AOIs data pre-selection and processing

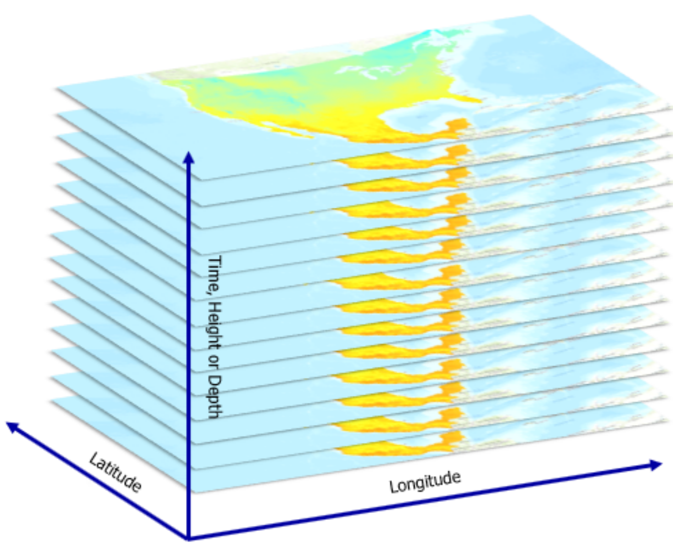
June 2021

The Areas Of Interest (AOI) feature of the Half-Earth Project aims to move the Half-Earth map into the direction of becoming more “tool-like” and promote local action. Their goal is that users understand the conservation priorities of their areas of interest and share these opportunities and challenges with key stakeholders whilst visualizing how protecting this area impacts on the Half-Earth Goal.

Currently the Half-Earth map shows spatial information about biodiversity, protection and human encroachment. This is completed by the social information that is found in the National Report Cards. Finally, geodescriber provides information on the environmental context. The AOIs will cover these topics using the available data at high spatial resolution.

High resolution data is available as raster information or as rasterized polygons. The rasters are processed to be stored as crf files in Azure storage, also referred as data cubes (more about multidimensional data in ESRI). The structure of the crf files will be determined by the specific questions that the AOIs will have to answer. For each data cube there is a lookup table. The lookup tables store further information of the variables that have been pre-calculated.

This document presents the available data and the workflow to create crf files.



Schematic representation of a Crf file. The datacube is made out of stacked 2D layers (longitude and latitude) each containing information of a variable (Time, Height, Depth, Species). Image credit: ESRI

Data pre-selection

The data to use in the AOIs must be of high spatial resolution (< 1 km^2) and global coverage. However, initially we will focus on the terrestrial realm.

Human pressures

The current Global Human Modification data will be served as a crf with four dimensions: Urban, Rangeland, Rainfed Agriculture and Irrigated Agriculture. MOL is recalculating the marine human pressures and eventually the marine and terrestrial layers will match their grid. The data cubes will be updated as the new data is produced.

Land cover use change:
https://www.arcgis.com/home/item.html?id=1453082255024699af55c960bc3dc1fe

Protection

The polygon and point data from the WDPA will be processed by MOL to match the input of the data in the prioritisation analysis and SPI calculations. The polygon layer will be rasterized as a binary layer. The protection will be stored as a data cube with a single dimension.

Biodiversity

Biodiversity information will be stored in datacubes where each dimension is a species range binary coded. For this first development of the AOIs, we will use range-refined data of terrestrial mammals, reptiles, amphibians and birds. Once the ranges derived from Species Distribution Models will be available the data cubes will be replaced.

The lookup tables of the biodiversity data cubes will store the binomial names of the species, their global protection target, their Species Protection Score per country and if they have been identified as a flagship species by McGowan et al (2020). Their global protection target and Species Protection Score will be updated as new analyses are performed by MOL.

Environmental

Environmental data will be used to describe the habitat in an Area of Interest. Chelsa contains daily temperature and precipitation information. Initially, the datacube could be the raw information, anomalies or summaries to reduce the number of dimensions.

The global location of cloud forest during the past 18 years has recently been published and is available. In this case the dimension would be each of the 18 years. The coverage is global, but the data only appears in the tropical areas where it exists.

Ecological Land Units inform on the geology, climate, vegetation cover and land formation. Initially this datacube will be of a single dimension.

State of habitats (Gosling et al. 2020) is a single dimension dataset that differentiates between Likely Natural Habitat, Potential Natural Habitat, Potential Modified Habitat and Likely Modified Habitat.

Socio-economic

Socio Economic information is often summarised at administrative level and it is a challenge when aiming at allowing any area of the world to be summarised. Esri's Population (2016) is an image raster that could be converted into a single dimension data cube.

Administrative information is available as polygons (GADM 3.6), the information can be rasterized and have a data cube with as many dimensions as administrative levels. Further information available at administrative level (e.g. SPI, GNI) needs to be explored and could be stored in different data cubes. This information wouldn't be precise information on the area of interest but contextual socio-economic information of the admin in which the area of interest is located.

Creating crf files for the Half-Earth Project

The Half-Earth Project Map uses crf files to source the Areas of Interest. Crf files are multidimensional raster files, they can be visualised as datacubes. The x and y dimensions are spatial whilst the z dimension is often used for depth or time. In the Half-Earth Project we use the z dimension in biodiversity data to store a unique species.

From the official docs:

If you have a series of images or rasters collected over time (and depth or height), you must:

1. first create a mosaic dataset and
2. use the Add Rasters To Mosaic Dataset tool to add the time series of images or rasters to the mosaic dataset.
3. Next, use the Build Multidimensional Info tool to generate variable and dimension information for the mosaic dataset.
4. Finally, you can use the Copy Raster tool, set the output format to CRF and check the box to process the data as multidimensional, and optionally build a transpose.

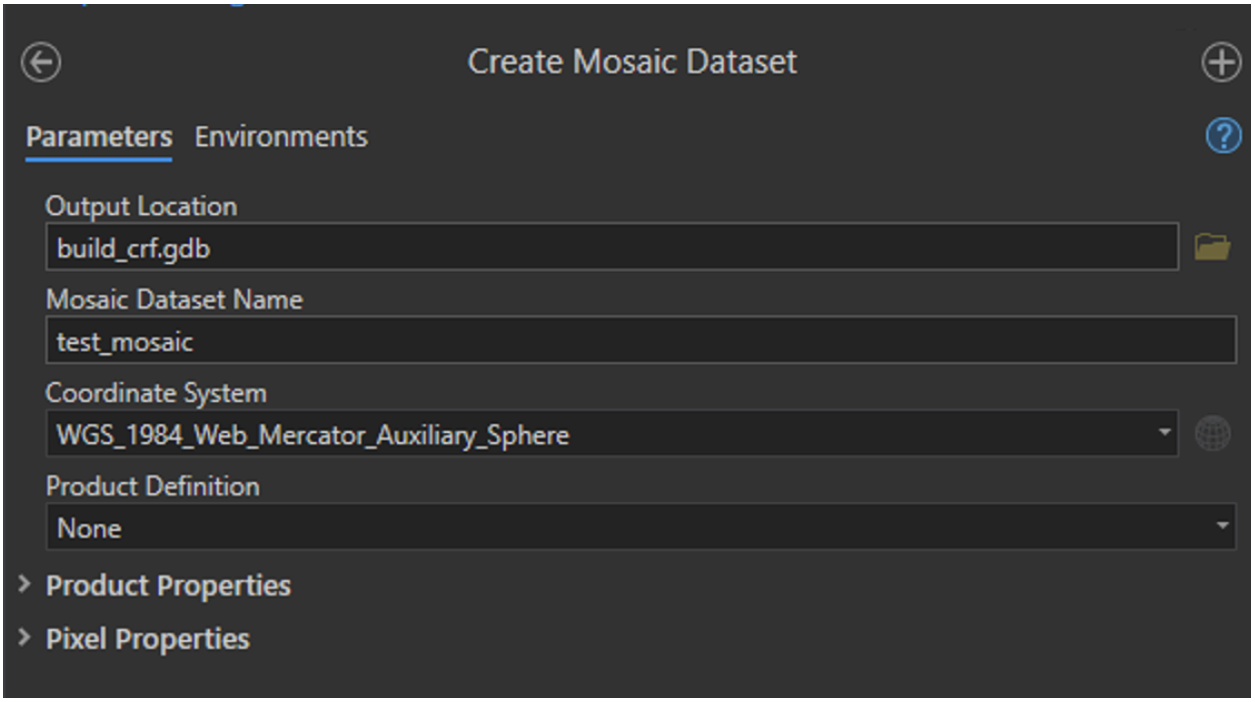
TIP: Do it with python once you understand the workflow. notebook

HE's specific workflow:

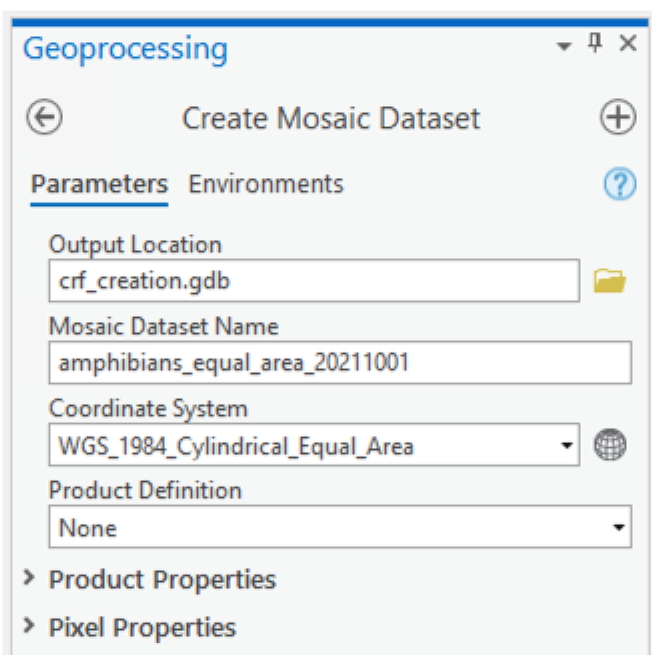
Set the environment to use Parallel Processing Factor at 90%.

Create a mosaic dataset

The tool `Create mosaic dataset` creates a container inside a gdb. The coordinate system across the application is WGS 1984.



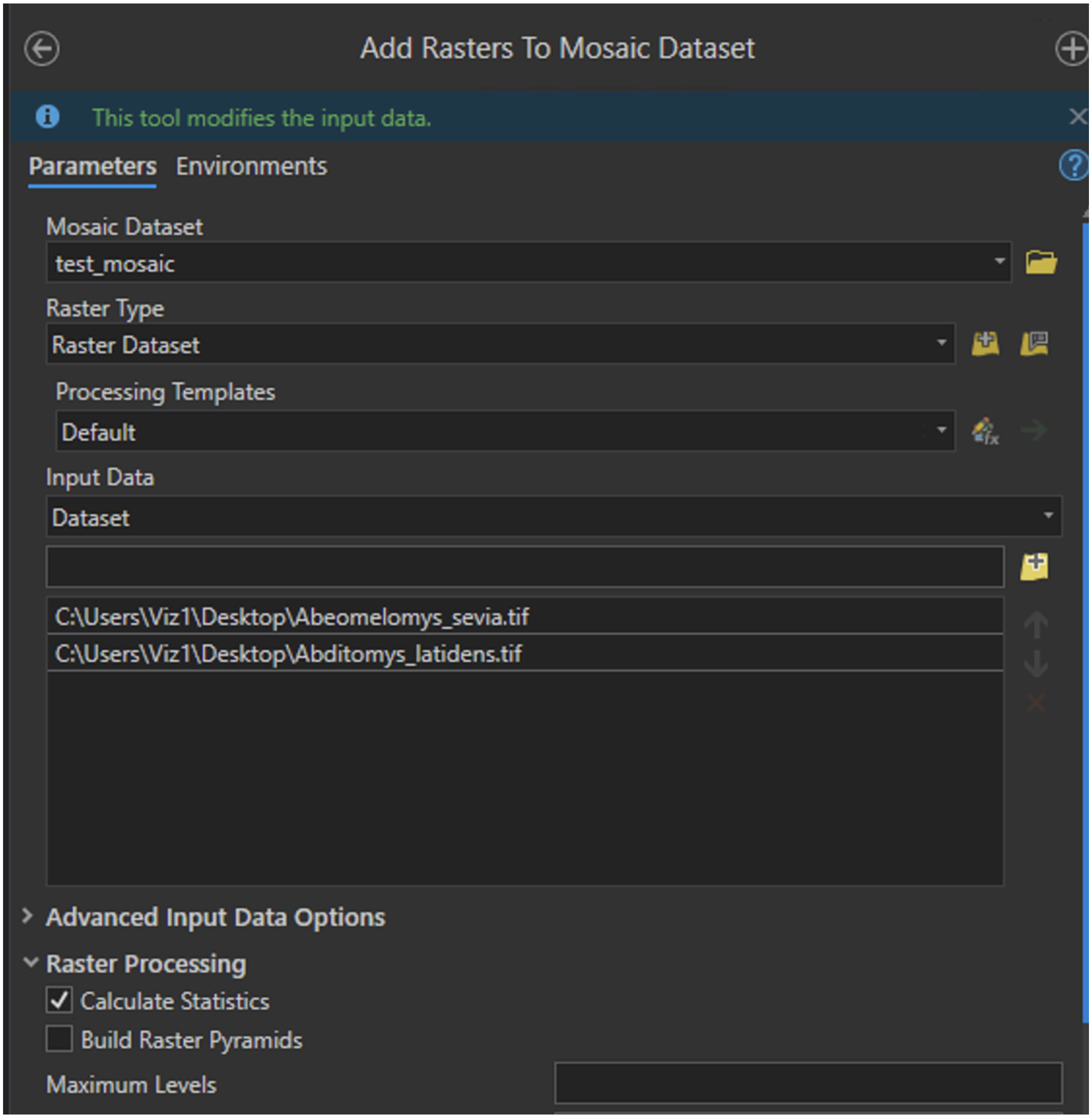
Changing the Projection of final crfs: Another version of the crfs was created with Equal Area Projection. In this case, we used one of the tifs that would be later added to the crf cube as template for the crf coordinate system (To do this, you first need to add the .tif to the map working space, and then selecting that tif in the Coordinate System box dropdown).



Add Rasters to Mosaic Dataset

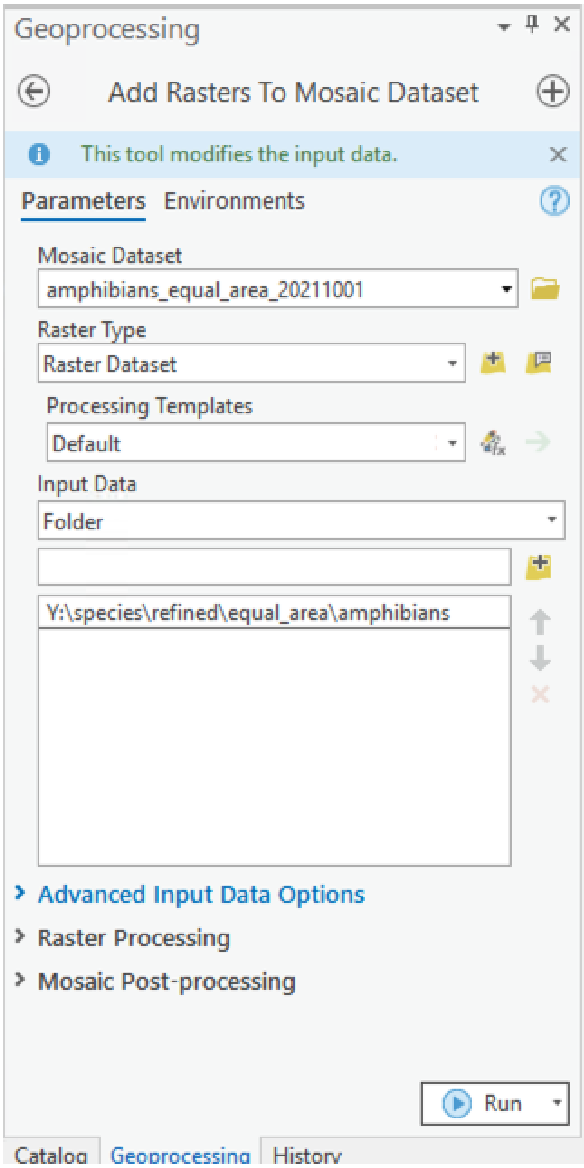
The tool `Add rasters to mosaic dataset` has the following parameters

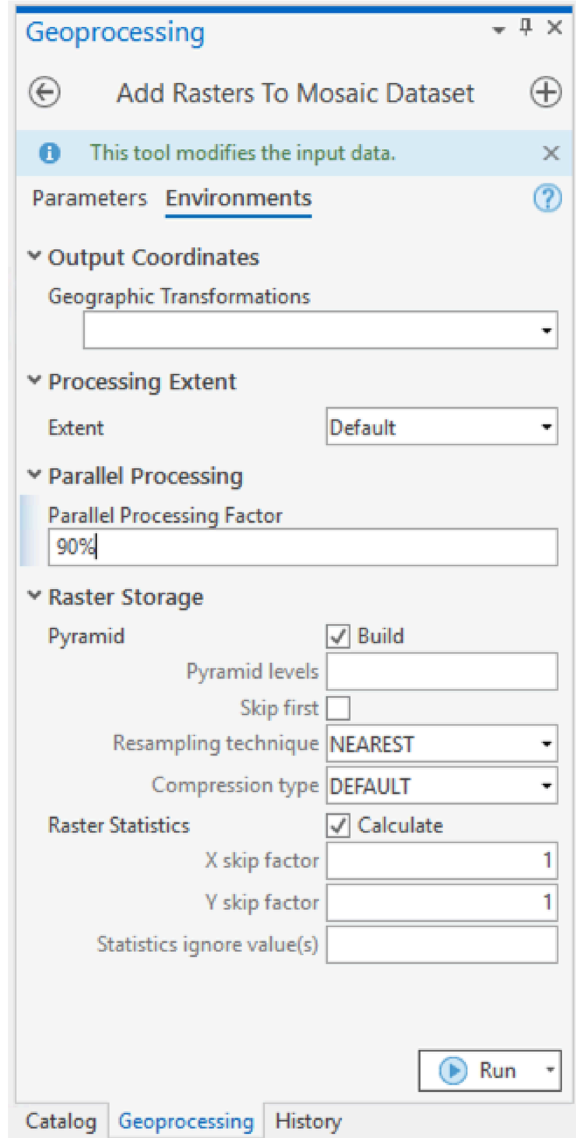
- Mosaic dataset: the name of Mosaic dataset we created in the previous step
- Raster Type: Raster Dataset
- Input data is “dataset” if they are rasters, select the rasters. If you don’t want to dump 1000 rasters in the list, you can choose to select a folder that contains the rasters. This will work recursively from folder to folder.
- Under Raster Processing, select “Calculate Statistics” (if you forget, that is ok, there is a tool to calculate it later). The rest is by default. What this tool does is to figure out the pixel size and spatial extent to create a table. That acts as an index.
- You can set the processing factor to 90% in the environment



You can open the mosaic as a table and see that it is a table with the name of the raster files (footprint table). The columns are used internally to figure out the display. This is not multidimensional at this point. We are going to add multidimensional information to this table. Normally the multidimensional information is time or depth.

The following screenshots show an example adding a folder that contains all the rasters inside:

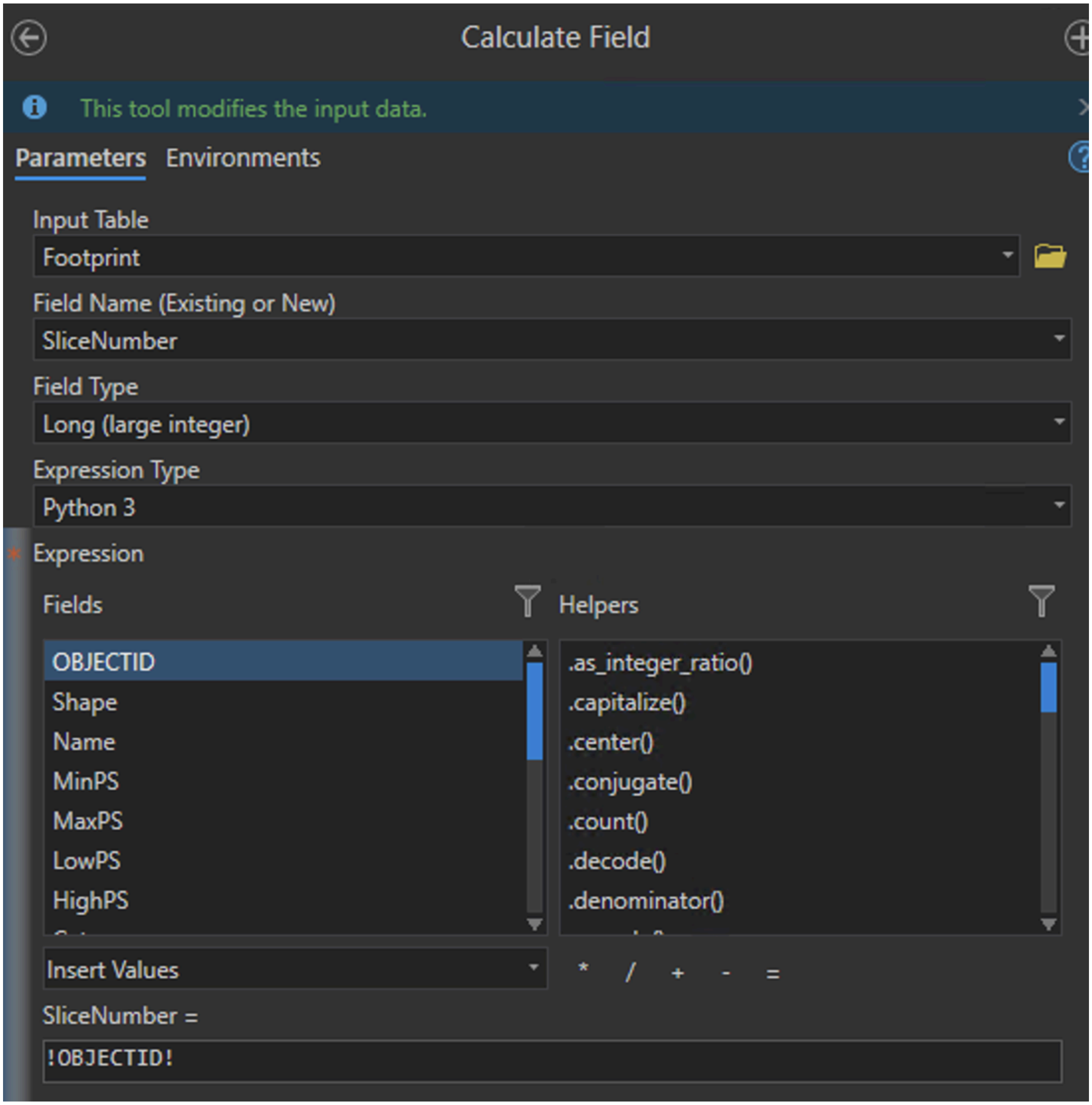




Add and calculate field in footprint table

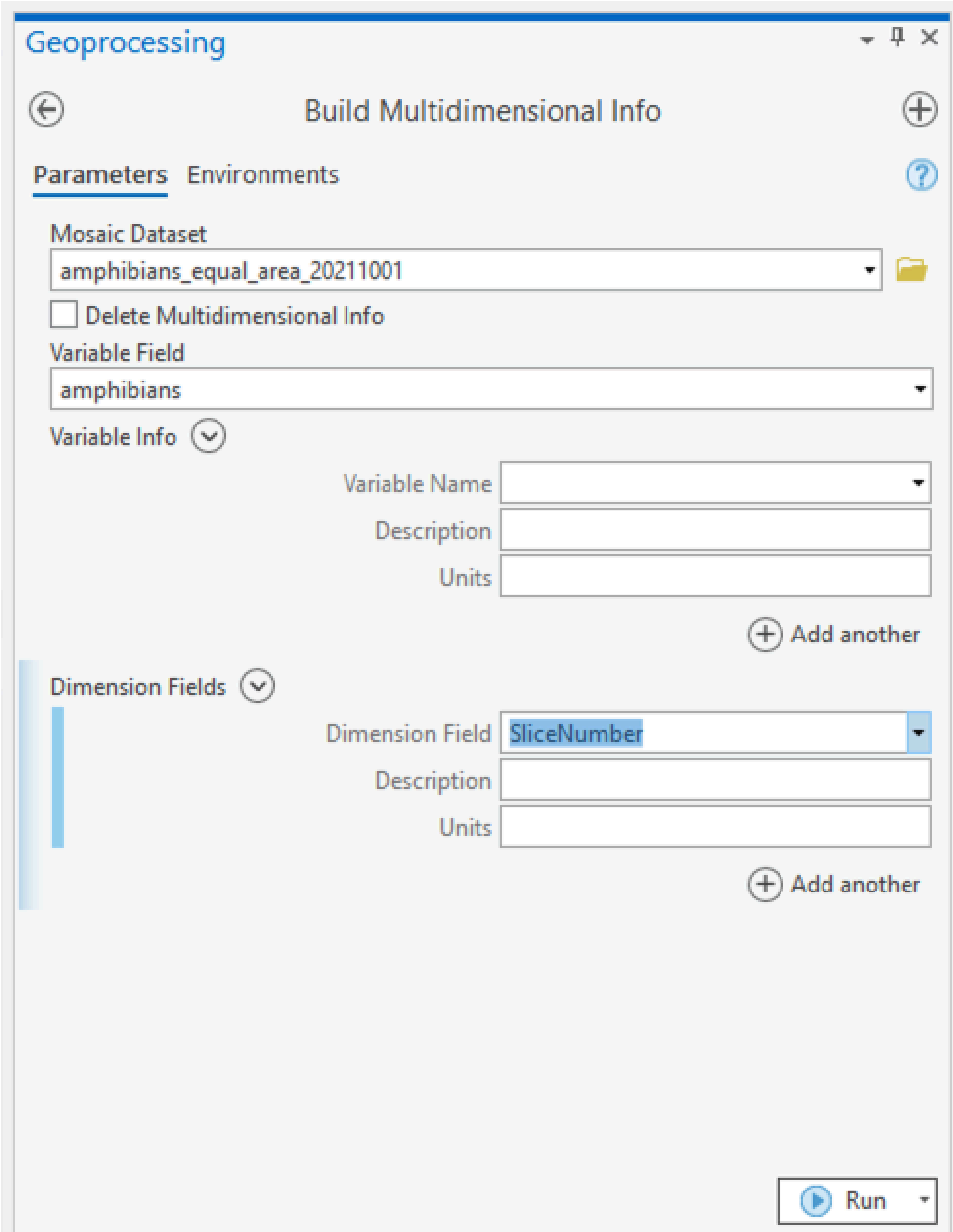
We are going to make up a field that acts as a multidimensional index in the footprint table. `Add field` of the type LONG that is numerical. `Calculate field` to provide a unique number using the Object id field. Both steps can be done with `Calculate Field`.

ATTENTION: It is possible that the tool does not keep the type of the field and it should be checked. Check by opening the Footprint table and checking that the added field is of type long



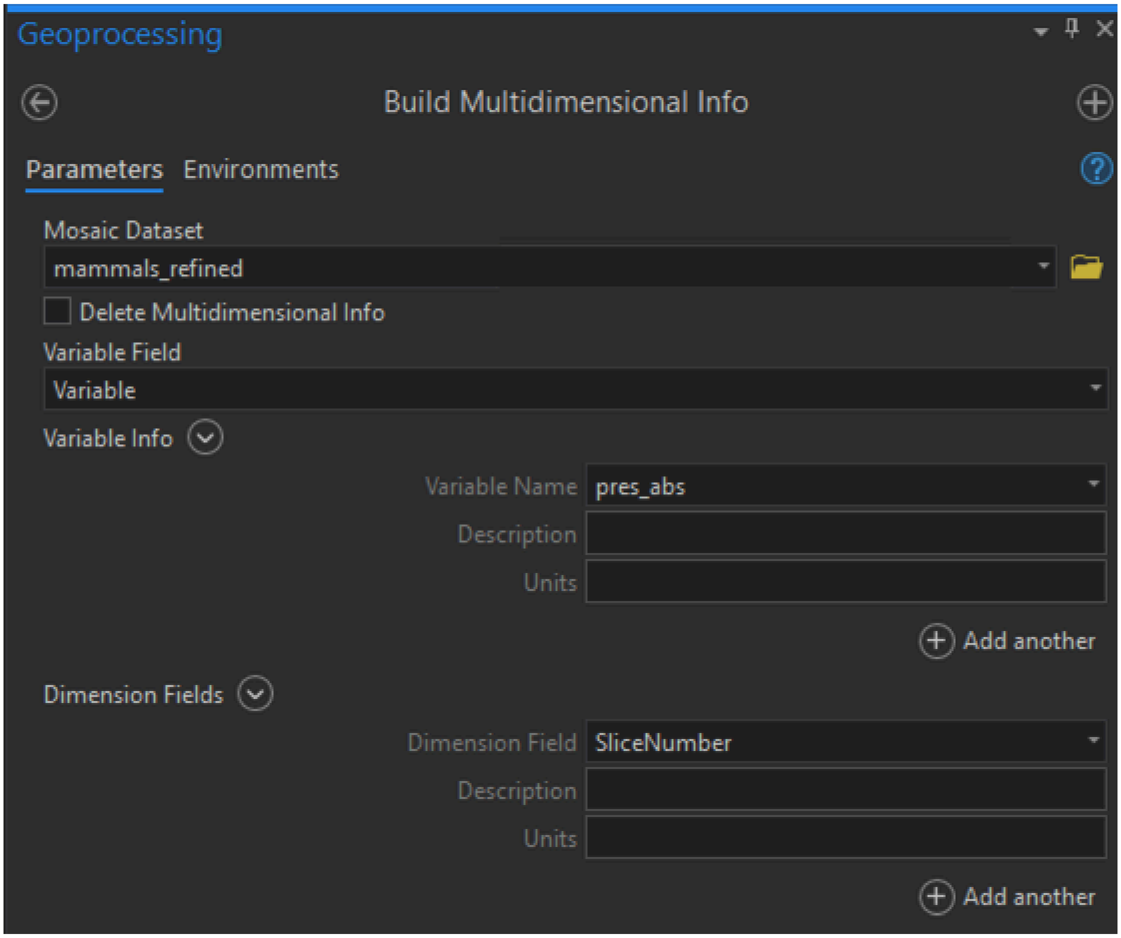
Build Multidimensional Info

The `Build Multidimensional Info` tool adds two fields to the mosaic dataset: Dimensions and Variable. We have to indicate a name for the Variable field (the field we want to use as variable) and a value for the dimension field: the variable we created in the previous step.

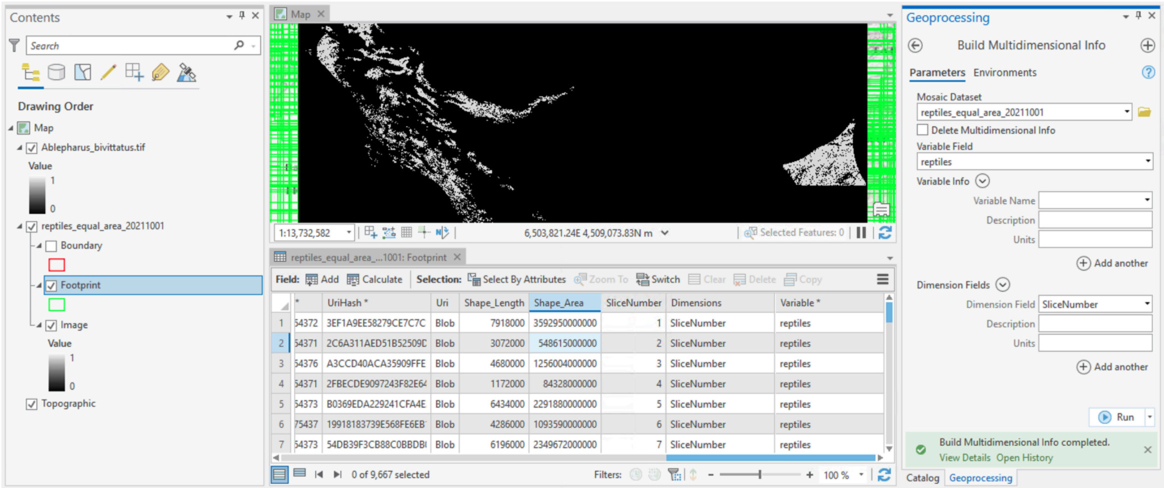


ATTENTION: when looking at the tool from the history, the fields have changed. The amphibians name we wrote gets overwritten to Variable. It may help to check the python snippet to revise what was done.

- `arcpy.md.BuildMultidimensionalInfo("mammals_refined", "Variable", "SliceNumber # #", "pres_abs # #", "NO_DELETE_MULTIDIMENSIONAL_INFO")`

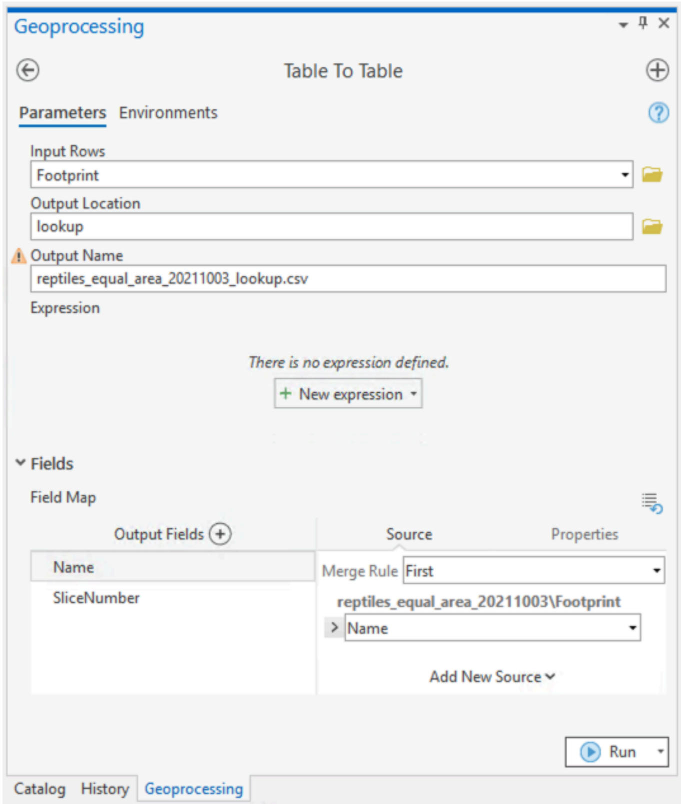


The output should look something like this screenshot. Note that the Geoprocessing panel is still open and therefore conserves the Variable name that we set (‘reptiles’). If we access this through History, it would appear as ‘Variable’



Create lookup table

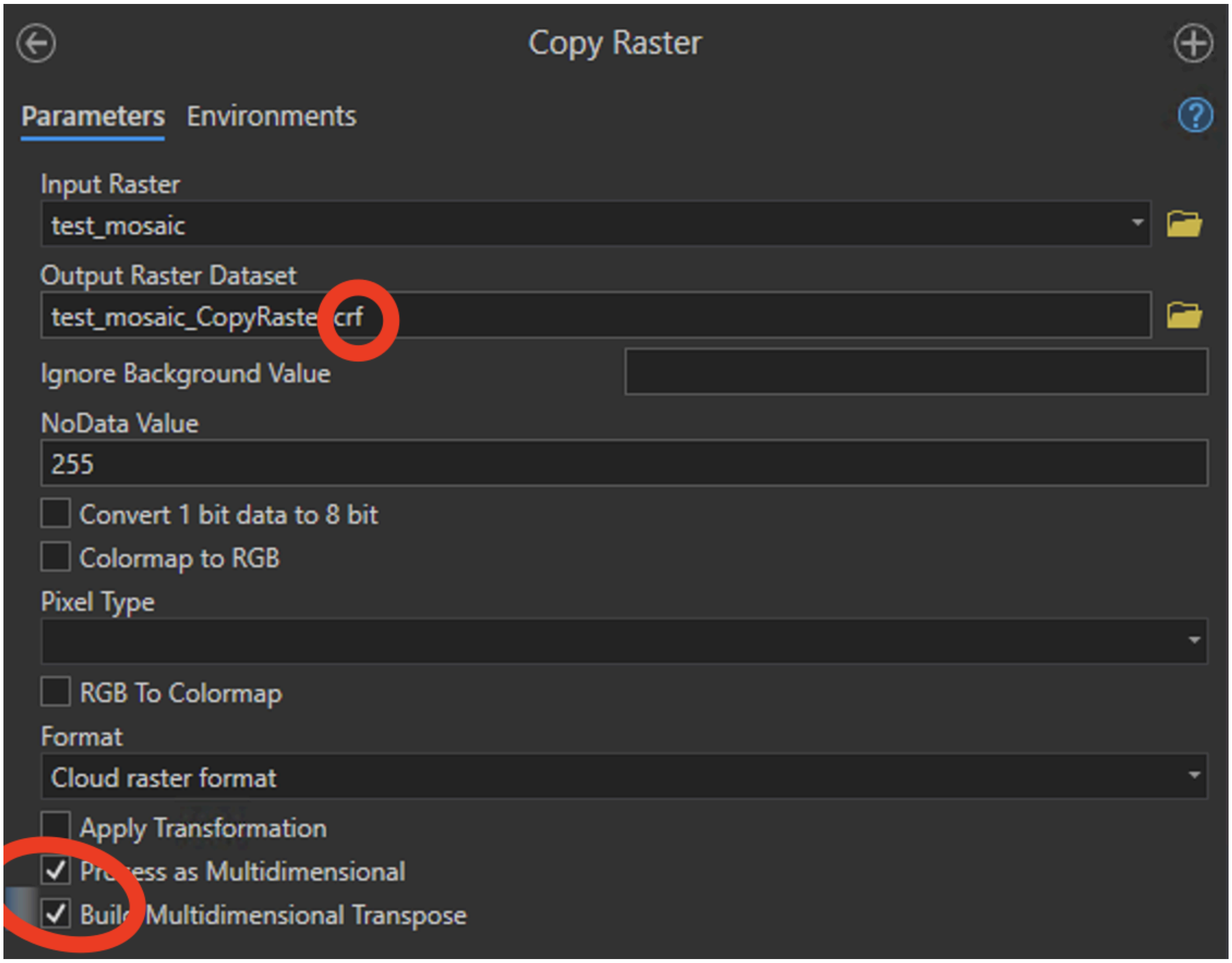
Use `Table to Table` to create a Table that has the species name with the Slice Number of the crfs. This will be used by the Front End to be able to show the name of the species. Use the Footprint table as input and in Fields, just keep the fields *Name* and *SliceNumber*



Copy Raster

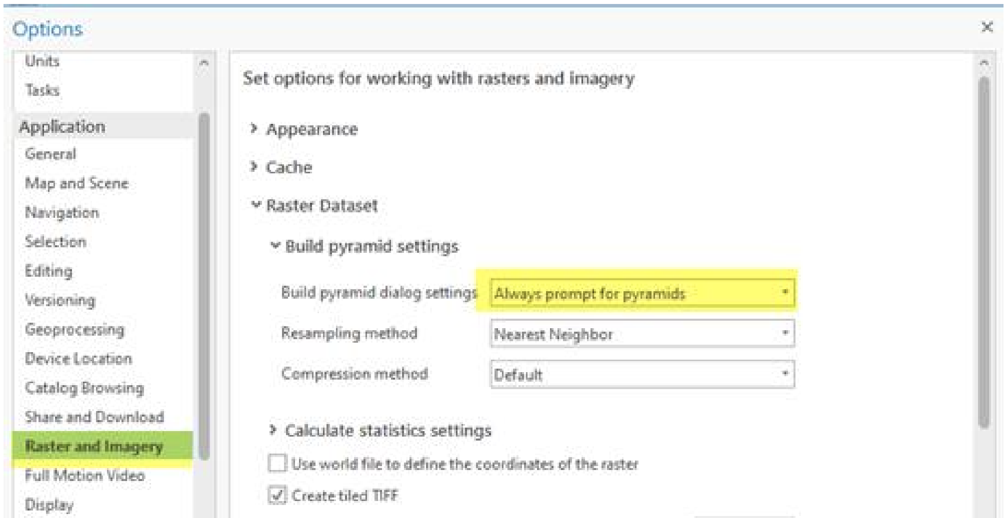
`Copy Raster` parameters:

- input raster is the mosaic dataset.
- When writing down the Output Raster Dataset name, you must specify the extension, which is .crf. (The name has to be written opening the folder dialog in order for the 'Process as Multidimensional options to appear)
- Once the extension is specified, the checkbox "Process as Multidimensional" appears and is available. Check the box.
- It is very important to "Build Multidimensional Transpose", check the box. What this builds is an internal index stored within the crf itself, it is what makes queries fast.

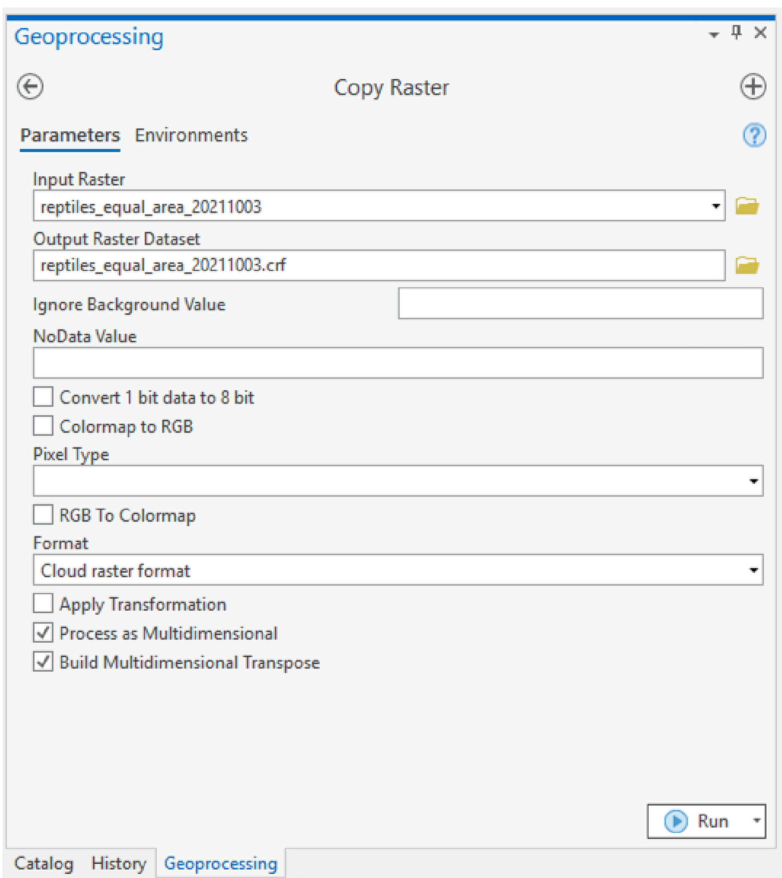


ATTENTION: Open the environments tab. Set the parallel processing factor to 90% for the Copy Raster tool. Setting it to 100% slows the tool down.

Make sure you are **NOT** building pyramids. This can add hours to the processing. Your Pro may be set to do this automatically. You can turn it off in Options:

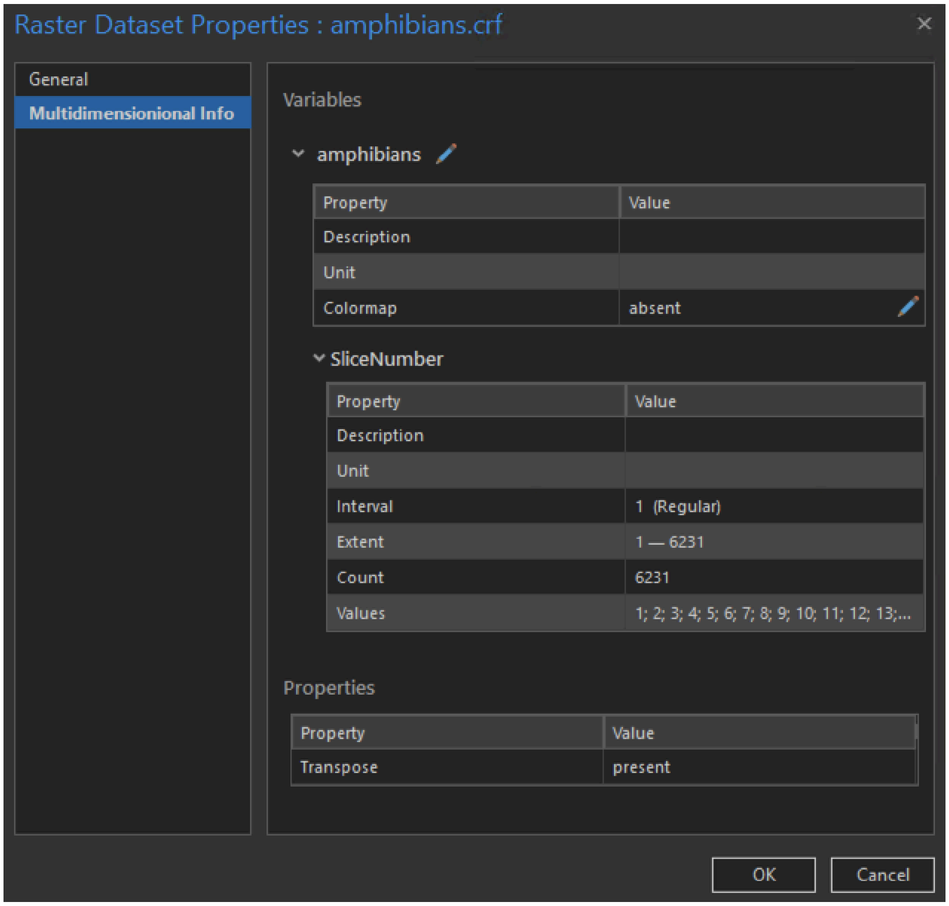


ATTENTION: No data for the crfs should be set as **empty** instead of using the default 255 that appears on the NoData Value box. If checked through the History, the 255 value is back on the NoData value Box.



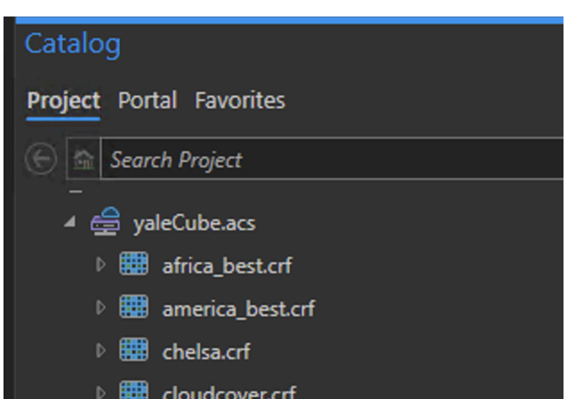
Resulting Crf

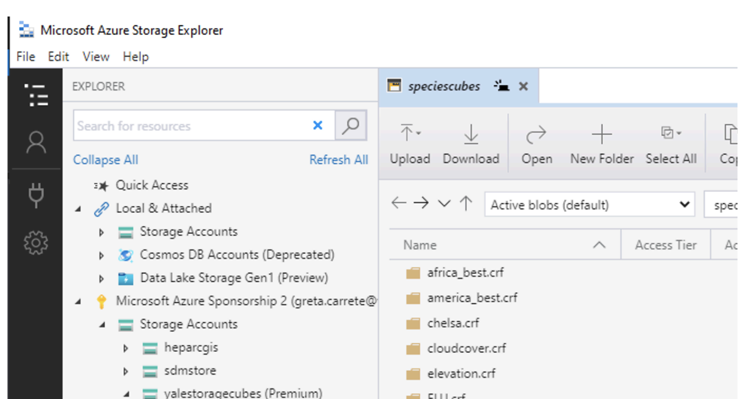
To check the properties, access the crf through the catalog, right click on the item and select ‘Properties’



Storing crf files for the Half-Earth Project

Crf files live in the azure storage set up at Yale University. To upload data is necessary to have a microsoft account and to be invited to the domain mapoflifeprojectoutlook.onmicrosoft.com (The Azure Storage is already mounted on the Viz2 and Viz3 machines, so no need to request a specific invitation unless MoL change these credentials or the location of the files). Data is uploaded using Microsoft Azure



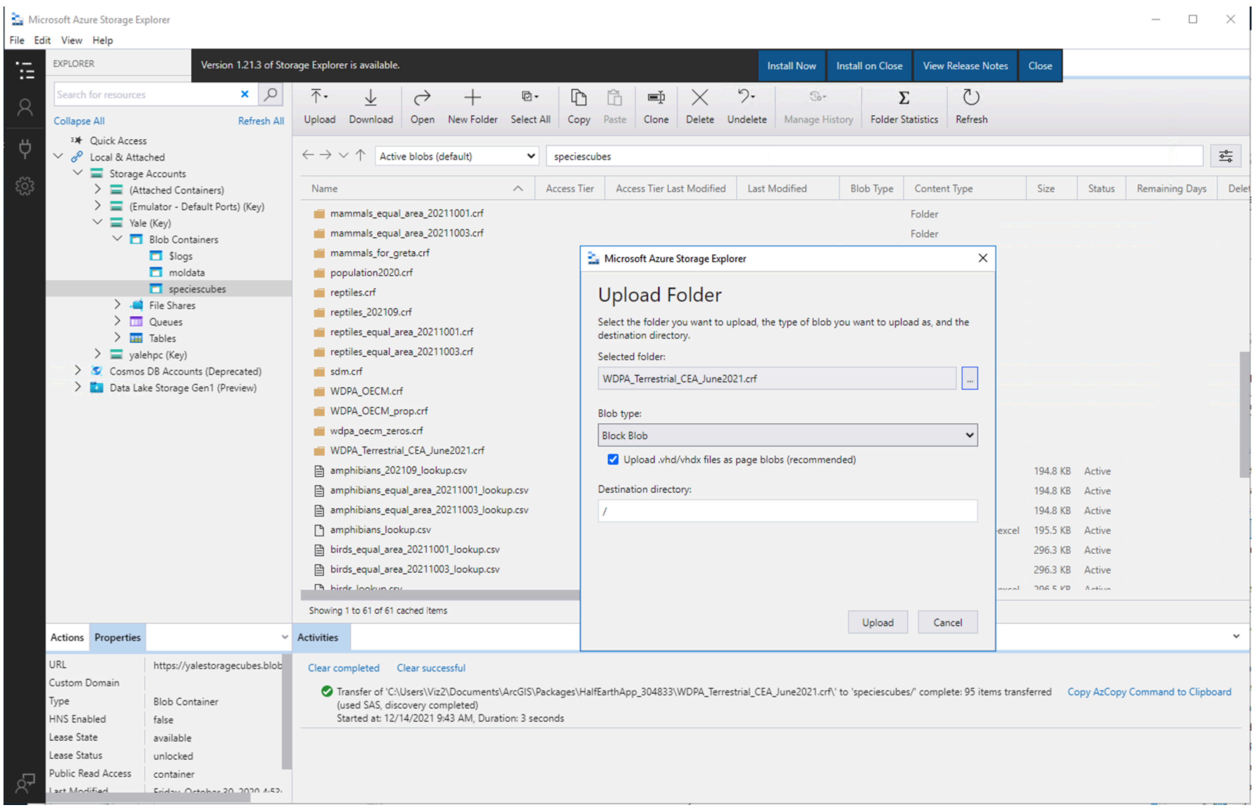


Storage Explorer. There are several accounts, the one containing the crf files is yalestoragecubes. The cubes are accessed via arcgis pro using an .acs file.

Accessing the crf files via Microsoft Azure Storage Explorer and the ArcGIS catalogue.

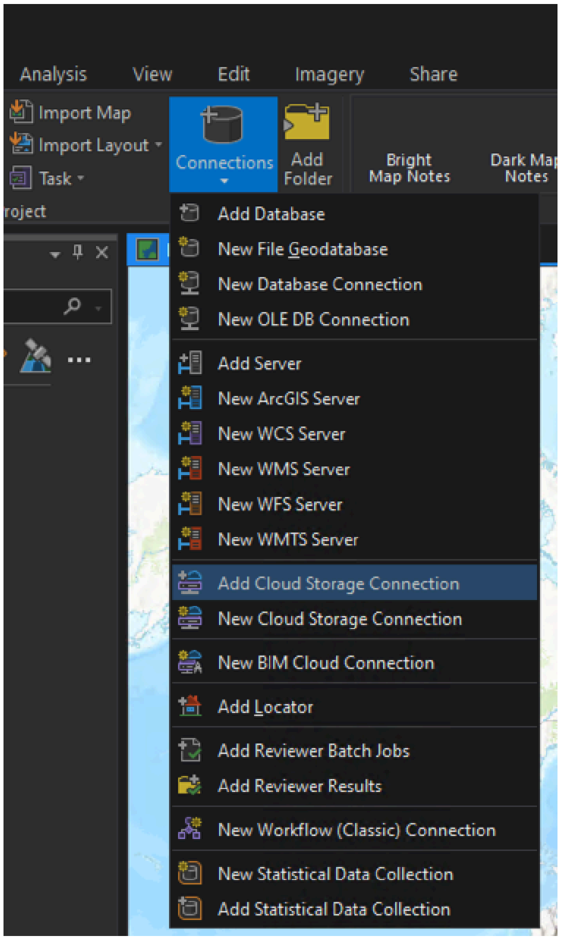
To upload a newly created crf inside ArcGis Pro, go to the Microsoft Azure Storage Explorer → Storage Accounts → Yale (Key) → Blob Containers → speciescubes

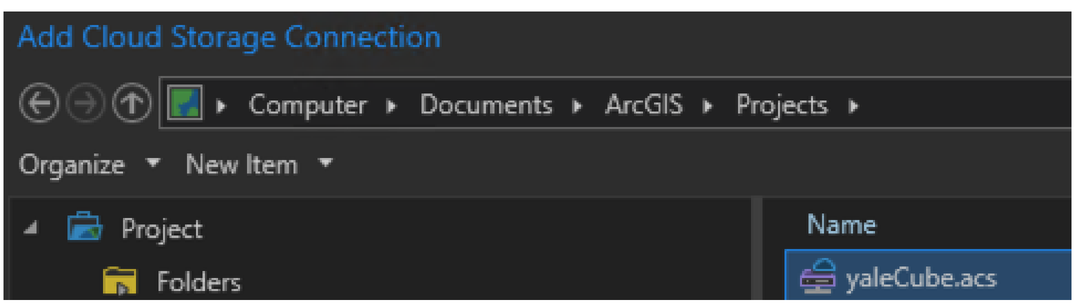
Click on the upper menu ‘Upload’, select ‘Folder’ and select the crf in the ArcGIS Project. Don’t add anything to ‘Destination directory’ and it will be added at the same level as the rest of the crfs.



How to connect an Arcgis Pro project to the Azure Cloud Storage (acs)

The Virtual machine has already the .acs file that makes the connection. When a new project is created a connection is done via the ribbon. The yaleCube.acs file is located in Documents/ArcGIS/Projects





Creating and publishing a webtool (WIP)

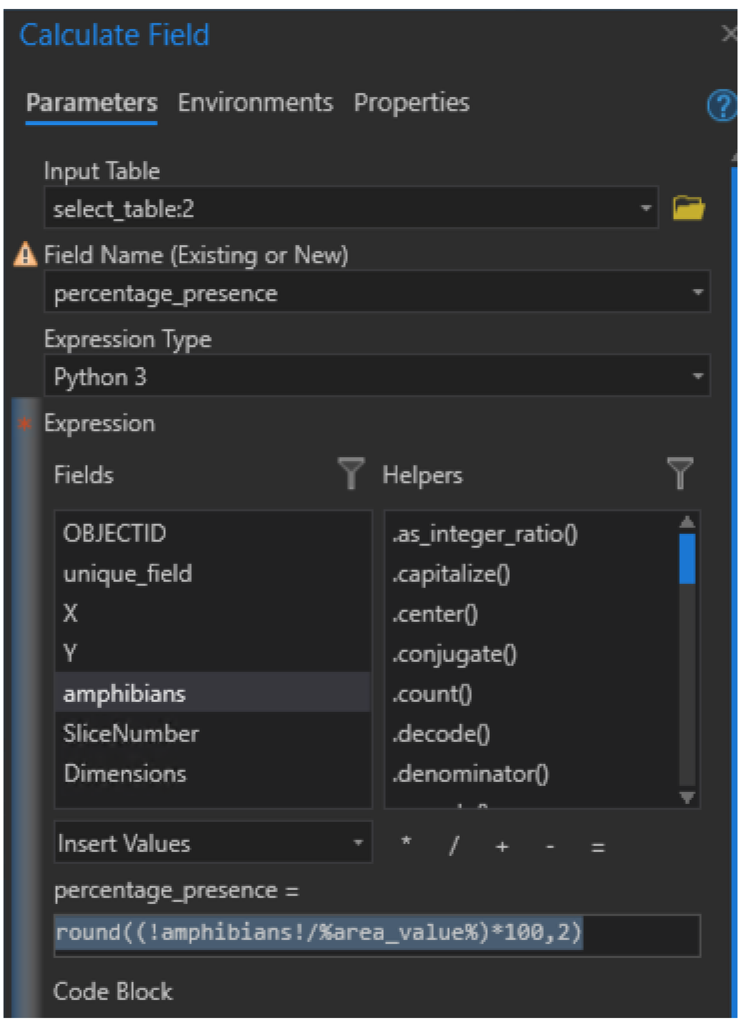
Create feature class for testing with unique_field

Create a subset of the crf you are using with the extent of the testing feature. Make sure you have turned off the automatic building of pyramids.

Steps in the model builder

```
with
arcpy.EnvManager(scratchWorkspace=r"C:\Users\Viz1\Documents\ArcGIS\Projects\handover\handover.gdb", parallelProcessingFactor="80%", workspace=r"C:\Users\Viz1\Documents\ArcGIS\Projects\handover\handover.gdb"):

    arcpy.sa.Sample("amphibians.crf", "pacuando", r"C:\Users\Viz1\Documents\ArcGIS\Projects\handover\handover.gdb\sample_output", "NEAREST", "unique_field", "ALL_SLICES", None, "SUM", None, None, "ROW_WISE", "TABLE")
```



Marine crfs

The rasters are in 4326, in a grid of 0.5 x 0.5 º. Each bixel has the value of the area, instead of binary presence/absence.

We keep this reference system and need to think that we need to get the summof pixels for the area