Finding Raster Data for GIS Projects

Ethan McIntosh, GIS & Data Associate, Brown University

June 21, 2022

Introduction

This tutorial describes how to locate and download free raster data from three widely-used online repositories: the Earth Explorer, provided by the United States Geological Survey (USGS), the World Environment Situation Room, provided by the United Nations Environmental Program (UNEP), and the National Map, which is also provided by USGS.

Repository	**Spatial Coverage**	**Types of Data**
Earth Explorer	Global	Environment, elevation, imagery
UN WESR (MapX)	Global	Environment, population, misc.
National Map	United States	Elevation & topography

In section 2, I'll describe what kinds of data are available in each of these repositories in more detail, and then demonstrate how to use each of their web interfaces to effectively search, filter, and download relevant data. In these demonstrations, I point out many of the "gotchas" that may not be intuitive for people who are new to these interfaces.

I conclude the tutorial by walking through how to combine several raster tiles for a given study area into a single file using either QGIS or ArcGIS Pro, since this is a common pre-processing step for visualizing or analyzing raster data.

1 Background and Terminology

Rasters are a type of spatial data in which space is represented as a grid of equally sized cells (also known as pixels), and each cell has a numeric value associated with it that represents some attribute of the geographic space bounded by that cell.

Rasters are commonly used to represent data that acts like a continuous surface, like temperature, precipitation, elevation, pollution, or land use & land cover.

Raster files are structured like other digital images, and may come in the same file formats. However, rasters are georeferenced, meaning that the grid of cells corresponds with a fixed geographic area in a specific coordinate reference system, and can be overlaid with other spatial data.

Raster file downloads, especially for high resolution data or data that covers a large geographic area, will often consist of multiple separate files, each covering a different rectangular portion of a given study area. These individual rectangular raster files are referred to as tiles.

2 Repositories of Free Raster Data

2.1 Earth Explorer (USGS)

Background and Types of Data Available

The Earth Explorer interface is widely used to access and download remotely sensed data, including:

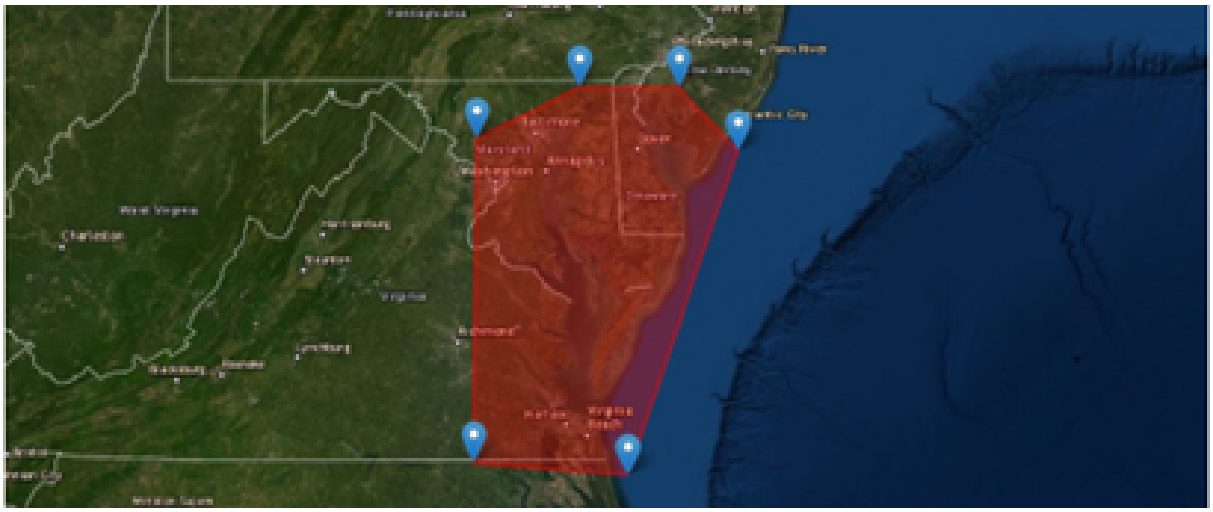
- multispectral satellite data and imagery (mostly from the LANDSAT mission, but there is a variety to choose from)
- aerial data and imagery
- global digital elevation models (DEMs) like GMTED2010 and SRTM
- raster maps of general land-use and land cover
- raster maps of specific environmental features like vegetation abundance, soil moisture, snow cover, burned area, etc

Most of the datasets on Earth Explorer are global in coverage, and many of them are collected repeatedly over time (in some cases, going back 30 or 40 years). Depending on the data, there may be options for both "raw" sensor readings and for more "ready-to-use" products on which additional corrections or processing steps have already been performed.

Using the Interface to Search, Filter, and Download Data

On the main Earth Explorer page, there is a map window covering most of the screen, as well as a pane on the left containing various menus and buttons. The first step in finding data is to use this pane to specify some search criteria. The options on the *Search Criteria* tab allow us to specify a geographic area, a range of dates, and other options for specific datasets (like cloud cover) on which to filter results. Selecting a geographic area will filter results to tiles that are either contained within or that intersect your area.

There is a USGS help document that goes into more detail on the various search criteria options. A common one is to have the *Polygon* tab selected and start clicking on the map window to create a polygon roughly corresponding to a study area:



Another option for choosing a study area is to zoom the map until it covers the full extent of your area, and then click on the *Use Map* option to generate a rectangle around the area that's displayed. There is an option on the *Search Criteria* tab for uploading a shapefile for study area boundaries, but it only works for shapefiles with fewer than 30 points, which rules out being able to upload an administrative boundary as the filter. For studies of specific countries or regions, it's easier to just draw a simple polygon that covers that area, and then if needed, you can clip the data to the area's boundaries after it's been downloaded.

Once a search area and other filters like dates or cloud cover have been applied, select one or more datasets from the *Data Sets* tab that you want to consider in the results. Click on the plus sign next to a category of dataset to expand the list of datasets within each category. Keep in mind that many categories also have subcategories.

The Data Sets help document describes some of the icons that are displayed for each dataset:

> Information about each data set is identified with the Information icon ( ).
>
> The coverage map icon ( ) displays the coverage area for the data set on the map interface.

Some of the datasets additionally have dollar sign icons, meaning they aren't free to download, and some of them have cloud icons, meaning you could filter those by cloud cover if desired.

Once search criteria and datasets have been selected, clicking on the *Results* tab will display a list of matching results in the pane. It's often useful to see what results come up, and then go back and tinker with the search criteria or the datasets and try again, until desirable results are identified. The USGS supplies a Results help document with more details on this part of the process. Each result in the *Results* tab has various options, as shown in the following screenshot from the help document:

Results - Include a thumbnail image and attribute information on each scene. Based on the data set and user profile, some controls may not be available (Figure 7).

Figure 7 – Example Scene Results

The icon menu for scene level result controls includes the following:

- Show Footprint
- Show Browse Overlay
- Compare Browse
- Show Metadata and Browse
- Download Options
- Bulk Download
- Order Scene
- Exclude Scene

For downloading a relatively small number of tiles, you can simply download data individually using the *Download Options* buttons on individual results. For projects requiring a large number of tiles, it makes more sense to open the result controls panel and add all the results from each page to a bulk download order. More details (and an instructional video) on Earth Explorer's Bulk Download Application can be found on the bulk downloads help page.

2.2 World Environment Situation Room (MapX)

Background

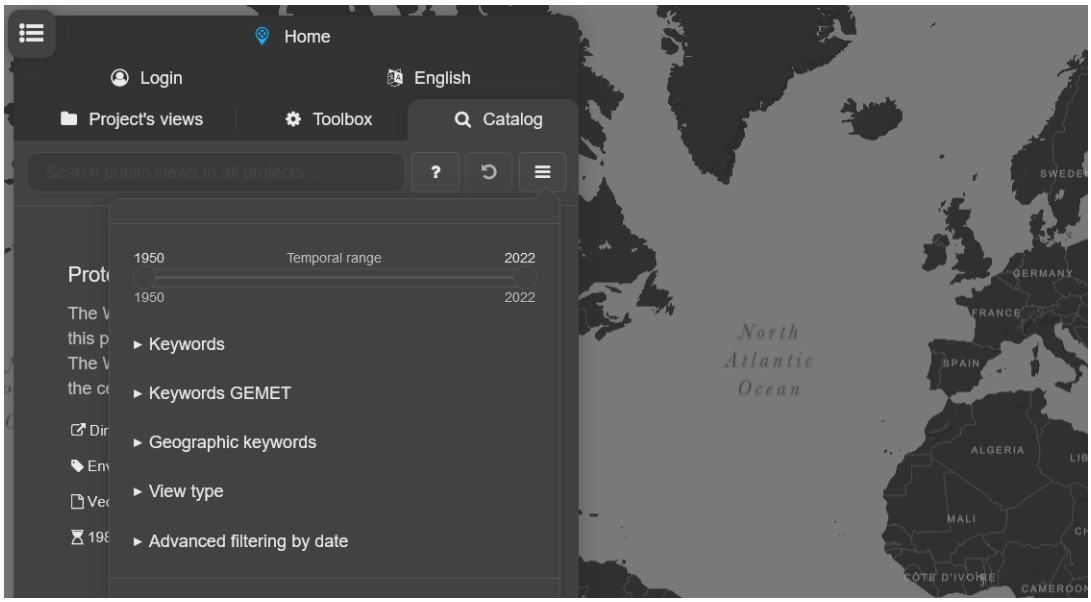
The United Nations Environmental Program (UNEP) publishes a huge variety of geospatial data from around the world through it's Global Resource Information Database (GRID) partnership with the Swiss government and the University of Geneva. UNEP-GRID advertises various data platforms, but in this guide, I will focus on the World Environment Situation Room (WESR), whose datasets cover a broad range of environmental topics, many of which also interface with population studies. Some examples of these topics include climate, pollution and air quality, agriculture, natural hazards, ecology and biodiversity, and land use/land cover.

Using the Interface to Search, Filter, and Download Data

The main WESR website is primarily geared toward browsing for articles, tables, reports, and story maps on the various thematic areas covered by WESR datasets. In this guide, I focus on using the WESR's map-based interface, MapX, to specifically search for georeferenced data in vector and raster formats.

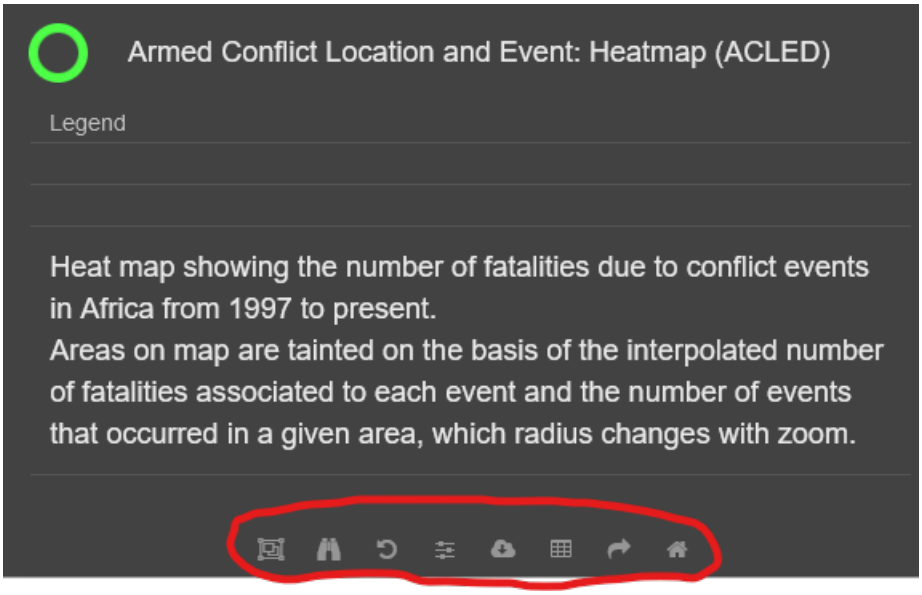
When the MapX interface is first opened, a list of sample datasets ("Project's views") are displayed on the menu on the left. Clicking on the circles next to each one will toggle them on or off as the visible layer on the map. Click on the *Catalog* tab on this menu to access the tools for searching the WESR's full collection of datasets.

Specific keywords or phrases can be typed into the search box on this tab. Additionally, specific topics (*Keywords*), time periods, geographies (*Geographic keywords*), and data formats (*View type*) can be specified by clicking on the button with the three lines and then clicking on a particular section to either expand or hide its available options:



There are two options for viewing individual dataset results: *Direct link* will open a new tab of MapX displaying the given dataset by itself. *Activate this view* will display the layer on the current map, overlaid on top of any other layers that are already being displayed. The *Metadata* link is also useful, usually containing a description of the dataset as well as links to the original data source and sometimes to related datasets.

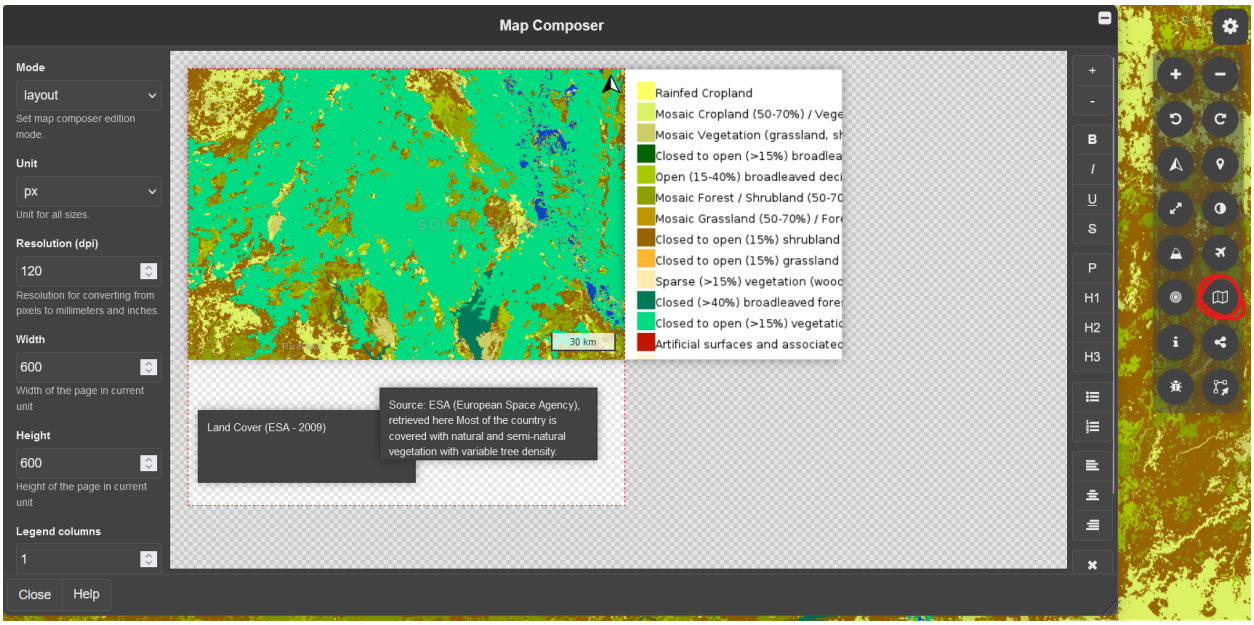
The *Project views* tab is where the layers being displayed can be turned on or off. Each layer will have a series of buttons along the bottom of its legend entry with various options:



The cloud-with-an-arrow button leads to a page with download options for a given dataset. However, the download button is only available if you make a (free) account by selecting the Login button. Note that not every dataset on MapX will be downloadable from MapX even with an account, due to differences in data licenses and availability. In many cases, however, it may be possible to download a dataset from its original source website, which can be accessed by opening the metadata for that dataset within MapX and following the appropriate link.

MapX also has options for manipulating and mapping datasets on its own web interface. Depending on the project, it may be sufficient to simply work with the data within MapX without having to download it. To the left of the download button, for example, is a filtering button which will display a menu with sliders and/or check boxes to filter the data that is displayed by certain attribute values. To the right of the download button is the attribute table button, which will load an attribute table for that layer and in some cases have an option for exporting the attributes as a csv file.

Additionally, clicking around on the map will yield pop-up menus identifying individual features and displaying their attributes. There's even an option called Map Composer, using the button on the far right of the screen that looks like a trifold poster board, where simple map layouts with legends and scale bars can be created and exported in common image formats.



For more detailed instructions and documentation of the MapX interface, explore the training materials available on the MapX knowledge base website.

2.3 The National Map (USGS)

Background and Types of Data Available

The National Map (TNM) maintains several web interfaces meant to help disseminate data produced by USGS, mostly on elevation within the United States.

The most recent and high-resolution data is available for download on the TNM downloader interface, but the National Map also maintains separate web interfaces for specific datasets. For example, TopoView is specifically devoted to publishing scanned & georeferenced images of historical topographic paper maps created by the USGS going back to the late 19th century.

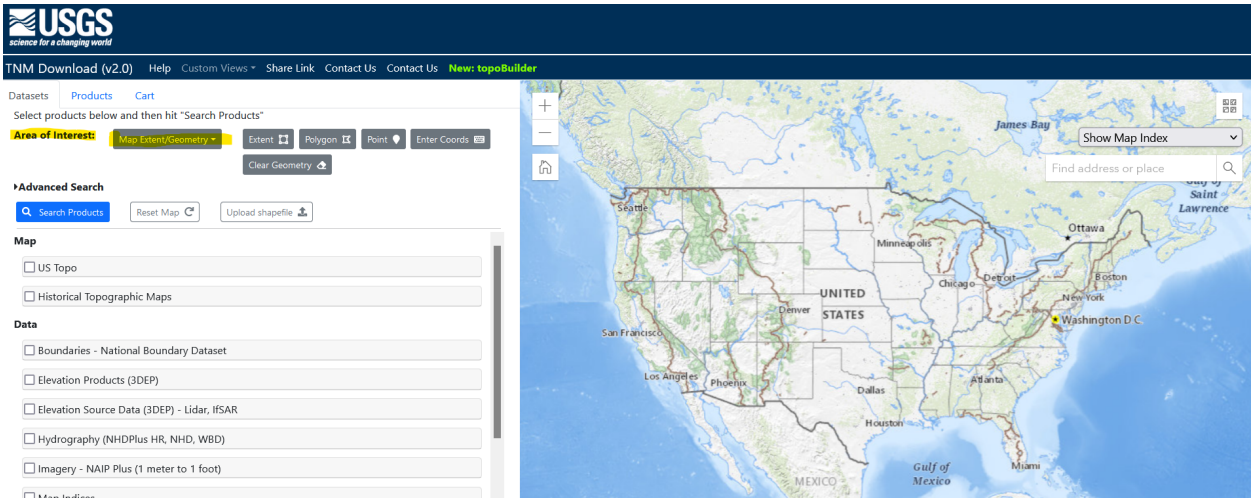
The main kinds of data available on the TNM downloader interface include:

- Digital Elevation Models, both current and historical, with spatial pixel resolutions ranging from 1 arc-second (roughly 30 meters) to 1 meter
- topographic maps, with map scales ranging from 1:250,000 to 1:24,000, both current and historical
- elevation contour line maps
- hydrography and bathymetry data for some areas
- geographic feature names (labelled points) for physical and cultural sites like mountains, water bodies, populated places, schools/hospitals/cemeteries, etc.
- state and county boundaries
- transportation (mostly road) lines derived from Census TIGER lines

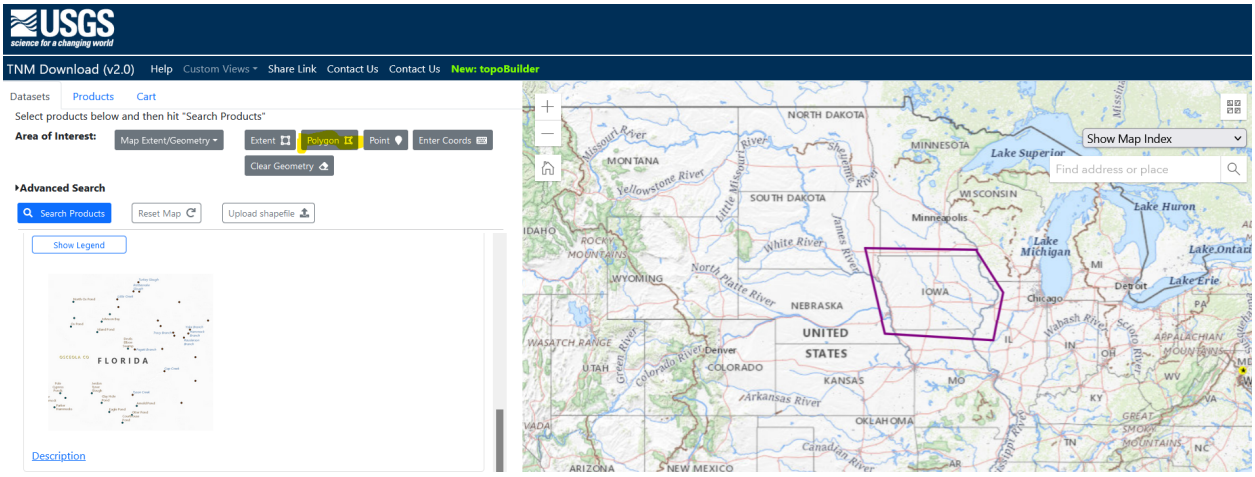
All of these datasets cover only the United States.

Using the Interface to Search, Filter, and Download Data

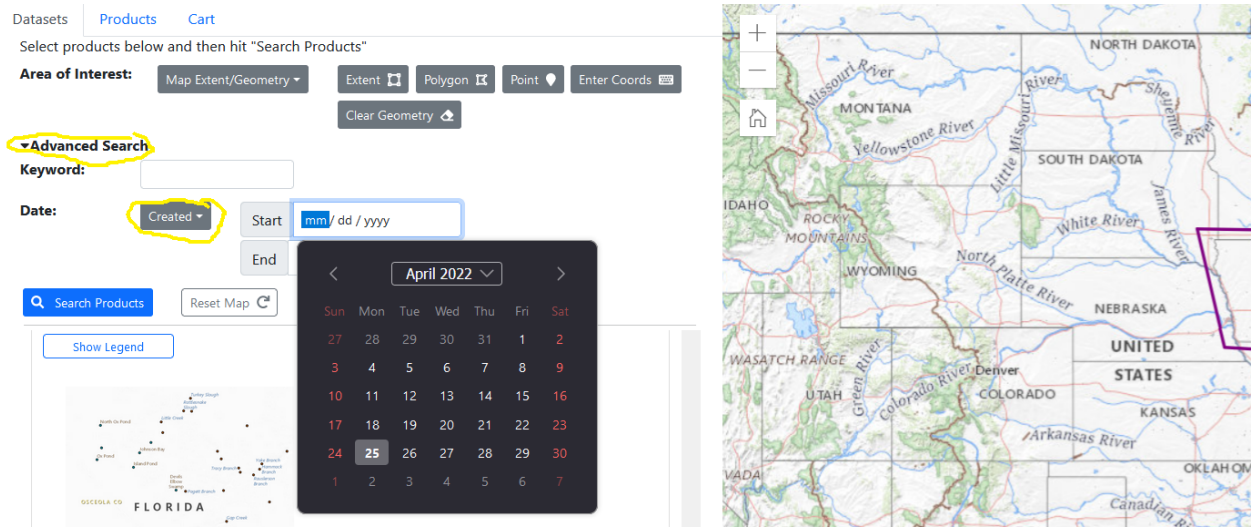
The National Map uses a map-based interface to display and help search for its data. By default, data search results will be filtered by the extent of the map, which can be altered by zooming in or out.



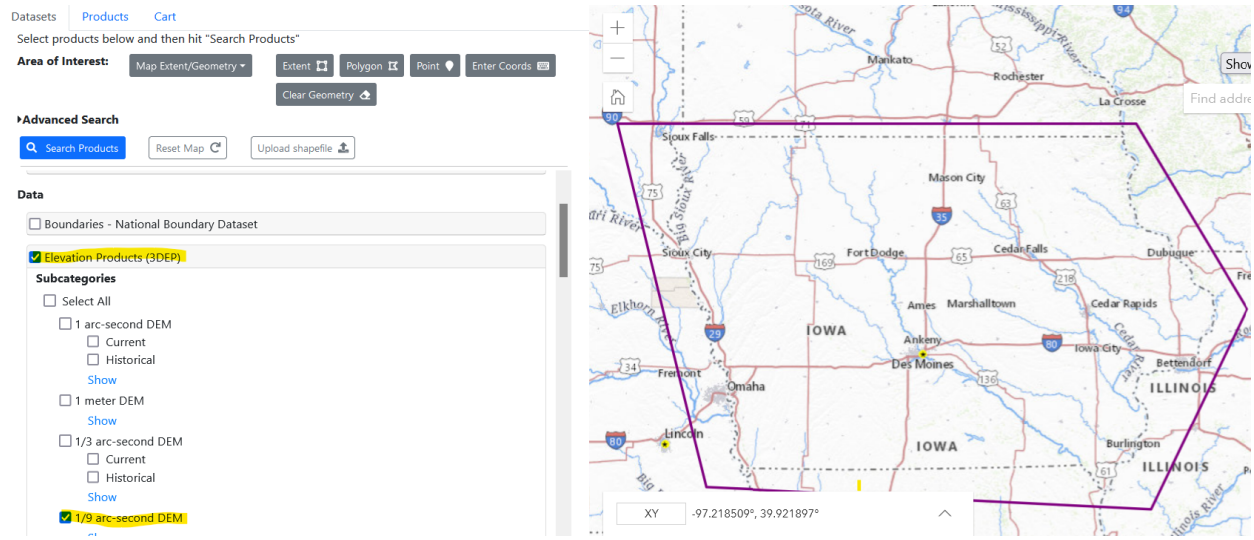
Another way of filtering results by geographic area is to use the *Extent*, *Polygon*, or *Point* buttons to draw or select an area of interest on the map.



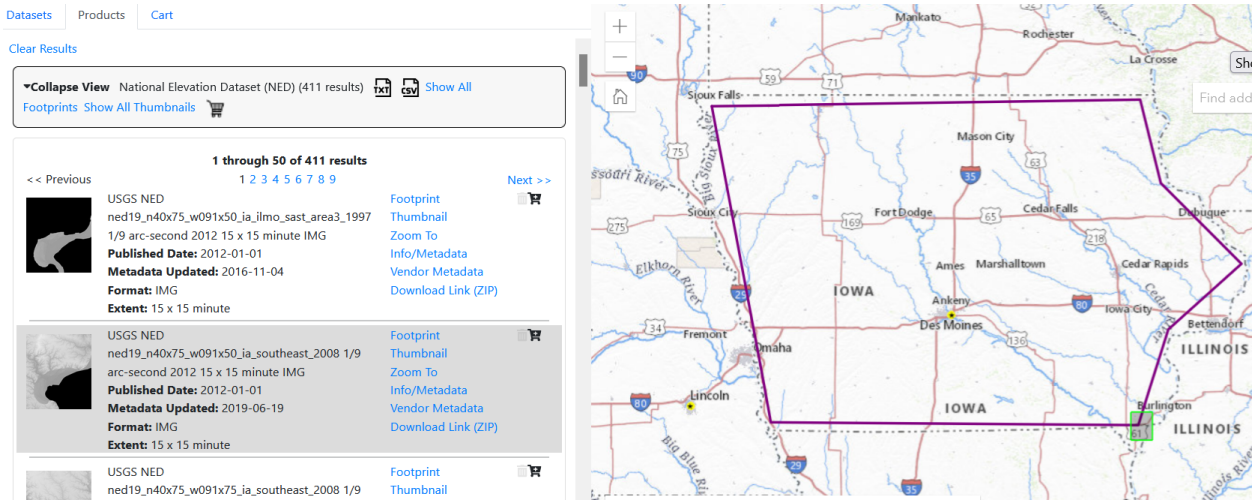
Some datasets will have multiple entries with repeated measurements over time. To filter data by date created, click on the *Advanced Search* bar and enter a range of dates:



After choosing a study area and, if applicable, a date range, select at least one of the check boxes corresponding to a dataset for which you want to search. In the example below, I've selected Elevation (3DEP) and then within that category, I've indicated that I specifically want 1/9 arc second (roughly 3 meter resolution) data.

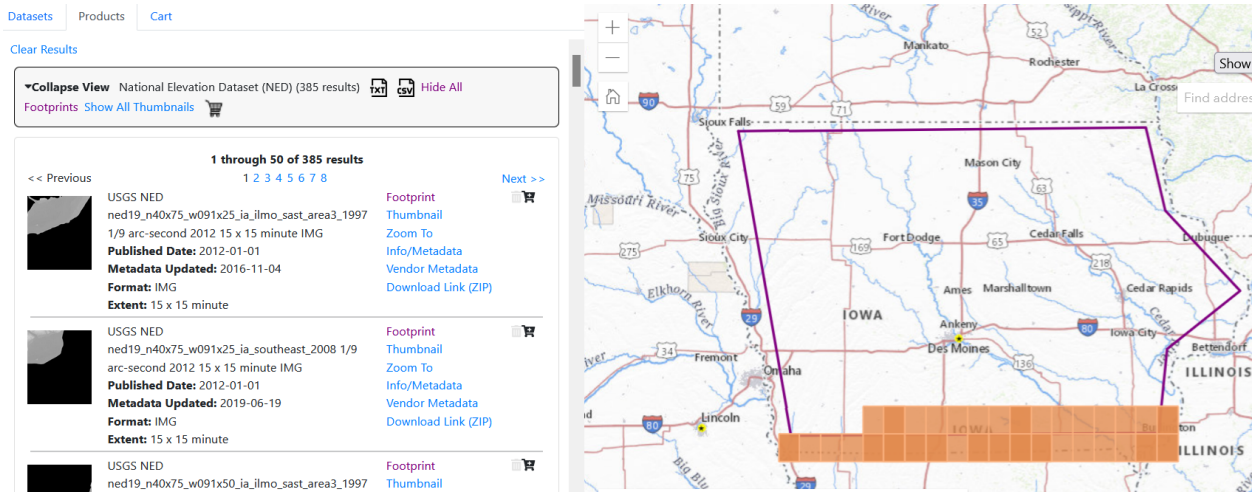


Click on *Search for Products*. This will take you to the *Products* tab, and displays the number of results that met your criteria. In this case, I got 411 results. To narrow down these results, I would alter the selections back on the *Datasets* tab and then search again.



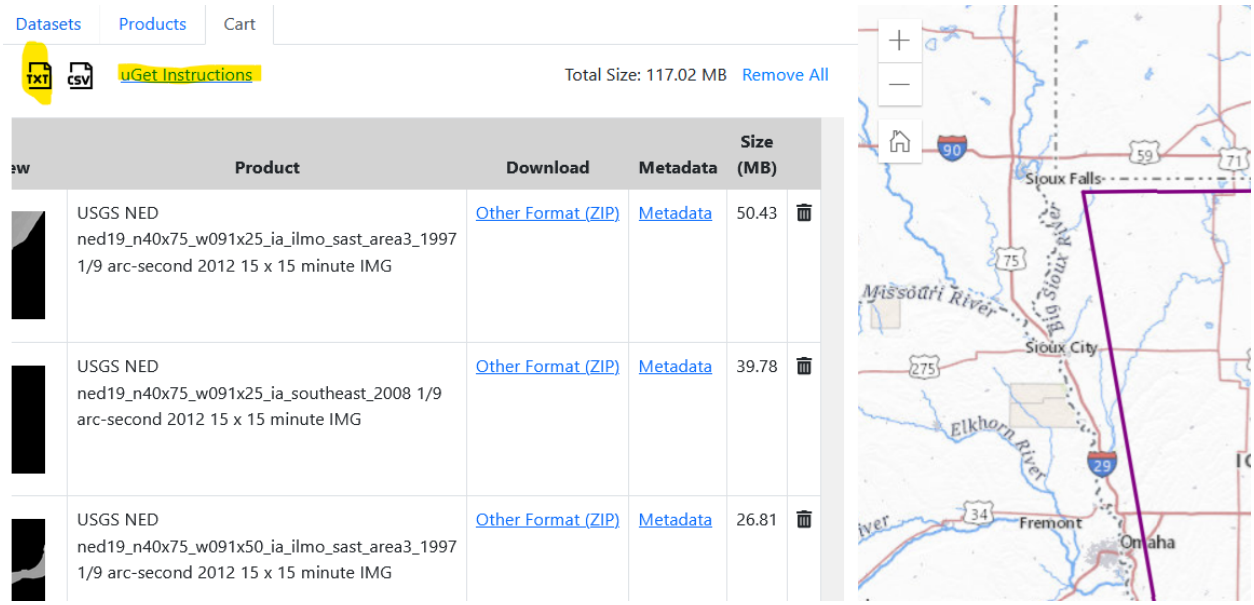
Hovering over each result or clicking on each result's *Footprint* button will display a rectangle on the map showing the area covered by the tile represented by that result. In this example, the highlighted result covers a small area in the southeastern corner of Iowa on the Illinois border.

Alternatively, clicking on *Show all footprints* will display the footprints of all the results on the current page of results.



For small study areas, data can be downloaded easily by clicking on individual download links. If a large number of tiles need to be downloaded, it's better to take advantage of the interface's option for downloading in bulk.

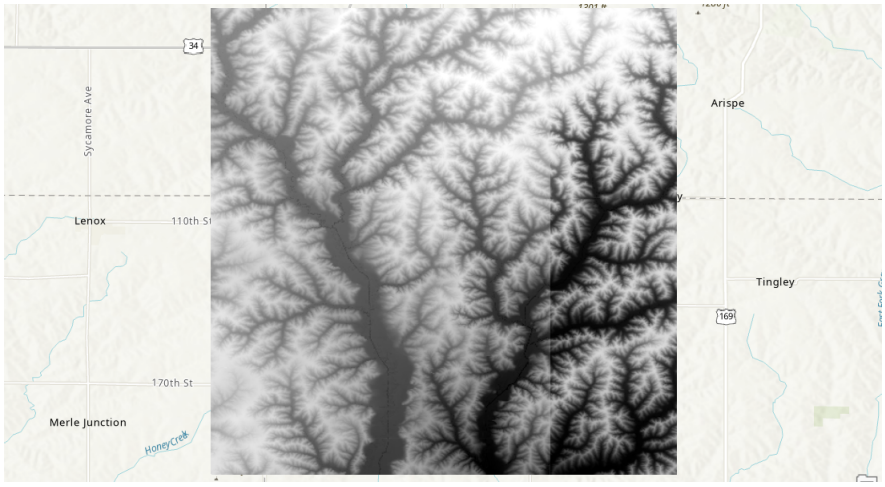
To download in bulk, go through every page of results that you want to download and click on the cart button at the top of the page to add all items from each page to the cart. Then, in the *Cart* tab, click on the TXT icon to download a text file with all the names of the files you want to download. Then, click on the *uGet Instructions* link, and follow the instructions for installing the uGet bulk download application. This application will take the text file you just generated and download all of the tiles you specified at once.



3 Combining Multiple Tiles into a Single Raster

A common operation when working with raster data is to combine multiple rectangular tiles into a single raster file. This step is not always necessary, as some raster analysis tools can be performed on collections of tiles, but other types of analysis may only work on one raster file input at a time. Stitching together multiple tiles can also be useful for simply visualizing data over a given study area. In the following subsections, I walk through how to use either QGIS or ArcGIS Pro to combine tiles together.

For these examples, I have downloaded two tiles of 1/9 arc-second elevation data from the National Map which cover a small area within the US state of Iowa.

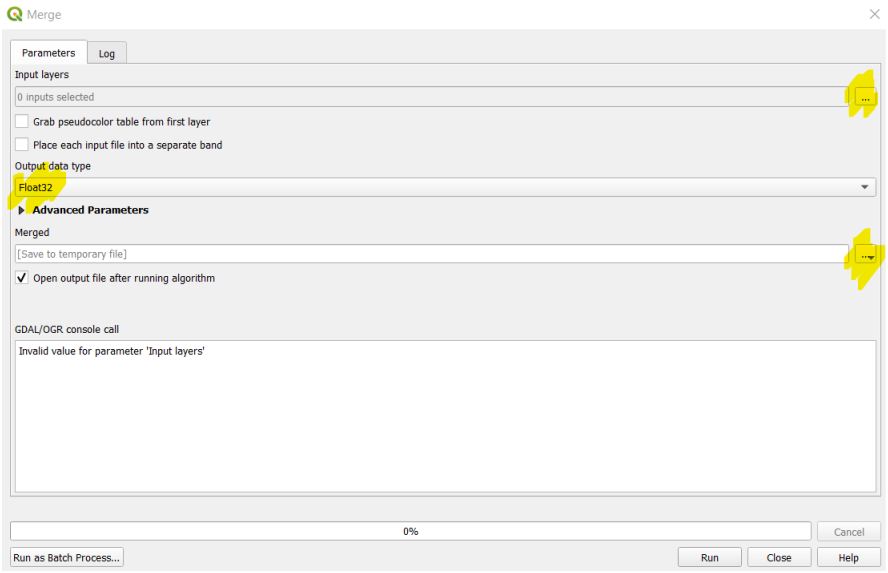


Note the visual discontinuity between the two tiles (the visible vertical line that visually separates the left tile from the right tile). This is completely normal when initially looking at multiple tiles in GIS software. The software is simply visualizing the two tiles on different color ramps based

on each one's distinct range of elevation values. The underlying elevation numbers are not actually distorted at that border between the two tiles, even though visually it will look distorted at first. If we were to manually alter the way that the two tiles are shaded to exactly match each other, it would eliminate the visual difference between the two tiles. Combining the two tiles into a single file will also eliminate this visual difference, as the combined image will be visualized on a single color ramp.

3.1 QGIS

In QGIS, the main tool for combining tiles into a single raster is called *Merge*. To access this tool, select *Raster -> Miscellaneous -> Merge*, or open the toolbox (*Processing -> Toolbox*) and search for "raster merge".



First, click on the three dots next to the *Input Layers* option. If you've loaded your raster tiles into a QGIS project as layers, they will show up on the next window and you can check the boxes next to the name of each layer you want to stitch together. However, loading in the tiles as layers is not necessary to run this tool, and not advisable if working with a large number of tiles. The other option is to use the *Add File(s)...* button to select the raster files that you want to stitch together directly from your file system.

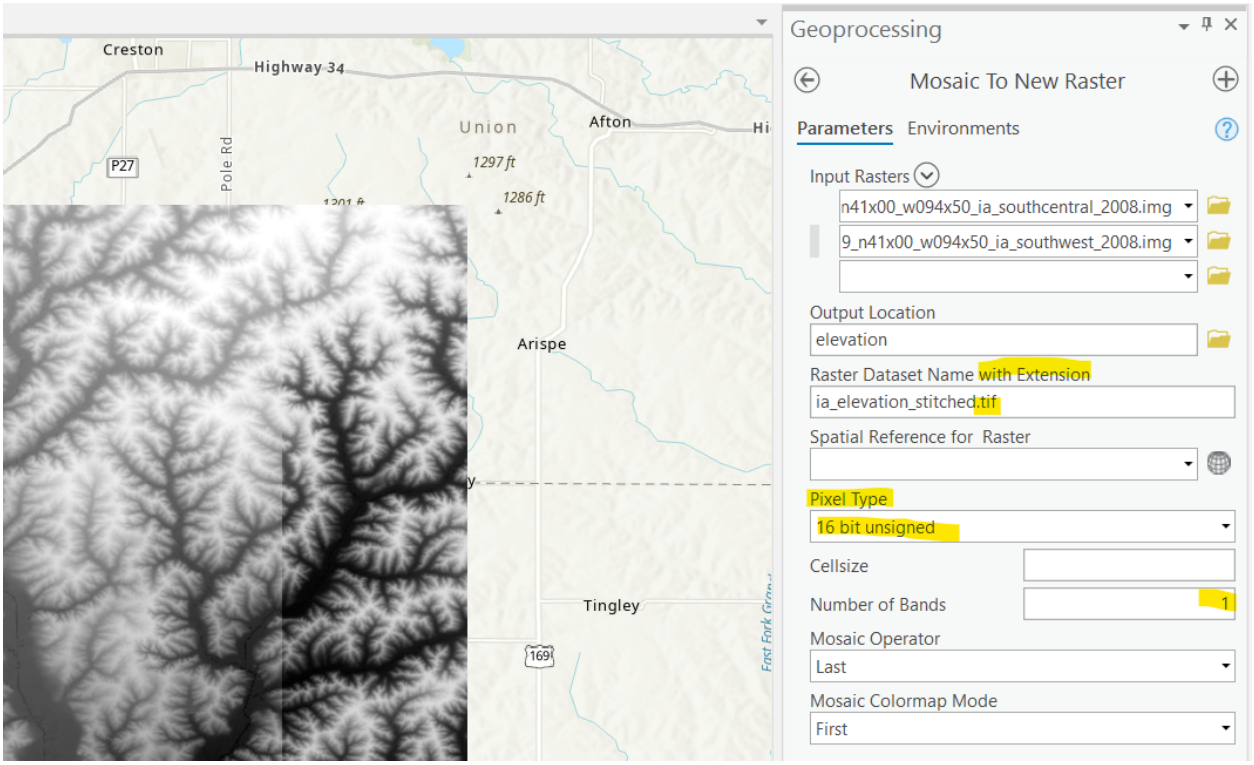
Next, select an option from the *Output data type* menu. Most projects can safely use the default option of Float32 - you would only need to consider other options (which would affect the range of possible values for each pixel) if you're working with very large amounts of data and are looking to control your file sizes, or if your data has extremely precise values that wouldn't be preserved with 32-bit float pixel values.

Finally, click on the three dots next to the *Merged* and click *Save to File* to save the output somewhere on your file system. On the next window, you can choose where your output gets saved, give it a name, and specify your desired image output format from the *Save as type* drop-down menu. If you don't click on these three dots, QGIS will execute the merge and simply display the result as a temporary layer within your project, which you could save later or discard, as desired.

3.2 ArcGIS Pro

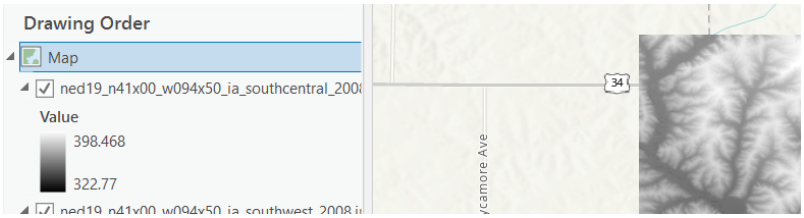
In ArcGIS Pro, the main tool for combining tiles into a single raster file is called *Mosaic to New Raster*. To access this tool, open the *Geoprocessing* tab and click on the *Tools* button, then search for "Mosaic to New Raster".

First, click on the folder icon next to *Input Rasters* to find the folder where you downloaded your raster tiles, and select all the tiles that you want to stitch together.



For *Output Location*, click on the folder icon and select the folder in your file system where you want the combined raster file to be saved. Then, give the output raster a name in the *Raster Dataset Name with Extension* box. An important note is that you must type in a specific file extension for the file format you want along with the file name. Typing in a file name with no file extension will cause the tool to fail. In the screenshot above, I have specified that I want a .tif file as my result, but I could have chosen some other common image or raster format instead.

Another important field to specify when using this tool is the *Pixel Type*. The choice of pixel type constrains the range of values that can be stored in a given pixel. An 8-bit pixel can only store 2^8 (256) possible values. For 8-bit unsigned, this means integers from 0 to 255, and for 8-bit signed, this means integers from -127 to 127. It's important to know what range of values your rasters represent and choose an appropriate pixel type to encompass its range of values.



In this example, I can see from the contents pane that the elevation values for each of these tiles are decimal numbers in the 300s. If I were to choose an 8-bit unsigned pixel type, my combined raster cells would all have a value of 255, since my data's values are all greater than this pixel type's maximum value. Instead, I have chosen to use a 16-bit unsigned pixel type, since this can represent integers from 0 to 65,535, and I am OK with my original data's decimal values being rounded to integers in the combined result. If my data had negative numbers, I would want to use a signed pixel type. If I wanted to preserve decimal values, I would use a 32-bit float pixel type.

Before running the *Mosaic to New Raster* tool, a *Number of Bands* needs to be specified. In most cases, this will be 1, but if you are working with multispectral satellite data or other data with multiple bands, like RGB images, type in as many bands as you want to preserve in the combined file. *Spatial Reference for Raster* and *Cell Size* do not need to be specified unless you want the output raster to be re-projected into a different spatial reference system or have a different cell size than the tile inputs.

The result of running *Mosaic to New Raster* on the Iowa elevation data looks like this: