



Environmental Analysis Program

Pomona College
Biogeochemistry Laboratory

SOP-L804

GIS and Spatial Data Management

Standard Operating Procedure

SOP Number	SOP-L804
Revision	0.1
Author(s)	Marc Los Huertos
Date	2026-06-26
Organization	EA Program, Pomona College

FOR INTERNAL USE ONLY

This document is the property of the Environmental Analysis Program at Pomona College.
Unauthorized distribution is not permitted.

Table of contents

1 Purpose	2
2 Scope	2
3 Responsibilities	2
4 Definitions	2
5 Safety and Hazards	3
6 Equipment and Materials	3
6.1 Software	3
6.2 Hardware	3
7 Procedure	3
7.1 Coordinate Reference Systems	3
7.2 File Formats	4
7.3 File Naming and Directory Structure	4
7.4 Spatial Data Acquisition	5
7.5 Spatial Data Processing in R	5
7.6 Map Production	6
7.7 Metadata Documentation	7
7.8 GPS Data Import	7
8 Quality Control and Acceptance Criteria	7
9 Waste Disposal	7
10 References	8
11 Training Record	8
12 Appendix A: Recommended R Packages for Spatial Analysis	8

1. Purpose

This SOP establishes procedures for acquiring, organizing, projecting, and archiving geospatial data used in EA Program research. It covers coordinate reference systems, file formats, metadata standards, and reproducible mapping workflows in R, ensuring that spatial data are interoperable across projects and analysts (Lovelace et al., 2019; Pebesma & Bivand, 2023).

2. Scope

- **Applies to:** All vector (point, line, polygon) and raster (satellite imagery, DEMs, interpolated surfaces) data used in EA Program projects
- **Does not apply to:** Drone image acquisition or photogrammetric processing (see [SOP-L805](#)); GPS receiver operation in the field (see relevant field SOPs)
- **Location:** EA Program laboratory (SGM 133) and RStudio Server

3. Responsibilities

Role	Responsibility
Lab PI	Approve spatial data management plan; ensure data-sharing compliance
GIS Lead	Maintain spatial data library; enforce projection and metadata standards
Analyst / Student	Follow naming conventions; document coordinate systems; archive final layers
Data Reviewer	Verify spatial metadata completeness before publication

4. Definitions

- **Coordinate reference system (CRS):** A framework that associates coordinates with locations on Earth; consists of a datum (reference ellipsoid), a projection (for 2D maps), and a coordinate system (e.g., latitude/longitude or meters); data layers must share the same CRS for spatial operations.
- **Geodatabase:** A structured container for storing multiple spatial datasets, attribute tables, and topological rules in a single file; the preferred format for complex GIS projects with many related layers.
- **GeoTIFF:** A TIFF image file with embedded geospatial metadata (CRS, extent, resolution); the standard raster format for spatial data exchange in remote sensing and GIS.
- **Raster:** A spatial data model representing continuous surfaces as a grid of equally sized cells (pixels), each with one or more attribute values; used for elevation, imagery, and interpolated surfaces.
- **Shapefile (.shp):** A common vector spatial data format introduced by Esri; stores geometry (points, lines, polygons) and attribute data in multiple associated files (.shp, .dbf, .shx,

.prj minimum); widely supported but limited by 2 GB file size and 10-character field name constraints.

- **Vector:** A spatial data model representing discrete features as points, lines, or polygons defined by coordinate pairs; used for sample locations, stream networks, and land parcel boundaries.
- **WGS84 (World Geodetic System 1984):** The global geodetic datum used by GPS; EPSG code 4326; the required CRS for all EA Program GPS-collected point data before projection into a local coordinate system.

5. Safety and Hazards

GIS work is a computational activity with no physical hazards. When using GPS receivers or conducting ground-truth fieldwork, follow [SOP-F001 Field Safety](#).

6. Equipment and Materials

6.1. Software

Item	Access	Notes
R (version 4.x)	RStudio Server	Primary analysis platform
sf package	CRAN	Vector data handling
terra package	CRAN	Raster data handling
tmap / ggplot2	CRAN	Map production
QGIS (3.x)	Free download	Visual inspection and digitizing
Google Earth Engine	earthengine.google.com	Satellite imagery processing

6.2. Hardware

Item	Location	Notes
GPS receivers (Garmin)	SGM 133 equipment cabinet	Field coordinate collection
High-resolution monitor	SGM 133	For digitizing and map review

7. Procedure

7.1. Coordinate Reference Systems

All EA Program spatial data must use one of the following coordinate reference systems (CRS). Document the CRS in every file, script, and metadata record.

CRS Name	EPSG Code	Use Case
WGS 84 (geographic)	4326	GPS data, web maps, data exchange
NAD83 / UTM Zone 11N	26911	Local analysis in Southern California
NAD83 / California Zone 5	2229	County and municipal data

Tip

When in doubt, store data in EPSG:4326 (latitude/longitude, WGS 84). Transform to a projected CRS for distance and area calculations.

When loading spatial data in R, always verify and set the CRS explicitly:

```
library(sf)

# Read a shapefile
sites <- st_read("data/spatial/sampling_sites.shp")

# Check CRS
st_crs(sites)

# Transform to UTM Zone 11N for area calculations
sites_utm <- st_transform(sites, crs = 26911)
```

7.2. File Formats

Format	Extension	Use Case	Notes
GeoPackage	.gpkg	Preferred for all vector data	Single file, open standard
Shapefile	.shp + .dbf + .shx + .prj	Legacy compatibility	Keep all component files together
GeoTIFF	.tif	Raster data	Include CRS in header
CSV with XY	.csv	Simple point data	Requires separate CRS documentation
GeoJSON	.geojson	Web applications	Always WGS 84

GeoPackage (.gpkg) is the preferred format for new projects because it stores geometry, attributes, and CRS in a single file without the limitations of shapefiles (10-character field names, 2 GB size limit) (Pebesma & Bivand, 2023).

7.3. File Naming and Directory Structure

Follow the conventions in [SOP-L803](#) with the addition of a `data/spatial/` subdirectory:

```

project_name/
  data/
    raw/
    processed/
    spatial/
      raw/          # Downloaded or collected spatial data
      processed/    # Clipped, projected, joined layers
      metadata/     # FGDC or ISO metadata files
    metadata/
  analysis/
    scripts/
    maps/           # Exported map figures
  README.md

```

7.4. Spatial Data Acquisition

Common data sources for EA Program research:

Source	URL	Data Types
USGS National Map	nationalmap.gov	DEMs, hydrography, boundaries
USGS EarthExplorer	earthexplorer.usgs.gov	Landsat, NAIP imagery
CA State Geoportal	gis.data.ca.gov	State and county layers
NRCS Web Soil Survey	websoilsurvey.sc.egov.usda.gov	SSURGO soil data
OpenStreetMap	openstreetmap.org	Roads, buildings, land use
Google Earth Engine	earthengine.google.com	Multi-temporal satellite data

When downloading external data, record the source URL, download date, and any subsetting parameters in the metadata file.

7.5. Spatial Data Processing in R

A standard workflow for processing spatial data:

```

library(sf)
library(terra)
library(tidyverse)

# --- 1. Load vector data ---
sites <- st_read("data/spatial/raw/sampling_sites.gpkg")
watershed <- st_read("data/spatial/raw/claremont_watershed.gpkg")

```

```
# --- 2. Verify and harmonize CRS ---
st_crs(sites) == st_crs(watershed)
sites <- st_transform(sites, crs = st_crs(watershed))

# --- 3. Spatial operations ---
# Clip sites to watershed boundary
sites_in_watershed <- st_intersection(sites, watershed)

# Buffer sites by 100 m
site_buffers <- st_buffer(sites_in_watershed, dist = 100)

# --- 4. Load and extract raster data ---
dem <- rast("data/spatial/raw/dem_10m.tif")
sites_in_watershed$elevation_m <- extract(dem, sites_in_watershed)$value

# --- 5. Save processed outputs ---
st_write(sites_in_watershed,
         "data/spatial/processed/sites_with_elevation.gpkg",
         delete_dsn = TRUE)
```

7.6. Map Production

All maps for reports and publications must include:

1. Title and date
2. Scale bar (use `tm_map::tm_scale_bar()` or `ggspatial::annotation_scale()`)
3. North arrow
4. Legend with clear labels and units
5. Data source attribution
6. CRS/projection noted in the caption or margin

```
library(tmap)

tm_shape(watershed) +
  tm_polygons(col = "lightblue", alpha = 0.3) +
  tm_shape(sites_in_watershed) +
  tm_dots(col = "elevation_m", size = 0.5,
         palette = "viridis",
         title = "Elevation (m)") +
  tm_scale_bar(position = c("left", "bottom")) +
  tm_compass(position = c("right", "top")) +
  tm_layout(
    title = "Sampling Sites: Claremont Watershed",
    title.size = 1.2,
    legend.outside = TRUE
  )
```

7.7. Metadata Documentation

Each spatial dataset must have a metadata record. At minimum, include:

Field	Description
Title	Descriptive name of the layer
Abstract	Brief description of content and purpose
CRS	EPSG code and full name
Extent	Bounding box coordinates
Source	Origin of data (URL, agency, field GPS)
Date created/acquired	When the data were generated or downloaded
Processing steps	Summary of transformations applied
Contact	Analyst name and email

For formal publications, use FGDC or ISO 19115 metadata standards. For internal projects, the table above saved as a text file alongside the data is sufficient.

7.8. GPS Data Import

When importing GPS waypoints or tracks:

1. Download data from the GPS unit to `data/spatial/raw/` in GPX format
2. Import into R using `sf::st_read("file.gpx", layer = "waypoints")`
3. Verify coordinates visually by plotting over a basemap
4. Assign the correct CRS (typically EPSG:4326 for consumer GPS units)
5. Document the GPS model, estimated accuracy, and collection conditions

8. Quality Control and Acceptance Criteria

- Verify CRS for every imported dataset before analysis
- Check that point locations fall within expected geographic bounds (e.g., all sites should be within the study area polygon)
- Validate topology for polygon layers (no gaps, overlaps, or self-intersections) using `sf::st_is_valid()`
- Compare GPS coordinates with known reference points when available
- Document the spatial accuracy and precision of each data source

9. Waste Disposal

Not applicable (computational procedure).

10. References

Lovelace, R., Nowosad, J., & Muenchow, J. (2019). *Geocomputation with R*. Chapman; Hall/CRC. <https://doi.org/10.1201/9780203730058>

Pebesma, E., & Bivand, R. (2023). *Spatial data science: With applications in R*. Chapman; Hall/CRC. <https://doi.org/10.1201/9780429459016>

11. Training Record

Training is documented in the EA Program training log. Before working independently with spatial data, personnel must:

1. Read this SOP in full
2. Complete the RStudio and GitHub orientation ([SOP-L801](#))
3. Demonstrate proficiency with **sf** and basic spatial operations in R under supervision
4. Produce a correctly formatted map with all required elements
5. Be signed off by the PI or GIS lead

12. Appendix A: Recommended R Packages for Spatial Analysis

Package	Purpose
sf	Vector data (points, lines, polygons)
terra	Raster data analysis
stars	Spatiotemporal raster/vector data cubes
tmap	Thematic maps (static and interactive)
leaflet	Interactive web maps
ggspatial	Spatial layers in ggplot2
mapview	Quick interactive map previews
exactextractr	Fast raster extraction by polygons
geodata	Download climate, soil, and elevation data
tidyterra	Tidy methods for terra objects