Data preparation for analytics

Alberto Meroni

Session agenda

Data Preparation

Data Transformation

Using a Mosaic dataset

STAC, using a STAC Catalog

Q&A

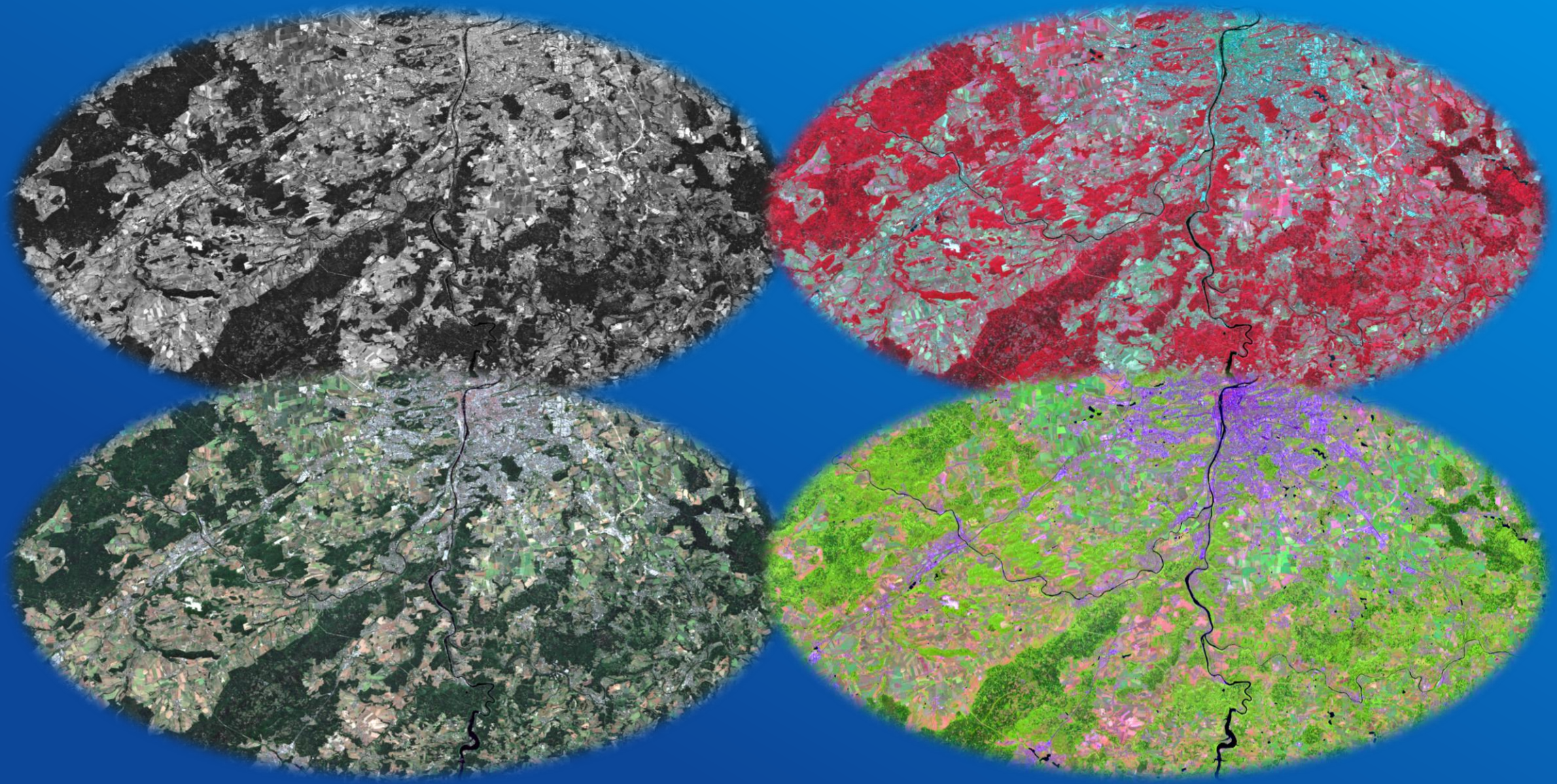
from Wayback, two images taken one year apart.

Data Preparation

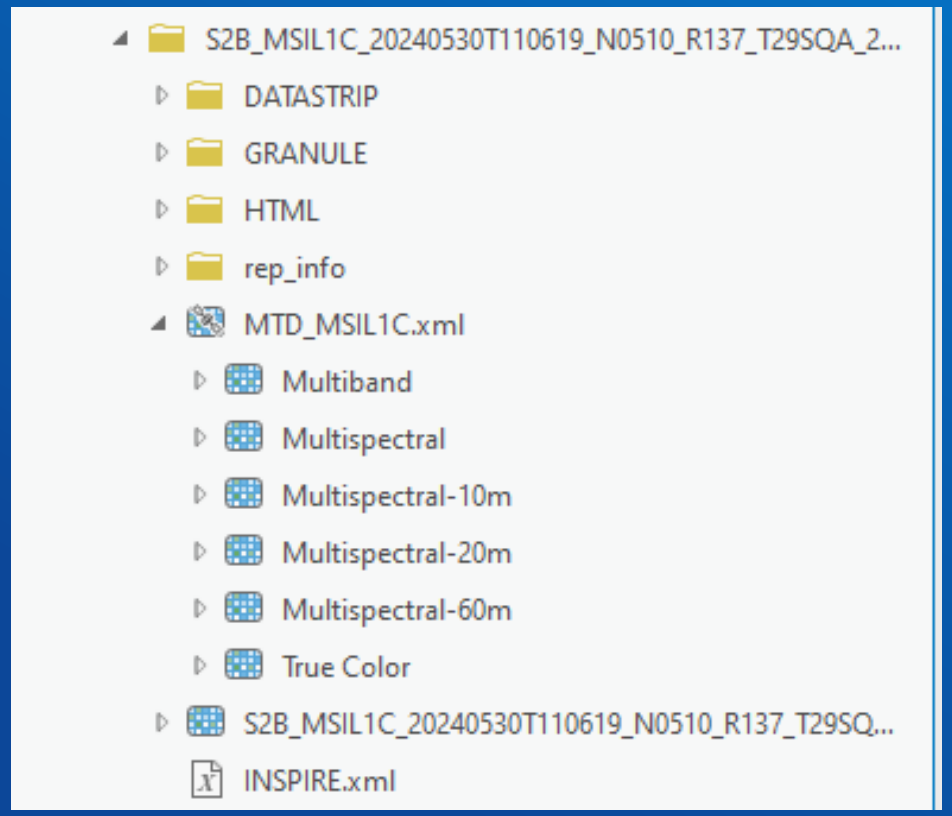
Data preparation calibrates and normalizes remote sensing data to ensure consistent, analysis-ready workflows.

- **Data Selection & Planning:** Identifying and selecting suitable datasets based on spatial, temporal, spectral, and quality requirements.
- **Pre-processing (Geometric & Radiometric Correction):** Removing atmospheric distortions, correcting sensor errors, and ensuring accurate spatial positioning.
- **Cloud & Noise Masking:** Identifying and masking clouds, shadows, and sensor artefacts to improve data quality.
- **Data Normalization & Scaling:** Standardizing pixel values (e.g., reflectance) to a common range for consistent analysis.
- **Data Transformation:** Applying indices, band combinations, textures, or temporal composites to enhance relevant information for analysis.

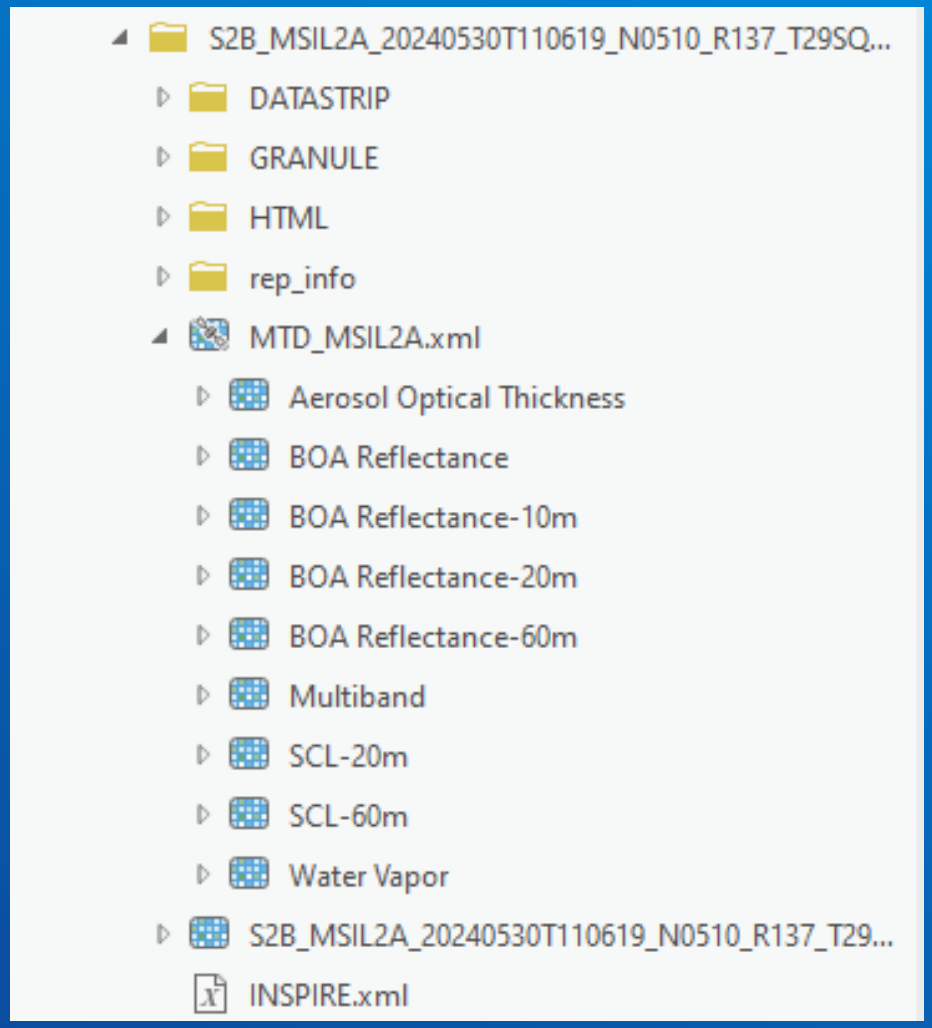
Imagery/Rasters a wide range



Raster products and raster types



Sentinel-2 L1C



Sentinel-2 L2A

Raster products and raster types

- Raster products are designed to simplify access the metadata available in sensor products.
- Raster products appear in the Catalog within the metadata files associated with specific vendor products.
- Raster products contain templates for one or more derived raster datasets (raster type) that can be added to the map view.
- The raster type identifies metadata, such as georeferencing, acquisition date, sensor type, and band wavelengths, along with a raster format. However, by adding raster data according to a raster type, the appropriate metadata is read and used to define any processing that needs to be applied.
- Each Raster Type has associated a Function Chain that describes the attributes of the data and how access the data within a Raster product.

Raster products and raster types

List of supported sensors

ArcGIS Pro 3.6 | Other versions | Help archive

Raster and image data can be supported in ArcGIS Pro as a raster dataset derived from a storage format, as a raster product deriv specific metadata files, or as a raster type. When data is supported as a raster type, it must be added to a mosaic dataset to be ha correctly in the application.

The following list encompasses all raster file formats, satellite sensors, aerial cameras, and product formats that ArcGIS Pro suppor

How each format is supported in ArcGIS Pro

Raster or image data format	Raster dataset	Raster type	Raster p
ADS		ADS aerial imagery	Raster pr
AIRSAR Polarimetric	AIRSAR raster dataset format		
Altum		Altum aerial imagery	
Applanix DSS		Applanix aerial imagery	
ARC Digitized Raster Graphic (ADRG)	ADRG raster dataset format		
ASCII Grid	ASCII raster dataset format		
ASTER		Aster satellite sensor	Raster pr
AVIRIS		AVIRIS aerial imagery	
Band interleaved by line (`.bil`), band interleaved by pixel (`.bip`), band sequential (`.bsq`)	Band interleaved raster dataset format		
Bathymetric Attributed Grid (`.bag`)	BAG raster dataset format		
BigTIFF	BigTIFF raster dataset format		
Binary Terrain (BT)	BT raster dataset format		
Bitmap (`.bmp`), device-independent bitmap (`.dib`), or Microsoft Windows bitmap	BMP raster dataset format		
BlackSky		Blacksky satellite sensor	Raster pr

Satellite sensor raster types

ArcGIS Pro 3.6 | Other versions | Help archive

There are many satellite sensor products to choose from, and each has properties you can refine to modify the processing applied to the imagery when it's added to a mosaic dataset. There are several defaults, but you can further modify these and save them for your own use at another time.

You can refine each raster type below by defining the product type. You can add all product types if you are unfamiliar with your data. This adds all your particular data, no matter how it was provided by the vendor. However, if you only want to add specific products to your mosaic dataset, choose the appropriate product type. For example, for QuickBird imagery, you can choose All, Basic, Standard, Standard Orthoready, and Orthorectified. Each of these options correlates to a product provided by the vendor.

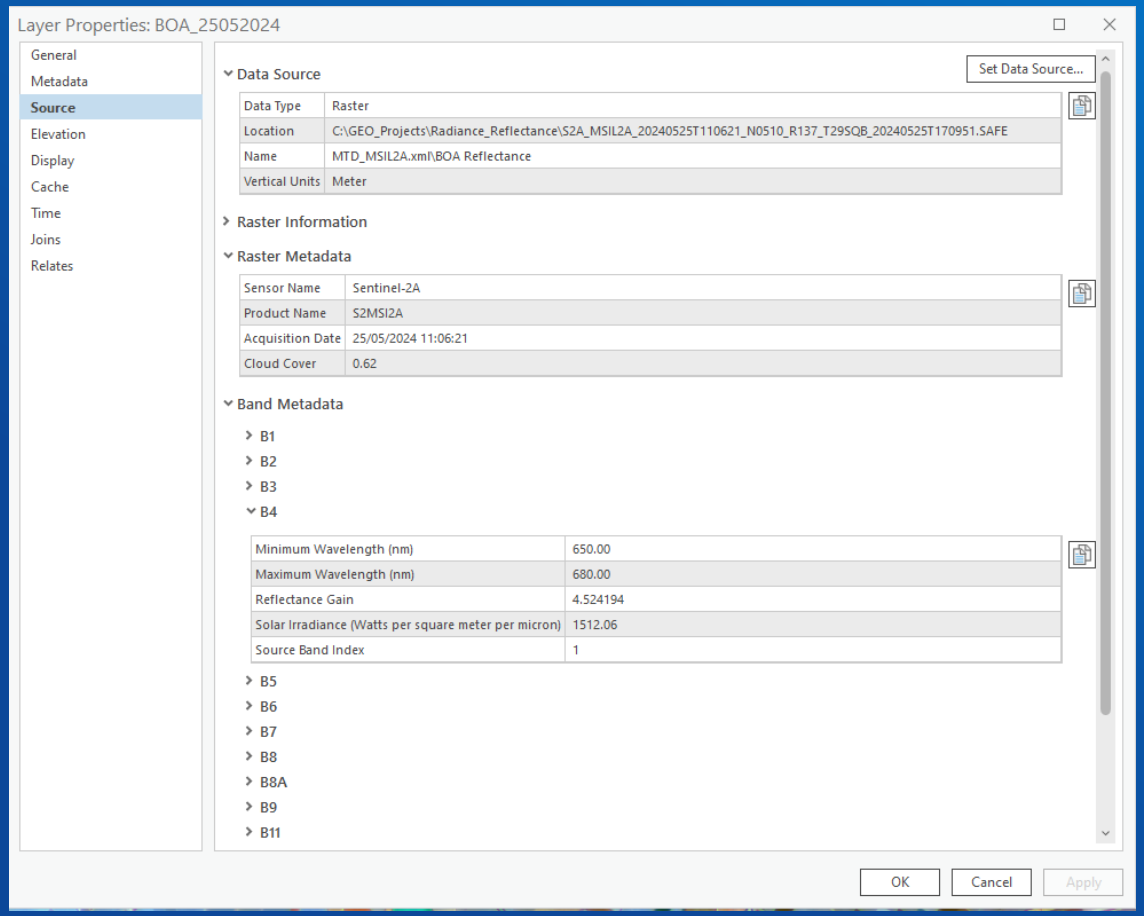
You can further refine how your data is added and processed by defining a processing template. These templates are specific to the type of data in each band. For example, in Landsat 7 ETM+ Level 1 products, there are nine bands. Band 8 is panchromatic and has the highest spatial resolution; bands 1 through 5 and 7 are multispectral bands (blue, green, red, and infrared) and have a lower spatial resolution than the panchromatic band; and bands 6a and 6b are thermal bands. For the Level 1 products, there are six processing templates:

- All Bands–The data is added and grouped in any or all of the processing templates that apply. For example, for Landsat 7, three entries will be created in the mosaic dataset for each scene: Panchromatic (band 8), Multispectral (bands 1 through 5 and 7), and Thermal (bands 6a and 6b). A field is added to the attribute table named `Tag`, which identifies the group each band belongs to, for example, `Pan` for panchromatic, `MS` for multispectral, and `Thermal` for the thermal bands.
- Panchromatic–Only the panchromatic band is added to the mosaic dataset. For example, for Landsat 7, only band 8 is added.
- Multispectral–The multispectral bands are added as a single raster dataset in the mosaic dataset. For example, for Landsat 7, there will be a single row in the mosaic dataset representing bands 1 through 5 and 7.
- Pansharpen–The bands used to generate the pan-sharpened product are added to the mosaic dataset. This generally results in a four-band mosaic dataset using the red, green, blue, infrared, and panchromatic bands.

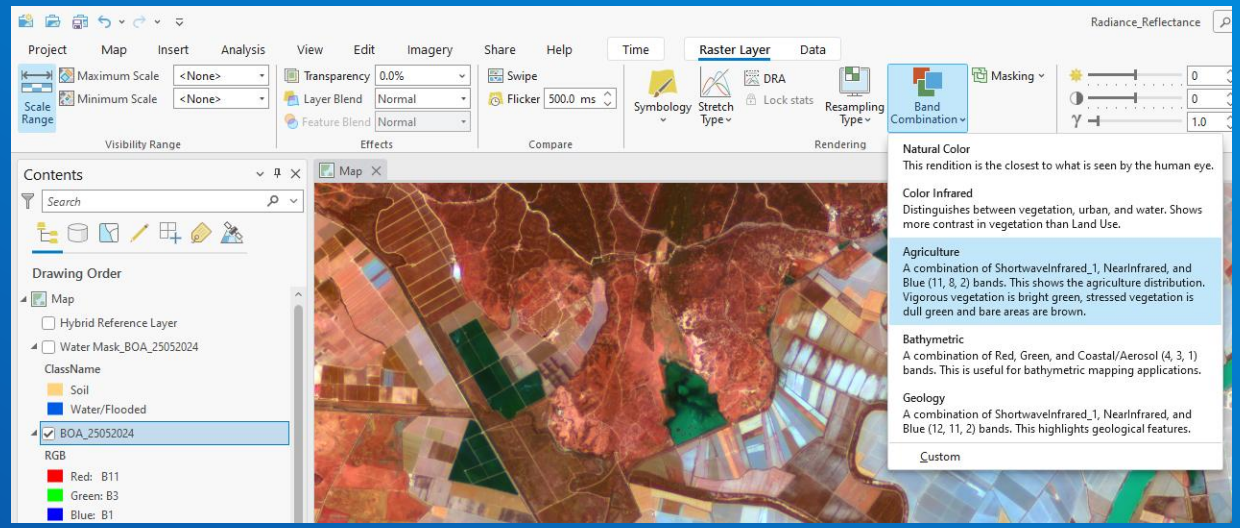
In this topic

Working with Bands and Wavelengths

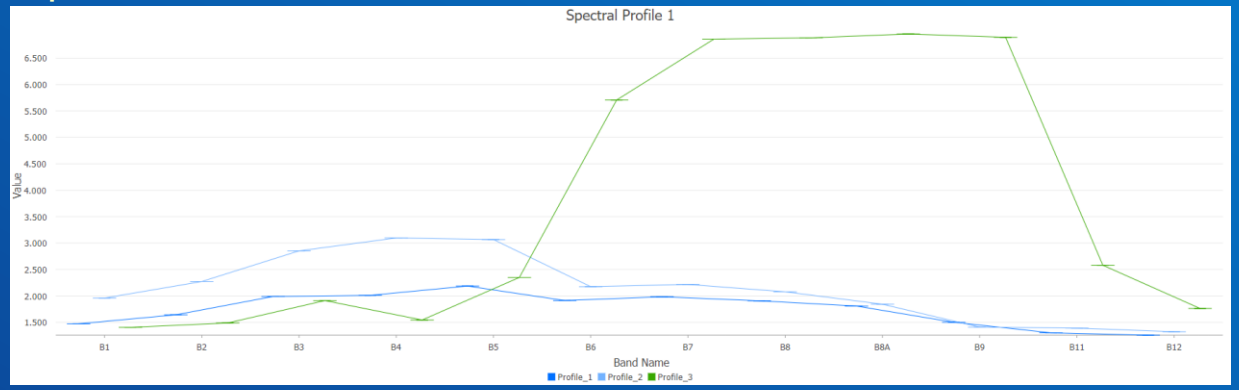
Metadata



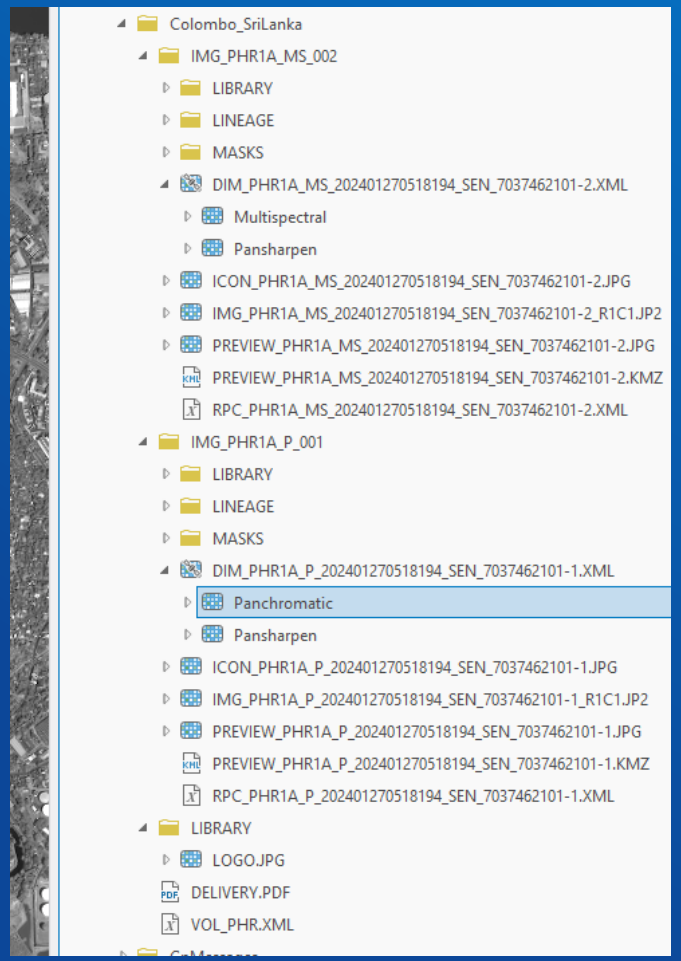
Band Combination



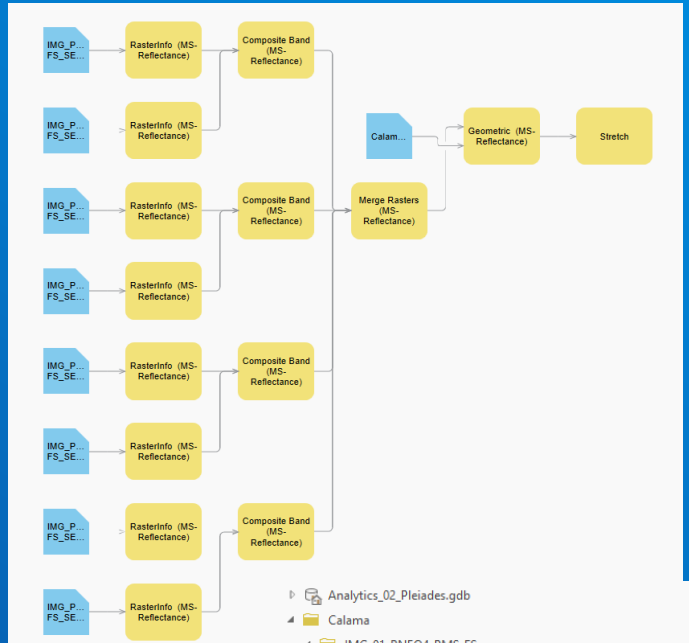
Spectral Profile

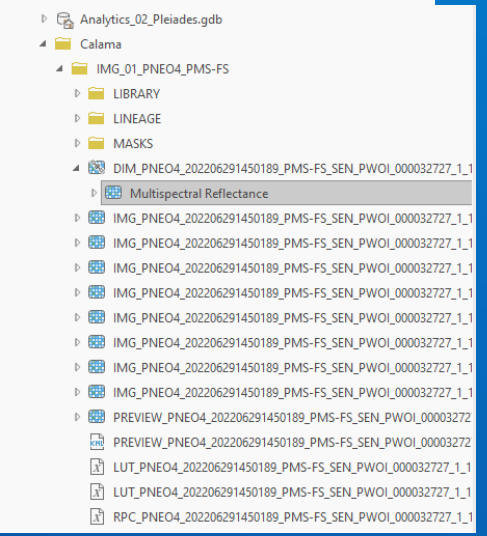


Pléiades Neo as example (Sensor)



Pléiades Neo as example (Multispectral Reflectance)



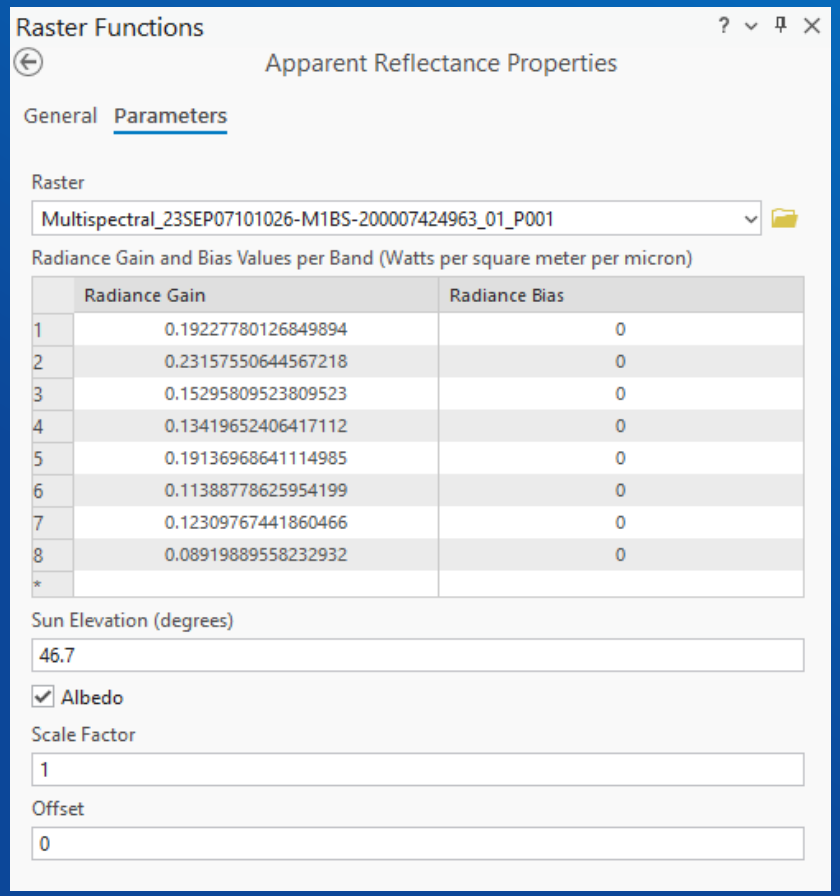


Calibration / Radiance vs. Reflectance

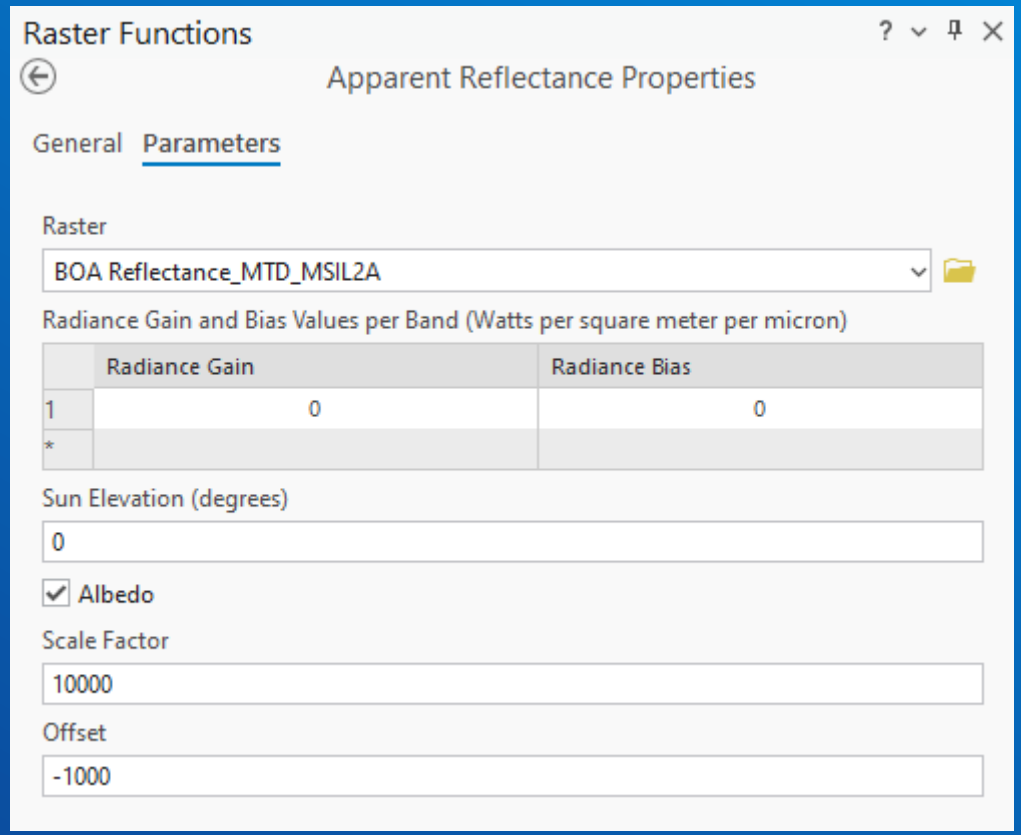
- Radiance is the variable directly measured by remote sensing instruments. We can think of radiance as how much light the instrument “senses” from the object being observed, influenced by illumination and atmosphere.
- Reflectance is the ratio of the amount of light leaving a target to the amount of light hitting the target. It has no units, but since it is a ratio, it ranges from 0 to 1
- Reflectance is a property of the material observed. Radiance, depends on the illumination both its intensity and direction), the orientation and position of the target and the path of the light through the atmosphere
- With effort, many of the atmospheric effects and the solar illumination can be compensated for in remote sensing data. This generate something which is called "apparent reflectance," and it differs from true reflectance in that shadows and directional effects on reflectance have not been dealt with. Many people refer to this as "reflectance.“
- For many applications, radiance, reflectance, and apparent reflectance can be used interchangeably. However, because reflectance is a property of the target material itself, it will yield the most reliable (and repeatable) measurements. Apparent reflectance is adequate in many cases.

Calibration in ArcGIS

In most of the case and we need to include also Landsat and Sentinel-2 data needs to be calibrated to represent "apparent reflectance"

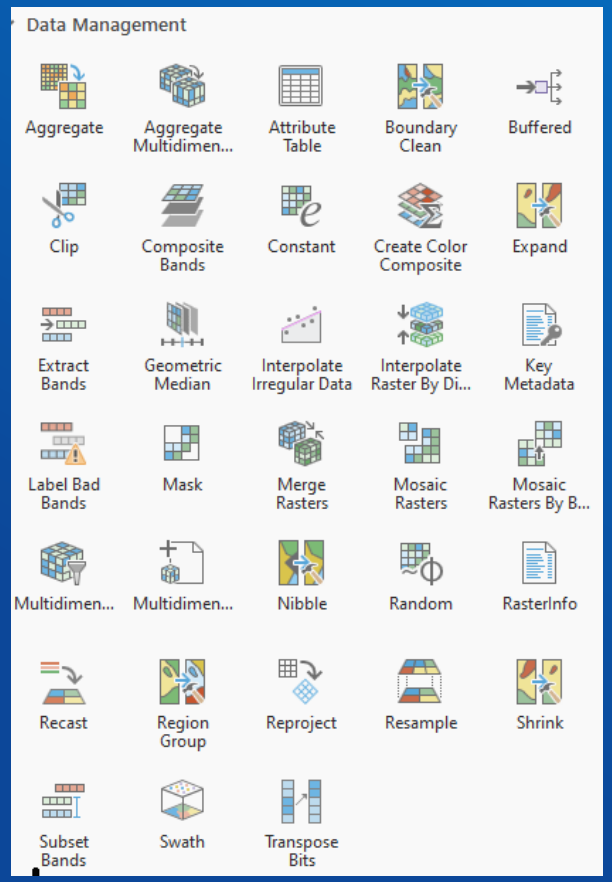


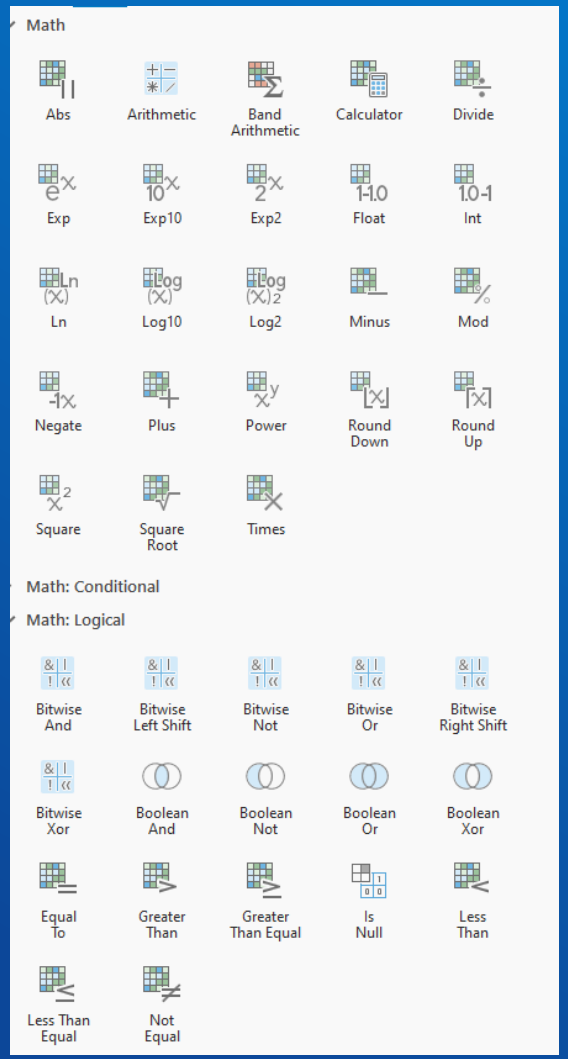
WorldView 2 – System Basic 1B

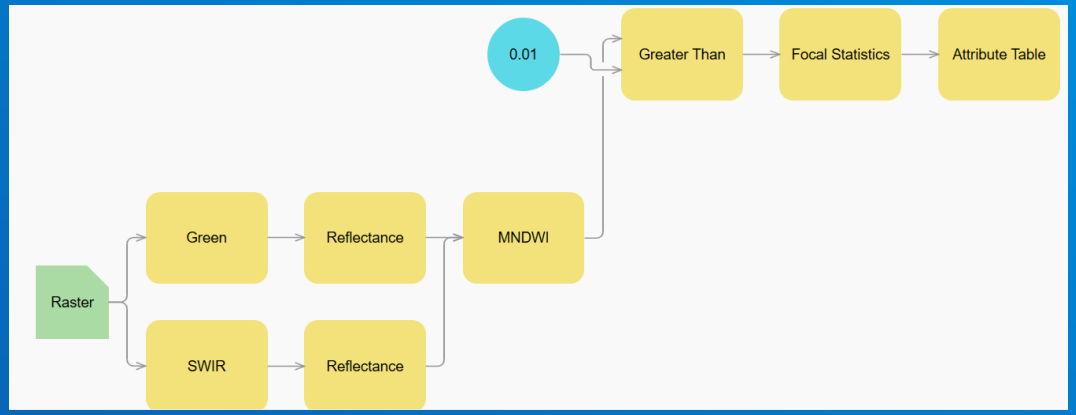


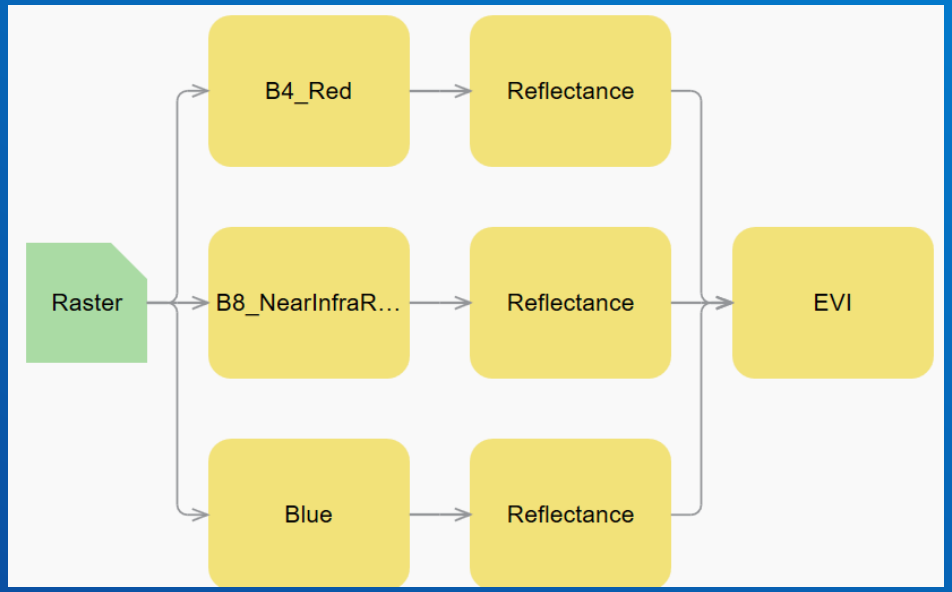
Sentinel 2 – Level 2A

Custom Calibration/Normalization

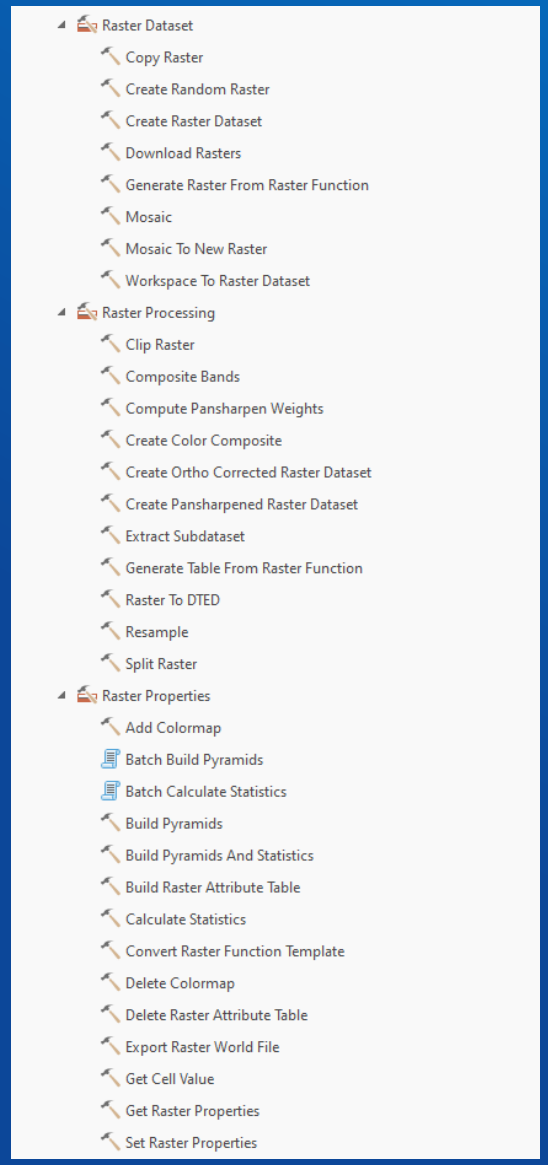


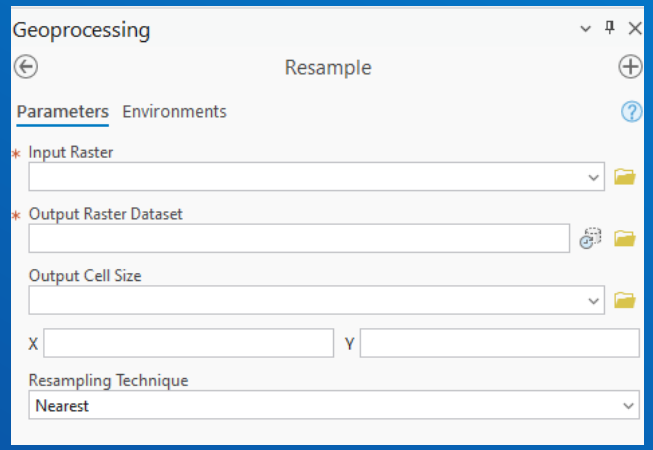


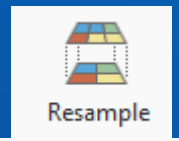


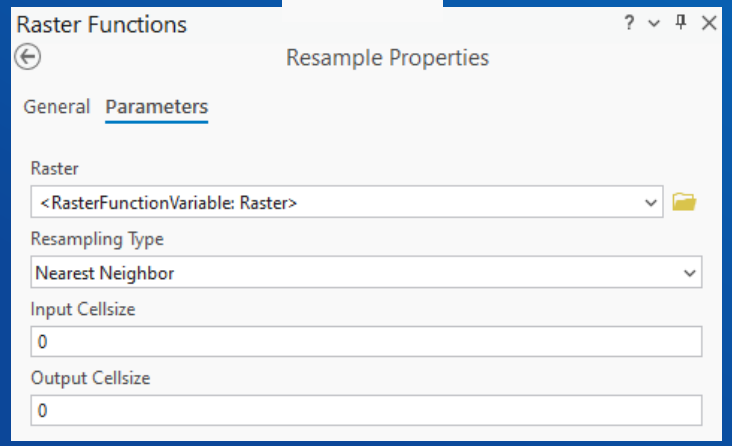


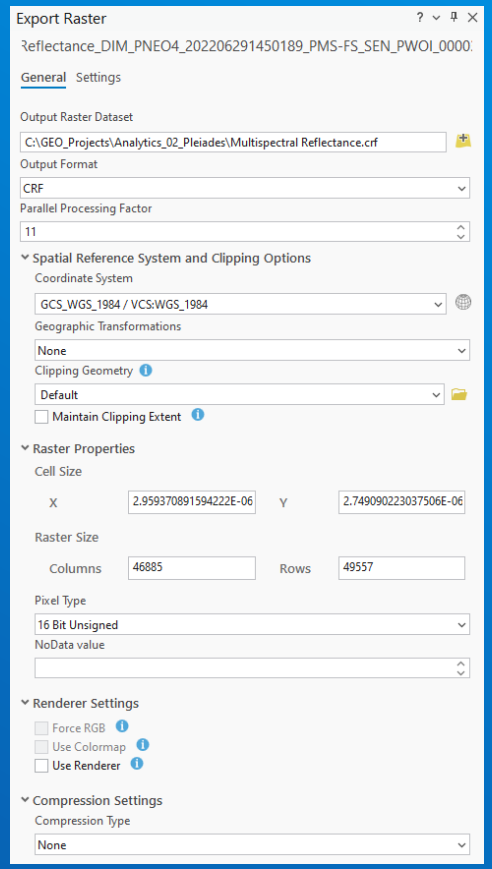
Custom Calibration/Normalization

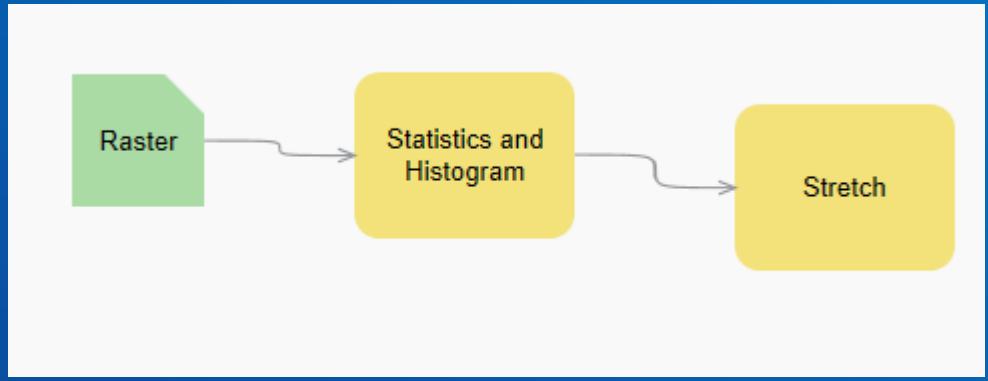












Landsat and other sensors

```
GROUP = LEVEL2_SURFACE_REFLECTANCE_PARAMETERS
  REFLECTANCE_MULT_BAND_1 = 2.75e-05
  REFLECTANCE_MULT_BAND_2 = 2.75e-05
  REFLECTANCE_MULT_BAND_3 = 2.75e-05
  REFLECTANCE_MULT_BAND_4 = 2.75e-05
  REFLECTANCE_MULT_BAND_5 = 2.75e-05
  REFLECTANCE_MULT_BAND_6 = 2.75e-05
  REFLECTANCE_MULT_BAND_7 = 2.75e-05
  REFLECTANCE_ADD_BAND_1 = -0.2
  REFLECTANCE_ADD_BAND_2 = -0.2
  REFLECTANCE_ADD_BAND_3 = -0.2
  REFLECTANCE_ADD_BAND_4 = -0.2
  REFLECTANCE_ADD_BAND_5 = -0.2
  REFLECTANCE_ADD_BAND_6 = -0.2
  REFLECTANCE_ADD_BAND_7 = -0.2
END_GROUP = LEVEL2_SURFACE_REFLECTANCE_PARAMETERS
GROUP = LEVEL2_SURFACE_TEMPERATURE_PARAMETERS
  TEMPERATURE_MAXIMUM_BAND_ST_B10 = 372.999941
  TEMPERATURE_MINIMUM_BAND_ST_B10 = 149.003418
  QUANTIZE_CAL_MAXIMUM_BAND_ST_B10 = 65535
  QUANTIZE_CAL_MINIMUM_BAND_ST_B10 = 1
  TEMPERATURE_MULT_BAND_ST_B10 = 0.00341802
  TEMPERATURE_ADD_BAND_ST_B10 = 149.0
END_GROUP = LEVEL2_SURFACE_TEMPERATURE_PARAMETERS
```

```
<Band_Radiance>
  <BAND_ID>B0</BAND_ID>
  <CALIBRATION_DATE>2017-04-01T16:00:00.000Z</CALIBRATION_DATE>
  <MEASURE_DESC>Raw radiometric count (DN) to TOA Radiance (L). Formulae L=DN/GAIN+BIAS</MEASURE_DESC>
  <MEASURE_UNIT>watt/m2/steradians/micrometers</MEASURE_UNIT>
  <MEASURE_UNCERTAINTY>Specification accuracy value</MEASURE_UNCERTAINTY>
  <GAIN>8.92</GAIN>
  <BIAS>0</BIAS>
</Band_Radiance>
<Band_Radiance>
  <BAND_ID>B1</BAND_ID>
  <CALIBRATION_DATE>2017-04-01T16:00:00.000Z</CALIBRATION_DATE>
  <MEASURE_DESC>Raw radiometric count (DN) to TOA Radiance (L). Formulae L=DN/GAIN+BIAS</MEASURE_DESC>
  <MEASURE_UNIT>watt/m2/steradians/micrometers</MEASURE_UNIT>
  <MEASURE_UNCERTAINTY>Specification accuracy value</MEASURE_UNCERTAINTY>
  <GAIN>8.99</GAIN>
  <BIAS>0</BIAS>
</Band_Radiance>
<Band_Radiance>
  <BAND_ID>B2</BAND_ID>
  <CALIBRATION_DATE>2017-04-01T16:00:00.000Z</CALIBRATION_DATE>
  <MEASURE_DESC>Raw radiometric count (DN) to TOA Radiance (L). Formulae L=DN/GAIN+BIAS</MEASURE_DESC>
  <MEASURE_UNIT>watt/m2/steradians/micrometers</MEASURE_UNIT>
  <MEASURE_UNCERTAINTY>Specification accuracy value</MEASURE_UNCERTAINTY>
  <GAIN>10.1</GAIN>
  <BIAS>0</BIAS>
</Band_Radiance>
<Band_Radiance>
  <BAND_ID>B3</BAND_ID>
  <CALIBRATION_DATE>2017-04-01T16:00:00.000Z</CALIBRATION_DATE>
  <MEASURE_DESC>Raw radiometric count (DN) to TOA Radiance (L). Formulae L=DN/GAIN+BIAS</MEASURE_DESC>
  <MEASURE_UNIT>watt/m2/steradians/micrometers</MEASURE_UNIT>
  <MEASURE_UNCERTAINTY>Specification accuracy value</MEASURE_UNCERTAINTY>
  <GAIN>15.31</GAIN>
  <BIAS>0</BIAS>
</Band_Radiance>
```

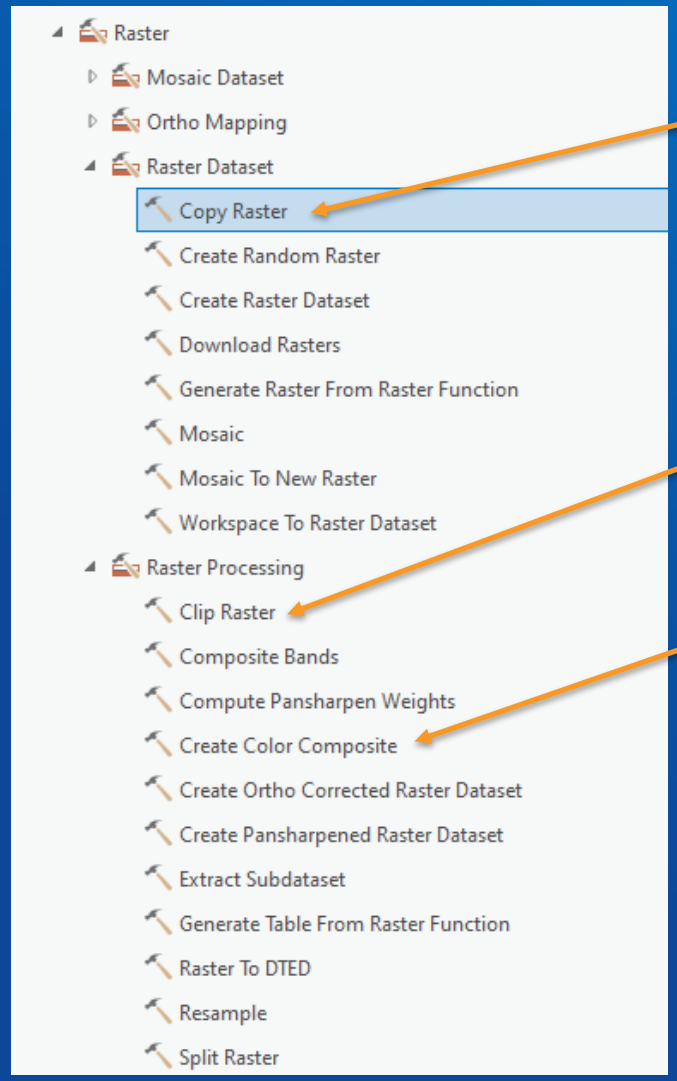
Data Transformation: *Making Data Analysis Easier*

Data transformation helps turn complex data (imagery) into useful information. Here are the main ways this is done:

- **Indices (Spectral Indices):** Indices, like NDVI for plants, combine certain data values to highlight important features and make them easier to find.
- **Band Combinations:** Mixing different data layers, such as visible and infrared, helps reveal patterns and details that might not be obvious in a single layer.
- **Texture Analysis:** Texture analysis looks at how data values change across an area. This can show differences in surfaces or materials that are hard to see otherwise.
- **Temporal Composites:** Combining data from different times helps track changes and spot trends over time.
- **Rescaling and Bit Depth Adjustment:** Rescaling changes the size of data pixels and adjusting bit depth reduces the amount of information stored. Both help make data easier to handle, faster to process, and keep important details.

Using these methods, you can focus on what matters most in your data and make better decisions based on clear information.

Toolboxes -> Data Management Tools

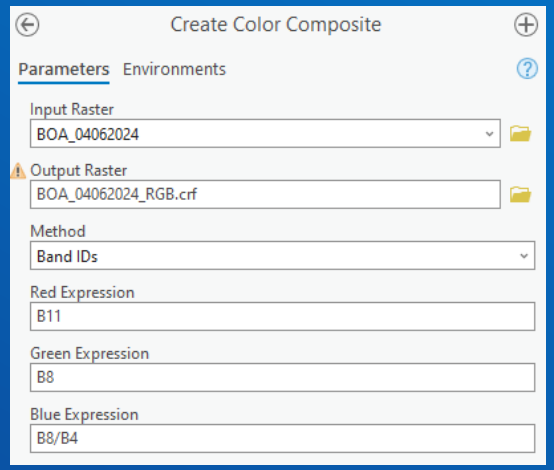


Saves a copy of a raster dataset or converts a mosaic dataset into a single raster dataset.

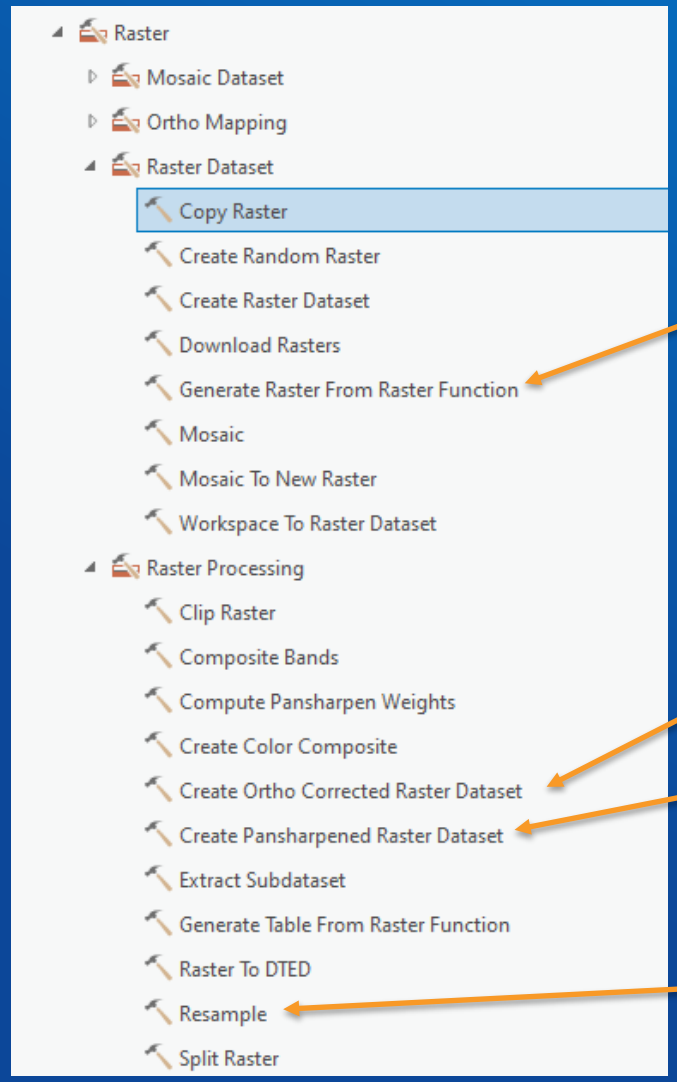
You can save the output to BIL, BIP, BMP, BSQ, COG, CRF, ENVI DAT, ERDAS IMAGINE, GIF, JPEG, JPEG 2000, MRF, NetCDF, PNG, TIFF, or Esri Grid format or to any geodatabase raster dataset.

Cuts out a portion of a raster dataset, mosaic dataset, or image service layer.

Creates a three-band raster dataset from a multiband raster dataset.



Toolboxes -> Data Management Tools



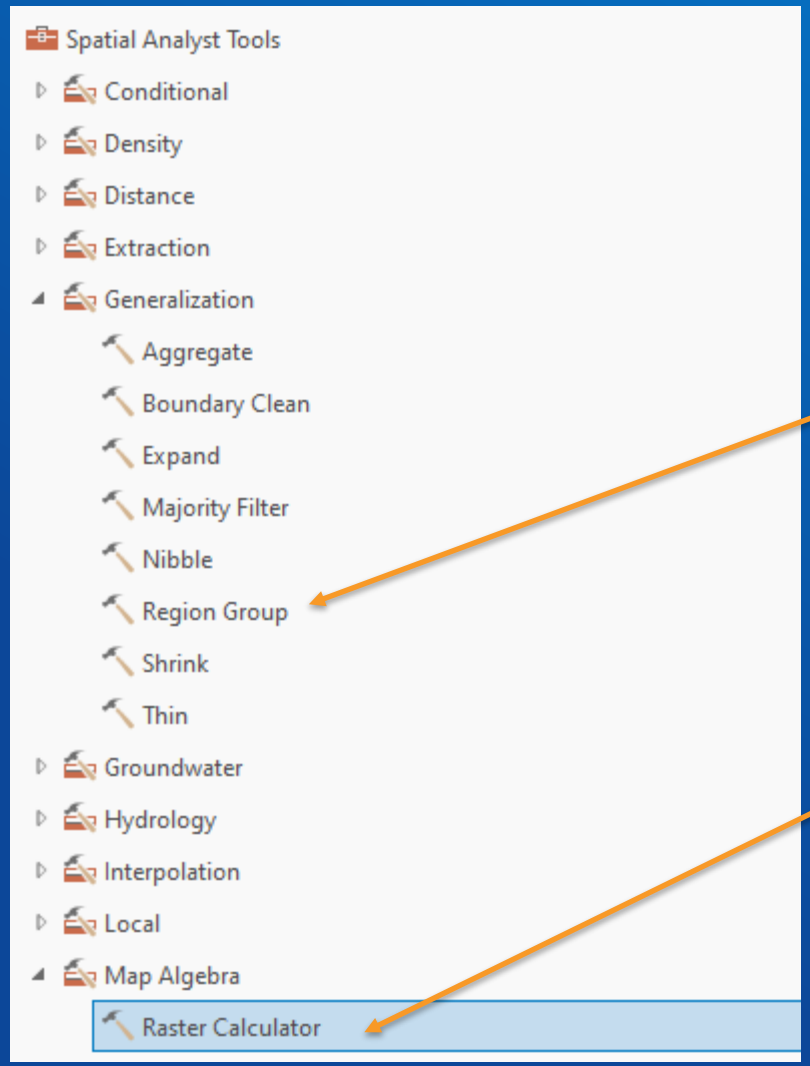
Generates a raster dataset from an input raster function or function chain.

Creates an Ortho corrected raster dataset using a digital elevation model (DEM) and control data to accurately align imagery.

Combines a high-resolution panchromatic raster dataset with a lower-resolution multiband raster dataset to create a high-resolution multiband raster dataset for visual analysis.

Changes the spatial resolution of a raster dataset and sets rules for aggregating or interpolating values across the new pixel sizes.

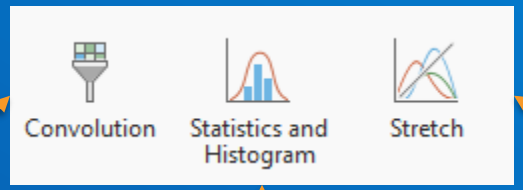
Toolboxes -> Spatial Analyst Tools



Generates a raster dataset from an input raster function or function chain.

Build and run a single map algebra expression using Python syntax.

Raster Functions

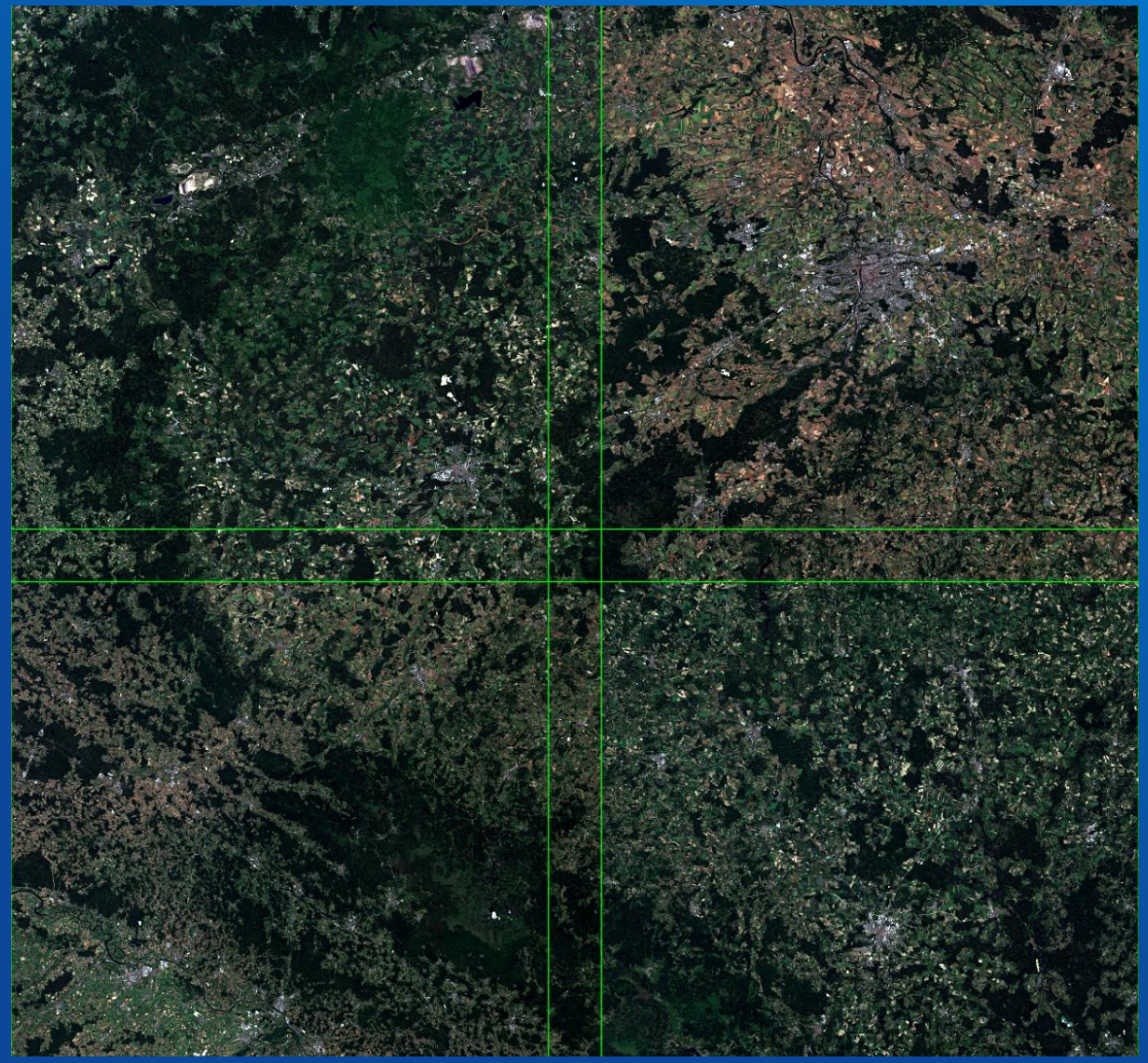


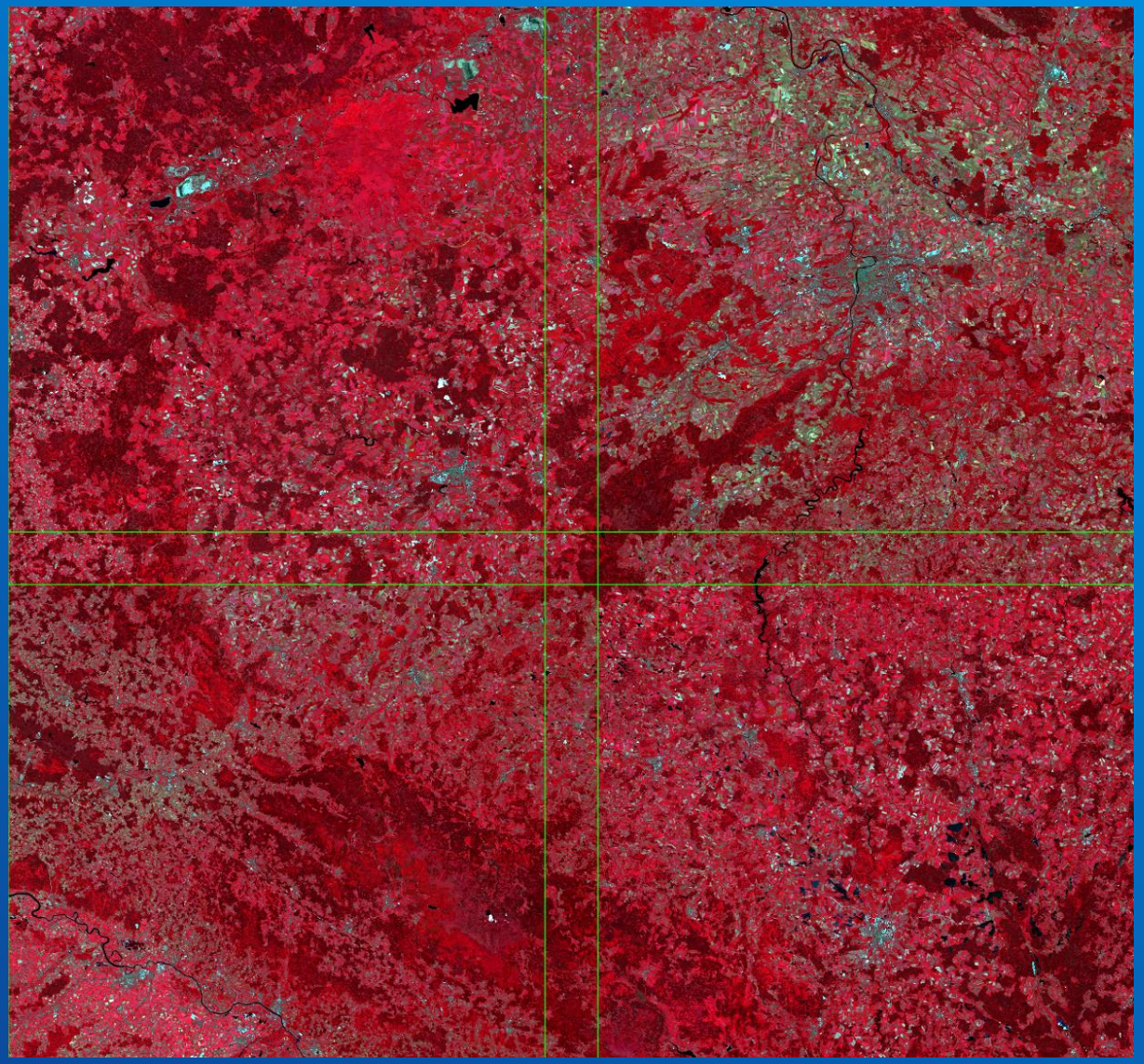
The Convolution function performs filtering on the pixel values in an image, which can be used for sharpening an image, blurring an image, detecting edges within an image, or other kernel-based enhancements.

The Stretch function enhances an image by changing properties such as brightness, contrast, and gamma through multiple stretch types. Additionally, it may be used to reduce the bit depth preserving the quality of the image

The Statistics and Histogram function is used to define the statistics and histogram of a raster. You can insert this function at the end of the function chain to describe the statistics and histogram of a raster function template (RFT). This may be needed to control the default display of the processing result, especially when defining a function chain that contains many functions.

Create a Sentinel-2 mosaic using Mosaic Dataset





STAC

Spatio Temporal Asset Catalog

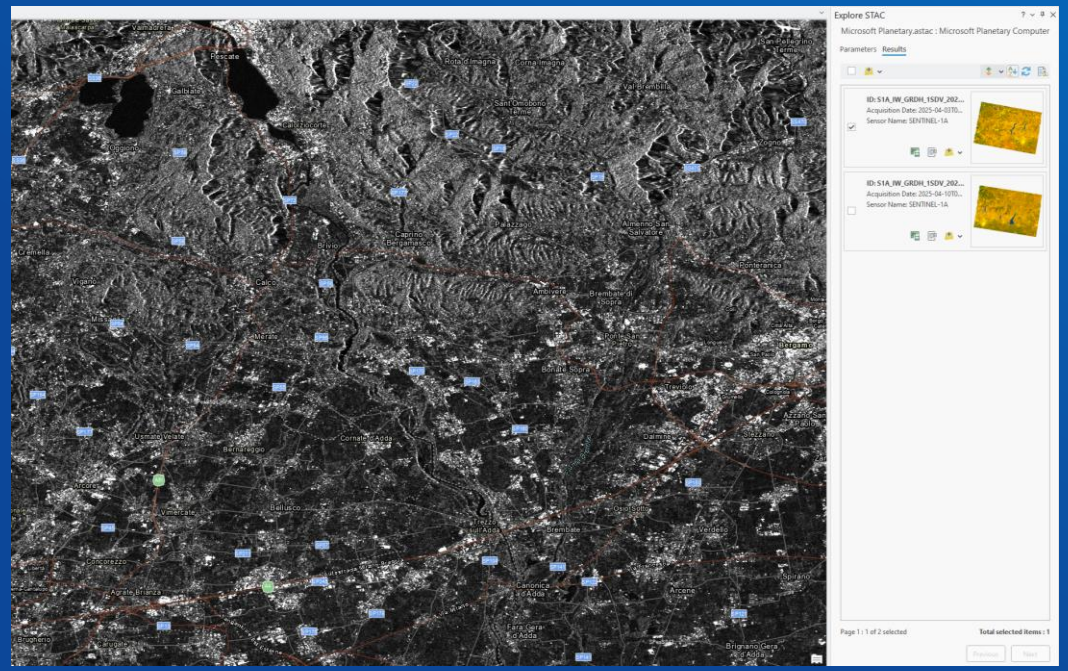
What is STAC?

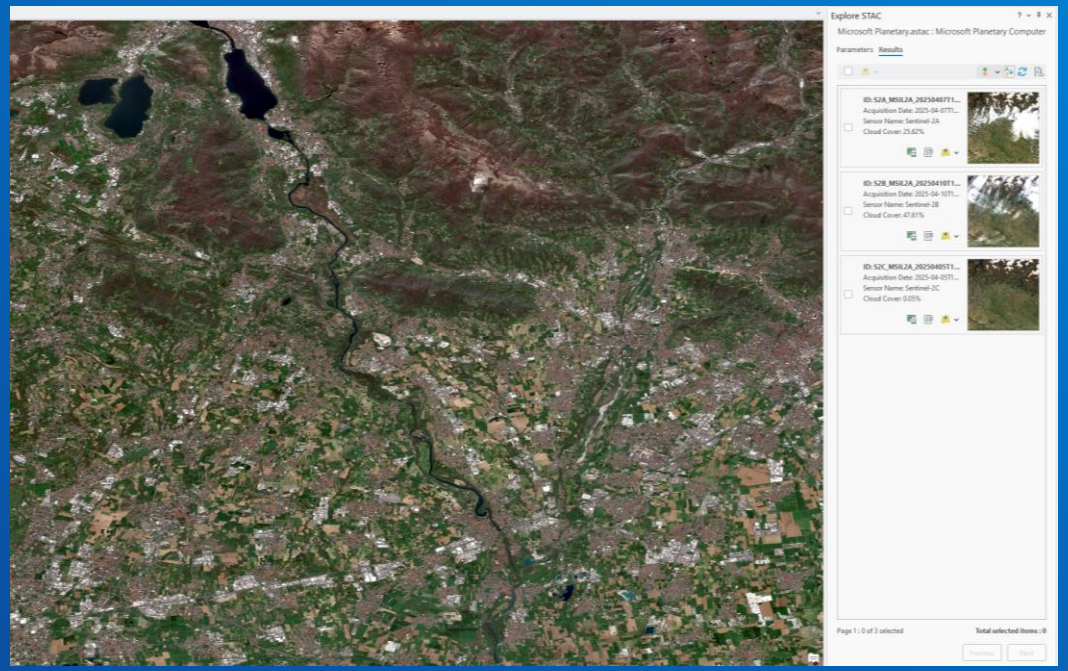
The SpatioTemporal Asset Catalog (STAC) is an open specification to increase the interoperability of searching for ~~satellite~~ imagery. When a user wants to search for all the imagery in their area and time of interest they can't make just one search—they must use different tools and connect to API's that are similar but all slightly different.
STAC aims to make that much easier, by providing common metadata to expose geospatial assets.

The goal of STAC is to enable a global index of all imagery (satellite, aerial, drone, etc), derived data products and alternative geospatial captures (LiDAR, SAR, Full Motion Video, Hyperspectral and beyond). STAC focuses on an easily implementable standard for organizations to expose their data in a persistent and reliable way.

SpatioTemporal Asset

A SpatioTemporal Asset is any file that represents information about the Earth captured in a certain place and at a particular time. Examples include spatiotemporal data derived from imagery (from satellites, airplanes, and drones), Synthetic Aperture Radar (SAR), point clouds (from LiDAR, structure from motion, etc.), data cubes, and full-motion video.





Sentinel-1 & Sentinel-2 from Microsoft Planetary STAC Catalog

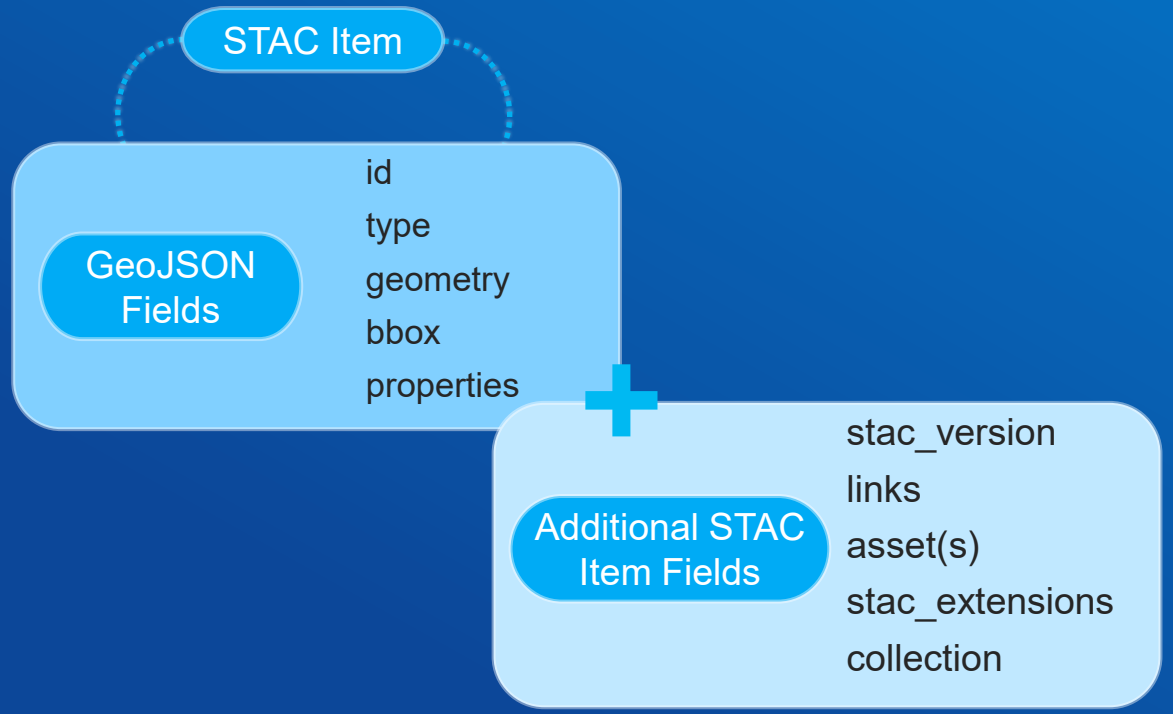
STAC Components

The core of the SpatioTemporal Resource Catalog specification consists of three components. These components are:

- **Item** - A STAC Item represents one or more spatiotemporal assets as GeoJSON so that it can be easily searched.
- **Catalog** - The STAC Catalog specification provides structural elements to group STAC Items and Collections.
- **Collection** - STAC Collections are catalogs that add more required metadata and describe a group of related items.

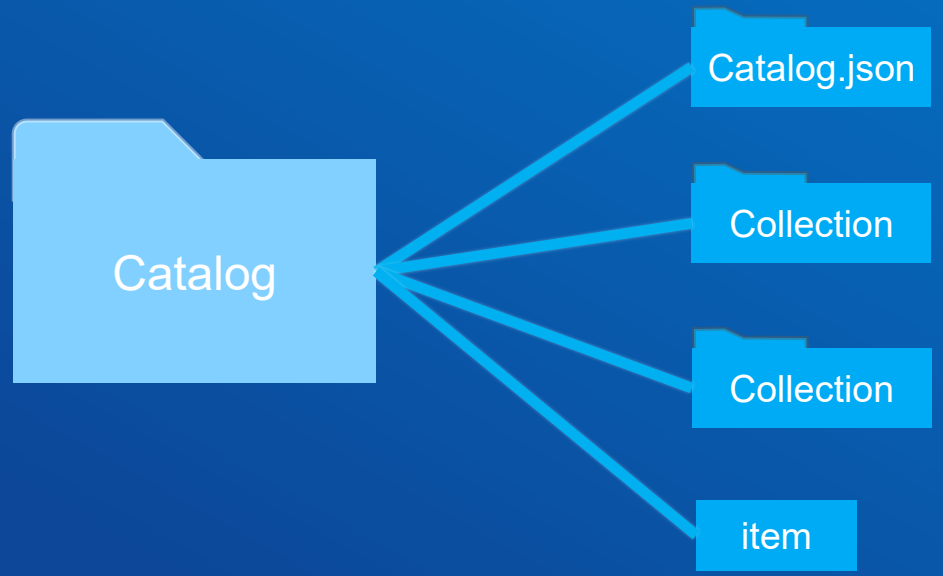
STAC Item

A STAC Item is the foundational building block of STAC. It is a GeoJSON feature supplemented with additional metadata that enables clients to navigate through catalogs. Since an item is a GeoJSON, it can be easily read by any modern GIS or geospatial library. One item can describe one or more SpatioTemporal Asset(s). For example, a common practice of using STAC for imagery is that each band in a scene is its own STAC Asset and there is one STAC Item to represent all the bands in a single scene.



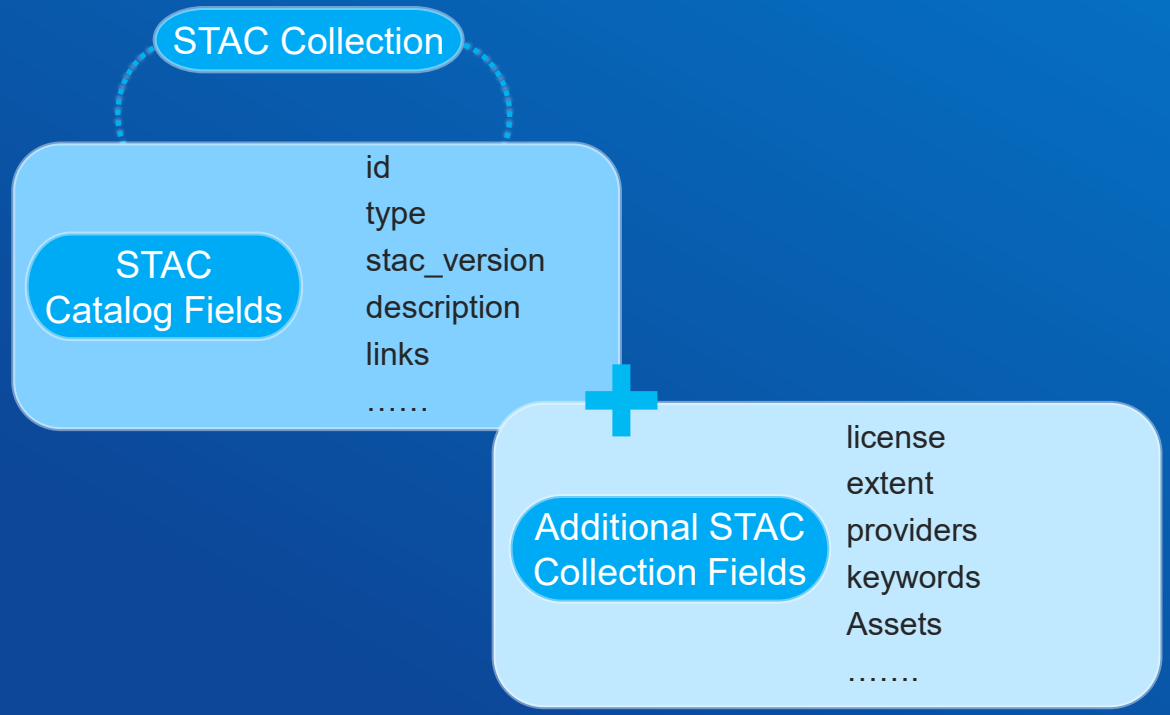
STAC Catalog

A Catalog represents a logical group of other Catalog, Collection, and Item objects. These Items can be linked to directly from a Catalog, or the Catalog can link to other Catalogs (often called sub-catalogs) that contain links to Collections and Items. The division of sub-catalogs is up to the implementor but is generally done to aid the ease of online browsing by people.
A Catalog object is typically the entry point into a STAC catalog. Their purpose is discovery: to be browsed by people or be crawled by clients to build a searchable index.

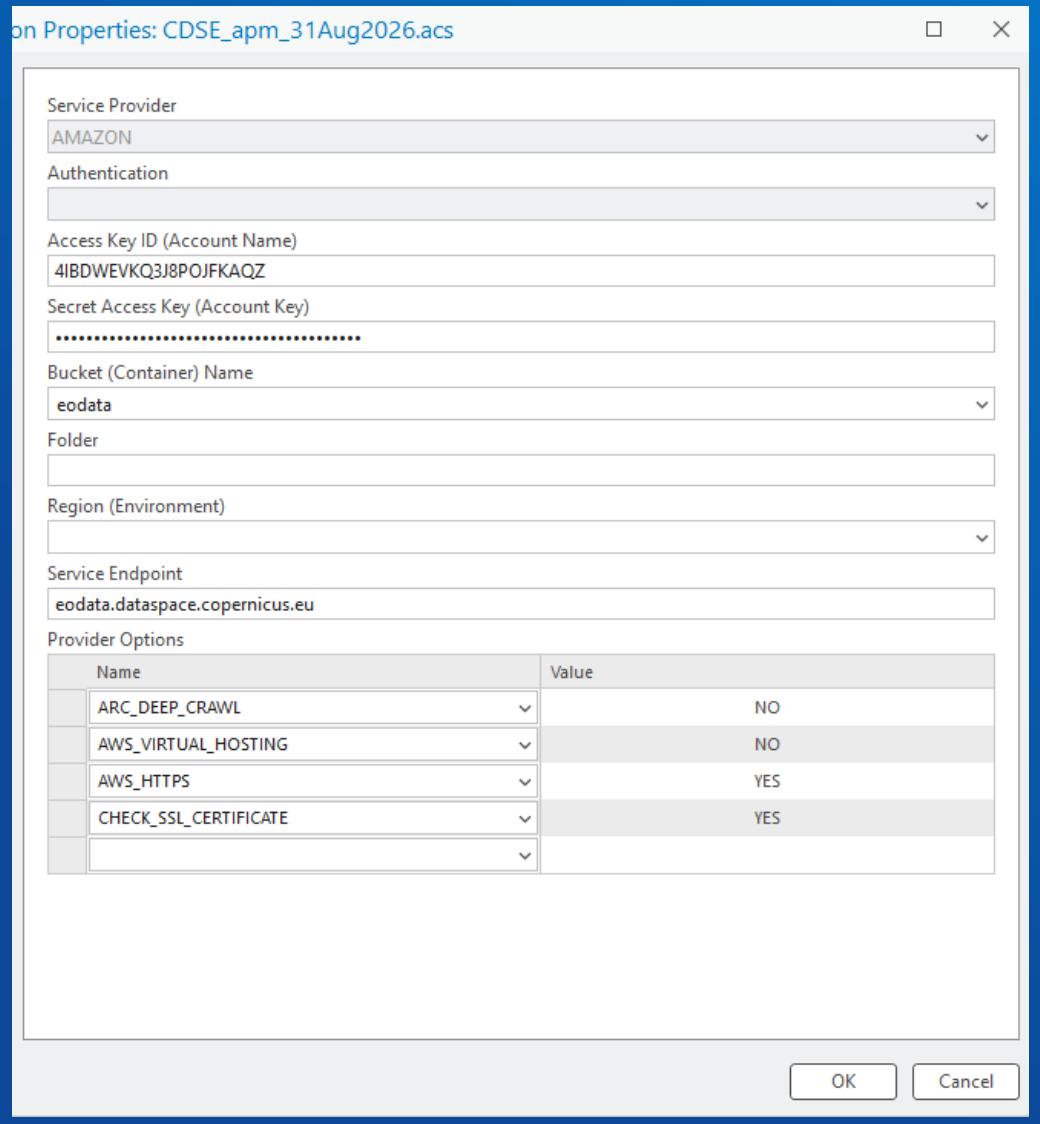


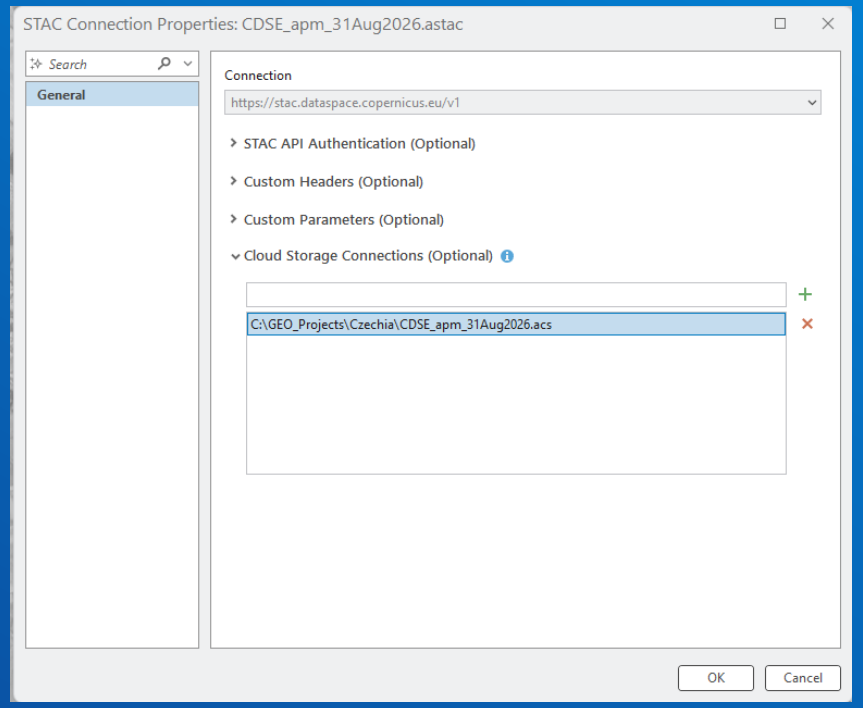
STAC Collection

A STAC Collection builds upon the STAC Catalog specification to include additional metadata about a set of items that exist as part of the collection. It extends the parent catalog by adding additional fields to enable the description of information like the spatial and temporal extent of the data, the license, keywords, providers, etc. Therefore, it can easily be extended for additional collection-level metadata.



DEMO – Copernicus Data Space Ecosystem





DEMO with STAC

A Time Series / Image cube


esri
THE SCIENCE OF WHERE