Chapter 4 Remote Sensing



Miguel Ángel Esbrí

Abstract In this chapter we present the concepts of remote sensing and Earth Observation and, explain why several of their characteristics (volume, variety and velocity) make us consider Earth Observation as Big Data. Thereafter, we discuss the most commonly open data formats used to store and share the data. The main sources of Earth Observation data are also described, with particular focus on the constellation of Sentinel satellites, Copernicus Hub and its six thematic services, as well as other private initiatives like the five Copernicus-related Data and Information Access Services and Sentinel Hub. Next, we present an overview of representative software technologies for efficiently describing, storing, querying and accessing Earth Observation datasets. The chapter concludes with a summary of the Earth Observation datasets used in each DataBio pilot.

4.1 Introduction

Remote sensing is one of the most common ways to extract relevant information about the Earth and our environment. It can be defined as "the acquisition of information about an object or phenomenon without making physical contact with the object and thus in contrast to on-site observation, especially the Earth, including on the surface and in the atmosphere and oceans, based on propagated signals (e.g. electromagnetic radiation)" [1]. The term "remote sensing" was first utilized in the early 1960s to describe any means of observing the Earth from afar, particularly as applied to aerial photography, the main sensor used at that time. Today, as a result of rapid technological advances, we routinely survey our planet's surface from different platforms: low-altitude unmanned aerial vehicles (UAVs), airplanes and satellites. The surveillance of Earth's terrestrial landscapes, oceans and ice sheets constitutes the main goal of remote sensing techniques [2]. Remote sensing acquisitions, done through both active (synthetic aperture radar, LiDAR) and passive (optical and thermal range, multispectral and hyperspectral) sensors, provide a variety of information about the

land and ocean processes. In a broader context, remote sensing activities include a wide range of aspects, from the physical basis to obtain information from a distance, to the operation of platforms carrying out the sensor system, and further to the data acquisition, storage and interpretation. Then, the remotely collected data are converted to relevant information, which is provided to a vast variety of potential end users: farmers, foresters, fishers, hydrologists, geologists, ecologists, geographers, etc.

The use of Earth observation data imposes a series of technological challenges to:

- Combine satellite data with in situ or enterprise data.
- Understand, select, download, conserve and process data.
- Harness a range of scientific and technical skills and manpower.
- Load and store petabytes of data.
- Deploy high-performance processing capabilities.

4.2 Earth Observation Relation to Big Data

Different types of Earth observation data have been produced over the last forty years, bringing significant changes in the context of the big data concept. Moreover, the precise and up-to-date worldwide Earth observation data are changing the way that Earth is interpreted. It is leading to the implementation of applications powered with humongous amounts of remote sensing information. In that regard, several of the remote sensing data characteristics allow us to consider remote sensing data as big data:

Volume

Among the various areas where big data sets have become common, the ones related to remote sensing and information and communication technology are foremost, since the datasets involved have reached huge dimensions. This makes exceptionally complex their visualization, analysis and interpretation [2]. Besides, just in 2010, the satellite observation networks around the world had more than 200 on-orbit satellite sensors [3], capturing several gigabytes of information per second [3]. Nowadays, with the advent of the Copernicus programme with its Sentinel and contributing missions' satellites and with the entering into the commercial market of the US satellite operator Planet, the observation capacities dramatically increased, adding several petabytes of annual observations. According to Open Geospatial Consortium (OGC), the worldwide observation information currently most likely surpasses one exabyte.

Variety

Variety refers to the number of types of data, and concerning remote sensing data, it is specifically linked to structured information such as images obtained by satellite

sensors. More specifically, in this context, variety depends on the different resolution (spectral, temporal, spatial and radiometric) of the captured data. Remote sensing data variety is enormous. There are approximately 200 satellite sensors with a huge variety of spatial, temporal, radiometric and spectral resolutions [3]. Thus, for instance, satellites have a wide range of orbital altitudes, optics, and acquisition techniques. Consequently, the imagery acquired can be at very fine resolutions (fine level of detail) of 1 m or less with very narrow coverage swaths, or the images may have much larger swaths and cover entire continents at very coarse resolutions (>1 km). In addition, the satellites are equipped with sensors capable of acquiring data from portions of the electromagnetic spectrum that cannot be sensed by the human eye or conventional photography. The ultraviolet, near-infrared, shortwave infrared, thermal infrared and microwave portions of the spectrum provide valuable information of critical environmental variables [1].

Velocity

Velocity refers to the frequency of incoming data and the speed at which is generated, processed and transmitted. In the case of remote sensing data, the orbital characteristics of most satellite sensors enable repetitive coverage of the same area of Earth's surface on a regular basis with a uniform method of observation. The repeat cycle of the various satellite sensor systems varies from 15 min to nearly a month. This characteristic makes remote sensing ideal for multi-temporal studies, from seasonal observations over an annual growing season to inter-annual observations depicting land surface changes [2].

4.3 Data Formats, Storage and Access

4.3.1 Formats and Standards

Nowadays, remote sensing images (both, currently acquired and historical images) are typically distributed in digital format. A digital image is a numeric translation of the original radiances received by the sensor, forming a 2D matrix of numbers. Those values represent the optical properties of the area sampled, where the pixel represents the minimum spatial unit of measurement within the sensor coverage [2].

The following are the file formats most generally accepted as standards for encoding and transferring the remote sensing images:

 HDF¹ is a self-describing and portable, platform-independent data format for sharing science data, as it can store many different kinds of data objects, including multi-dimensional arrays, metadata, raster images, colour palettes and tables in a

¹ https://www.hdfgroup.org/.

single file. There is no limit on the number or size of data objects in the collection, giving great flexibility for big data.

- *NetCDF*² is also a self-describing, portable and scalable format that is currently widely used by climate modellers.
- JPEG 2000³ is an image coding system that uses state-of-the-art compression techniques based on wavelet technology and offers an extremely high level of scalability and accessibility. Content can be coded once at any quality, up to lossless, but accessed and decoded at a potentially very large number of other qualities and resolutions and/or by region of interest, with no significant penalty in coding efficiency. Typically used for distributing Sentinel-2 images.
- GeoTIFF⁴ is a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file. The potential additional information includes map projection, coordinate systems, ellipsoids, datums and everything else necessary to establish the exact spatial reference for the file. More interestingly, "Cloud Optimized GeoTIFF" (COG)—a standard based on GeoTIFF—is designed to make it straightforward to use GeoTIFFs hosted on HTTP web servers, so that users/software can make use of partial data within the file without having to download the entire file. It is designed to work with HTTP range requests and specifies a particular layout of data and metadata within the GeoTIFF file, so that clients can predict which range of bytes they need to download.

These specially designed data formats work quite well when the amount of data is not very large. However, issues start to arise when data volumes increase. The most obvious problem is that it is not easy to find, retrieve and query the information needed.

A lot of effort has been spent during the last years for standardising many of the EO ground segment interfaces in the context of HMA (OGC)⁵ [4] and CEOS⁶ [5]. The interfaces for which widely accepted standards exist and are deployed include:

- EO dataset/product metadata [6].
- EO dataset/product discovery [7–9].
- Online data access [10–12].
- Viewing.
- Processing.

Further details concerning standards for EO metadata and discovery interfaces can be found in Chap. 2 "Standardized EO data platforms".

² https://www.unidata.ucar.edu/software/netcdf.

³ https://jpeg.org/jpeg2000/.

⁴ https://en.wikipedia.org/wiki/GeoTIFF.

⁵ Heterogeneous Missions Accessibility (HMA), https://wiki.services.eoportal.org/tiki-index.php? page=HMA+AWG.

⁶ https://ceos.org/.

Table 4.1 Sentinel missions⁷

SENTINEL-1	With the objectives of land and ocean monitoring, SENTINEL-1 is composed of two polar-orbiting satellites operating day and night and will perform radar imaging, enabling them to acquire imagery regardless of the weather	
SENTINEL-2	Its main objective is land monitoring, and the mission is composed of two polar-orbiting satellites providing high-resolution optical imagery. Vegetation, soil and coastal areas are among the monitoring objectives	
SENTINEL-3	Its primary objective is marine observation, with focus on studying sea surface topography, sea and land surface temperature, ocean and land colour. Composed of three satellites, the mission's primary instrument is a radar altimeter, but the polar-orbiting satellites will carry multiple instruments, including optical imagers	
SENTINEL-4	It is dedicated to air quality monitoring. Its UVN instrument is a spectrometer carried aboard Meteosat Third Generation satellites, operated by EUMETSAT. The mission aims to provide continuous monitoring of the composition of the Earth's atmosphere at high temporal and spatial resolution, and the data will be used to support monitoring and forecasting over Europe	
SENTINEL-5	It is dedicated to air quality monitoring. The SENTINEL-5 UVNS instrument is a spectrometer carried aboard the MetOp Second Generation satellites. The mission aims to provide continuous monitoring of the composition of the Earth's atmosphere. It provides wide swath, global coverage data to monitor air quality around the world	
SENTINEL-5P	A precursor satellite mission SENTINEL-5P aims to fill in the data gap and provide data continuity between the retirement of the Envisat satellite and NASA's Aura mission and the launch of SENTINEL-5. The main objective of the Sentinel-5P mission is to perform atmospheric measurements, with high spatio-temporal resolution, relating to air quality, climate forcing, ozone and UV radiation	

4.3.2 Data Sources

4.3.2.1 Copernicus Programme and Sentinel Missions

The Copernicus EO programme is a cooperation of the European Union (EU) and the European Space Agency (ESA). This agency is responsible for coordinating the satellite acquisition and delivery of the EO data. Since the launch in 2014 of Sentinel-1A, the fleet of Sentinel satellites is delivering data for environmental monitoring and civil security applications.

Copernicus is served by a set of dedicated satellites (the Sentinel families) and contributing missions (existing commercial and public satellites). The Sentinel satellites are specifically designed to meet the needs of the Copernicus services and their users (Table 4.1).

Thematic Services

Besides the Sentinel satellite constellation, Copernicus also provides access to specific services, which fall into **six main thematic categories**⁷: services for land management, services for the marine environment, services relating to the atmosphere, services to aid emergency response, services associated with security and services relating to climate change.

- Land Monitoring: Monitoring the Earth's land is useful for many fields, particularly agriculture, forestry, topography and land-cover and land-change studies. The data can be used to track current trends and predict future changes.
- Marine Monitoring: Information on the state and dynamics of the ocean and coastal zones can be used to help protect and manage the marine environment and resources more effectively, as well as ensure safety at sea and monitor pollution from oil spills and other events.
- Atmospheric Monitoring: Monitoring the quality and condition of our planet's atmosphere is important in that it helps us to understand how we may be affected and is an essential tool in forecasting weather events.
- Managing Emergency: When an emergency occurs, satellite data can prove essential in forming a response. Historical data can provide perspective on a situation, while current data can help to analyse and manage the emergency.
- Security: Surveillance and security can be difficult to manage from the ground.
 Observations from space can make monitoring borders and sea routes much easier and track developing situations.
- Climate Change: Satellites are a vital tool in monitoring our world's changing climate, providing wide-scale views of affected areas and contributing to growing archives of data for use in long-term studies.

Most of the data and information are delivered by Copernicus, and its services are made available via a "free, full and open" policy to any citizen and any organization everywhere on Earth.

For dissemination of level 0, level 1 and level 2 products, ESA provides access via the Copernicus **Open Access Hub**⁹ portal, providing access to Sentinel-1, -2, -3 and -5p data through an interactive graphical user interface. Additionally, there are the Collaborative Data Hub, International Access Hub and Copernicus Services Data Hub which are providing access to public authorities, European projects and Copernicus services.

⁷ https://sentinels.copernicus.eu/web/sentinel/thematic-areas.

⁷ https://sentinels.copernicus.eu/web/sentinel/missions.

⁹ https://scihub.copernicus.eu.

Name	Provider	Webpage
CREODIAS	Creotech instruments, Cloudferro	https://creodias.eu
Mundi	Atos Integration, DLR, e-GEOS, EOX, GAF, Sinergise, Spacemetric, Thales Alenia Space, T-Systems	https://mundiwebservices.com
ONDA	Serco, OHV	https://www.onda-dias.eu
Sobloo	Airbus, Capgemini, Orange	https://sobloo.eu
WEkEO	Eumetsat, ECMWF, Mercator Océan	https://www.wekeo.eu

Table 4.2 DIAS providers

4.3.2.2 DIAS

In order to facilitate the access of Earth observation products and the development of EO-powered applications for end users, five different **Data and Information Access Services** (DIAS) are available (see Table 4.2). The DIASes provide access to product repositories in cloud storage. They primarily are not thought to be used as "dissemination" hubs (download bandwidth is even lower than at Open Access Hub, and it is generally not free). The DIAS provides platforms for hosting processing in vicinity to the cloud storage. End users can bring their algorithms and run them with free and fast access to the product data (by combining simple access to curated petabyte-size collections of Copernicus, other satellite and third-party data). Eventually, the end user only needs to download the (typically low volume) processing results and not the (high volume) satellite input products.

4.3.2.3 Other

Other data access portals are available as well:

- Amazon Web Services (AWS) and Google Cloud Platform (GCP) offering storage and processing platforms services similar like the DIAS but differing in product offers and service pricing
- Sentinel Hub¹⁰ is a commercial data access and on-the-fly processing software instantiated on AWS and on two of the DIAS and exposing an application programme interface (API) to user applications for accessing Copernicus and Landsat products and derivatives.

¹⁰ https://www.sentinel-hub.com/.

4.4 Selected Technologies

The present section identifies information technology domains and contains further practically relevant insights (mainly from DataBio data access components) into these for builders of applications and systems using EO data and cloud-based environments.

4.4.1 Metadata Catalogue

As per the OGC definition¹¹: "Catalogue services support the ability to publish and search collections of descriptive information (metadata) for data, services and related information objects. Metadata in catalogues represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software. Catalogue services are required to support the discovery and binding to registered information resources within an information community".

In the case of Earth observation datasets, a series of specific EO metadata profiles have been defined in order to facilitate their description and findability. Chapter 2 "Standards and EO data platforms" provides further details about them. The following describes the concrete EO metadata catalogue implementations used in DataBio.

FedEO Gateway

This component [13] acts as a unique endpoint allowing clients to access metadata and data from different backend EO catalogues implementing different protocols. It supports access through OGC 10-032r8 and OGC 13-026r8 OpenSearch interfaces and provides atom responses with metadata in OGC 10-157r4 format (i.e. EO profile observations and measurements). Alternative response formats such as RDF/XML, Turtle, JSON-LD and GeoJSON (OGC 17-003) are available as well. SRU-style bindings and W3C linked data platform bindings are available as well.

FedEO Catalogue

This component [13] implements an EO catalogue server allowing to store EO (satellite) collections (series) and products (datasets) metadata. It offers an API to populate the catalogue and an API to search the catalogue.

Both components have been developed by Spacebel s.a.

¹¹ https://www.ogc.org/standards/cat.

4.4.2 Object Storage and Data Access

GeoRocket

GeoRocket¹² is a high-performance data store for geospatial files developed by Fraunhofer Institute for Computer Graphics Research IGD. It can store 3D city models (e.g. CityGML), GML files or GeoJSON data sets. It provides the following features:

- High-performance data storage with multiple back ends such as Amazon S3, MongoDB, distributed file systems (e.g. HDFS or Ceph), or your local hard drive (enabled by default)
- Support for high-speed search features based on the popular open-source framework elasticsearch. You can perform spatial queries and search for attributes, layers and tags.
- Its design and implementation (based on the open-source toolkit Vert.x), makes it perfectly suitable for being deployed in Cloud environments, making it reactive and capable of handling big files and larger numbers of parallel requests.

Rasdaman

Rasdaman¹³ is an array database system, which provides flexible, fast, scalable geoservices for multi-dimensional spatio-temporal sensor, image, simulation and statistics data of unlimited volume. Data are stored in a PostgreSQL database, thereby achieving full information integration (e.g. latitudes, longitudes, time coordinates, resolutions and other ancillary annotations.). Ad-hoc access, extraction, aggregation, as well as remix and analytics are enabled through a new SQL raster query language—the Rasdaman query language (RasQL)—with highly effective serverside optimization. The core features include—truly multi-dimensional—1D, 2D, 3D, 4D, and beyond—powerful, flexible query language for visualization, classification, convolution, aggregation and many more geospatial functions spatial indexing and adaptive tiling for fast data access—parallelization and for unlimited scalability from laptop to cluster and cloud—full information integration of raster data with all geo data in the PostgreSQL database—support for the raster-relevant OGC standards, reference implementation for WCS core and WCPS.

Data Cubes

EO data cubes are an advanced way how users interact with large spatio-temporal EO data [14]. Figure 4.1 illustrates the principle. The idea is to read incoming image tiles covering an area ("Dice") and arrange these in time series pixel stacks ("Stack"). This makes access to the time series of observations ("Use") much easier.

Data cubes implementations (such as Rasdaman or ADAM¹⁴) allow accessing a large variety of multi-year global geospatial collections enabling data discovery,

¹² https://georocket.io/.

¹³ https://rasdaman.com/.

¹⁴ https://adamplatform.eu.

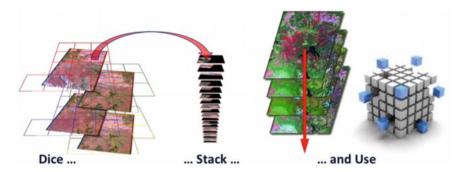


Fig. 4.1 Data cube (Credits Geoscience Australia)

visualization, combination, processing and download. They permit to exploit data from global to local scale (taken from distributed data sources are made accessible through the data cube layer that exposes OGC-standardized interfaces). On top of the data cube layer, platform-based interfaces (web application, mobile application, Jupyter Notebook and APIs) as well as third-party user interfaces can be deployed.

Another example is **Xcube**, ¹⁵ which is an open-source Python package for generating and exploiting data cubes. It comprises one of the core parts of the Euro Data Cube (EDC), ¹⁶ together with the Sentinel Hub. The EDC engine is able to technically serve custom raster data in addition to the freely available EO data archives like Sentinel, Modis or Landsat.

4.5 Usage of Earth Observation Data in DataBio's Pilots

A significant part of the 27 DataBio pilots uses Earth observation data as input for their specific purposes, in the context of efficient resource use and increasing productivity in agriculture [15], forestry [16] and fishery [17] (Table 4.3).

¹⁵ https://xcube.readthedocs.io/en/latest/.

¹⁶ https://eurodatacube.com.

 Table 4.3 Examples of use of EO datasets in DataBio pilots

EO dataset	Pilots	Common usage
Sentinel-2	Agriculture pilots A1.1, B1.2, C1.1	Time series and multiple statistics of EO-based indicators that describe various agri-environmental conditions and are assigned to each agricultural parce (e.g. corrected products, vegetation indices like NDVI and NDWI)
	Agriculture pilots A1.2, A1.3, B1.1, B1.3, C1.2, C2.1 Forestry pilots 2.3.1, 2.3.2- AIS, 2.3.2-FH, 2.4.1	A time series of Sentinel-2 L1C images (both A and B satellites) are used to cover a growing season
	Agriculture pilot B1.4	Scenes covering vegetation period of cereals and meeting cloud cover criteria
Sentinel-3	Fishery pilots A1 and B1	Sentinel-3 SLSTR for sea surface temperature Sentinel-3 SRAL/MWR for altimetry (anomalies) Sentinel-3 OLCI for chlorophyll
Landsat	Agriculture pilot B1.1 Forestry pilot 2.3.2-FH, 2.3.2-AIS	Despite the resolution being lower than the Sentinel mission, it has been TRAGSA-TRAGSATEC reference data for years. It is used as reference, contrast and on those dates or areas with no Sentinel coverage
Landsat 8	Agriculture pilot B1.4	Scenes covering vegetation period of cereals and meeting cloud cover criteria
Proba-V	Agriculture pilots A1.2, A1.3, B1.3, C1.2	Used for long-term time series, which are not available for Sentinel-2 yet
Meteorological data	Agriculture pilots A1.2, A1.3, B1.3, C1.2	Temperature and rainfall data from national providers and/or ECMWF
CMEMS products	Fishery pilots A1 and B1	The Copernicus Marine Environment Monitoring Services of interest cover many products/variables used in the fishery pilots

References

- 1. https://en.wikipedia.org/wiki/Remote_sensing. Last accessed September 3, 2019.
- 2. Chuvieco, E. (2016). Fundamentals of satellite remote sensing: An environmental approach (2nd ed.). Boca Raton, FL, USA: CRC Press Inc.
- NASA. (2010). On-orbit satellite servicing study. https://sspd.gsfc.nasa.gov/images/NASA_S atellite%20Servicing_Project_Report_0511.pdf
- Heterogeneous Missions Accessibility (HMA), design methodology, architecture and use of geospatial standards for the ground segment support of earth observation missions. European Space Agency, ESA TM-21, April 2012, https://esamultimedia.esa.int/multimedia/publicati ons/TM-21/TM-21.pdf, 2012, ISBN 978-92-9221-883-6.
- CEOS OpenSearch Best Practice. Issue 1.2, 13/06/2017. https://ceos.org/document_manage ment/Working_Groups/WGISS/Interest_Groups/OpenSearch/CEOS-OPENSEARCH-BP-V1.2.pdf.
- 6. OGC 10-157r4, Earth observation metadata profile of observations & measurements, Version 1.1, June 9, 2016. https://docs.opengeospatial.org/is/10-157r4/10-157r4.html.
- OGC 10-032r8, OGC OpenSearch geo and time extensions. https://www.opengeospatial.org/ standards/opensearchgeo.
- 8. OGC 13-026r8, OGC OpenSearch Extension for Earth Observation Products. https://docs.opengeospatial.org/is/13-026r8/13-026r8.html.
- OGC 17-047, OGC EO OpenSearch Response GeoJSON(-LD) Encoding Standard. https://docs.opengeospatial.org/is/17-047r1/17-047r1.html.
- OGC 13-043, Download Service for Earth Observation Products, Version 1.0, January 31, 2014.
- OGC 14-055r2, OGC OWS Context GeoJSON Encoding, May 30, 2016. https://docs.opengeospatial.org/is/14-055r2/14-055r2.html.
- OGC 12-084r2, OGC OWS Context Atom Encoding Standard. https://portal.opengeospatial. org/files/?artifact_id=55183
- 13. Coene, Y., Gilles, M. et al. (2020). "D5.1 EO component specification", Deliverable D5.1 of the European Project H2020-732064 Data-Driven Bioeconomy (DataBio), June 13, 2020 [Online].
- Triebnig, G. (2020). D3.9 guidelines on EO-ICT support improvement to stakeholders v1.
 Deliverable D3.9 of the European Project H2020-821940 Bringing together the Knowledge for Better Agriculture Monitoring (EO4AGRI) July 2, 2020. [Online].
- Estrada, J., Rogotis, S., Miliaraki, N., Mastrogiannis, K. et al. (2020). "D1.3 agriculture pilot final report", Deliverable D1.3 of the European Project H2020-732064 Data-Driven Bioeconomy (DataBio), July 15, 2020 [Online].
- Miettinen, J., Tergujeff, R., Seitsonen, L. et al. (2020). "D2.3 Forestry Pilot Final Report" Deliverable D2.3 of the European Project H2020-732064 Data-Driven Bioeconomy (DataBio), July 15, 2020 [Online].
- 17. Fernandes, J. A., Quincoces, I., Reite, K.-J. et al. (2020). "D3.3 Fishery Pilot Final Report", Deliverable D3.3 of the European Project H2020-732064 Data-Driven Bioeconomy (DataBio), July 15, 2020[Online].

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

