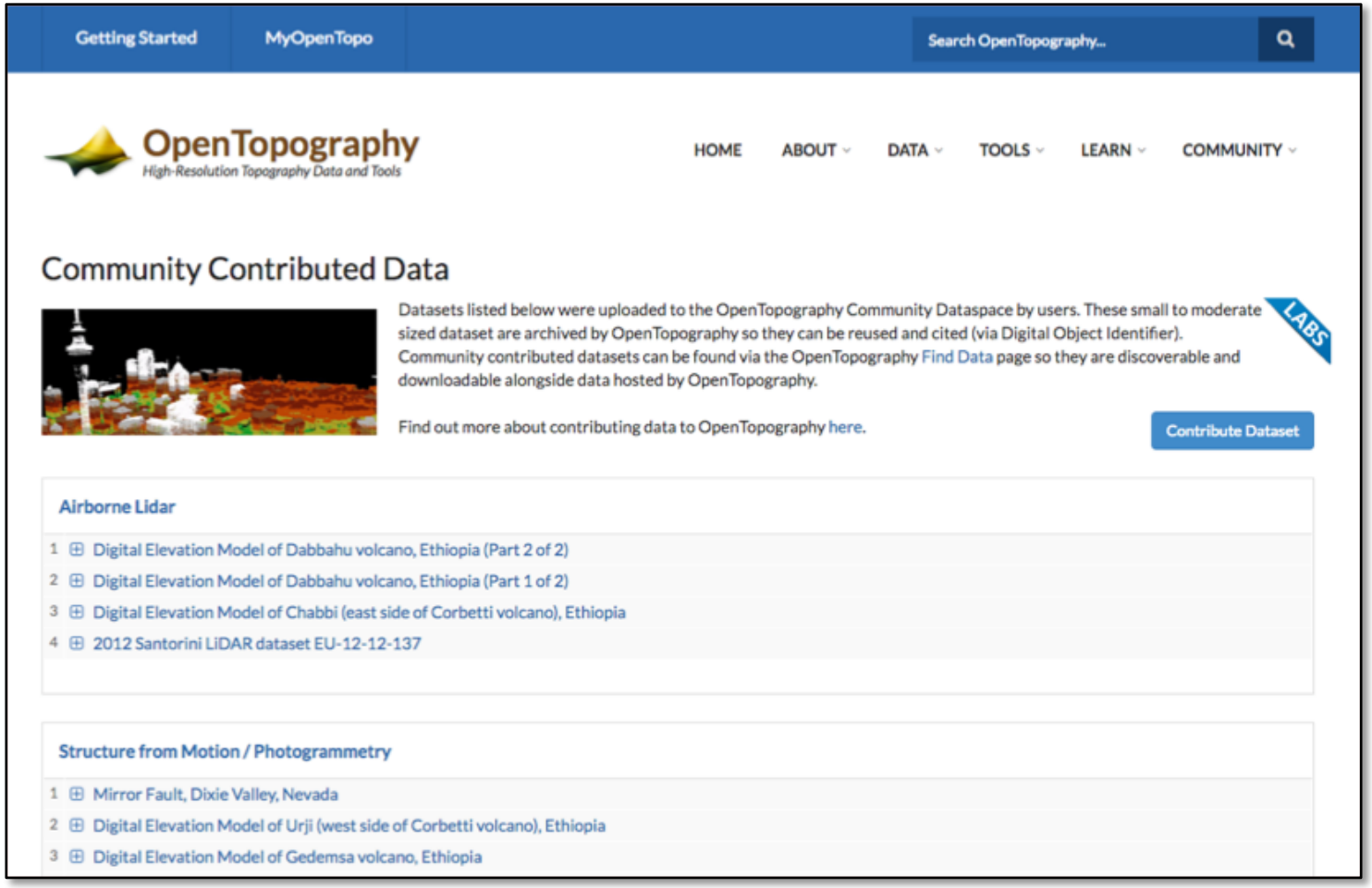
Community Dataspace Tutorial

The OpenTopography Community Dataspace allows users who are producing small to moderate sized topographic datasets to:

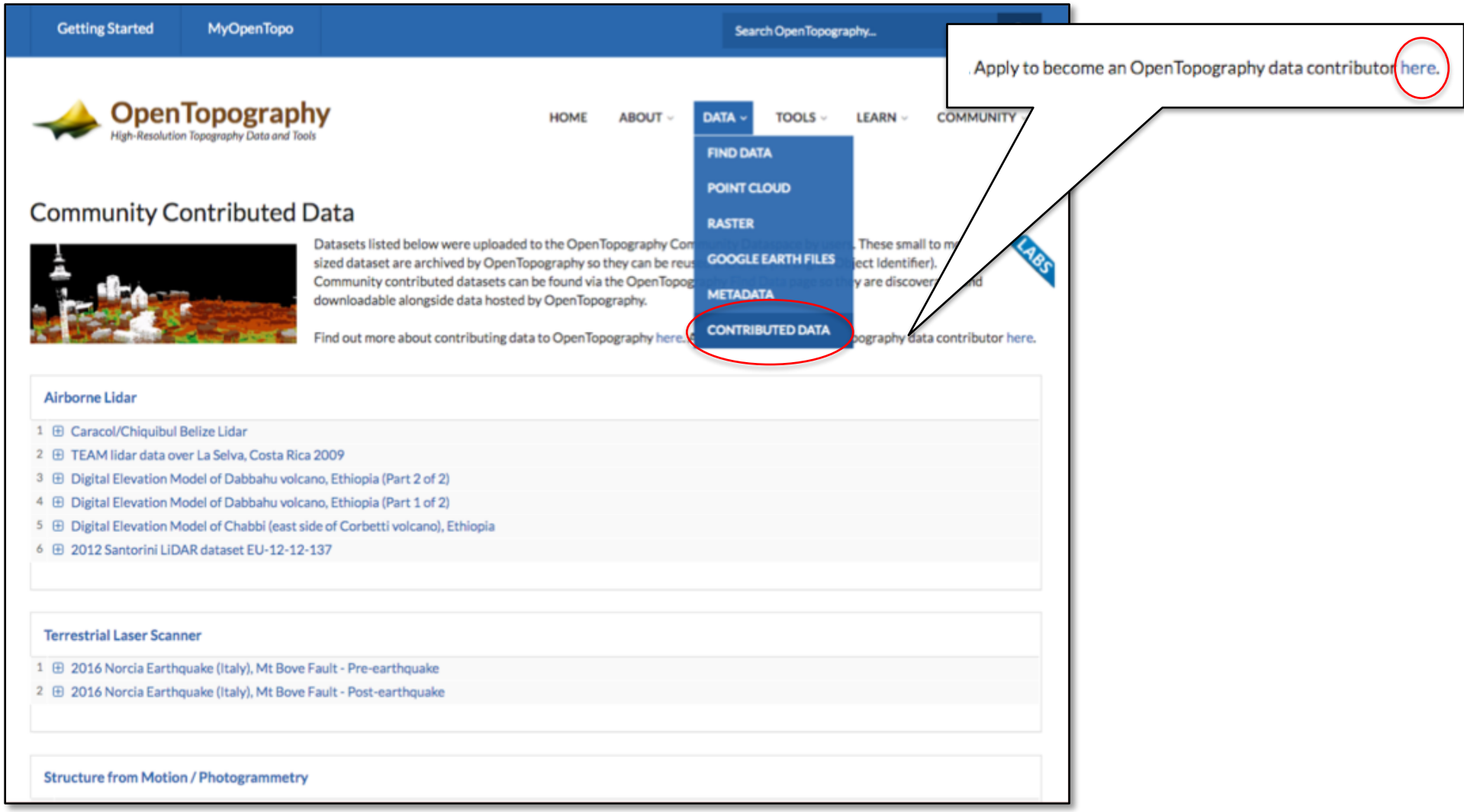
- Archive their data with OpenTopography
- Establish proper dataset citation
- Enable dataset sharing and reuse
- Provide educational applications
- Grid, subset, and visualize their data with standard OpenTopography tools (only if data is provided in a projected coordinate system)



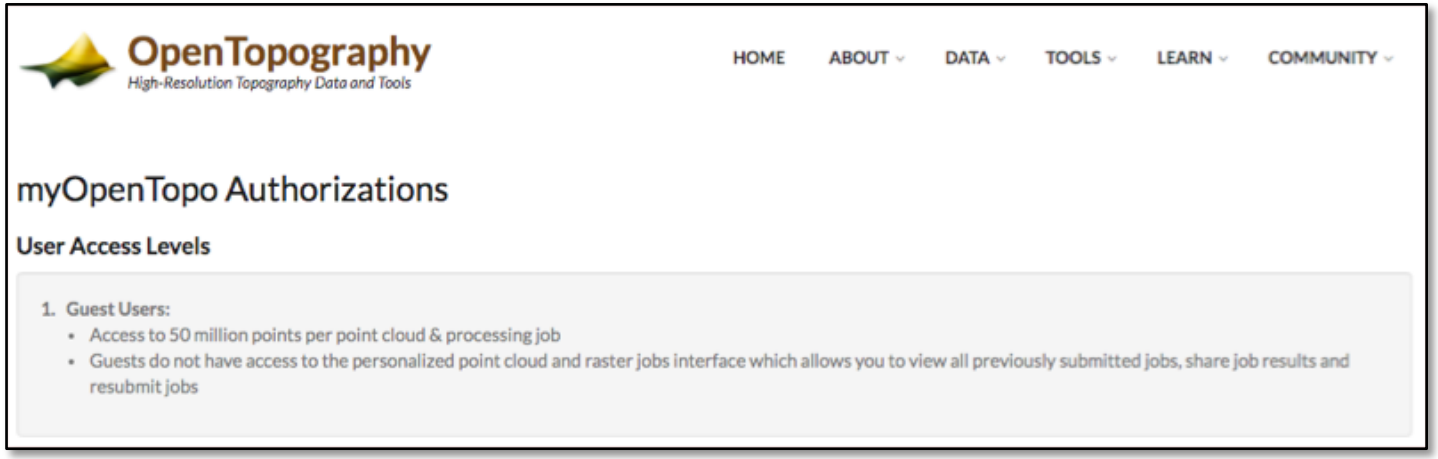
Community Dataspace Tutorial

If you do not have a Community Dataspace account, you can apply for one:

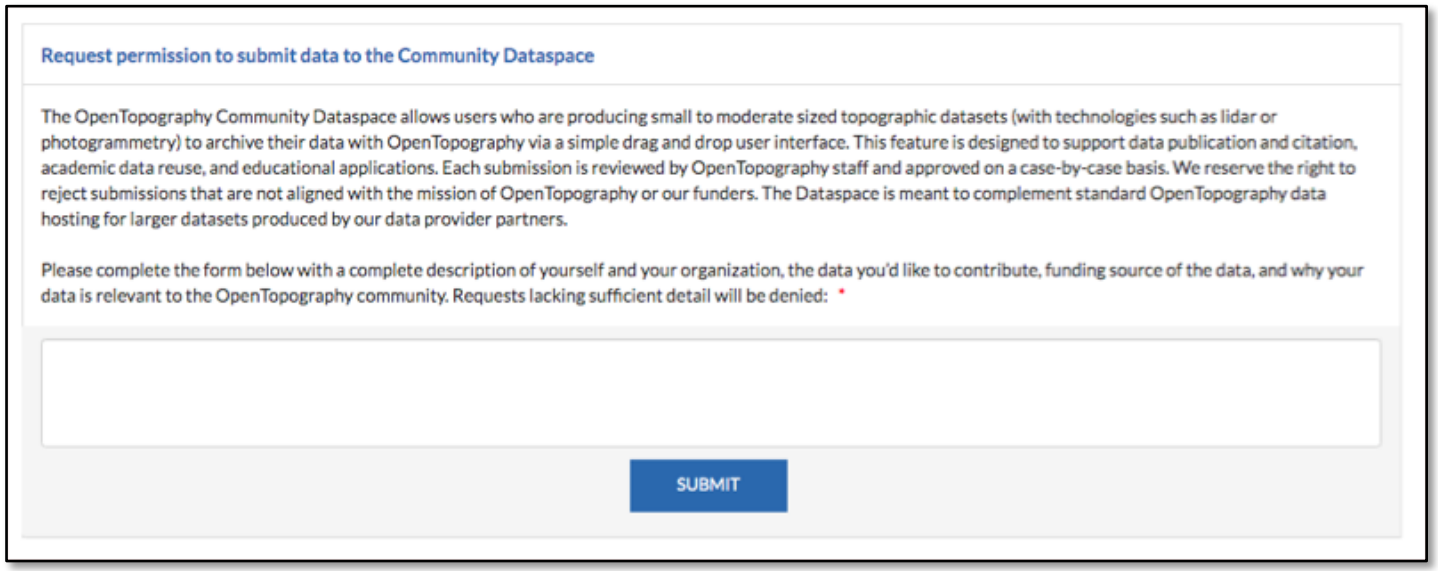
- Go to Data→ "Contributed Data". From that page there will be a link to click in order to apply to become a data contributor.



Users will be directed to their MyOpenTopo Authorizations page:

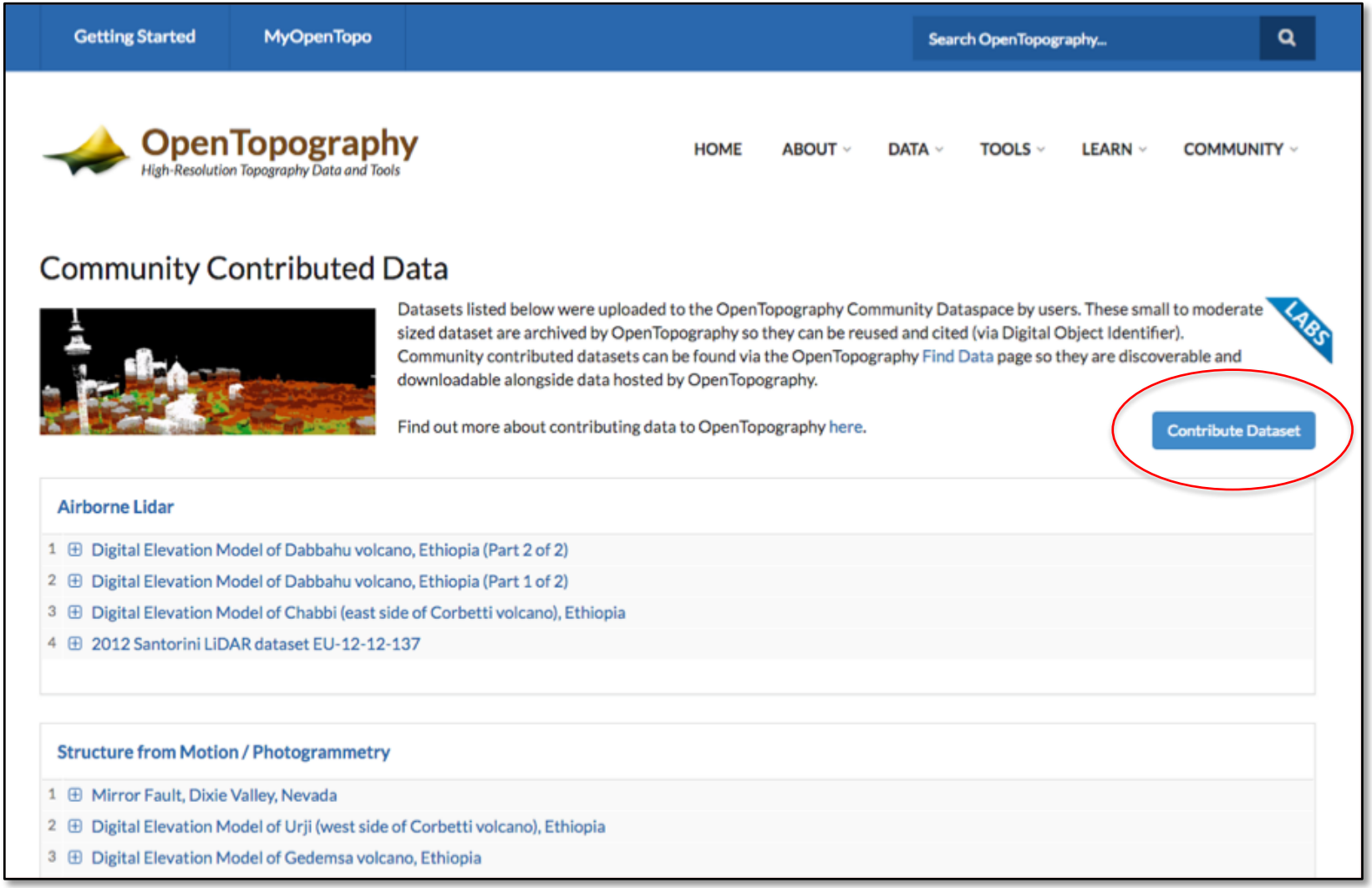


Scroll down to the bottom to request permission to submit data to the Community Dataspace:



Once the account is set up, log into your OpenTopography account:

- Click on the “Contribute Dataset” button to start the process.



Preparing your datasets for upload:

- We **strongly recommend** that all datasets be in a projected coordinate system. Only data in a projected coordinate system will be available for downstream processing (i.e. gridding of point clouds, 3D point visualization, etc.). Please make sure that the coordinate system info is included in the file header. In addition, make sure that the same coordinate system is used for all uploads (i.e. don't mix different coordinate systems in a single submission).
- Point cloud data should be compressed into LAZ format to reduce file sizes. This can be done with LAStools, PDAL, and other point cloud processing software packages. Also, do not submit a single Large LAZ file. **If a file is larger than 1 GB, it probably should be tiled into smaller, tiled LAZ files.**
- For compressing DEMs, we recommend the "deflate" algorithm. It provides more compression than LZW (although LZW is also fine). To create a compressed COG (Cloud Optimized GeoTIFF) in GDAL (V3.1 and later) you would use a command like this:

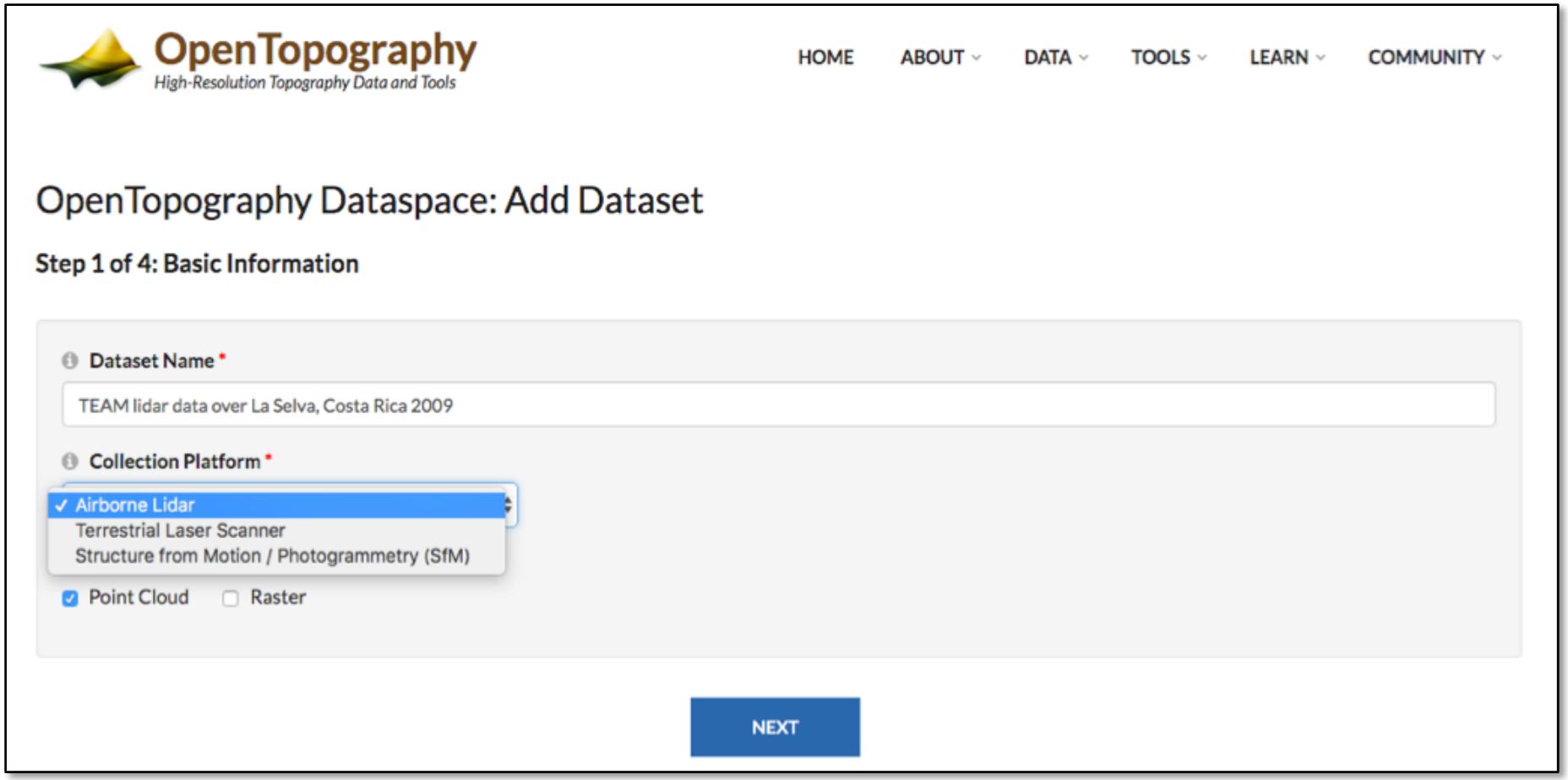
  gdal_translate -of COG -co "COMPRESS=DEFLATE" inputDEM.tif outputDEM_COG.tif
- For orthoimagery, using JPEG compression REALLY helps, so we recommend the following GDAL command (V3.1 and later) to compress your ortho imagery:

  gdal_translate -of COG -co "COMPRESS=JPEG" -co "PHOTOMETRIC=YCBCR" input.tif output_COG.tif

  The JPEG compression drastically reduces file sizes, but only use it on orthoimagery. **Don't use JPEG compression on DEMs**. More details on orthophoto compression are in Paul Ramsey's blog post here: http://blog.cleverelephant.ca/2015/02/geotiff-compression-for-dummies.html

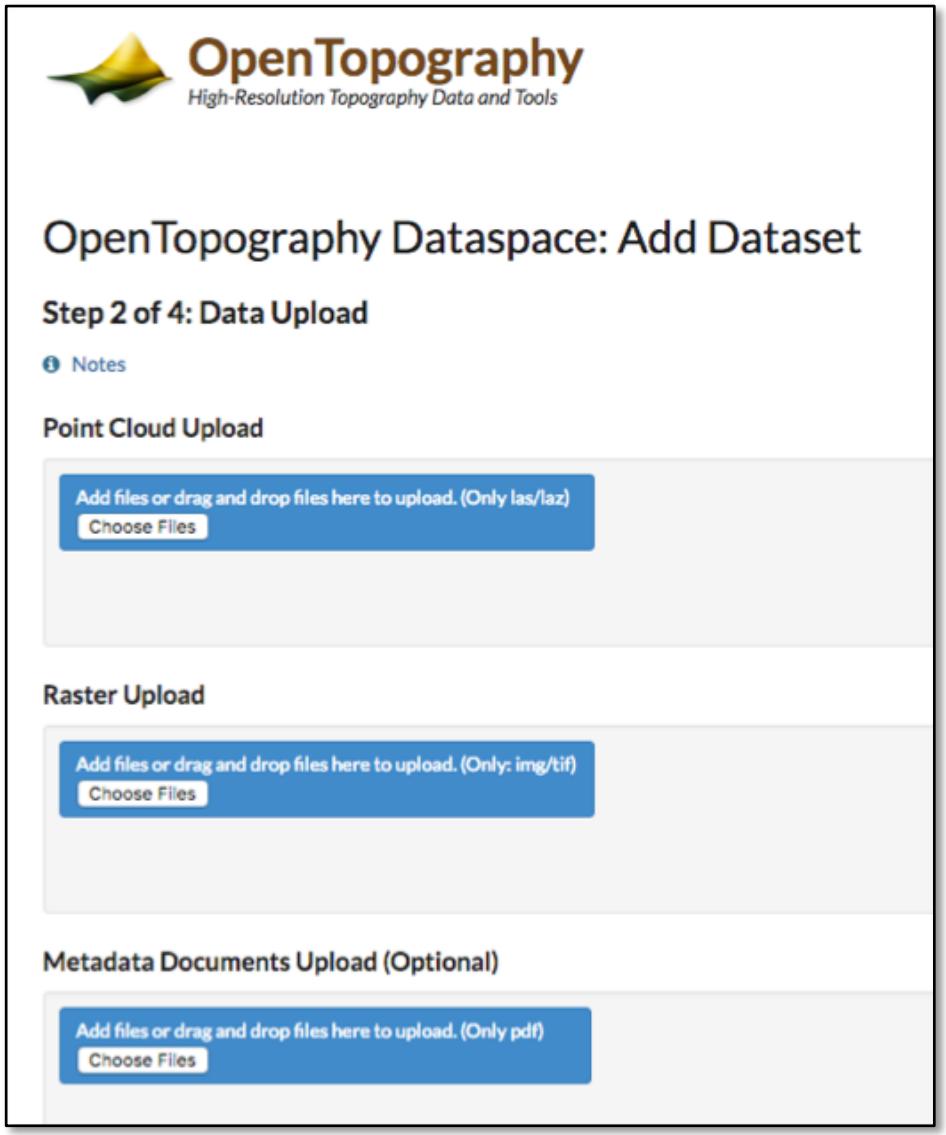
Step 1:

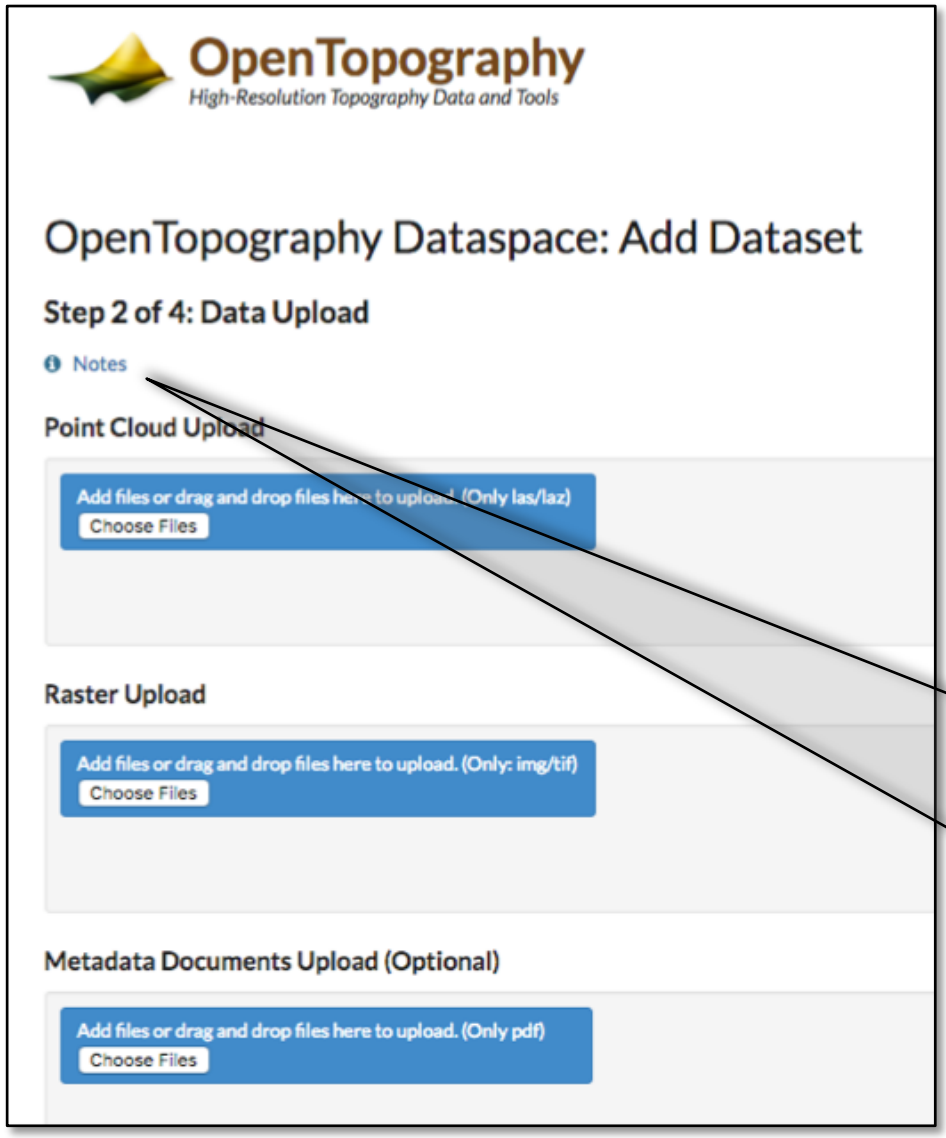
- Give your dataset a name. **This title will be used in your data citation**, so make it unique and descriptive. It's best to follow a template of "*Measured Feature, Location, D*ate" For example: "*Survey of Ozaukee South Beach Bluff, Wisconsin, May 2018*"
- Choose the collection platform that was used to collect your dataset.
- Check the appropriate box if you dataset contains Point Clouds, Rasters, or both.
- Select "Next" when done.



Step 2:

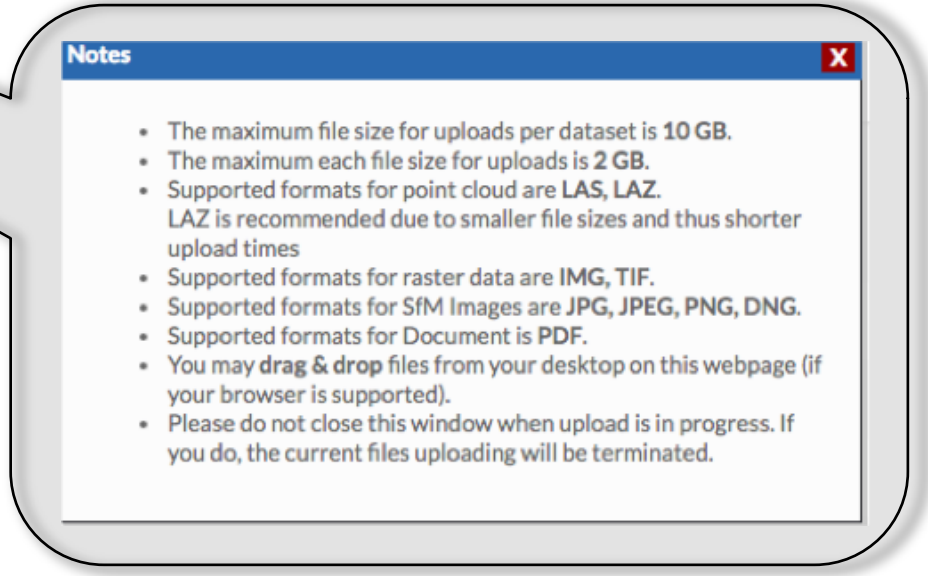
- For each file type, upload the appropriate files.
- Point Clouds: We currently only accept LAS or LAZ formats. LAZ formats are preferred as they reduce file size drastically, and will reduce upload times. **We strongly recommend that all datasets be in a projected coordinate system. Only data in a projected coordinate system will be available for downstream processing** (i.e. gridding of point clouds, 3D point visualization, etc.) **Please make sure that the coordinate system info is included in the header**.
- Rasters: We currently only accept IMG or TIF files. **Please make sure that the coordinate system info is included in the header**. If feasible, use the same coordinate system for both the raster and point cloud products.
- Metadata: The metadata document must be in PDF format. Users are free to use whatever format they wish, but it is preferrable if the OpenTopography template is used. The template is available on the data upload pages.





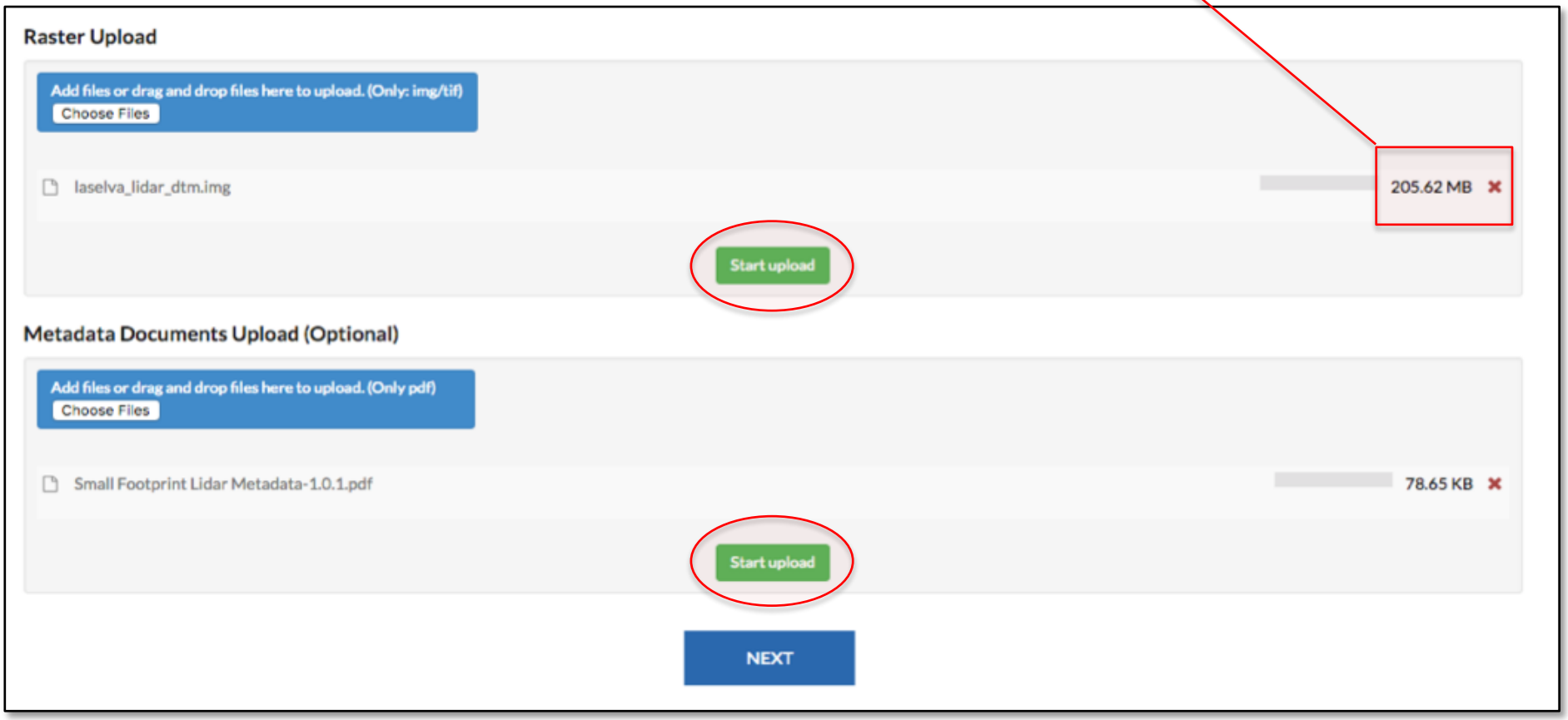
Notes:

- Click on the "Notes" link for guidance on upload requirements.
- Users can upload multiple files through the "Choose Files" button, or by dragging and dropping files.
- Although there is a 10GB limit on a single submission, if a project is larger than 10GB, it should be broken into parts and submitted with part numbers designated clearly.



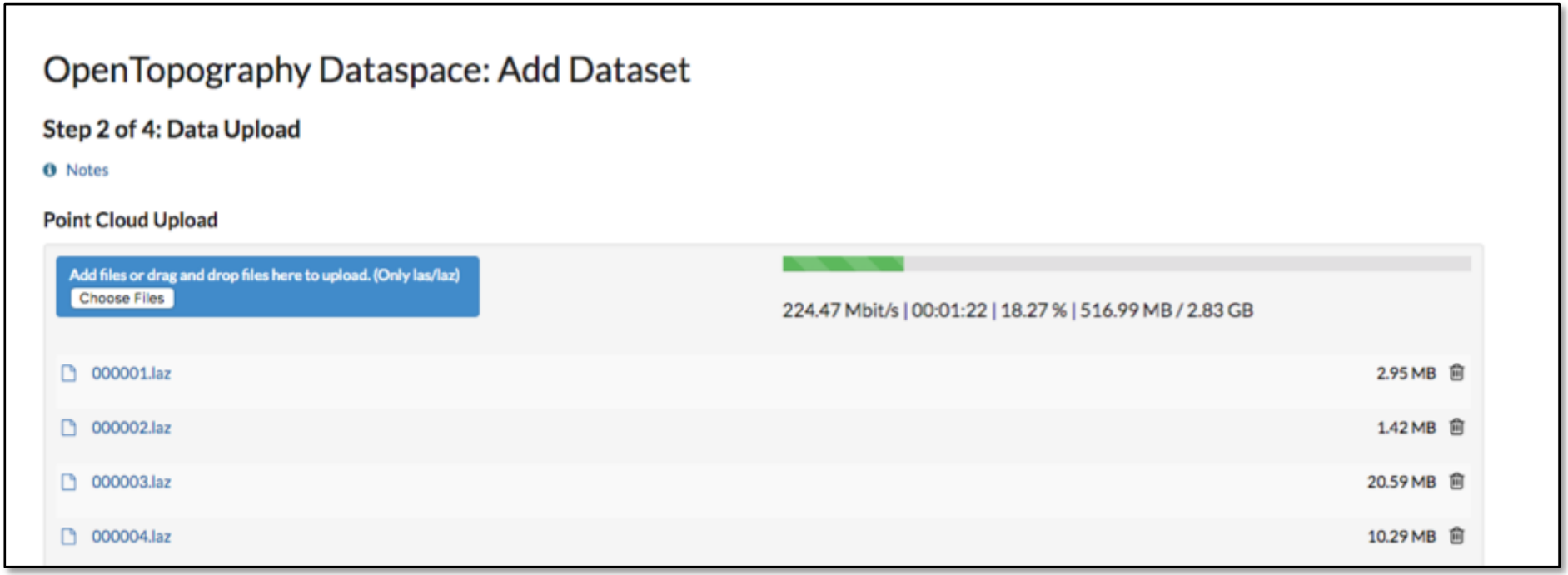
Step 2 (continued):

- Add the files by using the “Choose Files” button or by simply dragging the files into the designated area of the window.
- When all files have been added, select the green, “Start Upload” button to begin the upload process.
- There is a “Start Upload” button for each category of file, so be sure to perform this operation for each of the sections you added.
- Note that if you have not uploaded successfully, there will be a small “x” next to the file size.



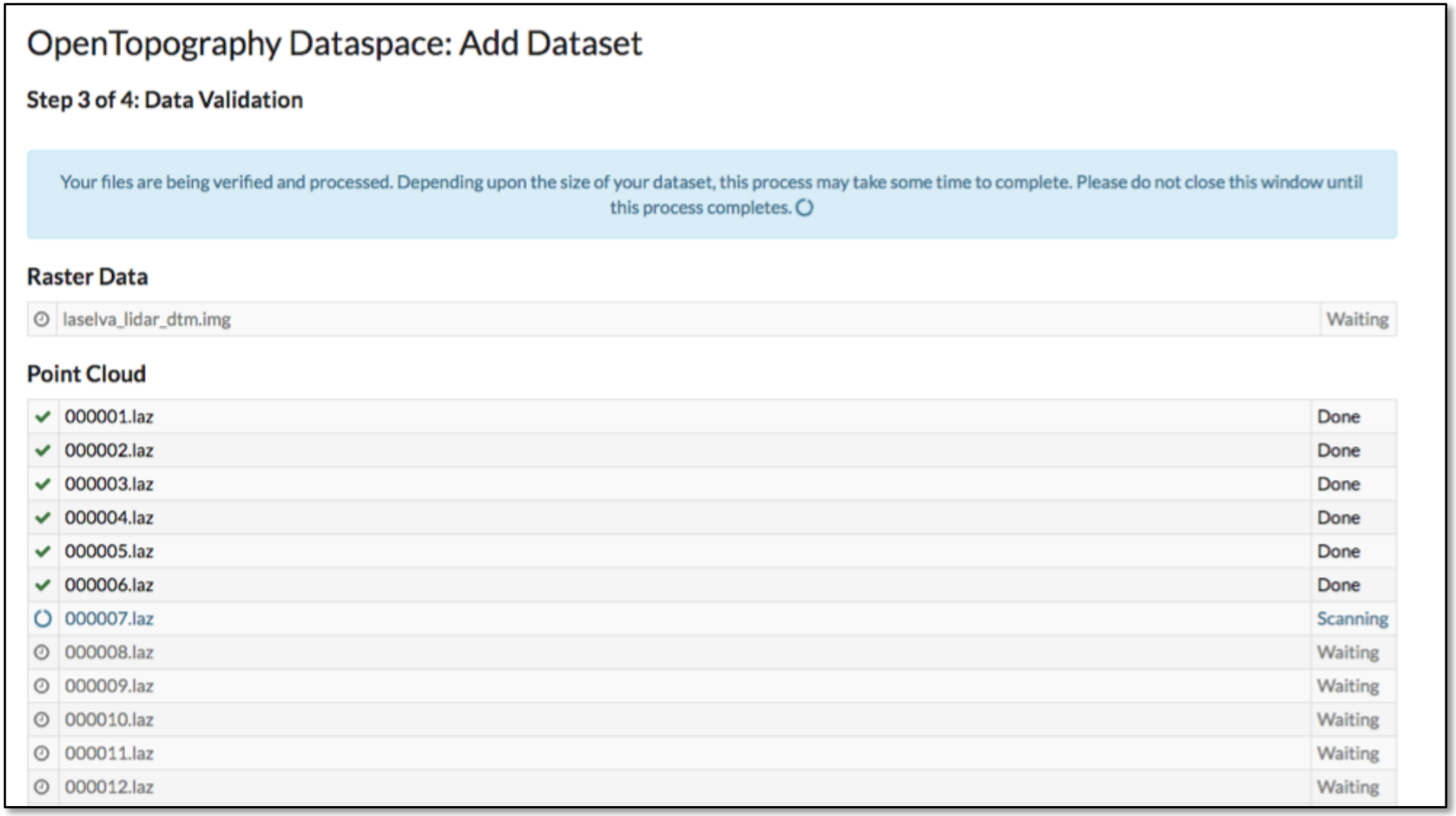
Step 2 (continued):

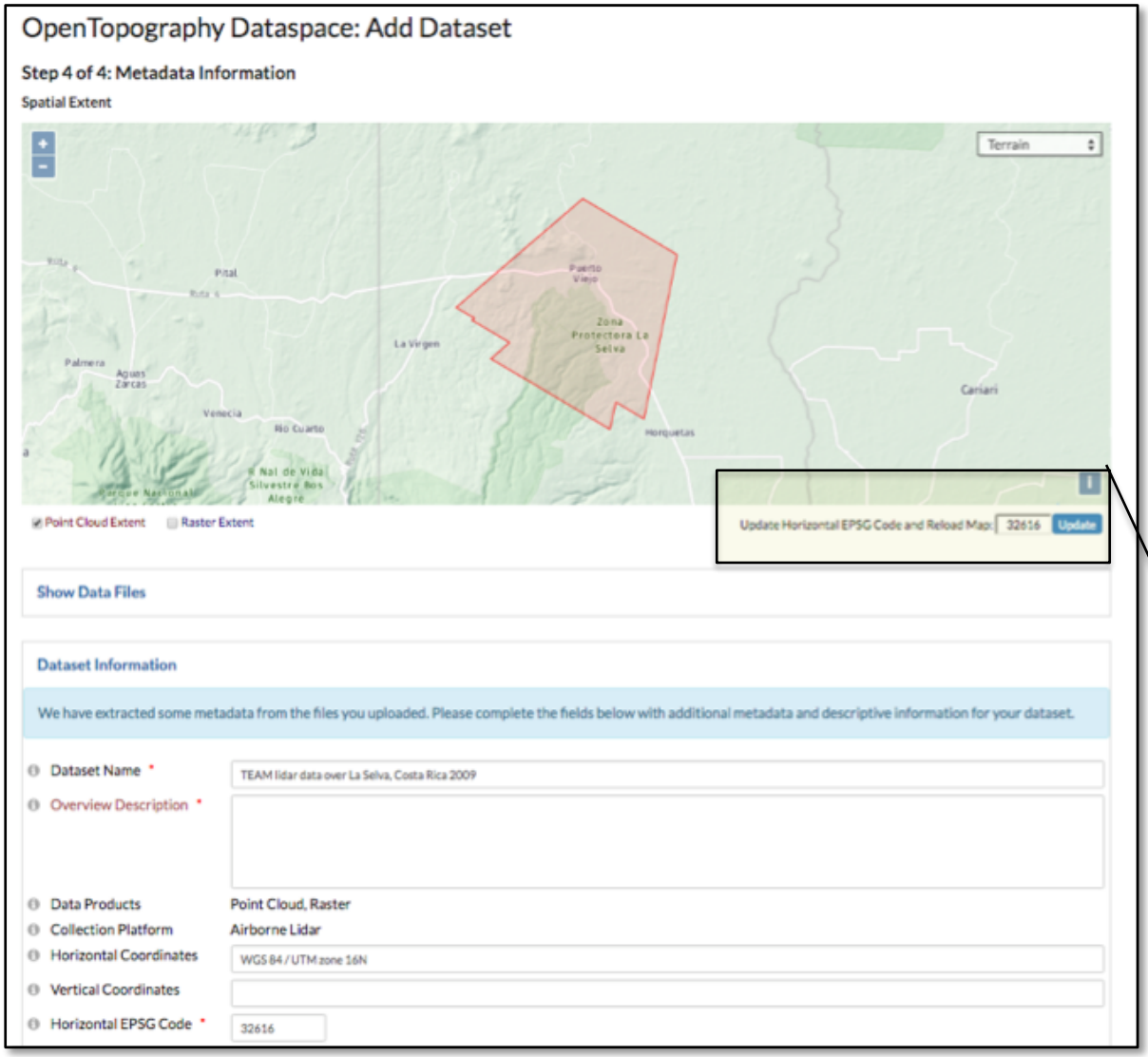
- Once you select the "Start Upload" button, you will see the progress bar as each file is uploaded to the San Diego Super Computing (SDSC) server.
- When all the files have completed the upload process, select the "Next" button at the bottom of the page.



Step 3:

- This step validates the input data. The amount of time this requires will depend on the size, and amount files uploaded.
- The progress of the validation is displayed on the website as seen below.



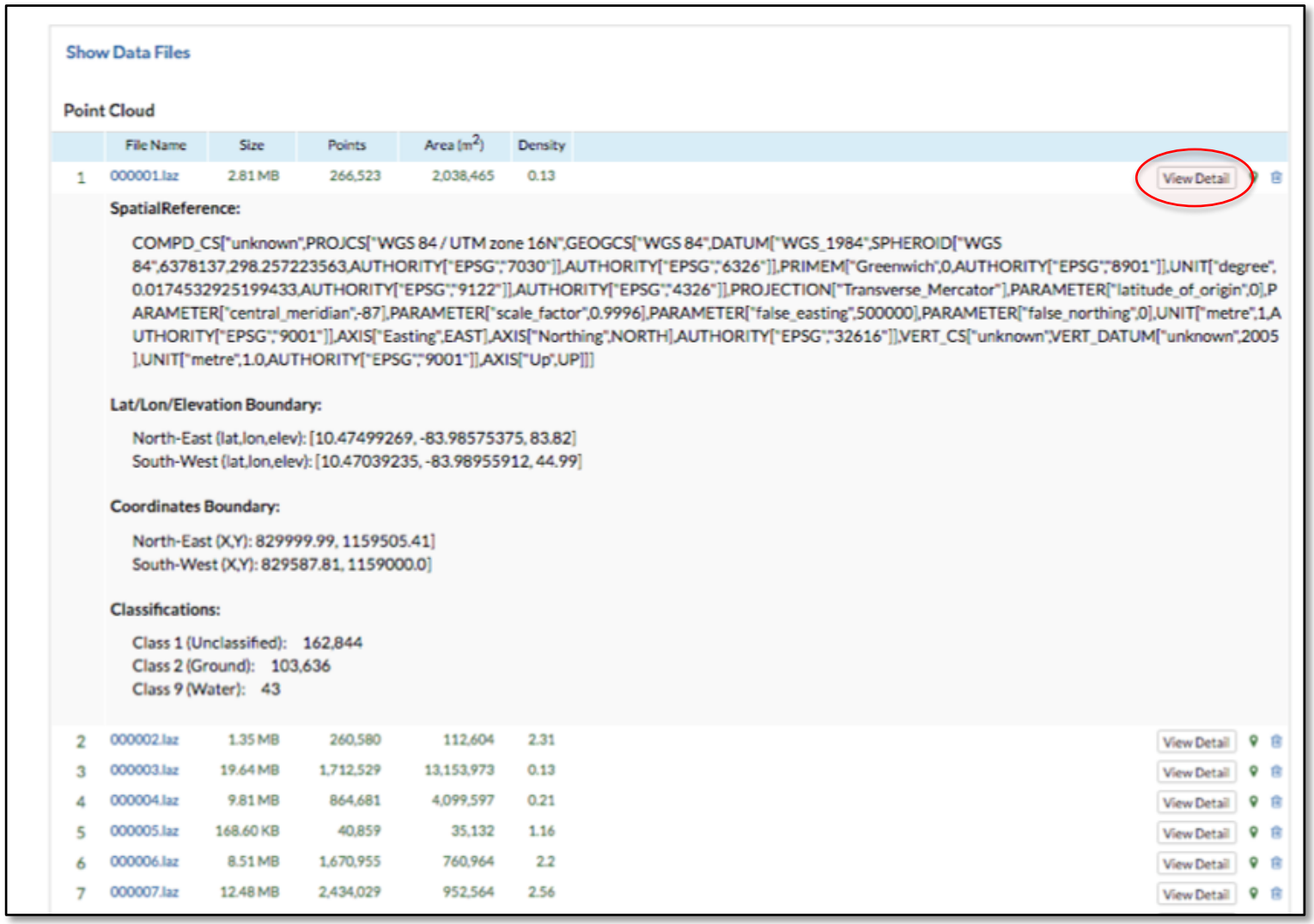


Step 4:

- If the validation completes successfully, a page will load that shows a map with the total spatial coverage of all the datasets in red.
- There will also be a series of text boxes to add additional information about the datasets.
- Required fields are marked with a red asterisk.
- If the map does not show the footprint locations, it could mean that either the data did not have proper coordinate system information, or it was not recognized on ingest. If you know the EPSG code of the data, you can enter this code in the box below the map to update the coordinate reference system.

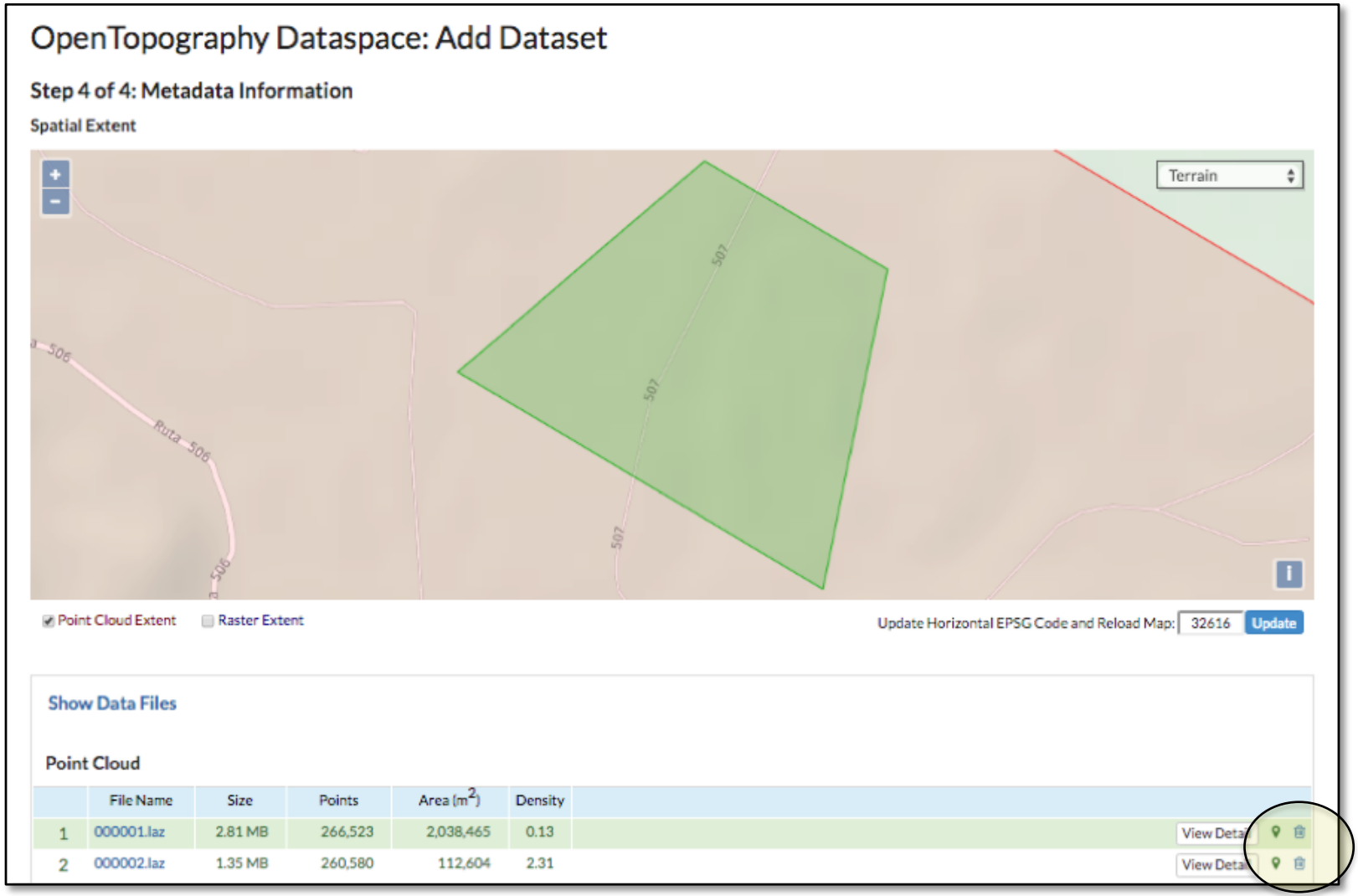
Step 4 (Continued):

- Clicking on the "Show Data Files" bar will show a list of the files that were ingested and some basic statistics on each file.
- Clicking on the "View Detail" button will display additional metadata for each file.



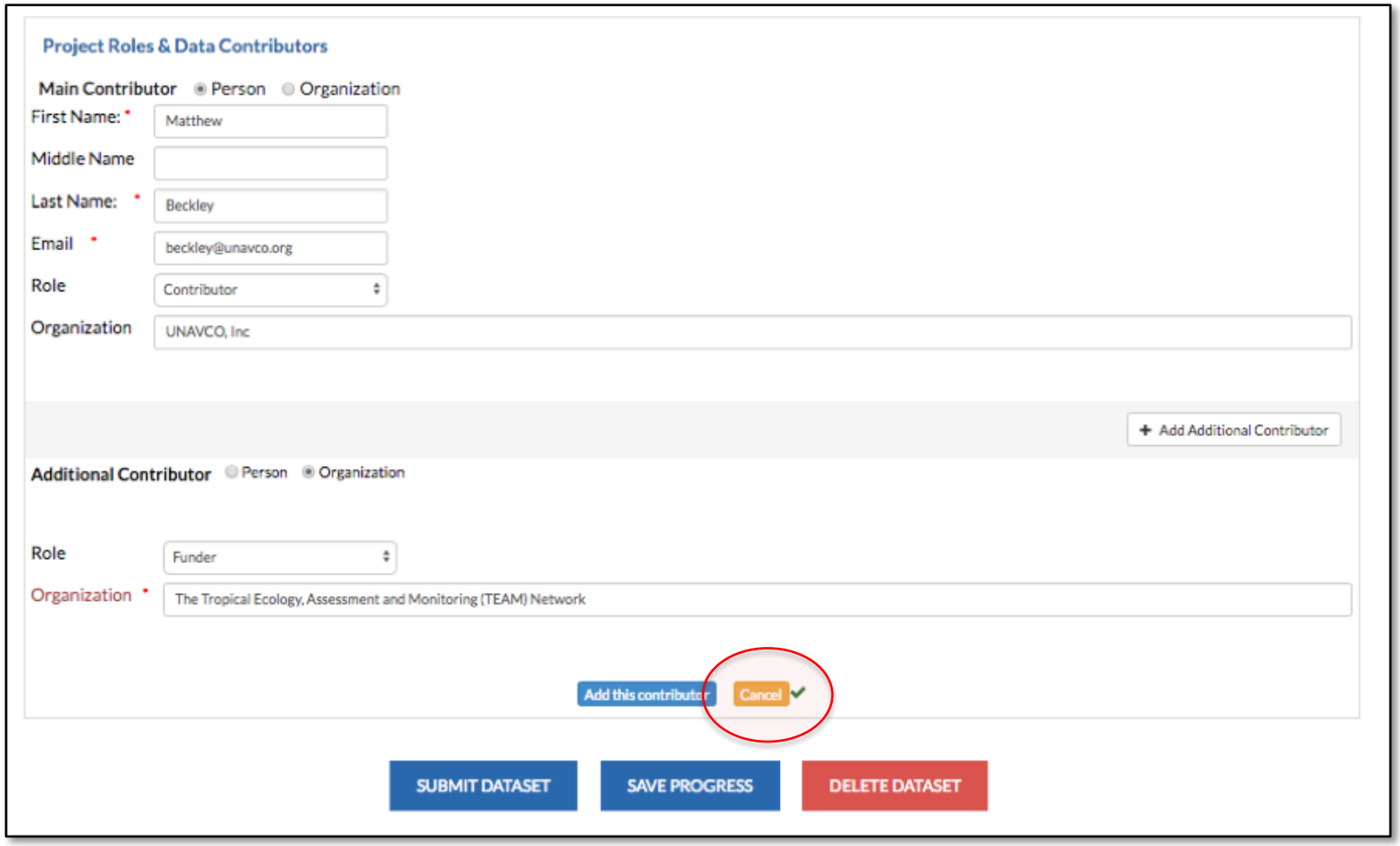
Step 4 (Continued):

- Clicking on the map pin icon on the right hand side of each file listing will update the map to show the spatial extent of that file colored in green.
- Clicking on the garbage can icon will delete a file from the collection.



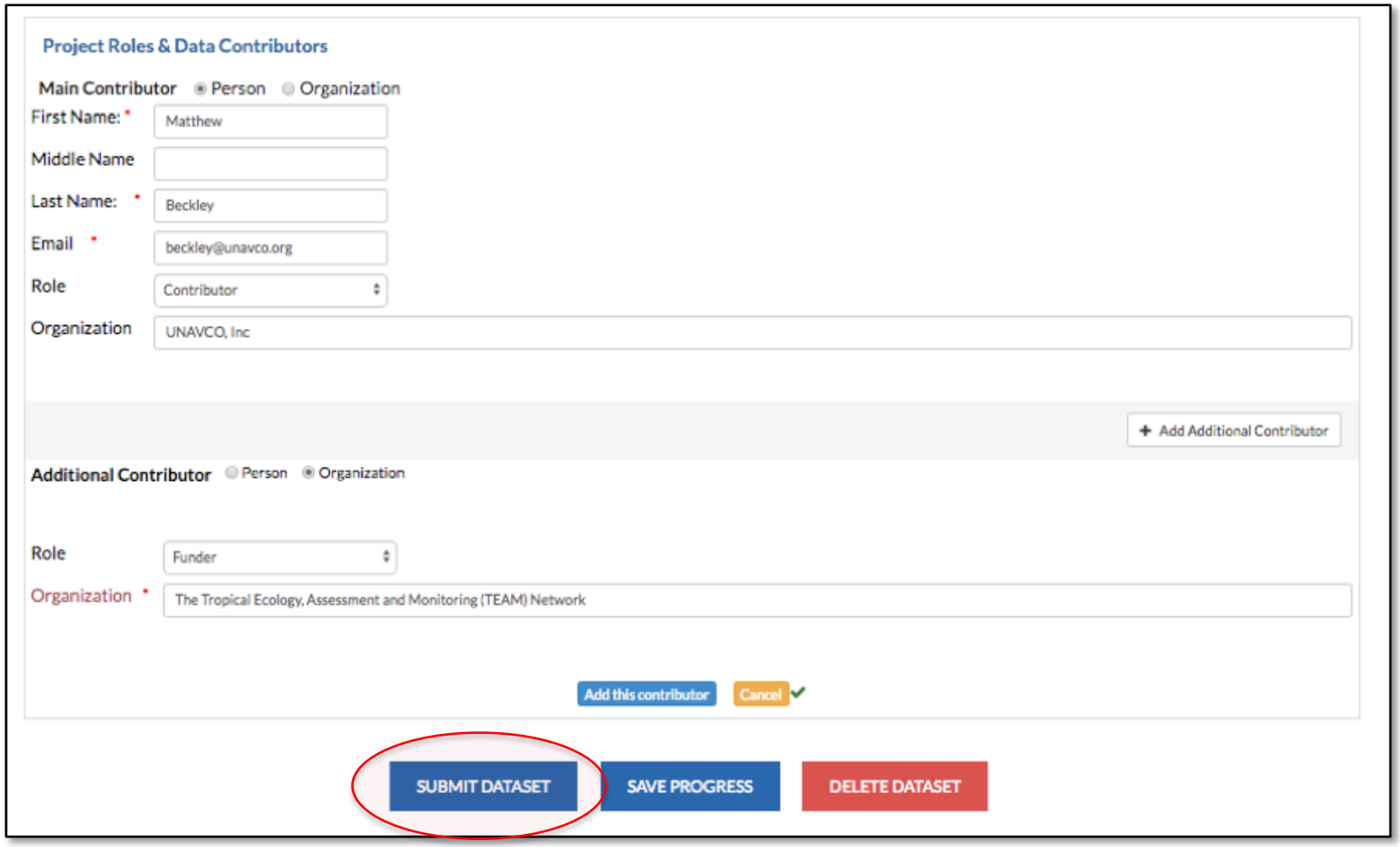
Step 4 (Continued):

- **It is best practice to provide as much information about the dataset as possible**, so make an effort to fill in all applicable fields.
- Project Roles and Contributors can be added to help provide contact information. After adding a contributor, a green check mark will display to show that the user or organization has been added successfully.



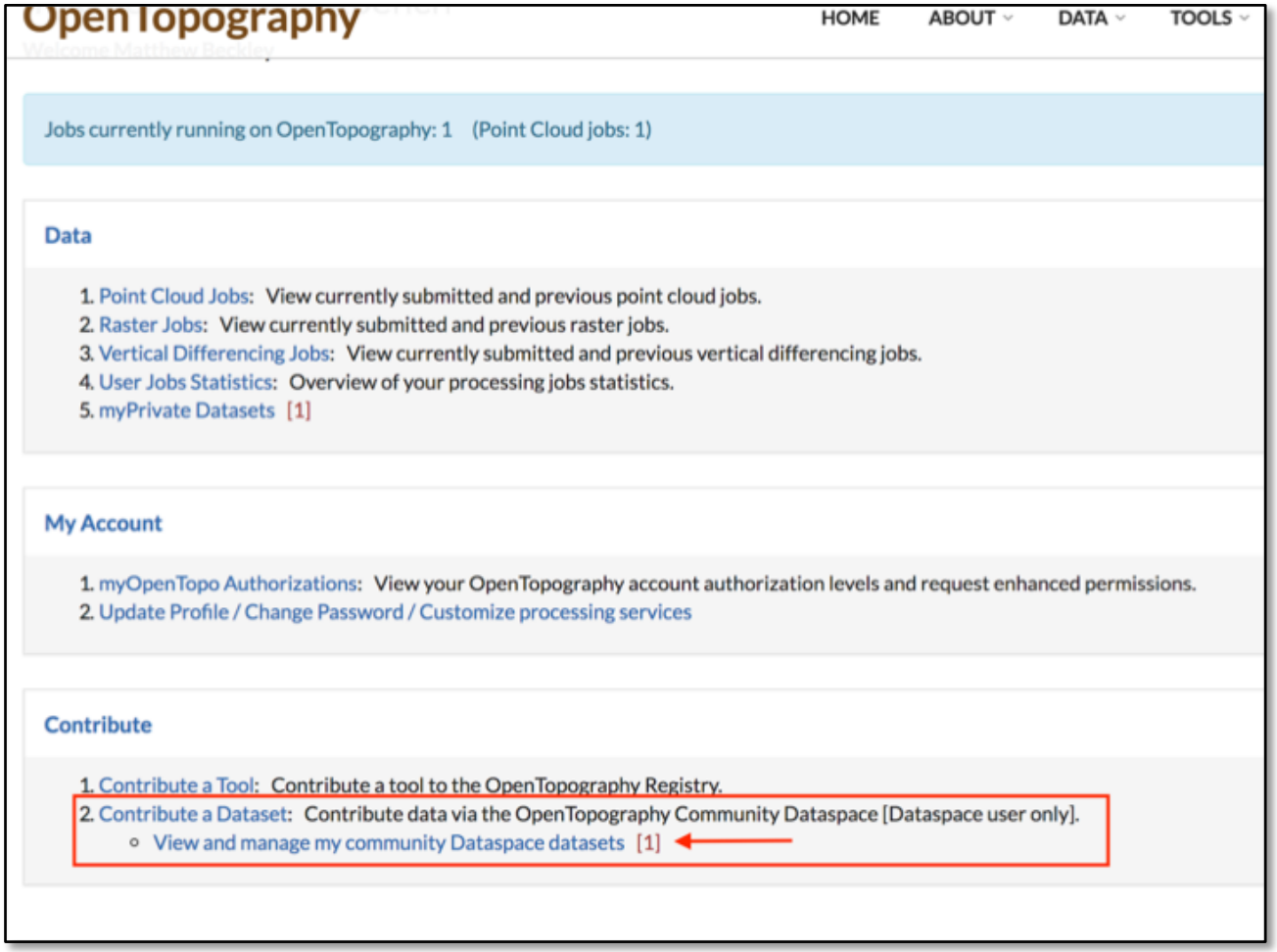
Step 4 (Continued) Final Submission:

- Use the "Save Progress" button to save work, but delay submitting the dataset for another time.
- When all the data entry is completed, and the dataset is ready for final submission, select the "Submit Dataset" button. The page will refresh and display a banner at the top that states that the dataset has been successfully submitted and is pending OpenTopography approval.
- You are done! Once an OpenTopography team member reviews the submission, an email will be sent to your account with a link to verify the data submission and display the data home page.



Tips and Troubleshooting:

- Use LAZ files when possible. The compression drastically reduces filesizes, and will speed up the whole process.
- When appropriate, verify that files have coordinate system information in the header.
- If uploads are taking a long time, or have stalled, the process may have still completed. Users can check the status of the upload by going to MyOpenTopo → Contribute → "View and manage my community Dataspace datasets" (see image below). Select the edit button next to the project of interest, and verify that all the files are uploaded and have correct statistics.



Tips and Troubleshooting (continued):

- When working with TLS data, sometimes errors can occur in the upload because the distribution of points between the files is very uneven – e.g. files containing points near the scan-center will have MUCH larger point counts than files containing points far from the scanner. This large discrepancy in file sizes can cause errors when uploading to the Dataspace. The easiest solution right now is to split the data using lassplit from the LAStools package (https://rapidlasso.com/lastools/). The following command can be used:

  lassplit -v –i input.laz –o output_tile.laz -digits 2 -split 25000000

  This splits the the laz file into files with 25,000,000 points in each file. The "digits" keyword is just telling the software to increment the filenames by 2 digits, so output filenames will be in the form: output_tile00.laz,output_tile01.laz, etc.

- Note that if there is still an uneven distribution of points between the files, it may be necessary to manually split the files, with the XY location of the scanner as the split location. This can be accomplished by using the las2las command, and manually specifying the XY range of the data to include in each file. Using the "-keep_xy" keyword enables the ability to cut out data for a specified XY range. For example:

  las2las -i input.laz –o output.laz -keep_xy 350601 4752322 351021 4752696

Tips and Troubleshooting (continued):

- For citation of datasets contributed to the Community Dataspace, a citation will be automatically generated in the following format:

  PI's lastname, PI's first/middle initials. (public release year): Title of dataset collection. Distributed by OpenTopography. link to DOI. Date of Data Access.

  Examples:
  Hiscock, A. (2019). 2019 ZigZag Landslide Reconnaissance, Morgan County, Utah. Distributed by OpenTopography. https://doi.org/10.5069/G96D5R4Q. Accessed: 2020-01-22

  Amatulli, G., McInerney, D., Sethi, T., Strobl, P., Domisch, S. (2020). Geomorpho90m - Global High-Resolution Geomorphometry Layers. Distributed by OpenTopography. https://doi.org/10.5069/G91R6NPX. Accessed: 2020-01-22

  Gruetzner, C. (2019). Suusamyr Basin, eastern 1992 surface ruptures, Kyrgyzstan. Distributed by OpenTopography. https://doi.org/10.5069/G9V69GRT. Accessed: 2020-01-22

  Arrowsmith, J R., Pisciotta, F. and CERG-C (2018): Photogrammetric model of a portion of the Faraglione--a rock outcropping in the Vulcano Town area, Vulcano Island, Sicily, Italy (point cloud [17M points] and DEM [6 cm/pix]). Distributed by OpenTopography. . https://doi.org/10.5069/G9WD3XPD Accessed October 29, 2018