

Geospatial and Image Data from the “When Computers Dream of Charcoal: Using Deep Learning, Open Tools and Open Data to Identify Relict Charcoal Hearths in and Around State Game Lands in Pennsylvania” Paper



WESTON CONNER 

BENJAMIN CARTER 

JEFF BLACKADAR 

**Author affiliations can be found in the back matter of this article*

]u[ubiquity press

ABSTRACT

These data were used to build an object detection model to locate Relict Charcoal Hearths (RCH) as described in the paper “When Computers Dream of Charcoal: Using Deep Learning, Open Tools and Open Data to Identify Relict Charcoal Hearths in and around State Game Lands in Pennsylvania” [1]. This is the first grouping of data for the paper above. The second grouping is also available in this journal, see “Object detection model, image data and results from the “When Computers Dream of Charcoal: Using Deep Learning, Open Tools and Open Data to Identify Relict Charcoal Hearths in and around State Game Lands in Pennsylvania” paper”.

These files consist of georeferenced Digital Elevation Model, Hillshade and Slope files derived from LiDAR for the State Game Lands (SGL). Included is a Shapefile and GeoJSON of State Game Land borders as well as the program used for downloading the LiDAR files. These data are stored on [Zenodo.org](https://zenodo.org).

CORRESPONDING AUTHOR:

Weston Conner

archaeologist, author, data collection and verification, GIS expert, Lehigh University, US
westonconner96@gmail.com

KEYWORDS:

Relict Charcoal Hearth; RCH; Mask R-CNN; Pennsylvania; Digital Elevation Model; Iron Production; Deep Learning

TO CITE THIS ARTICLE:

Conner W, Carter B, Blackadar J 2021 Geospatial and Image Data from the “When Computers Dream of Charcoal: Using Deep Learning, Open Tools and Open Data to Identify Relict Charcoal Hearths in and Around State Game Lands in Pennsylvania” Paper. *Journal of Open Archaeology Data*, 9: 7, pp. 1–6. DOI: <https://doi.org/10.5334/joad.80>

(1) OVERVIEW

CONTEXT

Spatial coverage

Description: State Game Lands within the State of Pennsylvania, United States.

(WGS84):

North_Bounding_Coordinate: 42.269479

South_Bounding_Coordinate: 39.719860

East_Bounding_Coordinate: -74.689583

West_Bounding_Coordinate: -80.519349

TEMPORAL COVERAGE

AD1700-AD1945

(2) METHODS

INTRODUCTION

This dataset consists of LiDAR derivatives covering the State Game Lands (SGLs) of Pennsylvania. It was created with the objective of automatically detecting all relict charcoal hearths (RCHs) in these SGLs. For details on these results, please see “When Computers Dream of Charcoal: Using Deep Learning, Open Tools and Open Data to Identify Relict Charcoal Hearths in and around State Game Lands in Pennsylvania” [1].

Herein we detail the steps taken to assemble the raw data and process it into a usable format. The unprocessed LiDAR data used for this project was obtained through Pennsylvania Spatial Data Access (PASDA) [2]. The process included identifying the data needed to encompass the SGLs and a 1 kilometer buffer around them, downloading the raw LiDAR files in .las format, organizing these files, and processing them into the following LiDAR derivatives: digital elevation map (DEM), hillshade, and slope analysis. The text below and [\(Figure 1\)](#) describe the detailed steps.

INITIAL STEPS

Pennsylvania’s LiDAR data is organized in two grids for the north and the south of the state, which overlap slightly [3, 4]. These grids were merged into a single shapefile via the “Merge Vector Layers” tool in QGIS. Shapefiles for these grids contain attributes such as direct links to download data for each grid tile. Next, we downloaded a shapefile containing SGL boundaries [5] provided by Pennsylvania and added a one-kilometer buffer around each SGL, which was given the same identifying number as the SGL it surrounded. By overlaying this buffer with the unified grid and using the “Join Attributes by Location” function, any grid tile that touched an SGL or its buffer was given that SGL’s number as an attribute. This provided a single shapefile of the grid with two central attributes for each tile, one with the SGL/buffer number and one with a download link for the LiDAR data for that tile.

Exporting the attribute table for the state’s tile grid into Microsoft Excel allowed us to manipulate the data more easily. We deleted all grids that were not associated with a SGL/buffer. This resulted in a selection of 3,925 unique tiles to download out of the 13,797 (28.4%) total tiles in the grid. One field given for each grid tile was a direct link to download the .las file, the raw LiDAR point cloud. Copying this list of URLs into a text document allowed us to use simple bash scripting and the *wget* command to automate the downloading of these files. Note that the list of tiles downloaded was actually over 4,000, not exactly 3,925. This is because tiles were downloaded consecutively by each individual SGL and some tiles overlapped more than one SGL.

In order to use bash scripting on a Windows computer, we enabled the Linux Subsystem and download a Linux Distro, which is described below. These instructions are current for Microsoft Windows 10 Home, Version 10.0.19041 Build 19041.

Enabling a Linux Subsystem on Windows

Enable the Linux Subsystem by following these steps:

- On a computer running Windows, open “Settings” from the Start Menu
- Click “Apps” | “Apps & features” | “Optional features”
- Click “More Windows features” under the “Related settings” submenu
- Scroll down to “Windows Subsystem for Linux” and check the box
- Click “OK”
- When prompted, restart computer

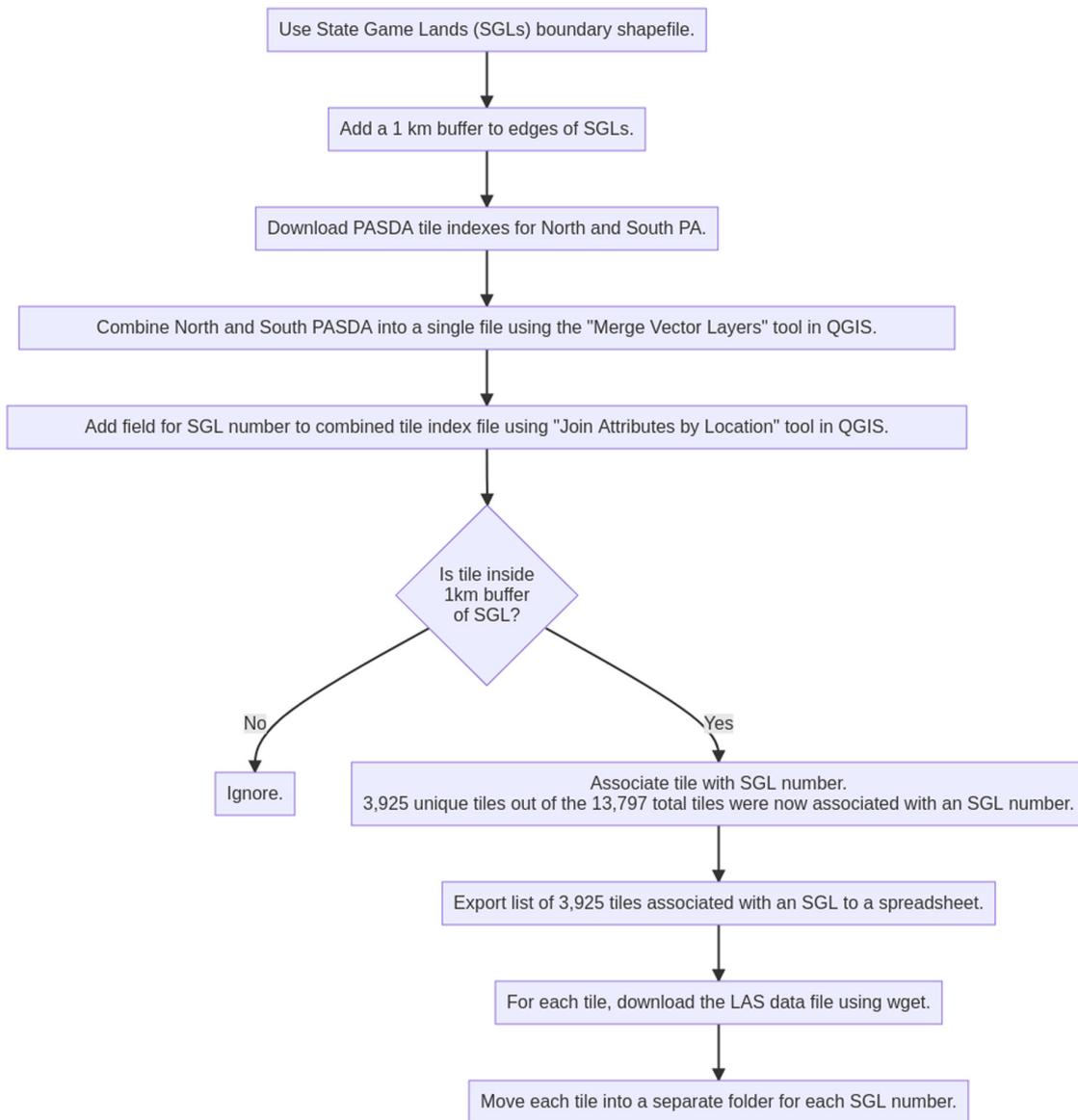
Second, download a Linux distribution through the Microsoft Store app. We used Ubuntu, published by Canonical Group Limited, and our instructions assume you are using the 20.04 LTS version [6]. Once you have downloaded Ubuntu, open the app and follow the prompts to create a username and password.

It is important to note that files saved through this system will be saved to the Linux Subsystem. It can be somewhat difficult to locate these files if you are not familiar with the program. To find where your files will be saved, open the File Explorer and follow these steps:

- Navigate to %localappdata% in the File Explorer search bar
- Open the “Packages” folder
- Housed within should be the folder: “CanonicalGroupLimited.UbuntuonWindows_79rhk1fndgsc” or the equivalent for your Linux distro
- Within this particular folder, navigate down through these subfolders:
 - “LocalState” - “rootfs” - “home” - *Username*

Steps to download and process LiDAR data files for State Game Lands.

Identify and download LiDAR LAS tiles.



Process LiDAR LAS tiles

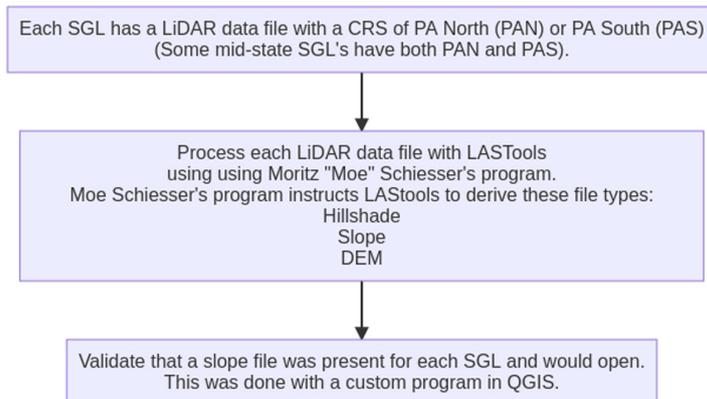


Figure 1 Steps to download and process LiDAR data files for State Game Lands.

This *Username* folder is where you must store and place files for your Linux subsystem to read and edit. It is also where files you download using Linux will be stored.

Downloading .las Files with *Wget*

To download the .las files using this Linux subsystem, we used the *wget* command. This command tells the computer to download a file from a specific URL or to download multiple files from a list of URLs in a text file. For this dataset, we copied the column from Excel containing the URLs of .las files for each SGL and pasted it into a text document. There was a single URL per line—this is critical. Blank lines separated each SGL into clusters within the text document due to Excel formatting, but this did not impact the functionality of using *wget*.

The *wget* command was formatted with several modifiers. To run *wget* as an administrator, which is necessary to execute the command, you must use *sudo* which means “superuser do.” Using the modifier *-i* allowed us to specify a text file to be read, and using the modifier *-P* allowed us to specify where the downloads should be saved. The final command executed in Ubuntu is:

```
sudo wget -i SampleList.txt -P sample_file_folder
```

In this example, *SampleList.txt* is the name of the text file containing the URLs and *sample_file_folder* is the name of the folder where downloads should be saved. Both this file and this folder should be in the ‘Username’ folder discussed earlier prior to running the command.

After pressing enter to run the command, *wget* will download the files from the list and save them in the specified folder. The Linux app will display progress and the files will appear in the specified folder.

This process may take some time, and we left *wget* running overnight. Using this process, over 4,000 LAS files for a total of about 350 gigabytes of data in zip files was downloaded. The files contained the LiDAR data for every tile touching a State Game Land, or their buffer, in Pennsylvania.

ORGANIZING FILES

Although we had downloaded all of the data, it was all stored within a single folder. Separating these files into separate folders by SGL was accomplished with a program (ProjectLambda) created by Moritz “Moe” Schiesser. Moe Schiesser’s program read an Excel sheet containing two columns – Column A: the current file location, Column B: desired file location – and copied each file to the appropriate folder. (See [Table 1](#)). Note that the file structure in Column B had been created prior to running this program.

A	B
F:\New LAS Files\26001780PAS.las	E:\Pennsylvania Iron and Charcoal\LAS Files\73\26001780PAS.las

Table 1 Excel sheet sample input for ProjectLambda program.

PROCESSING LIDAR DATA

The .las files, which were now separated into folders by SGL, were processed through LAsTools into three LiDAR derivatives: a digital elevation model (DEM), hillshade, and slope analysis [7]. This process was automated using another program (ProjectKappa) written by Schiesser. This program read the default file names for Pennsylvania’s .las files to determine which coordinate reference system (CRS) to use (State Plane 83 PA South or State Plane 83 PA North) and saved the three LiDAR derivative outputs to the same SGL folder. All tiles for each SGL were combined into one image (for each derivative) in TIF format. For those SGLs that contained portions in both the North and South CRS, it processed the files for each CRS separately, which means some SGLs had two DEMs, hillshades, and slope analyses. In the end, the total dataset, including all .las files and LiDAR derivatives was 1.3 terabytes.

SAMPLING STRATEGY

The project studied all SGLs in Pennsylvania. We added a 1km buffer outside of the SGL boundary for each file in order to ensure objects on the edges of SGLs were fully visible. All LAS tiles of which any portion fell within the SGL or buffer were included.

QUALITY CONTROL

We wrote a program to load each file into QGIS according to the list of SGLs in order to validate each file was present and opened correctly. An example of this program loads the layers Northeast Pennsylvania: https://github.com/jeffblackadar/charcoalhearths/blob/master/qgis_3_load_rasters_north_east.py.

CONSTRAINTS

The files for SGLs 264 and 258 could not be processed due to problems in the source files.

(3) DATASET DESCRIPTION

OBJECT NAME

Since the dataset contains multiple objects, filenames are displayed in Zenodo.

DATA TYPE

Primary data: Digital Elevation Model, Slope, Hillshade files. Secondary data: SGL boundaries.

FORMAT NAMES AND VERSIONS

TIFF, Shapefile, Geojson

CREATION DATES

01/01/2020 – 31/08/2020

DATASET CREATORS

Weston Conner, archaeologist, author, data collection and verification, GIS expert, Lehigh University. <https://orcid.org/0000-0001-9906-3762>

Moritz “Moe” Schiesser, Programmer.

LANGUAGE

English

LICENSE

CC-BY.

REPOSITORY LOCATION

Documentation: <https://zenodo.org/record/4766351>

Data for:

1. State Game Lands with a 1 km buffer
 - a. Shapefile: <https://zenodo.org/record/4593518>
 - b. GeoJSON: <https://zenodo.org/record/4593540>
2. Grid tiles with SGL information included
 - a. Shapefile: <https://zenodo.org/record/4593491>
 - b. GeoJSON: <https://zenodo.org/record/4593572>
3. Description for downloading LiDAR tiles (in LAS format): <https://zenodo.org/record/4596905>
4. Program for arranging files downloaded with the above, known as ProjectLambda: <https://zenodo.org/record/4572993>
5. Program for processing original LiDAR data (in LAS format) to yield Digital Elevation Model, Hillshade and Slope, known as ProjectKappa: <https://zenodo.org/record/4573004>
6. Files resulting from the above for State Game Lands (in GeoTIFF format) [2]
 - a. SGL 12–50: <https://zenodo.org/record/4596903>
 - b. SGL 51–80: <https://zenodo.org/record/4597896>
 - c. SGL 81–110: <https://zenodo.org/record/4597943>
 - d. SGL 111–160: <https://zenodo.org/record/4597941>
 - e. SGL 161–200: <https://zenodo.org/record/4597994>
 - f. SGL 201–250: <https://zenodo.org/record/4598045>
 - g. SGL 251–260: <https://zenodo.org/record/4598089>
 - h. SGL 261–280: <https://zenodo.org/record/4598096>
 - i. SGL 281–335: <https://zenodo.org/record/4598100>

PUBLICATION DATE

11/03/2021

(4) REUSE POTENTIAL

This dataset is useful for further study of Pennsylvania’s SGLs. Generally, SGLs are relatively undeveloped today,

preserving archaeological evidence such as RCHs, building foundations and roadbeds. The Shapefile and GeoTIFFs may be used for further study of the SGLs for RCHs or other types of objects. The method and programs are useful for the collection and processing of LiDAR files to study of areas of Pennsylvania outside of the SGLs. The process may also be adapted for use with LiDAR datasets for geographic areas outside of Pennsylvania.

ACKNOWLEDGEMENTS

Thank you to Moritz “Moe” Schiesser for writing the programs to organize LAS tiles and create Hillshade, Slope and DEM files.

FUNDING STATEMENT

We would like to acknowledge the support of the Social Sciences and Humanities Research Council of Canada. We thank the Provost’s Office and the Department of Sociology and Anthropology of Muhlenberg College for financial and logistical support.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Weston Conner  orcid.org/0000-0001-9906-3762

Archaeologist, author, data collection and verification, GIS expert, Lehigh University, US

Benjamin Carter  orcid.org/0000-0002-7464-0989

Archaeologist, author, data collection and verification, GIS expert, Muhlenberg College, US

Jeff Blackadar  orcid.org/0000-0002-8160-0942

Object recognition programmer, author, Carleton University, CA

REFERENCES

1. **Carter, B. P., Blackadar, J. H. and Conner, W. L. A.** (2021) “When Computers Dream of Charcoal: Using Deep Learning, Open Tools, and Open Data to Identify Relict Charcoal Hearths in and around State Game Lands in Pennsylvania.” *Advances in Archaeological Practice*. Cambridge University Press, pp. 1–15. DOI: <https://doi.org/10.1017/aap.2021.17>
2. **PAMAP Program, PA Department of Conservation and Natural Resources, Bureau of Topographic and Geologic Survey. PAMAP Program LAS Files (LiDAR Data of Pennsylvania).** n.d. [Online]. Available: <https://www.pasda>.

- psu.edu/uci/FullMetadataDisplay.aspx?file=pamap_lidar_LAS.xml [Accessed 30 April 2020].
3. **PAMAP Program, Bureau of Topographic and Geologic Survey, PA Department of Conservation and Natural Resources. PAMAP Tile Index - North 2006.** 2006. [Online]. Available: <https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=266> [Accessed 30 April 2020].
 4. **PAMAP Program, Bureau of Topographic and Geologic Survey, PA Department of Conservation and Natural Resources. PAMAP Tile Index - South 2006.** 2006. [Online]. Available: <https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=267> [Accessed 30 April 2020].
 5. **Pennsylvania Game Commission., PGC State Game Lands.** 2018. [Online]. Available: <http://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=86> [Accessed 30 April 2020].
 6. **Canonical Group Limited. Ubuntu.** n.d. [Online]. Available: <https://www.microsoft.com/store/productId/9NBLGGH4MSV6> [Accessed 30 April 2020].
 7. **Rapidlasso. LASStools.** n.d. [Online]. Available: <https://rapidlasso.com/lastools/> [Accessed 30 April 2020].

TO CITE THIS ARTICLE:

Conner W, Carter B, Blackadar J 2021 Geospatial and Image Data from the “When Computers Dream of Charcoal: Using Deep Learning, Open Tools and Open Data to Identify Relict Charcoal Hearths in and Around State Game Lands in Pennsylvania” Paper. *Journal of Open Archaeology Data*, 9: 7, pp. 1–6. DOI: <https://doi.org/10.5334/joad.80>

Published: 05 November 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Archaeology Data is a peer-reviewed open access journal published by Ubiquity Press.