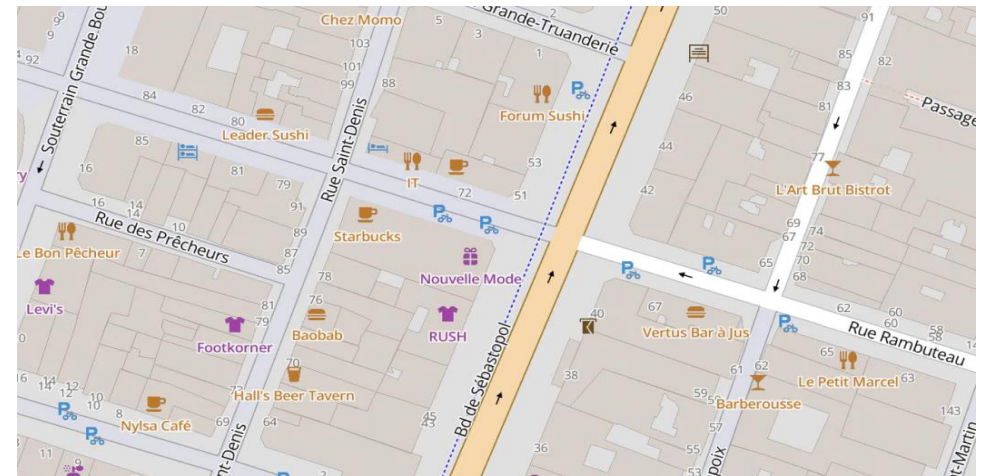
Distance between principal point of adjacent photographs is known as the "air base"



Aerial photography

# SOURCES OF GEOSPATIAL BIG DATA

- GPS enabled devices (in vehicles, smart phones, wildlife collars);
- IoT and distributed Sensors (monitoring air quality, water chemistry, energy consumption); Smart devices,
- Crowd sourcing/Volunteered Geographic Information (VGI): producing volunteered geospatial data (e.g. Open Street map),





# HOW BIG IS GEOSPATIAL BIG DATA (VOLUME)?



Geospatial Big data (volume): Zettabytes of data are collected from different sources of geospatial data: GPS, EO satellites, UAVs, Drones, LiDAR, IoT, etc.



Earth Observation Satellites in orbit: High-resolution; Medium-resolution and Low-resolution collect hundreds of petabytes of data.



1 Petabyte = 1,000 terabyte

1 Laptop = 1 terabyte

1, 000 Laptops = 1 Petabyte

# HOW BIG IS GEOSPATIAL BIG DATA (VOLUME)?

## High Resolution Satellites

WorldView-3 (30cm)



(Launched in 2014)

WorldView-1&2 (0.5m)



(Launched in 2007/2009)

GeoEye-1 (0.5m)



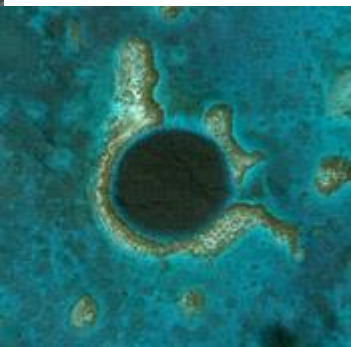
(Launched in 2008)

QuickBird (1m)



(Launched in 2001)

IKONOS (1m)



(Launched in 1999)

SPOT-7 (1.5m)

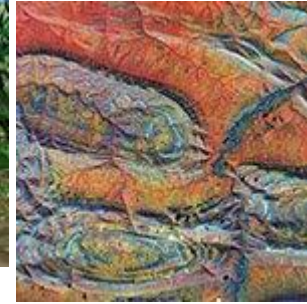


## Medium Resolution Satellites

SPOT-5 (2.5m-5m)



RapidEye (5m)



Sentinel-2A (10m)



ASTER (15 -90m)  
(Terra)



Landsat 7 ETM+ (15m)

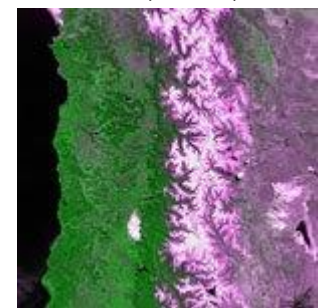


Landsat 8 (15m)



Landsat 1, 2, 3, 4, 5, 6; first Launched in 1972

CBERS-2 (20m)



# GEOSPATIAL BIG DATA PLATFORMS



Due to exponential increase in geospatial big data, It became difficult to store, manage, process, analyze, visualize and extract useful information from geospatial big data using traditional approaches on local machines



The capability of high-performance computing is being required, for processing and analyzing geospatial content



Number of Geospatial Big Data platforms have been developed for processing geospatial big data so that users can implement big data analytics software very easily on a distributed, cloud computing platform, e.g. Amazon Web Services (AWS)



# GEOSPATIAL BIG DATA PLATFORMS

- **DigitalGlobe's Geospatial Big Data Platform (GBDX)** cloud-based platform, Launched in 2015, provides direct access to immense quantities of satellite imagery in the cloud, as well as a number of added services and machine learning and artificial intelligence tools to extract meaningful insights from the image library
  - Feature extraction
  - Object detection
  - Change detection
- integrates the GBDX platform with Esri's (Industry leading GIS software provider) **ArcGIS Enterprise** platform and Harris Corporation's (image analysis Software developer) **ENVI** remote sensing analytics portfolio
- Esri users can get access to more than 100 petabytes of highest-quality commercial satellite imagery and most advanced analytic and machine learning tools
- Esri subscribing customers licensed for **ArcGIS Image Server** have access

# GEOSPATIAL BIG DATA PLATFORMS

- **Esri's ArcGIS Online:** Cloud-based software to create and share interactive web maps
  - hosts huge amount of authoritative location-based data and map layers
  - Integrates web app builders, and data science models
  - Allows hosting, streaming, and analysis of Imagery and raster collection and point-clouds/LiDAR data in the cloud computing environment through the **ArcGIS Image for ArcGIS Online**
- **ArcGIS Living Atlas of the World:** collection of big data hosted in Esri ArcGIS Online platform; provides access to massive ready-to-use maps and datasets, including base maps and imagery

# WHAT IS GEOSPATIAL DATA SCIENCE?

➤ **Geospatial Data Science** is the subset of Data Science which focuses on the unique characteristics of spatial data, using modeling to know where and why things happen

Major Knowledge Domain in Data Science:

- **Computer Science**
- **Mathematics**
- **Statistics**
- Artificial Intelligence/**Machine Learning**
- **Geospatial Information Science & Technology/**  
Data Visualization/
- Domain Expertise, ...

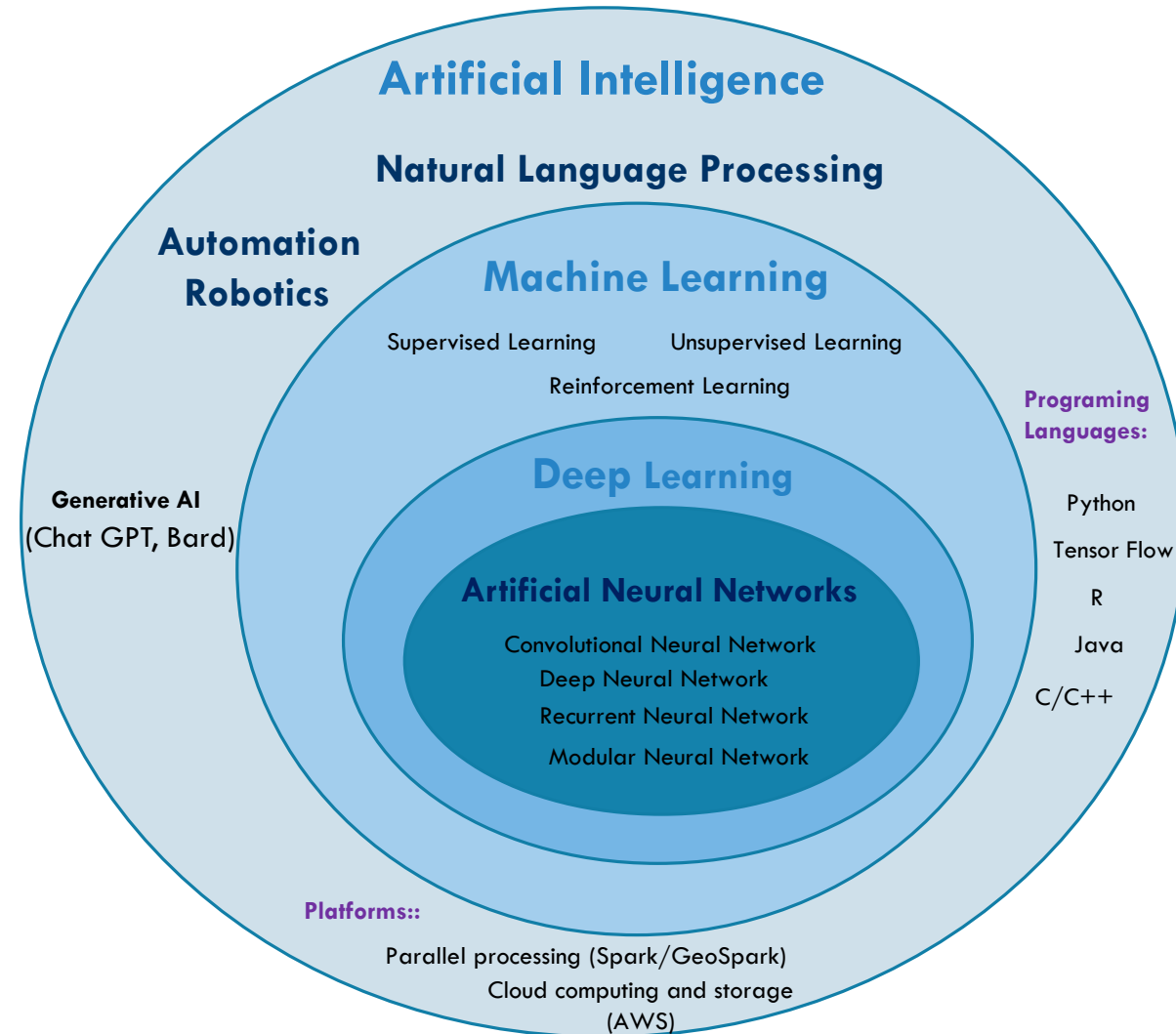
\*No single person or single domain can be perfect data scientist; we need to build teams





# WHAT IS GEOSPATIAL DATA SCIENCE

- **Artificial Intelligence (AI)** is computer software that mimics the ways that humans think in order to perform complex tasks, such as analyzing, reasoning, and learning
- **Machine learning (ML)** is a subset of AI that uses algorithms trained on data to produce adaptable models that can perform a variety of complex tasks such as analyzing big data
  - ML algorithms analyzes data and discovers patterns by learning from training datasets
  - Machines can then use the patterns to recognize unknown instances
- **Deep learning:** A subset of ML, in which artificial neural networks (AANs) that mimic the human brain are used to perform more complex reasoning tasks without human intervention



# WHAT IS GEOSPATIAL ARTIFICIAL INTELLIGENCE (GEOAI)?

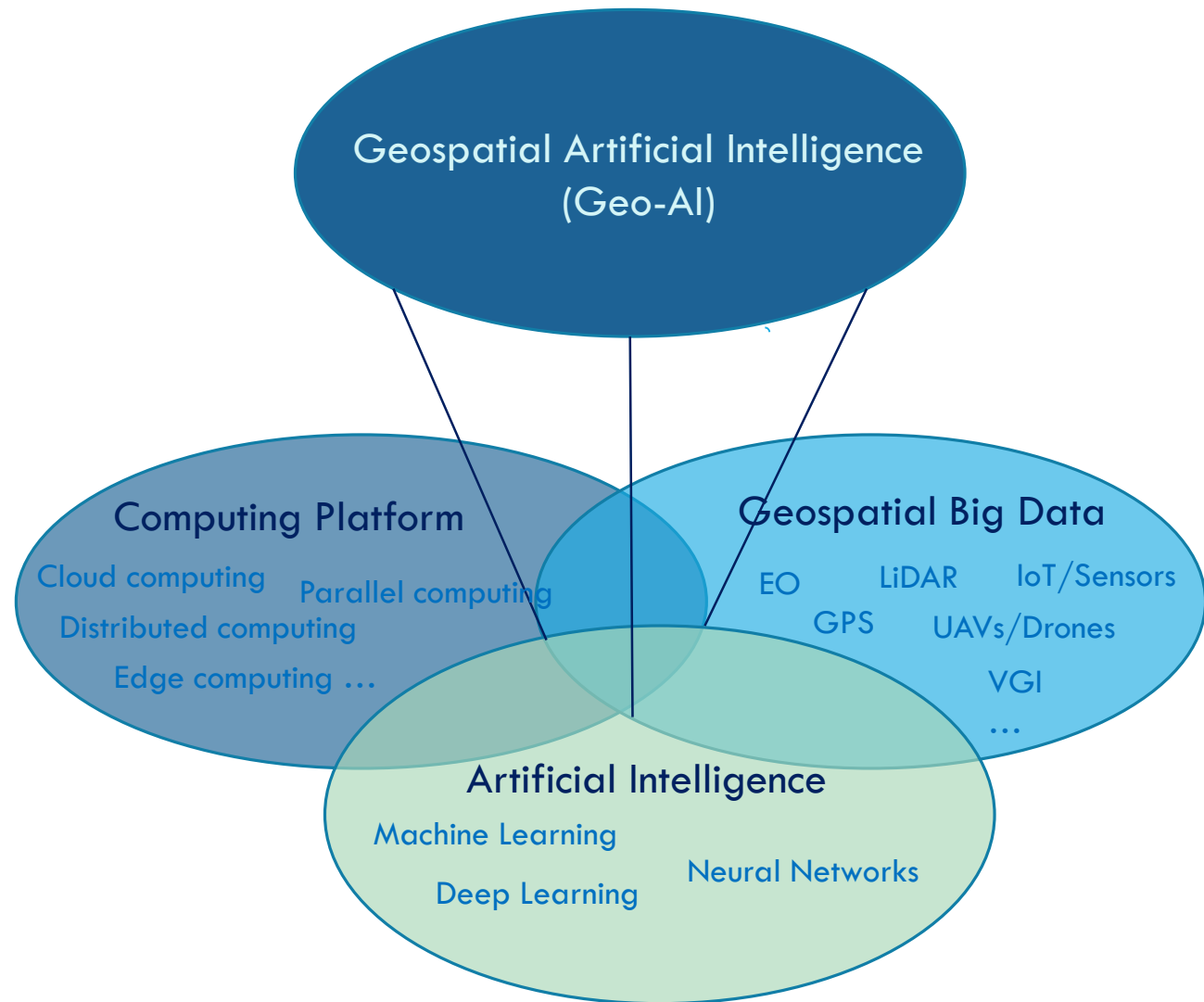
**GEO-AI:** A subset of AI that deals with geospatial datasets, where data possess spatial and/or temporal characteristics and helps solving spatial problems

The integration of geospatial and AI, or Geo-AI, provides novel approaches for addressing a variety of problems in the natural environment and human society



# WHAT IS GEOSPATIAL ARTIFICIAL INTELLIGENCE (GEOAI)?

**GEO-AI: WHERE  
MACHINE LEARNING  
AND BIG DATA  
CONVERGE IN  
GEOSPATIAL  
SCIENCE**



# GEOAI AND BIG DATA ANALYTICS

## Geo-AI and Geospatial Big Data Analytics:

- Transform massive spatial data into manageable information;
- Use distributed architecture/distributed computing to analyze and display large volumes of data;
- Ingest real-time data from sensors, social media feeds, and IoT systems; Analyze real-time data as it streams and trigger alerts as assets move or change.
- Distributed processing and deep learning for object detection, classification, terrain analysis, and change detection on a global scale from massive satellite imagery collections



# GEOAI AND GEOSPATIAL BIG DATA ANALYTICS TOOLS

- Data science tools and methods and Artificial intelligence / machine learning algorithms, such as deep convolutional neural networks, are used with geospatial big data
- Solve large-scale geospatial problems efficiently by using geospatial cloud computing platforms.
- E.g. Amazon Web Service (AWS) provides analytics, repository and storage, AI and ML services
- DigitalGlobe developed commercial GBDX Notebooks
- GBDX Notebooks leverages the open source Jupyter Notebooks environment to provide access to DigitalGlobe's library of high-resolution imagery, as well as machine learning and remote sensing algorithms.

# GEOAI AND GEOSPATIAL BIG DATA ANALYTICS TOOLS

Artificial Intelligence: Machine Learning and Deep Learning tools and algorithms used by remote sensing software for image analysis for satellite images (e.g. ERDAS Imagine):

## Image Classification

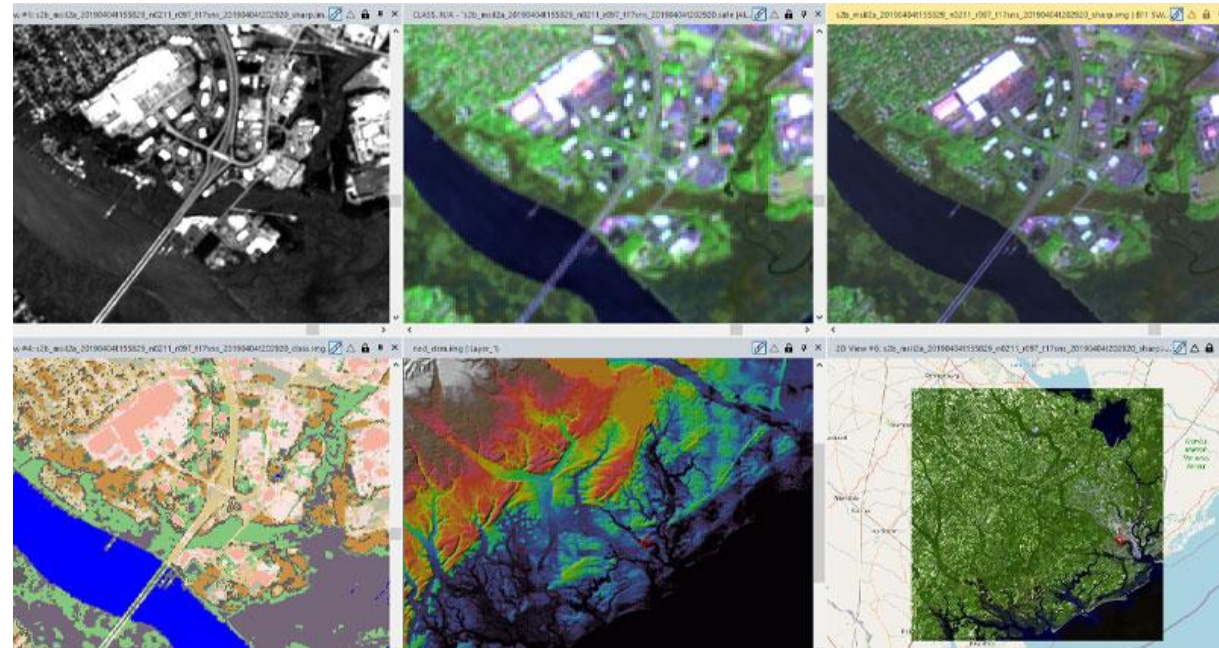
- Uses training data to identify/categorize specific features
- ML algorithms used to classify data according to category

## Object Detection

- Identifies and locates specific objects, (e.g. feature extraction: buildings/building footprint)
- Used to identify changes (building changes: urban expansion; disaster assessment)

## Image Segmentation

- Used to identify and analyze specific areas of change



# GEOAI AND GEOSPATIAL BIG DATA ANALYTICS TOOLS

- GIS software (e.g. Esri's ArcGIS Enterprise) provide AI and ML tools:
- Esri provide solutions for geospatial analytics, ArcGIS Enterprise software
  - **ArcGIS GeoAnalytics Server:** distributed computing framework, **process and analyze big data**, identify patterns, identify hot spots and clusters
  - **ArcGIS GeoEvent Server:** **Analyze streaming data in real time** and Generate geo-enabled alerts; tracks dynamic assets that are constantly changing location (vehicles, aircraft, and vessels); weather and environmental monitoring sensors and IoTs
  - **ArcGIS Image server:** **Serves large collections of imagery** and remote sensing data from different sensors, allowing users to generate useful information products from massive datasets
  - **ArcGIS Notebook Server:** a **spatial data science system** within ArcGIS Enterprise
    - Allow building machine learning and deep learning models within integrated Jupyter notebooks that combine Esri's spatial tools with open-source Python libraries
    - Support machine learning and deep learning workflows such as **object detection, pixel classification, and image segmentation**

# GEOAI AND GEOSPATIAL BIG DATA ANALYTICS TOOLS



## ArcGIS Velocity

### Add-on in **ArcGIS Online**

Analyze real-time and big data:

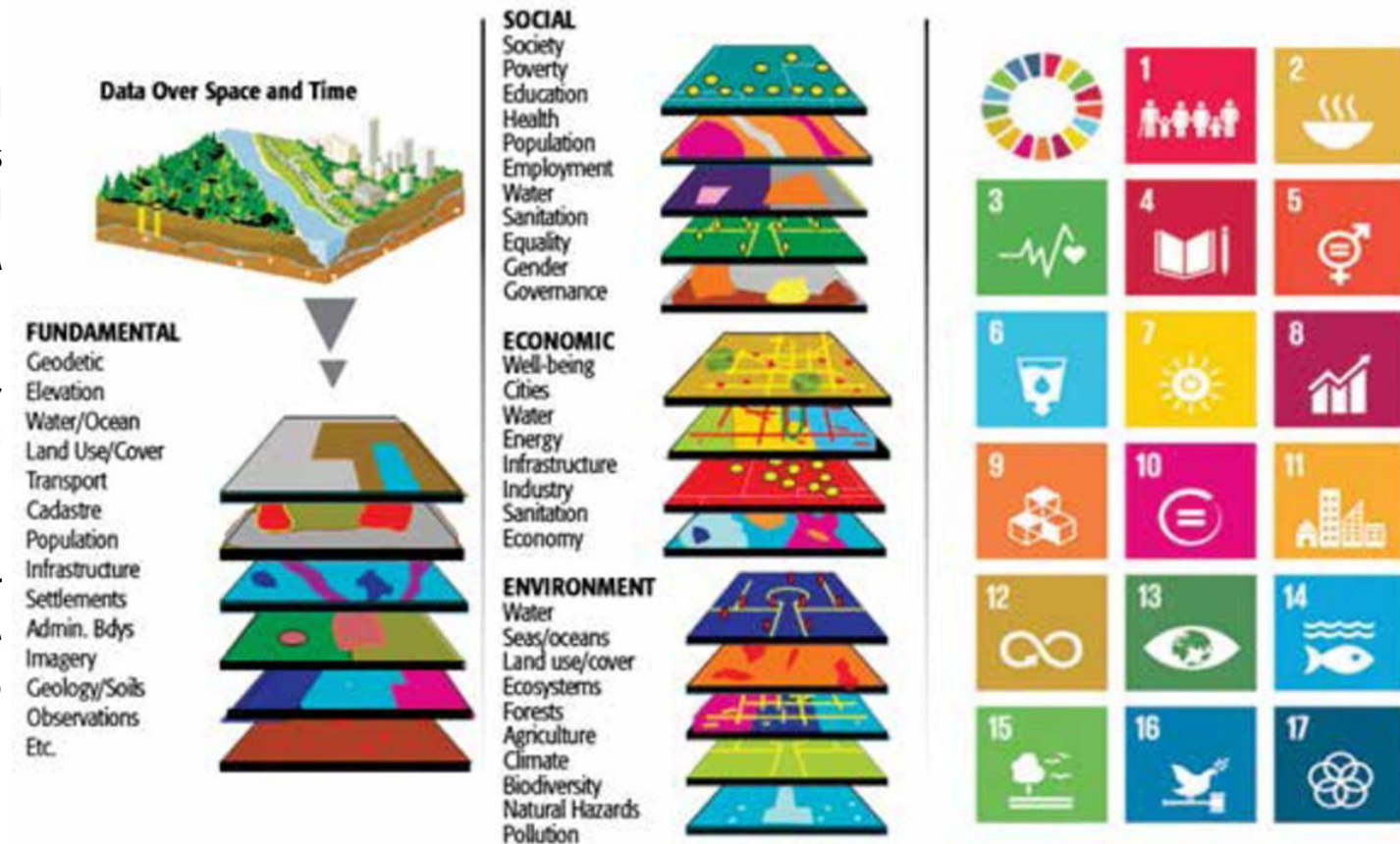
- Perform real-time analysis of streaming data
- enables users to ingest data from the Internet of Things (IoT) platforms
- helps users process, visualize, and analyze real-time data feeds;
- enables the processing of massive volumes of spatial data at very high speeds



# APPLICATIONS OF GEOSPATIAL BIG DATA & DATA SCIENCE

## Monitoring Sustainable Development Goals:

- Many of the SDGs benefit from geospatial big data: 'Inventory of geospatial datasets for monitoring SDGs' identified geospatial datasets relevant for SDGs (14 UNGGIM Geospatial Data Themes)
- Availability of geospatial datasets for monitoring SDGs in Africa: low availability of geospatial datasets
- E.g. ML/Automatic extraction to extract features from satellite images to produce Fundamental Geospatial Datasets will fill gap in availability of geospatial datasets



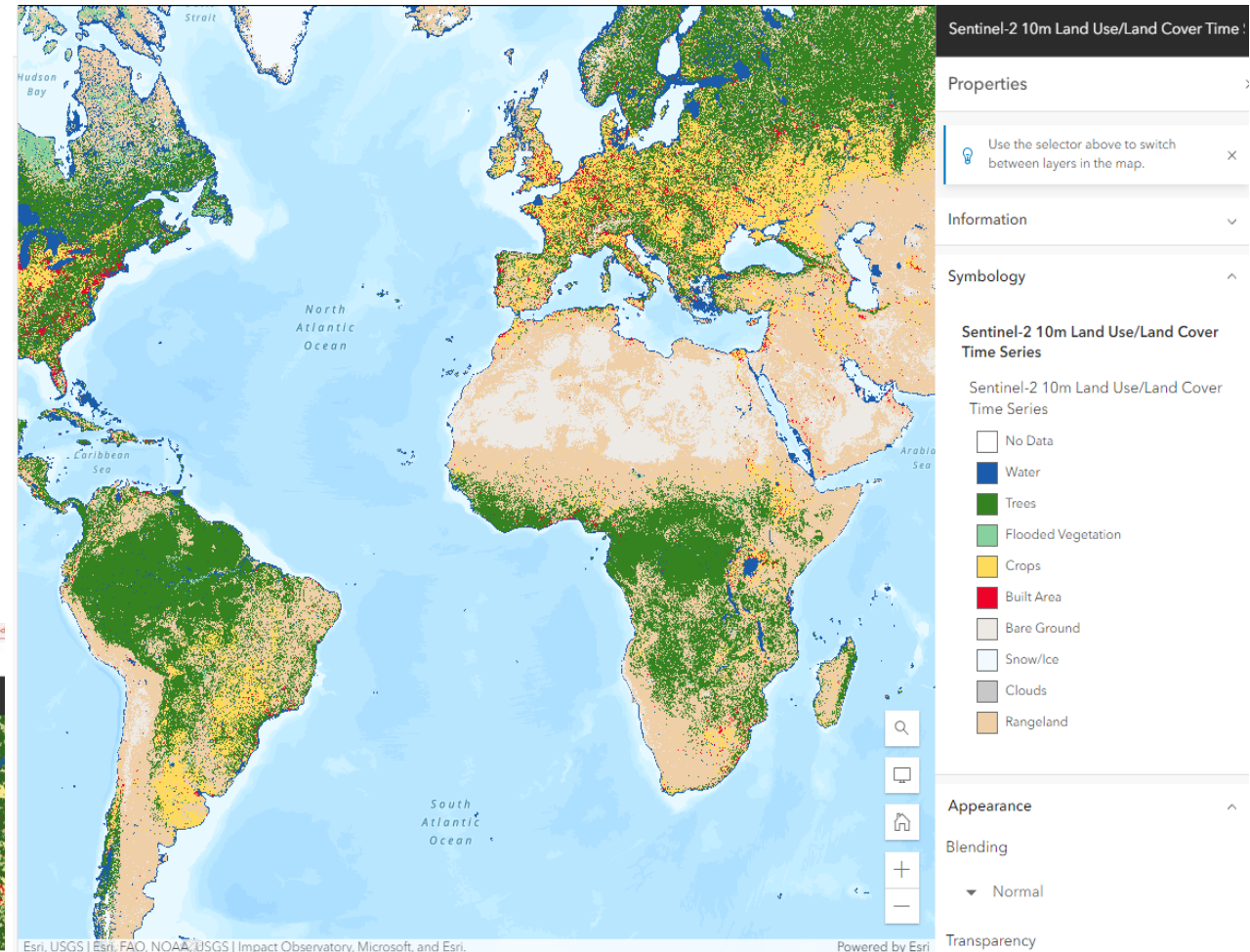
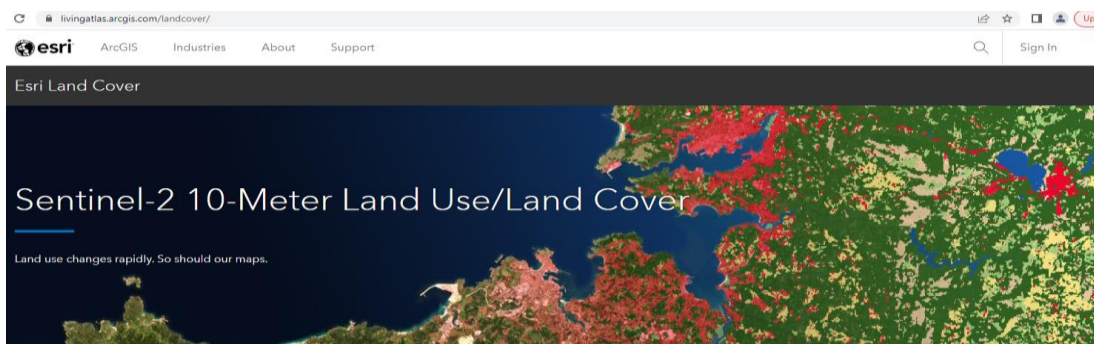
# APPLICATIONS OF GEOSPATIAL BIG DATA & DATA SCIENCE

## Land Use /Land Cover Map:

Esri, with Impact Observatory and Microsoft, developed a Global Land use/Land cover map using Sentinel 2 data (10m) :

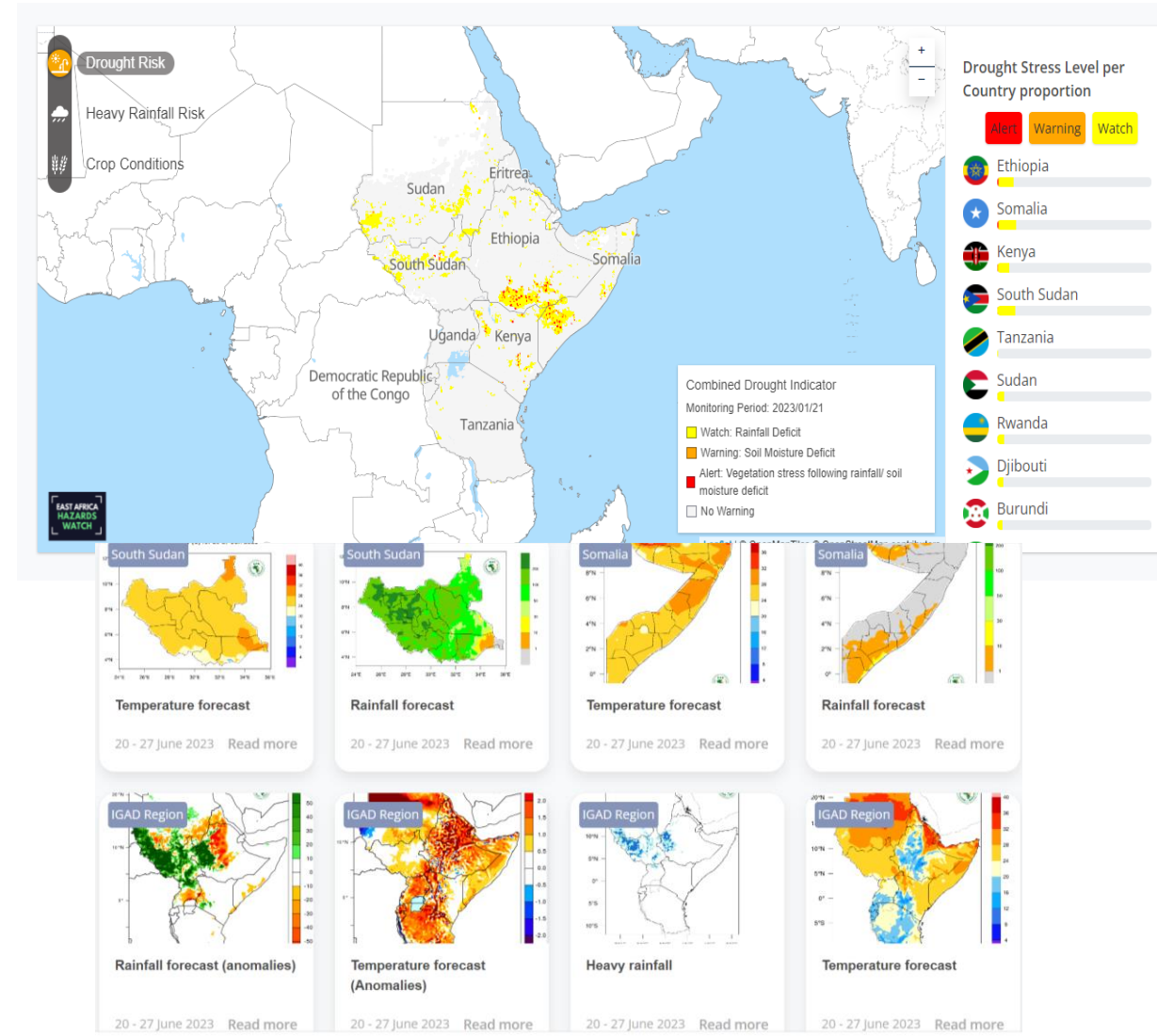
- AI land classification models used to a massive training dataset
- The models were applied to the entire Sentinel-2 scene collection for each year from 2017 to 2022 to produce the maps
- The output classification provides 9 class map of the Earth surface

## [ArcGIS Living Atlas of the World](https://livingatlas.arcgis.com/landcover/)



# APPLICATIONS OF GEOSPATIAL BIG DATA & DATA SCIENCE

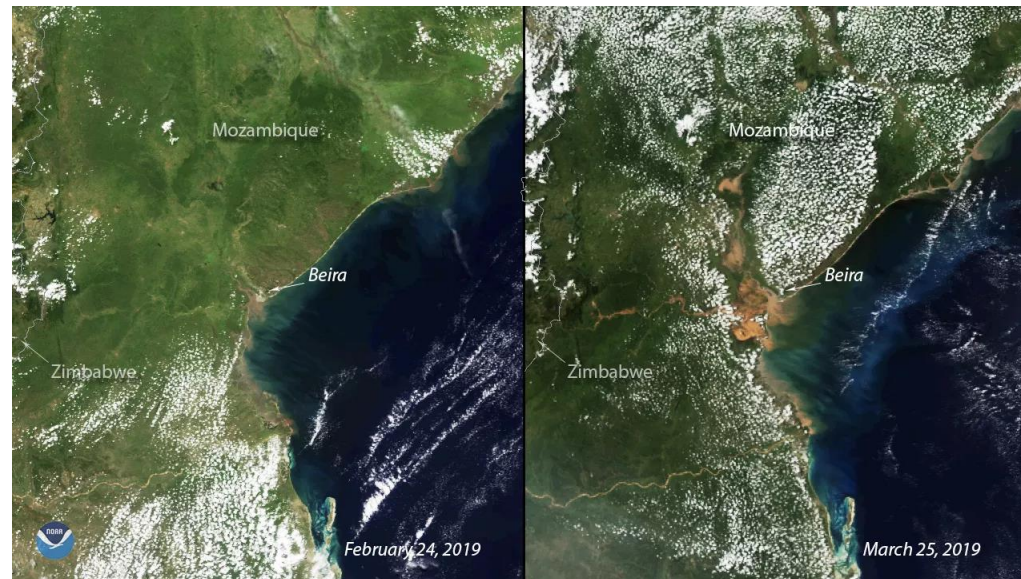
- **Weather Forecasting and Climate Monitoring:** Advances in imaging and weather observations have resulted in a dramatic increase in the accuracy and precision of weather forecasts
- **ICPAC (IGAD Climate Prediction and Application Center):** a center devoted to delivering Climate Services to Eastern Africa
  - Climate monitoring, prediction, and early warning;
  - Dissemination of information and warnings and Advocacy for Climate Action
  - EO-based applications and services to monitor the environment
  - Food Security and Agricultural monitoring in Eastern Africa; promote smart agriculture
  - Drought forecasting





# APPLICATIONS OF GEOSPATIAL BIG DATA & DATA SCIENCE

- **Disaster Risk Reduction and Management:** Disaster Recovery / damage assessment
- Machine learning plays important role in disaster risk reduction (DRR)
  - forecasting of extreme events
  - development of hazard maps
  - detection of events in real time,
- ML used for extraction of building footprint (using before and after disaster images) for damage ;
- Map extent of disaster/floods



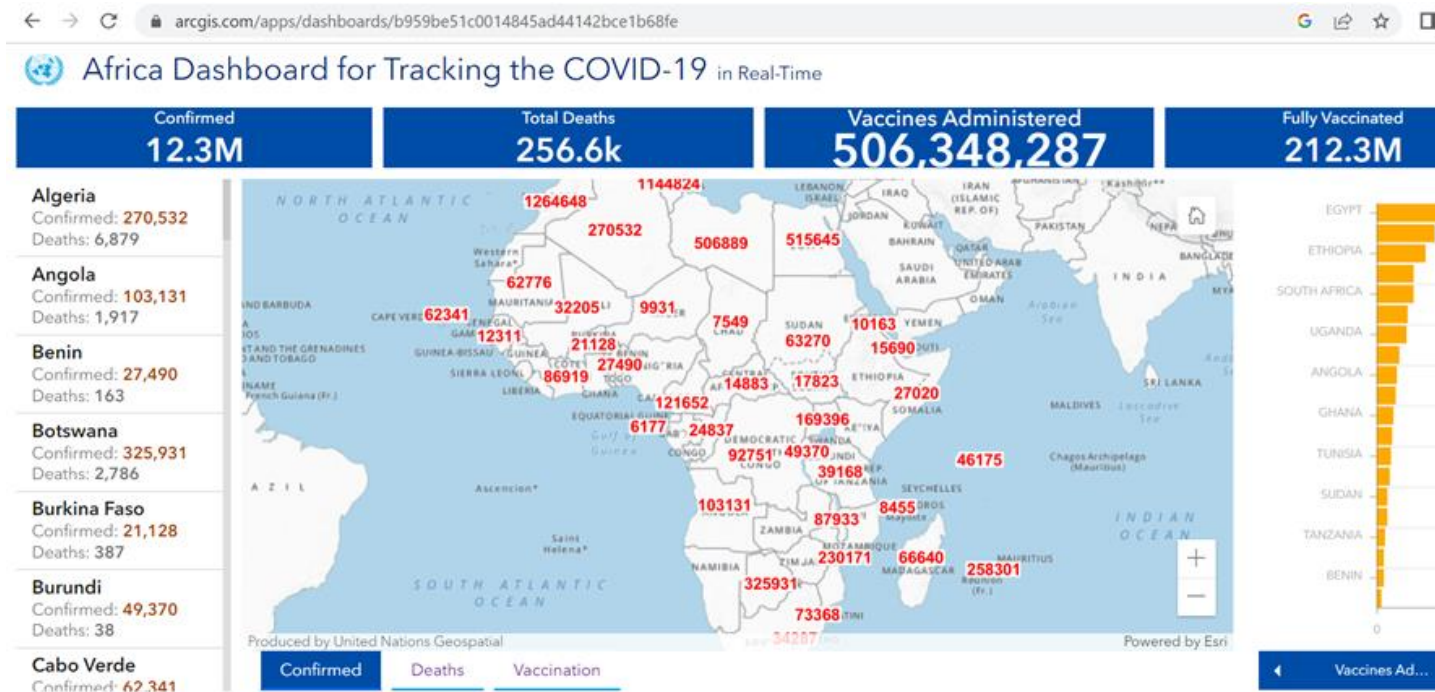
Mozambique flooding After Cyclone Idai



# APPLICATIONS OF GEOSPATIAL BIG DATA & DATA SCIENCE

## Health Care:

- COVID-19 tracking; Vaccination campaigns



COVID-19 tracking Dashboard, JHU

Africa COVID-19 tracking Dashboard, developed by ECA

# APPLICATIONS OF GEOSPATIAL BIG DATA & DATA SCIENCE

## Data Science Campus Pilot Projects:

- Estimating livestock population from EO data, South Sudan
- Used Low resolution (low cost) and high resolution
- Machine Learning Model used with ground truthing (training data) to train the model
  
- Use of AI and EO data to identify crop type at national level, **Senegal**
- Use of Satellite image to identify metal roofs as a proxy-indicator for poverty, **Uganda**



# APPLICATIONS OF GEOSPATIAL BIG DATA & DATA SCIENCE

ECA support to Burundi in Census mapping:

- Esri's Official statistics modernization program provide GIS tools and training for 2020 census undertaking
- Use of ML to extract building footprint
- Point location of each building center point is generated
- To be used by enumerators to conduct census
- Esri ArcGIS Online platform is used to map and disseminate



# ROLE OF ECA IN PROMOTING BIG DATA AND AI

- ECA need to develop a comprehensive strategy on how ECA can sustainably use technologies, tools and innovative data sources such as Big Data, Geospatial information and AI to support its vision and mandate.
- Demystify these terms for common understanding within ECA and our Member States by preparing blogs and conducting webinar series/StatsTalk Africa launched on the 18th of April 2023
- ACS/ECA to continue working with countries and partners to promote the use of Big data, data science and AI in Africa



# ROLE OF ECA IN PROMOTING BIG DATA AND AI

- ECA to continue encouraging member States in the use of new technologies: Big data, data science and AI as a means to the transformation and modernization of national statistical systems and official statistics
- ECA to continue supporting member States in building their capacity in the use of new technologies including AI (Improve Infrastructure, develop skills/human capital)
- ECA need to value and leverage the use of geospatial data in its work including formulating policies, research and publications.

