Copernicus and openEO

OER4SDI

Open Educational Resources for
Spatial Data Infrastructures

In this technical tutorial on the Copernicus data infrastructure, you will learn about the Copernicus Data Space Ecosystem and OpenEO.

Tobias Krumrein, Albert Remke
Institute for Geoinformatics, University of Münster

September 2024

ein Kooperationsprojekt,
empfohlen durch:

DIGITALE HOCHSCHULE NRW

gefördert durch:

Ministry of Culture and Science
of the State of
North Rhine-Westphalia

1.Overview

In this technical tutorial on the Copernicus data infrastructure, you will learn about the **Copernicus Data Space Ecosystem** and **OpenEO**. After working through the tutorial, you will be able to

- understand the purpose and scope of the Copernicus program
- understand the role of the Copernicus Data Space Ecosystem and OpenEO for accessing and processing Copernicus data
- use Python to create information products from Earth observation data using the Copernicus Data Space Ecosystem and OpenEO.

This tutorial is structured as follows:

1. Overview
2. Background
    2.1. The Copernicus Program
    2.2. The Copernicus Data Space Ecosystem
    2.3. Processing of EO-Data
3. Practical Exercise
    3.1. Exercise Overview
    3.2. Preparation
    3.3. Exercise
4. Summary and Discussion

This Tutorial takes about 30 minutes for reading and viewing the provided materials, downloading the software and for conducting the hands-on exercises and tasks

This OER is primarily designed to be used by students in Geoinformatics, Geomatics and similar study programs. It is also useful for students of other study programs and for practitioners who want to enhance their understanding of SDI concepts and technologies. Some basic knowledge of web technologies such as HTTP and Web Services is required. However, you will be able to follow and find links to further resources if needed. Your computer should have 8 GB of usable RAM and 2 GB of usable disk space to download and use the software for this tutorial.

This Tutorial has been developed at the Institute for Geoinformatics, University of Münster. Authors are Tobias Krumrein and Albert Remke.

The OER4SDI project has been recommended by the Digital University NRW and is funded by the Ministry of Culture and Science NRW.

Let's Get Started!

Begin by familiarizing yourself with the background concepts in the next section. Understanding these will be crucial for successfully completing the hands-on exercises.

---

2. Background (15 min)

Before we can dive into the exercises, we need to explain some terms and technical concepts. Note that in a 30 to 45 minute tutorial we cannot go into all the details and have to limit ourselves to certain picked out approaches. Therefore, we will focus on the basic functionality of the software components.

2.1 The Copernicus Program

The **Copernicus Program**, previously known as Global Monitoring for Environment and Security (GMES), is one of the biggest earth observation programs in the world. It was started 1998 by the European commission and the European Space Agency (ESA). The program provides an infrastructure for earth observation and geoinformation. Since 2013 the Copernicus services are freely and open for everybody to use.

2.1.1 Introduction to the Copernicus Program

To get started, please watch this brief introduction video that outlines the Copernicus Program's goals, challenges, and operators. After viewing the video, we'll discuss these elements in more detail.

Goals, Components and Operators

The goals and challenges of the Copernicus Program are mainly **environmental and climate monitoring** as well as **security**. With the continuous, accurate and current data which comes from the different components it's possible to monitor the environment, including land, oceans and atmosphere. The data collection and real time data also enable climate change observation and are helpful for **disaster management**. But the data can also be used in other areas like for security and economic reasons.

To reach these goals the Copernicus program has three different components. There are of course the **Sentinel-Missions** and its satellites which some of you might have already heard off. They consist of six Sentinel missions with different focused areas. The next component is the **ground segment** of the Copernicus program. It's the infrastructure on the ground which is responsible for the satellite control, data processing, data archiving and distribution to the users. The last component is the **services** which the program also provides. These include services that use the data and other resources to provide specific information about land, sea, atmosphere, climate, disaster management or security.

The operators of the program are the **European Space Agency (ESA)**, the **European commission** as well as the **member states**, **EUMETSA**T and **other private companies**. But the biggest operators are the first two. The ESA is responsible for the development of the Sentinel satellites, the technical support and the operation of some parts of the ground segment. The European commission coordinates the whole program. They also fund the program and manage the services. EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) is operating meteorological satellites and hosts some Sentinel Mission (Sentinel 3 and 4) on them. The member states and other private companies participate in other aspects like giving additional data or implementation of research and development projects.

2.1.2 Earth Observation through Copernicus

The Copernicus program has different satellite missions with different monitoring use cases. In the table below they are shortly described. We will use only data of the Sentinel 2 satellite in this tutorial, so we will look at this one in more detail. It would push way above the limits of this module to describe every mission in detail. If you're interested the links with more information are provided in the table below.

Satellite Mission	Monitoring task
Sentinel 1	**Type**: satellite for radar images (with CSAR) **Satellites**: 1A (2014), 1B (2016), 1C & 1D (planned) **Use Case**: disaster management, infrastructure monitoring
Sentinel 2	**Type**: satellite for optical and infrared earth observation (with MSI) **Satellites:** 2A (2015), 2B (2017) **Use Case**: monitoring of vegetation, land use and inland waters
Sentinel 3	**Type**: multipurpose satellite (with OLCI, SLSTR, SRAL) **Satellites**: 3A (2016), 3B (2018) **Use Case**: monitoring global sea and land surface for climatic and environmental applications
Sentinel 4	**Type**: measuring instrument that flies with satellite from EUMETSAT (Geostationary satellite over Europe (with UVN-Spectrometer)) **Satellites**: 4 (planned) **Use Case**: air quality monitoring for Europe
Sentinel 5	**Type**: instrument on MetOp-SG satellite from EUMETSAT (low flying atmospheric monitoring satellite)

	Satellite: 5 (2022) **Use Case**: global air quality and atmospheric monitoring
[Sentinel 5P]()	**Type:** satellite (with TROPOMI) **Satellite:** 5P (2017) **Use Case:** Continuity of atmospheric data
[Sentinel 6]()	**Type:** satellites with tracking instruments **Satellite:** 6 (2021) **Use Case:** observing sea level and surface

Tab. 1: Sentinel Missions

Sentinel 2

Sentinel 2 Bild muss hier noch hin

Fig:1 Sentinel 2 satellite

The Sentinel-2 mission consists of two identical satellites, Sentinel-2A and Sentinel-2B, launched in 2015 and 2017 respectively. These satellites scan every point on Earth every 5 days, providing high-resolution optical imagery.

Sentinel-2 uses a Multispectral Instrument (MSI), which is special because it captures images in 13 different spectral bands, ranging from visible and near-infrared to short-wave infrared. This allows for detailed and diverse observations of the Earth's surface. Unlike radar systems, Sentinel-2 relies on optical and infrared sensors, which means its data is influenced by daylight and weather conditions. The MSI captures high-resolution images with spatial resolutions of 10 meters for visible and near-infrared bands, 20 meters for red-edge and short-wave infrared bands, and 60 meters for atmospheric correction bands.

The Sentinel-2 data is used for various tasks, primarily environmental and land monitoring. It's key applications are:

- **Vegetation Monitoring**: The MSI's spectral bands are optimized for monitoring plant health, chlorophyll content, and vegetation growth, making it essential for agricultural management.
- **Land Use and Land Cover Mapping**: High-resolution imagery helps map and monitor changes in land use and land cover, which is crucial for urban planning, forestry, and natural resource management.
- **Water Quality Monitoring**: Sentinel-2 provides data on turbidity, chlorophyll concentration, and the extent of water bodies, supporting the monitoring of inland and coastal waters.
- **Disaster Management**: Frequent revisit times enable timely monitoring of areas affected by natural disasters such as floods, wildfires, and landslides, aiding in emergency response and recovery efforts.

Sentinel-2's capabilities in capturing detailed optical and infrared images under various conditions make it an invaluable tool for continuous observation and analysis of the Earth's surface, supporting a wide range of environmental, agricultural, and scientific applications.

2.1.3 How to use Copernicus

The Copernicus Program has several ways for interacting and accessing the data and services. As mentioned before the data is freely available on many platforms. You can access them as an individual, an organization or as a company. One way to get access to the data is through an online portal. For example the **Copernicus Open Access Hub**. There you can download the data for free. Another approach is the **Copernicus Data and Information Access Services (DIAS)**. There are five platforms which provide a simplified access to Copernicus data and services as well as processing tools in the cloud. They have the function for users to access large datasets without the need to download it. But there are not only platforms directly from the Copernicus program. Another similar approach to the DIAS platforms is the use of a standardised API like **OpenEO**.
As you can see there are many different ways to interact and access the data and services of the Copernicus Program. Unfortunately, when there are so many possibilities to interact, not everything is available on one platform. So in this module we will focus on a relatively new platform to get access to the data, the services and tools. The **Copernicus Data Space Ecosystem**. Here they try to combine all the functionalities of the Copernicus Program in one.

2.2 The Copernicus Data Space Ecosystem

The Copernicus Data Space Ecosystem is a digital platform. It was developed to facilitate the use of Copernicus data through improved access, processing and analysis. It combines physical data resources with virtual services and tools. This optimizes the use of data for a wide range of users. Therefore the main purpose of the Copernicus Data Space Ecosystem is to **maximiz**e the **availability** and **usability of Earth observation data** from the Copernicus Program. It aims to facilitate data integration and utilization to support research, innovation and commercial applications not only for Europe but also globally.

Since the Copernicus Data Space Ecosystem was developed for the Copernicus Program it is not that surprising that two of the main operators are the ESA and the European Commission. In addition to these the DIAS platforms and some companies for cloud services and special analysis tools are also operators. Together they provide the needed infrastructure, to make the access of such big datasets possible.

So now that you have a first general overview about the platform, let’s analyze its components a bit more. The name of the platform can be divided in three terms. *Copernicus*, *Data Space* and *Ecosystem*. Obviously *Copernicus* stands for the Copernicus Program. The term *Data Space* refers to the digital infrastructure which is required to store, manage and access data. So in the combination of the first two terms, this includes the storage, processing and also hosting of the big sentinel data. The last term *Ecosystem* describes the whole network that is necessary to make the usage of Copernicus data work. This includes the technologies, services, operators and also the users. Together *Copernicus* forms the basis, the *Data Space* provides the necessary infrastructure and the *Ecosystem* connects all the stakeholders together.

But what is the advantage of using the Copernicus Data Space Ecosystem? The Copernicus Data Space Ecosystem promotes **data accessibility** and **interoperability**. And how is it doing that? There are several important mechanisms to these goals.

1. **Standardized Interfaces and Data Formats**: Ensures seamless interaction between different platforms, tools, and user applications.
2. **Centralized and Cloud-Based Services**: Enables users to access and process large datasets without needing to download them, making it easier to handle vast amounts of data.
3. **Open Access Policy**: Allows everyone to access and use the data for diverse applications.
4. **Integration of APIs and Toolkits**: Tools like OpenEO facilitate the development of custom solutions by providing direct access to Copernicus data.
5. **Community Collaboration**: Encourages a community of users and developers to share resources, knowledge, and experience, enhancing interoperability and innovation.

That all sounds very nice. But how can you work with it and how can you get access to the Copernicus Data Space Ecosystem not only through platforms of the Copernicus program like the Copernicus Open Access Hub. But that wouldn't let us try out the full potential of this powerful tool. We can also use an API like the OpenEO API. You want to know how? Let's see! In our last theorie block we learn a bit more about OpenEO and after that the theoretical part is finished and we can get more into practical work.

2.3 Processing EO-Data with OpenEO

When working with Earth Observation (EO) data, such as satellite images from the Copernicus Program, it can be challenging to manage large datasets and navigate the complexities of various data platforms. **OpenEO** is a tool designed to make this process simpler, providing a user-friendly way to access and analyze this data.

But what exactly is OpenEO, and how does it work?

2.3.1 Introduction to OpenEO

Think of OpenEO as a **bridge** between you and large amounts of satellite data that are stored in different locations worldwide. Normally, you might have to download large files and figure out how to process them using complex software. OpenEO solves this problem by offering a **single interface** that allows you to access and analyze data, without needing to worry about the technical details of how that data is stored or processed.

OpenEO is built around an **API** (Application Programming Interface). This API is essentially a set of tools that allow you to request specific data, perform calculations, and get results, without downloading or processing the data locally on your computer.

2.3.2 How does OpenEO work?

The typical workflow with OpenEO consists of some simple steps:

1. **Accessing Data:** Instead of manually downloading satellite data from multiple platforms, OpenEO allows you to access data directly through its API. Whether you're looking for data from Sentinel-2 or other satellites, you can easily retrieve what you need.
2. **Processing in the Cloud**: OpenEO allows you to process large amounts of data in the cloud, which means you don't need a powerful computer to handle complex analyses. You can send requests to OpenEO, and the actual data processing happens on remote servers. This frees up your computer's resources and avoids the need for downloading massive datasets.
3. **Standardized Interface**: OpenEO provides a standardized way to work with data from different sources. Whether you're working with Copernicus data, NASA data, or others, you can use the same commands and methods across all platforms. This simplifies your workflow, as you don't need to learn new systems for each data provider.
4. **Python Client Library**: In this tutorial, we will use the Python client for OpenEO. This client allows us to connect to the OpenEO API, search for data, and perform analyses—all using Python code. This is especially convenient because Python is a widely used programming language, and many tools for data analysis are built around it.

2.3.3 Why use OpenEO?

OpenEO provides several advantages that make working with satellite data easier:

- **No Local Storage Needed**: Instead of downloading huge datasets to your computer, OpenEO allows you to access and process the data directly in the cloud. This saves time, storage space, and computer power.
- **Simplified Workflow**: OpenEO offers a consistent interface for working with different types of EO data. Once you learn how to use it, you can apply the same skills to multiple data sources, making the process of accessing and analyzing data much smoother.
- **Efficient for Large Datasets**: Working with large datasets can be challenging, especially on a regular computer. OpenEO allows you to process data on powerful cloud servers, so even complex analyses can be done quickly and efficiently.

Now that you have a good grasp of the **Copernicus Program**, the **Data Space Ecosystem**, and how **OpenEO** can simplify processing large satellite datasets, it's time to apply these concepts. In the following **Hands-On Exercise**, you'll work with real **Sentinel-2 data** from the Copernicus platform.

Using the **OpenEO API** and Python, you will analyze this data to calculate the **NDVI** and assess flood damage. This practical exercise will help you see how the tools and methods you've just learned about can be applied to real-world environmental monitoring.

Let’s move on to the practical part and get hands-on with satellite data!

---

3. Hands-On Exercise (15 min)

In this chapter, you'll apply the concepts you've learned so far by engaging in a practical exercise. This exercise is designed to give you hands-on experience with Copernicus data, specifically using Sentinel-2 satellite imagery, within the Copernicus Data Space Ecosystem. You will also learn how to process this data using the OpenEO API and Python.

3.1 Exercise Overview

Scenario:
Imagine you are working for a governmental disaster management agency. A recent flood has caused significant damage in a low-lying region near a major river in Central Europe. Your task is to assess the extent of the flooding and its impact on both agricultural land and urban infrastructure. You will use Sentinel-2 data to analyze the flooded areas, identify changes in land cover, and calculate the NDVI to measure damage to vegetation.

Challenge:
Floods can quickly devastate large areas, affecting agriculture, infrastructure, and communities. Timely assessment of flood impact is crucial for disaster response, recovery planning, and minimizing future risks. The challenge is to quickly determine the extent of the flood damage, prioritize areas for intervention, and support recovery efforts.

Objective:
Your objective is to use Sentinel-2 data to identify flooded areas, assess the extent of damage to vegetation using NDVI, and analyze the changes in land cover before and after the flood. This information will help authorities direct resources to the most affected areas and plan for long-term recovery.

Advantages of Using OpenEO and Copernicus Data Space Ecosystem:

- **Rapid Data Access**: Quickly access recent satellite imagery from the Copernicus Data Space Ecosystem to assess flood damage in near real-time.
- **Accurate Analysis**: Use OpenEO to efficiently process large datasets and calculate NDVI, providing a clear picture of the extent of the damage to agricultural land and vegetation.

- **Scalable Processing**: Leverage the cloud-based resources of OpenEO to handle large-scale environmental data, especially in areas covering hundreds or thousands of hectares

3.2 Preparation

For the exercise you need to register for Copernicus Data Space Ecosystem and also need to install Docker for using a Python notebook.

3.2.1 Copernicus Data Space Ecosystem Account

First you need to register and create an account for Copernicus Data Space Ecosystem. You can register by following this link: https://dataspace.copernicus.eu/ and clicking on Login in the top right corner.
After you register you need to verify your email address before continuing.

3.2.2 Installing Docker, Jupyter and a Python Notebook

The software environment that we need for this exercise is Docker:

1. Install Docker Desktop: Go to the official web site https://docs.docker.com/get-docker/ and follow the guidance which is provided there to install docker on your computer (8GB usable RAM recommended).
2. Download the GitHub Repository of this Tutorial (https://github.com/oer4sdi/OER-CopernicusAndOpenEO) and unzip it in your wanted working directory
3. Make sure Docker Desktop is up and running
4. Open a terminal and change to your working directory
5. Use "docker-compose up --build" to build and run the docker container

Once this is complete, your terminal output should look similar to this:

```
    To access the notebook, open this file in a browser:
        file:///home/jovyan/.local/share/jupyter/runtime/nbserver-7-open.html
    Or copy and paste one of these URLs:
        http://177a0a74ba1b:8888/?token=3a2e5088e85cde486ab00cfbf5a5ef7346e51092980f43ce
     or http://127.0.0.1:8888/?token=3a2e5088e85cde486ab00cfbf5a5ef7346e51092980f43ce
```

Fig. 2: Example of a correct response in the terminal after the docker container is running

To access the python notebook look out for the link http://127.0.0.1:8888/?token=... in your output (it should look like the one on the bottom in the picture above). Copy the link and the token (the latter will be created new every time you start the container). Paste the link to your preferred browser to access the Jupyter server which displays the notebooks that can be used.

3.3 Exercise

Now that you have set up your environment, you are ready to begin the hands-on exercise.

1. **Open the Jupyter Notebook**: Navigate through the Jupyter interface to find the notebook provided in the tutorial repository.
2. **Download Sentinel-2 Data**: Follow the steps in the notebook to download Sentinel-2 data for the affected flood region. You will retrieve data from two different time periods: one before the flood and one after.
3. **Identify Flooded Areas**: Use OpenEO and the Sentinel-2 data to analyze land cover changes and detect the flooded areas.
4. **Calculate NDVI**: The exercise will guide you through calculating the **Normalized Difference Vegetation Index (NDVI)** for both time periods. This will help you assess damage to agricultural areas and vegetation by comparing pre- and post-flood NDVI values.
5. **Analyze Results**: Compare the NDVI values and land cover maps to identify the most severely affected areas. You can use these insights to help direct recovery efforts, such as focusing on areas where agricultural fields have been flooded or where significant vegetation loss has occurred.

4. Summary and Discussion

4.1 Additional Resources

So I integrated your feedback and I’m currently still working on the practical exercise of my module. When I’m ready with that I will get in touch with you