Big Data in Earth Observations – challenges and opportunities

Earth and Planetary Research Centre GEOPLANET

Dr inż. Sebastian Aleksandrowicz (CBK PAN)

What is a Big Earth Data?

Big data refers to data that is so large, fast or complex that it's difficult or impossible to process using traditional methods.

SAS Institute Inc.

Big Earth data refers to big data associated with the Earth sciences, derived from but not limited to Earth observation. Big Earth data is characterized as being massive, multi-source, heterogeneous, multi-temporal, multi-scalar, highly dimensional, highly complex, nonstationary, and unstructured. It consists of all data related to the Earth, including the Earth's interior, surface, atmosphere, and near-space environment. Following the lead of big data in other disciplines, big Earth data serves as the basis of data-intensive scientific discovery in the Earth sciences, and gradually becomes a new engine of knowledge innovation.

International Society for Digital Earth, 2017

How do we characterize it?

The characteristics of Big Earth Data systems are typically classified as the **five Vs**

Velocity

- The rate at which new data are being generated or the rate at which data needs to be processed to be useful.

Variety

- Number of unique types of datapresent in the Big Data project

Volume

- Amount of data associated with a Big Data project

Value

- Refers to the concept that data must provide some usable output or knowledge in order to have made collecting and storing it worthwhile

Veracity

- The overall quality of the data available

Big Earth Data sources

Remote Sensing

- Observations obtained from satellite, airborne, unmanned aerial vehicles and ground-based instruments

In-situ and laboratory analysis

- Data collected from observation stations, networks, laboratory analyses, surveys, expeditions, field experiments and so on

Social sensing

- Refers to data relating to human activities

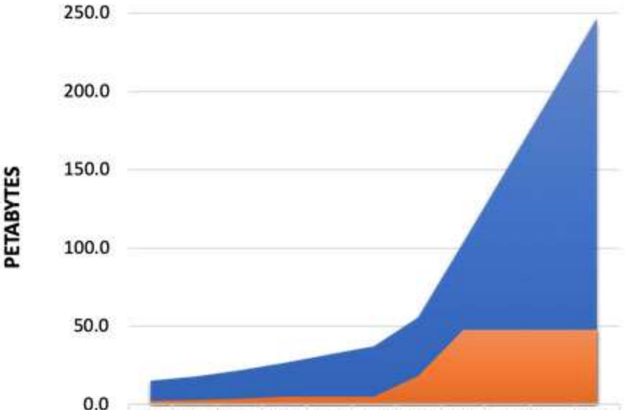
Simulation and reanalysis

- Data produced from simulations and reanalysis (data assimilation) of Earth systems and their interactions (e.g. Copernicus Servces)

How much EO data do we have?

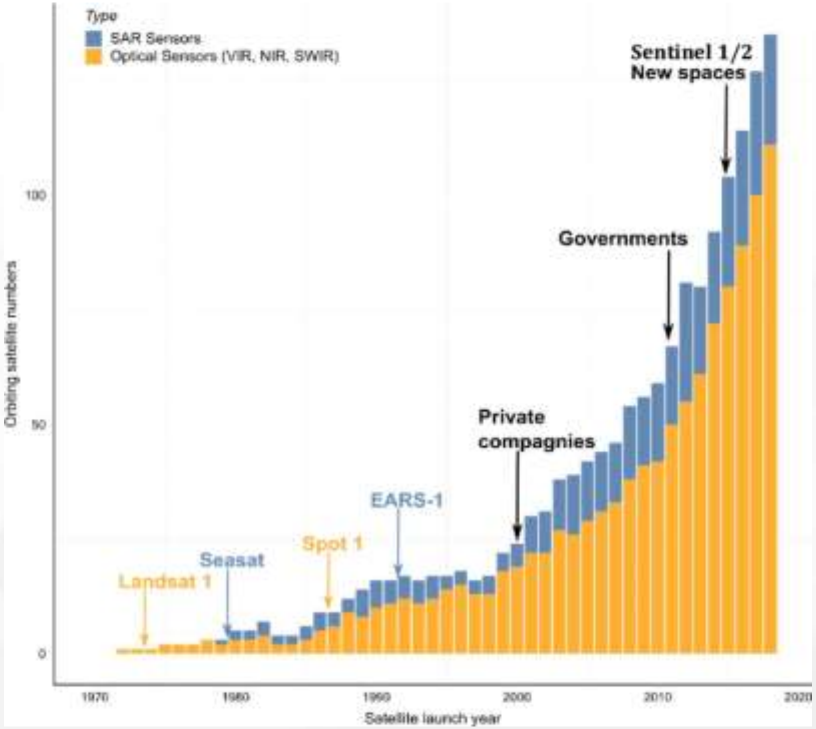
Free to access EO programmes such as Copernicus (Sentinel satellites), as well as new constellations privately owned have multiplied the amount of data well beyond our capacity to process it manually.

- Sentinels 1,2&3 ~20TB/day,
- DigitalGlobe constellation ~80TB/day,
- NASA EOSDIS: ~50TB/day,
- Maxar: ~80TB.



	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
Archive Size (PB)	15.0	17.7	21.6	26.8	32.0	37.2	55.6	103.4	151.1	198.9	246.6
Annual Growth (PB)	2.6	2.8	3.9	5.2	5.2	5.2	18.4	47.7	47.7	47.7	47.7

Source: [1]



Source: [2]

Key challenges

Data

- Volume
- Quality
- Heterogenity
- Accessibility
- Unavailability of labelled data

Infrastructure

- Storage limitations
- Computational limitations
- Cost

Algorithms

- Requirement for specialized tools
- Black-box
- Uncertainty Quantification

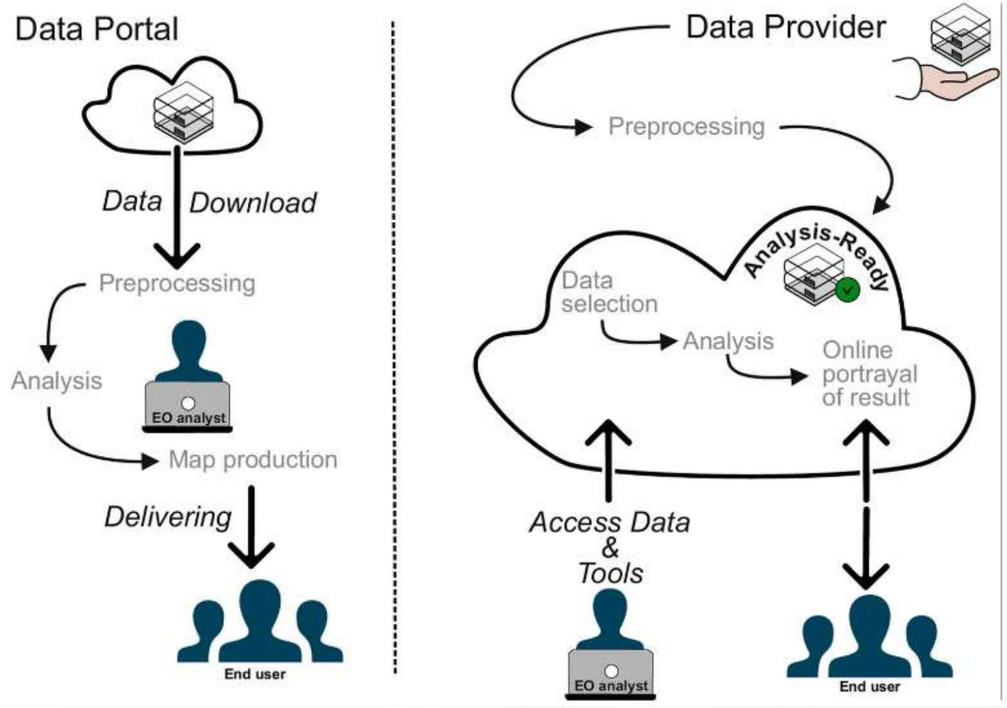
Other

- Contextual Understanding
- Diverse Stakeholders
- Ethical and Security Concerns

Earth and Planetary Research Centre GEOPLANET

Than and now: changes in EO data analysis workflow

Data-centric approach:
Different EO data providers (e.g. satellite-, model-based and in-situ data) are responsible for managing the data and disseminating it via a download service



Source: [3]

User-centric approach:
with dedicated platforms for EO data management and analysis

Earth and Planetary Research Centre GEOPLANET

How do we deal with Big Earth Data?

Data cubes and cloud services in action...

mundi WEB SERVICES
WWW.MUNDIWEBSERVICES.COM

CREODIAS
WWW.CREODIAS.EU

aws

Swiss Data Cube

ONDA
WWW.ONDA-DIAS.EU

Google Earth Engine

WEKEO
WWW.WEKEO.EU

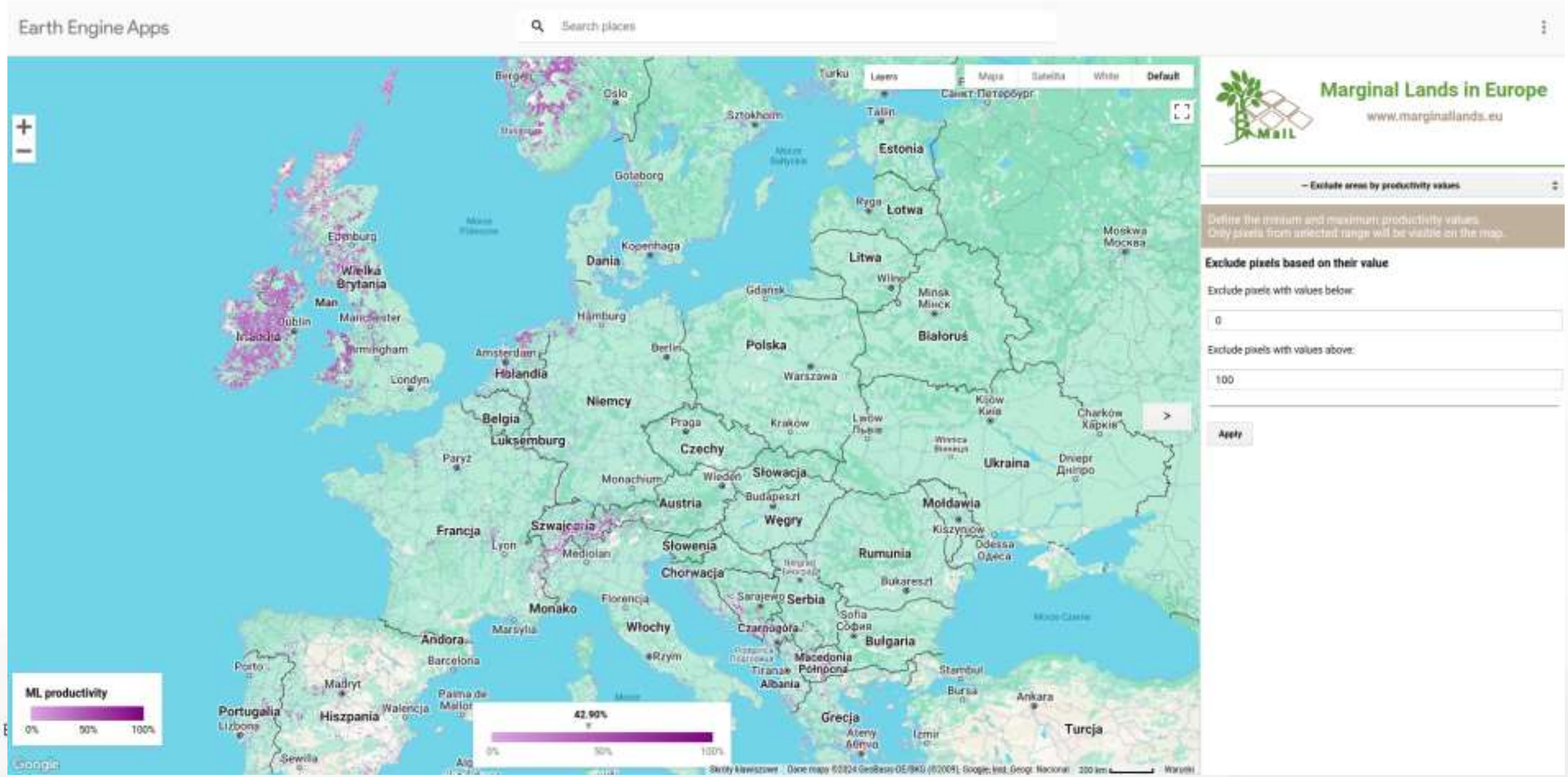
Digital Earth AUSTRALIA

EarthServer

sobloo
WWW.SOBLOO.EU

Earth and Planetary Research Centre GEO

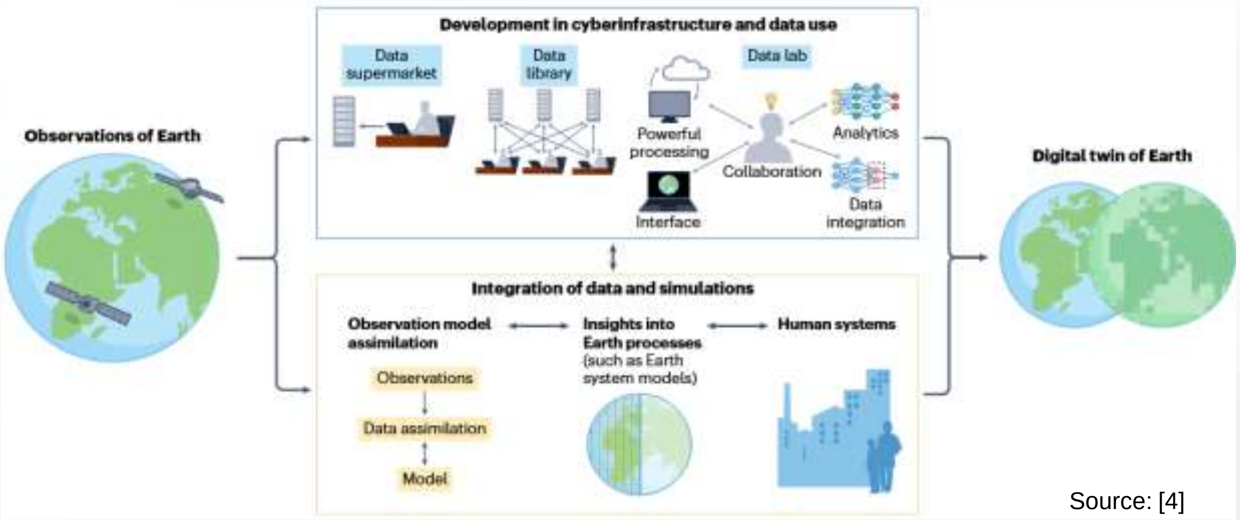
How do we deal with Big Earth Data?



The future: Digital Twin of Earth

Digital Twin main goals:

- To develop information system frameworks that provide continuous and accurate representations of systems as they change over time.
- To mirror various Earth Science systems and utilize the combination of Data Analytics, Artificial Intelligence, Digital Thread and state-of-the-art models to help predict the Earth's response to various phenomena.
- To provide the tools to conduct "what-if" investigations that can result in actionable predictions.



Source: [4]

Earth and Planetary Research Centre GEOPLANET

The future: Digital Twin of Earth

Destination Earth (DestinE)

- BioDT (biodiversity)
- Intertwin (digital twin engine)
- GEO-DT (geophysical extremes)
- Digital Twin of the Ocean (DTO)

Earth System Digital Twins (ESDT) - NASA

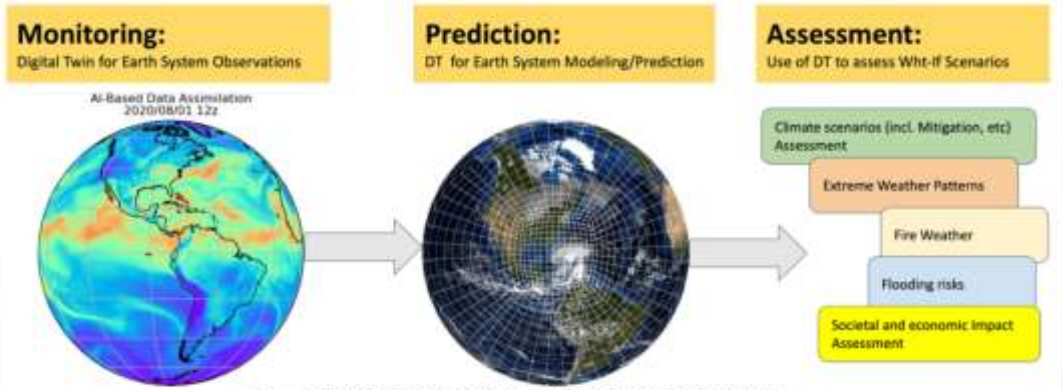


Figure 3. ***NOAA Vision for Earth System Digital Twin.***

Source: [5]

References

[1] https://nasadaacs.eos.nasa.gov/eosdis/cloud-evolution
[2] Tonneau, J., Bégué, A., Leroux, L., Augusseau, X., Faure, J., Mertens, B., Pinet, C., Tomasini, L. (2020). Geospatial Information For African Agriculture: A Key Investement For Agricultural Land Policies. Conference: 2020 World Bank Conference on Land and Poverty
[3] Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., … Blaschke, T. (2019). Big Earth data: disruptive changes in Earth observation data management and analysis? International Journal of Digital Earth, 13(7), 832–850. https://doi.org/10.1080/17538947.2019.1585976
[4] Earth Observation Cloud Platform Concept Development Study Report, Open Geospatial Consortium, http://www.opengis.net/doc/PER/EO-Cloud-Platform-CDS-ER
[5] Le Moigne J., Smith B.,Advanced Information Systems Technology (AIST) Earth Systems Digital Twin (ESDT) Workshop Report https://esto.nasa.gov/files/ESDT_Workshop_Report.pdf