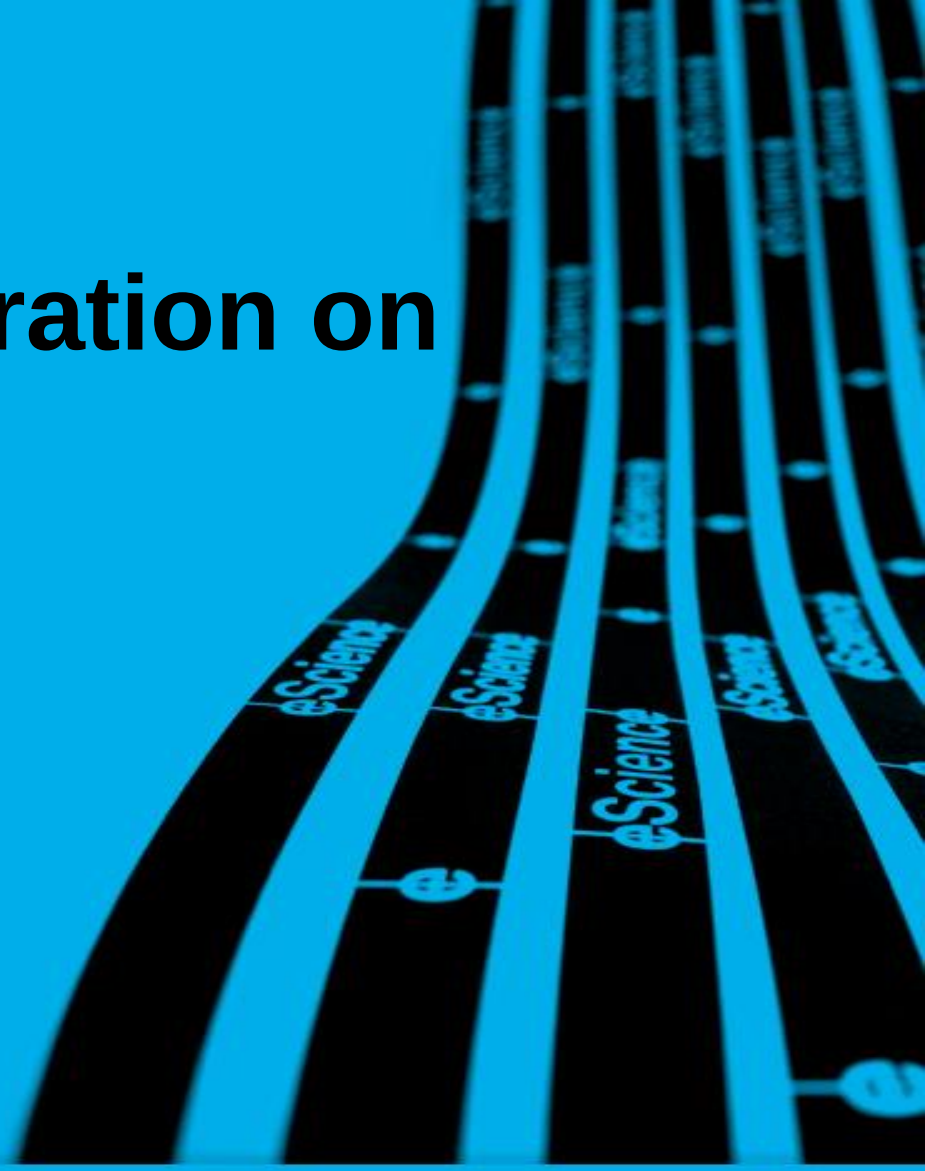
Technology to enhance exploration on space and field data

Romulo Goncalves,
r.goncalves@esciencecenter.nl
Netherlands eScience Center

Fieldwork from Space Symposium

Space Expo in Noordwijk, the Netherlands



netherlands eScience center

by SURF & NWO

New era…

- **Today:** Data exploration (e-science)
  - **Synthesizing theory, experiment and computation with advanced data management and statistics**

a)

b)

Images from: a) http://community.arm.com/servlet/JiveServlet/showImage/38-4971-14238/square-kilometre-array-2024-large.jpg
b) http://www.czechspaceportal.cz/files/files/images/EO/2012_12_11_15_40_47.jpg

Opportunities

- **Source of information for different applications**
- **Practical way to obtain data from regions such as Antarctica and Amazon forest**
- **Work at the lab, reduce field work**
  - **But for this we need new technology**
- **Let's navigate through few examples to identify the requirements.**

Images from: a) http://pre11.deviantart.net/2504/th/pre/f/2013/203/8/c/a_guy_working_on_his_computer__colored_edition__by_retronator-d6eplbx.png
b) http://wallpoper.com/images/00/38/13/87/birds-penguins_00381387.jpg

Digital soil mapping

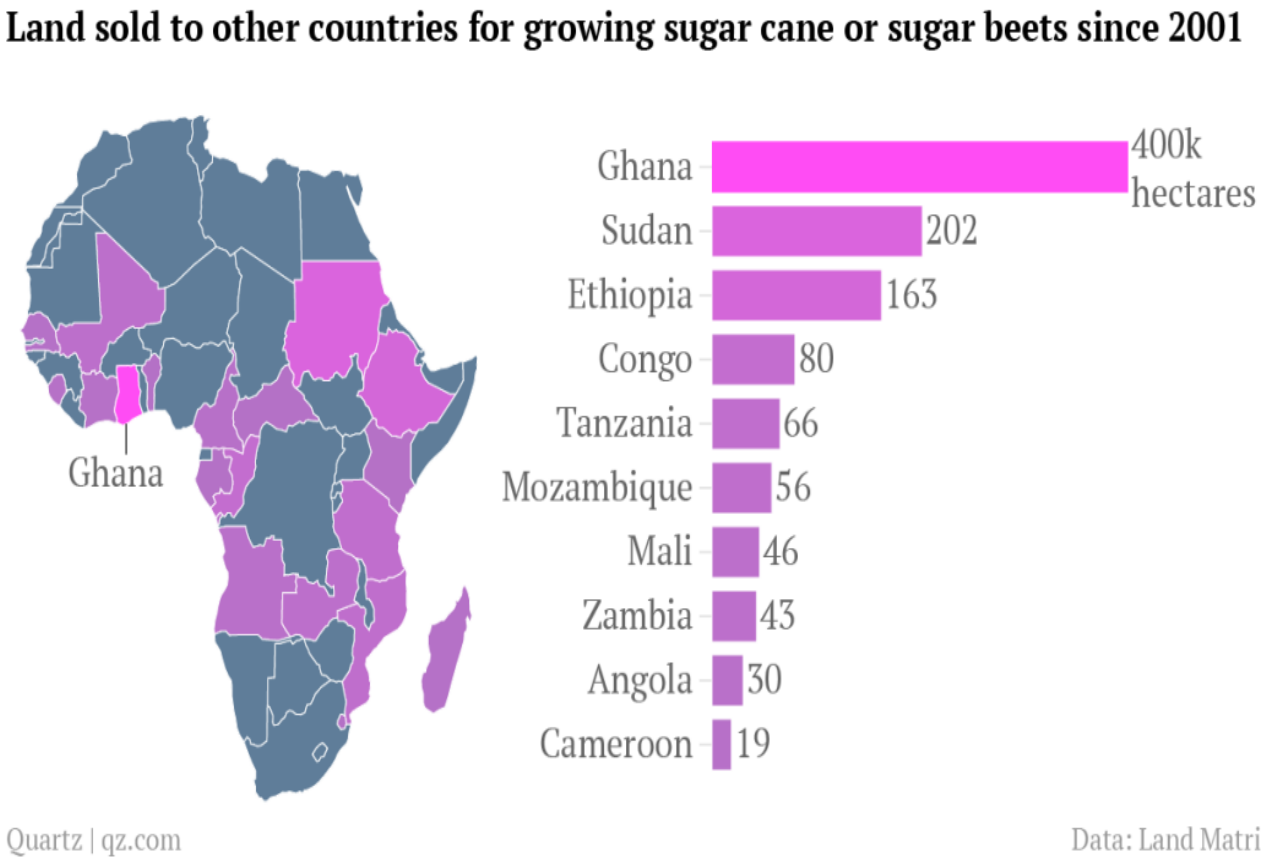
- **“Remote sensing provides important coverage, mapping and classification of landcover features, such as vegetation, soil, water and forests.”**
- **Soil classification**
  - **It provides information on the productivity of forests, meadows, wildlife habitat conditions, land-use and recreational suitability**
- **In regards to available covariates, it can be beneficial to use some type of data mining**
  - **e.g., multi-temporal spectral data of vegetation can indicate a lot about subsoil conditions**
- **Do not ignore crowd-sourcing and already established governmental data sources**
  - **Air photos and field work measurements provide additional information**
  - **The more field data you have, the better the models can be calibrated**

http://ucalgary.ca/geog/Virtual/remoteintro.html
http://ucalgary.ca/geog/Virtual/Remote%20Sensing/rssoils.html

Precision Agriculture (PA)

- **“It is a whole - farm management approach using information technology, satellite positioning (GNSS) data, remote sensing and proximal data gathering.”**

- **“These technologies have the goal of optimizing returns on inputs whilst potentially reducing environmental impacts.”**



http://www.europarl.europa.eu/RegData/etudes/note/join/2014/529049/IPOL-AGRI_NT(2014)529049_EN.pdf
Images from: https://qzprod.files.wordpress.com/2013/10/sugar-land-grab-africa.png?w=1024&h=583

Technology requirements

- **Multiple layers and different resolutions**
- **Data integration**
- **Spatiotemporal analysis and data mining**

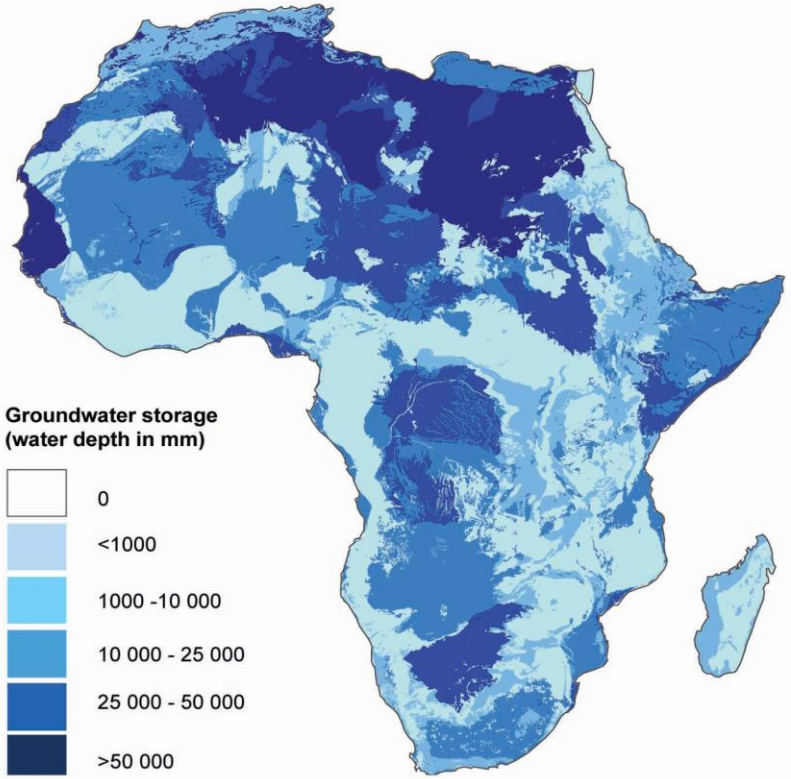
Not only Copernicus

- **Soil Moisture Mapping Satellite (SMAP) by NASA**
  - **“SMAP now begins a three-year mission that will figuratively scratch below Earth's surface to expand our understanding of a key component of the Earth system that links the water, energy and carbon cycles driving our living planet.”**
  - **Accurate soil moisture maps obtained from space.**
    - **SMAP’s combined radar and radiometer instruments will peer into the top 2 inches (5 centimeters) of soil, through clouds and moderate vegetation cover, day and night, to produce the highest-resolution (9km).**
  - **The mission will help improve:**
    - **climate and weather forecasts.**
    - **Allow scientists to monitor droughts and better predict flooding caused by severe rainfall or snowmelt.**
    - **Better forecast crop yields and assist in global famine early-warning systems.**

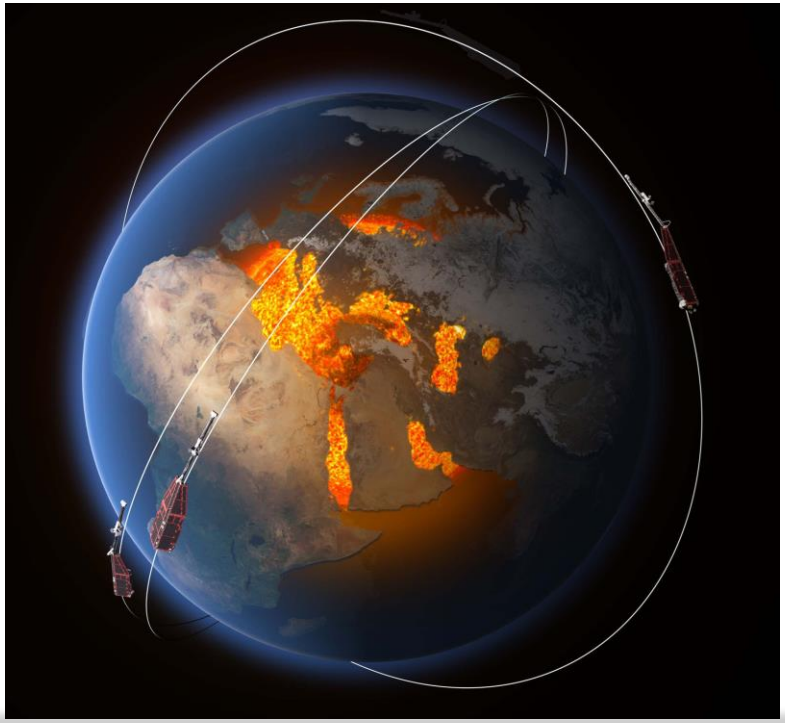
https://www.nasa.gov/press/2015/january/nasa-launches-groundbreaking-soil-moisture-mapping-satellite

The interior of the Earth

- **Geological surveys**
  - **Groundwater storage**



- **"The European Space Agency's three Swarm satellites observe magnetic signals that originate in Earth's core, mantle, crust and oceans…"**



a) http://www.bgs.ac.uk/research/groundwater/international/africanGroundwater/images/Fig2_groundwaterstorageSML.jpg
b) http://www.space.com/34264-ocean-monitoring-satellites-study-earth-interior.html

Technology requirements

- **Multiple layers and different resolutions**
- **Spatiotemporal analysis and data mining**
- **Data integration**
- **Data fusion**
- **Deployment at different geo-locations**

*“**Data integration** involves combining data residing in different sources and providing users with a unified view of these data.”*

*“**Data fusion** is the process of integration of multiple data and knowledge representing the same real-world object into a consistent, accurate, and useful representation.”*

Source: www.wikipedia.org

European data relay system (EDRS)

- **January 2016**
  - **“EDRS is designed to transmit data between low earth orbiting satellites and the EDRS payloads in geostationary orbit using innovative laser communication technology.”**
  - **“The SpaceDataHighway system will provide high-speed laser communication in space at up to 1.8 gigabits per second.”**
  - **“….up to 50 terabytes per day can be transmitted securely in near-real-time to Earth, as opposed to the delay of several hours currently experienced.”**
  - **“The European Commission is EDRS’s anchor customer through its Sentinel-1 and -2 missions.”**
- **June 2016:**
  - **“ESA today unveiled the first Sentinel-1 satellite images sent via the European Data Relay System’s world-leading laser technology in high orbit”**
- **Near-real time data access, great for risk assessment and management**

http://www.esa.int/Our_Activities/Telecommunications_Integrated_Applications/EDRS
https://airbusdefenceandspace.com/newsroom/news-and-features/successful-launch-of-the-first-laser-telecommunication-satellite-of-the-spacedatahighway-3/

Technology requirements

- Multiple layers and different resolutions
- Spatiotemporal analysis and data mining
- Data integration
- Data fusion
- Deployment at different geo-locations
- Stream processing & near real-time interaction

- EO Exploitation Platforms (EPs)
  - A set of R&D activities which aims to create an ecosystem of interconnected Thematic Exploitation Platforms (TEPs)
    - Coastal
    - Forestry
    - Hydrology
    - Geohazards
    - Polar
    - Urban themes

Interactive data centric approach

- **Interaction through different interfaces**
  - **Shift on the search parameters**
  - **New re-computations leading to new intermediates**
- **Step-wise exploration and avoid pre-computations**
  - **Computations should be pushed down all the way to the raw data**
  - **It does not only to improve resource utilization, but also to cope with user diversity**
  - **Increases the search space**
- **Drop the step by step scripted solution**
  - **Use workflows and eliminate pre-computed layers by turning them into on-demand parametrized computations**

Technology to enhance...

- **Data exploitation**

WMS, WFS, WCS & WPS

- **Data exploration**

SQL
Database Language

jupyter

R python Scala

CWL

- **Data processing & Orchestration**

APACHE Spark

Flink

docker
containers swarm compose

MESOS

- **Storage & Infrastructure**

SURF SARA

Google Cloud Platform

amazon web services

Microsoft Azure

Towards a new paradigm

- **Open-source state-of-the-art technology**
  - **The EU Joint Research Center is in the process of defining a “Data and Processing Platform”**
- **Tools and information validation**
  - **Standards**
  - **Provenance**
- **Infra-structure independent**
  - **The Earth Observation Data Centre (EODC)**
  - **The Copernicus Integrated Ground Segment consists of distribution service and Data and Information Access Services (DIAS)**
  - **…**

Netherlands eScience center

- **Website**
  - https://www.esciencecenter.nl/
- **Calls**
  - https://www.esciencecenter.nl/project-calls
- **Projects**
  - https://www.esciencecenter.nl/projects

Questions & Ideas?

Image from: http://aboutinterviews.com/5-questions-to-not-ask-in-a-job-interview/
http://www.atlantainjurylawblog.com/files/2016/08/bright-idea.jpg