Geospatial Project Data

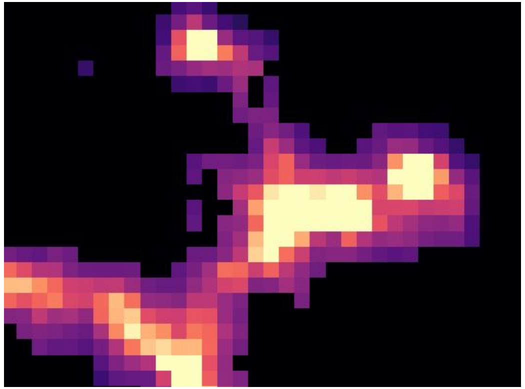
Guidance and examples for building location (geospatial) information from administrative and project data to be used in a geospatial impact evaluation.

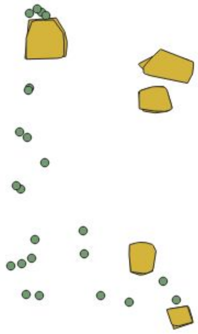
Topics include:

- Geospatial data basics
- Units of analysis
- Preparing project geospatial data
- Project metadata

AIDDATA
A Research Lab at William & Mary

Geospatial Data Basics





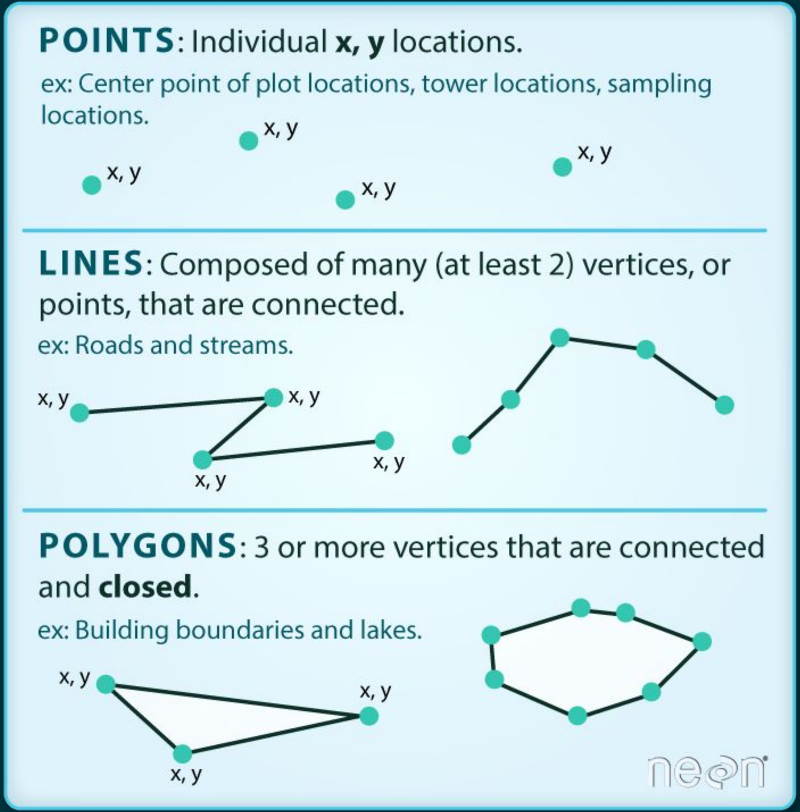
What is Geospatial Data?

- Any observation connected to specific location
  - E.g., at latitude Y and longitude X the temperature is 30 degrees
- Because all geospatial data have locations, you can always merge different datasets together (though it can be difficult).
- This locational merging process enables various types of analysis (e.g. targeting, monitoring, and evaluation) that would otherwise not be feasible

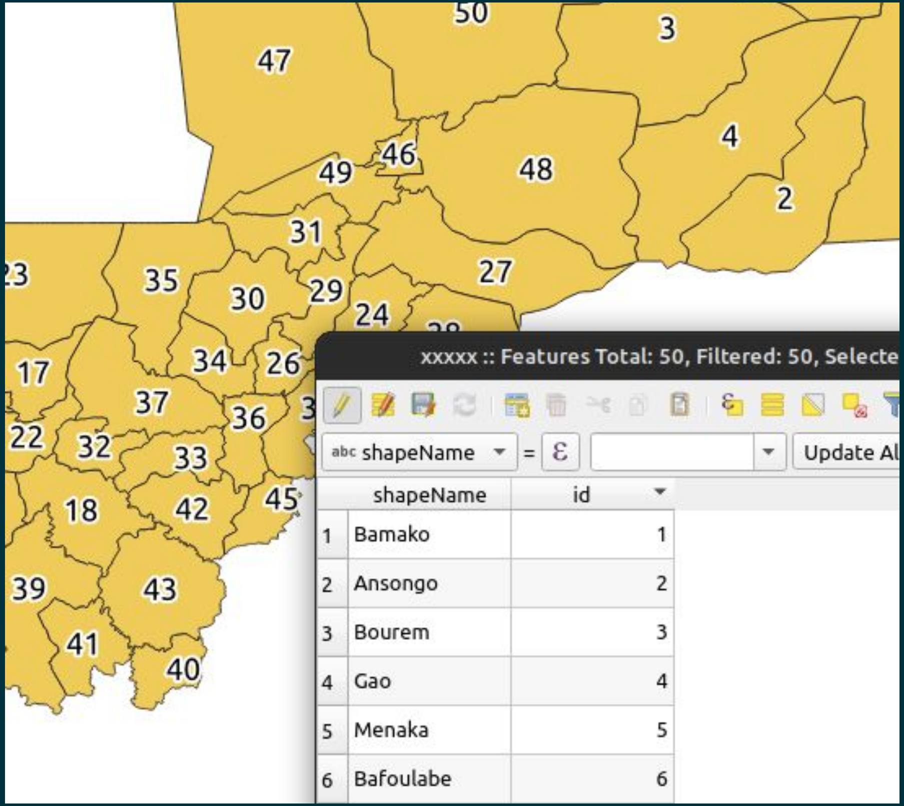
Examples of Geospatial Data

- GPS coordinates of household where a survey was conducted
- A description of villages that received improved crop seeds
- Series of coordinates forming a line which represents canal that was built
- Coordinates at 4 corners of a field where harvest data was collected
- A copy of a paper map showing fields growing specific crop types
- Government administrative boundaries of a district in which a agricultural education program was implemented
- Start and end points of a road project
- A construction firm's drawing of site plans for an irrigation system
- Drone imagery of the site of a newly developed small earth dams

Vector Data

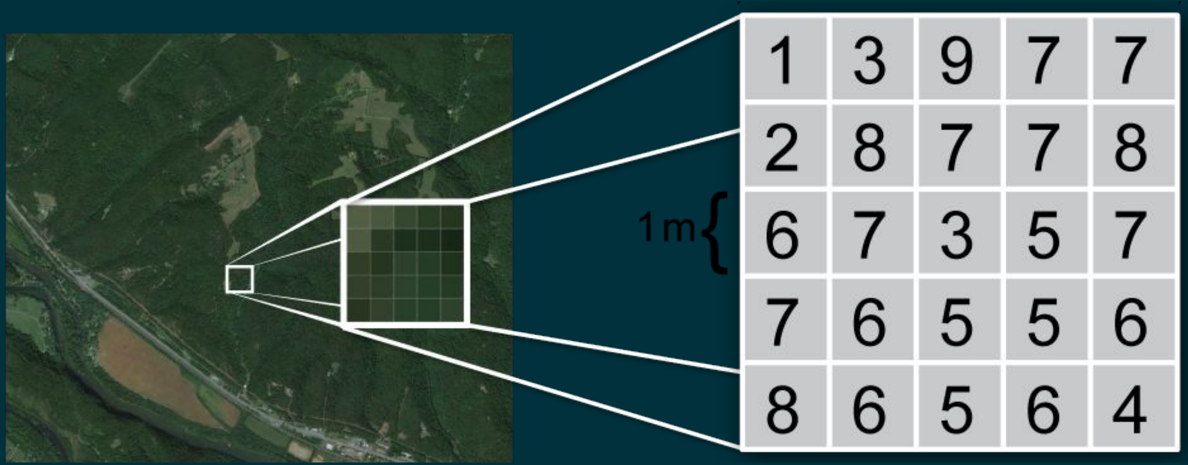


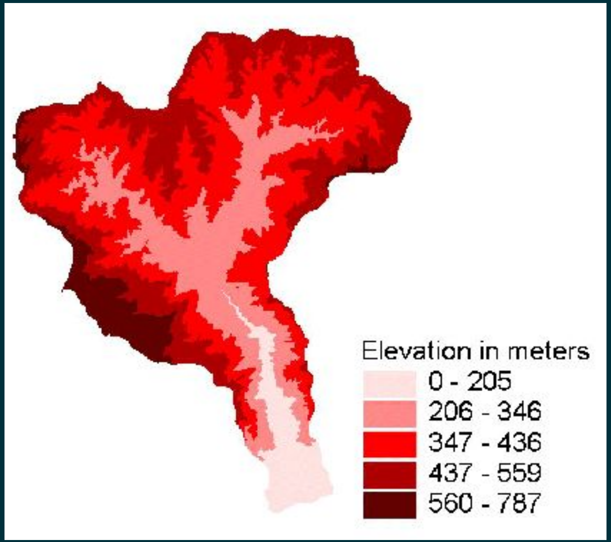
POINTS: Individual x, y locations.
ex: Center point of plot locations, tower locations, sampling locations.
x, y
x, y
x, y
x, y
LINES: Composed of many (at least 2) vertices, or points, that are connected.
ex: Roads and streams.
x, y
x, y
x, y
x, y
POLYGONS: 3 or more vertices that are connected and closed.
ex: Building boundaries and lakes.
x, y
x, y
x, y
neon

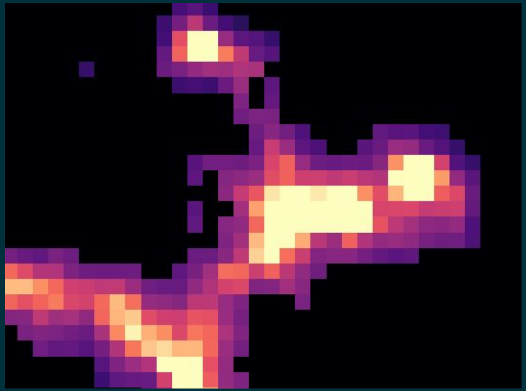


50
3
47
46
49
48
4
2
31
35
30
29
27
24
17
34
26
37
36
22
32
33
18
42
45
39
43
41
40
xxxxx :: Features Total: 50, Filtered: 50, Selecte
abc shapeName
=
ε
Update Al
shapeName
id
1 Bamako 1
2 Ansongo 2
3 Bourem 3
4 Gao 4
5 Menaka 5
6 Bafoulabe 6

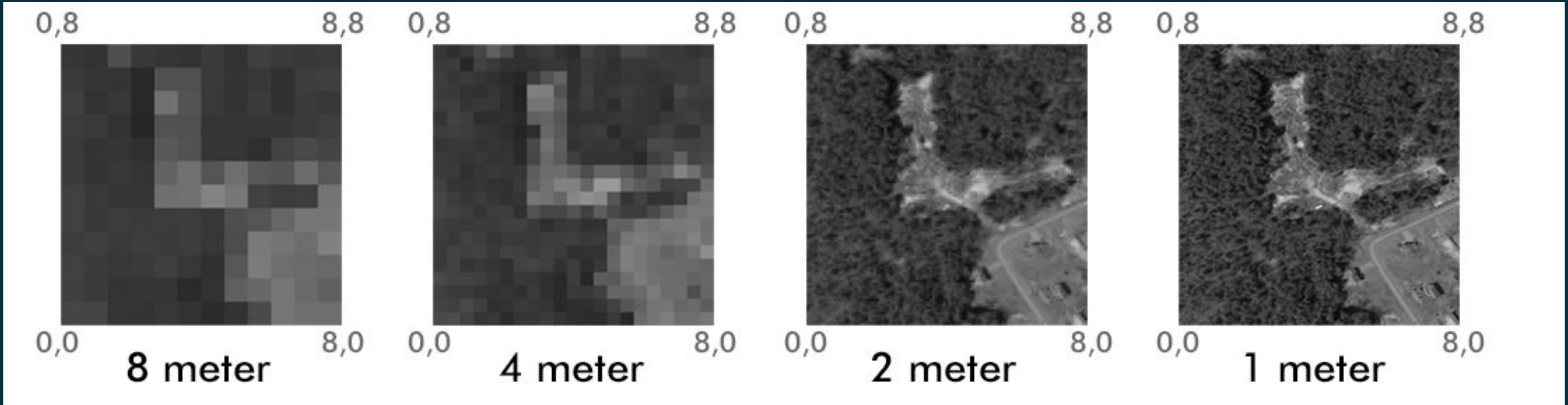
Raster Data


1 3 9 7 7
2 8 7 7 8
1m
6 7 3 5 7
7 6 5 5 6
8 6 5 6 4

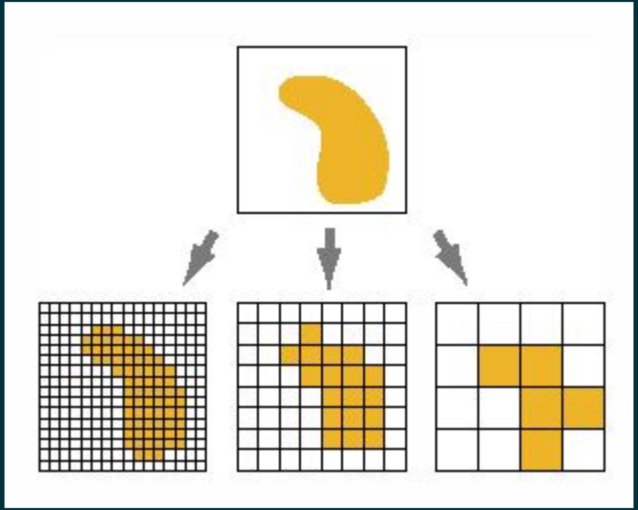

Elevation in meters
0 - 205
206 - 346
347 - 436
437 - 559
560 - 787

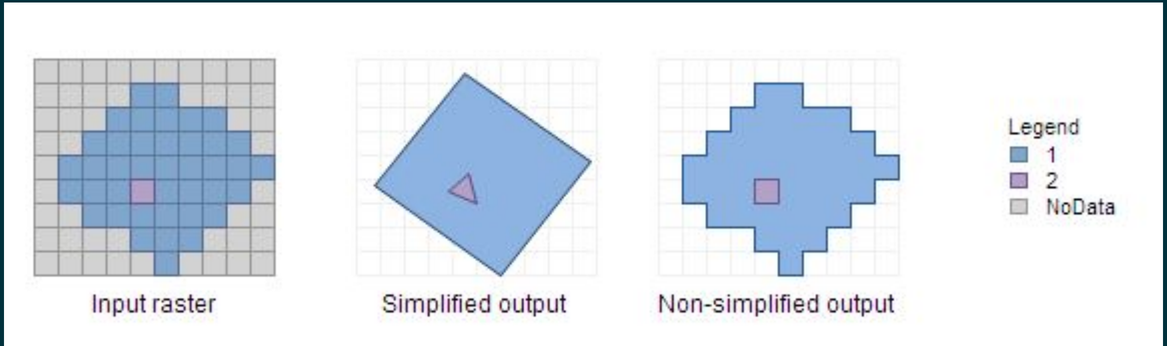


Raster Data Resolution

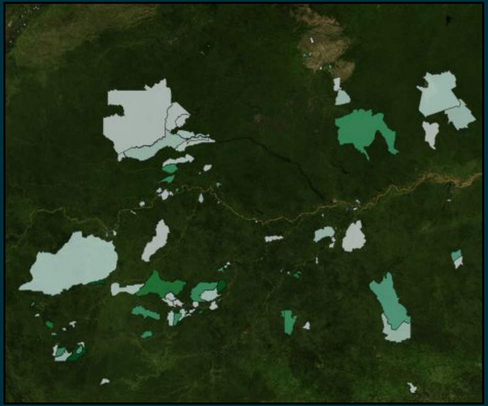


Vector / Raster Transformation

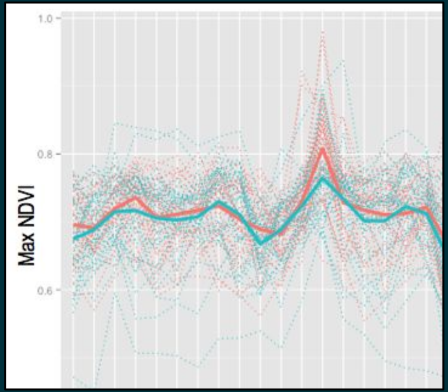




Vector / Raster Integration



Unit	NDVI 1990	NDVI 2005
A	0.15	0.10
B	0.13	0.12
C	0.16	0.17
D	0.20	0.19



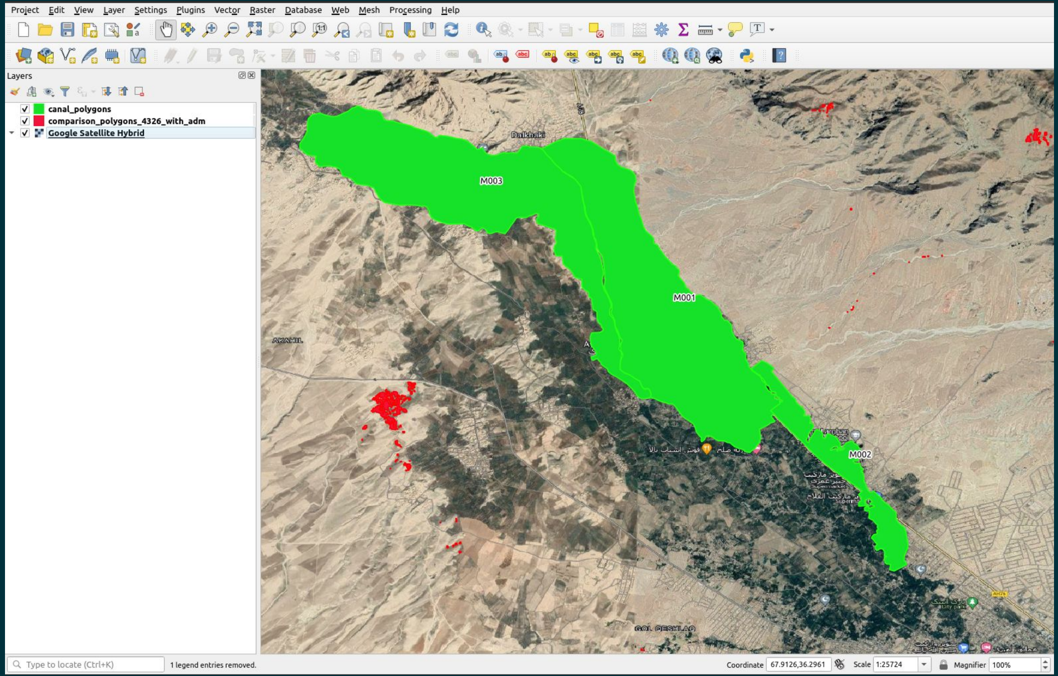
Measurement Data (Raster)

Geospatial Units (Vector)

Tabular

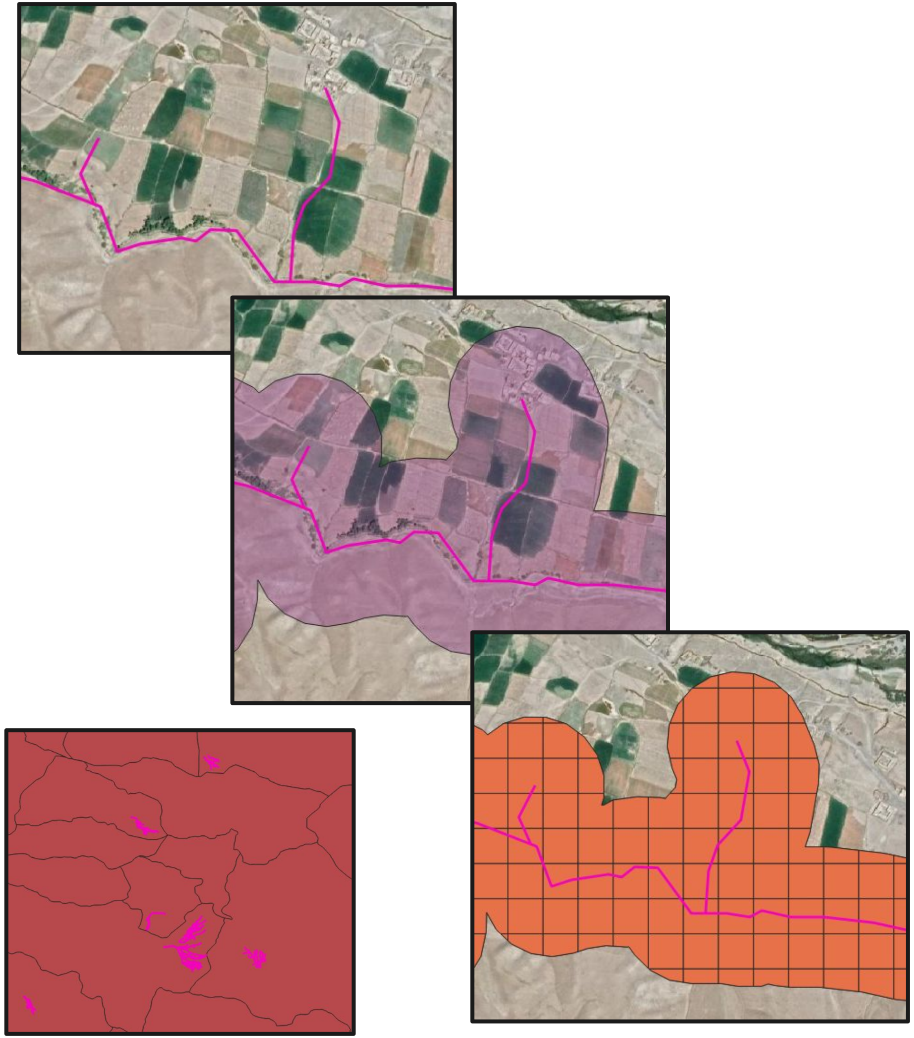
Analysis

Geospatial Data Tools




```python
from pathlib import Path
import pandas as pd
import geopandas as gpd

base_path = Path('/home/user/Desktop')
geojson_path = base_path / 'SEZS_Ethiopia.geojson'

raw_gdf = gpd.read_file(geojson_path, driver='GeoJSON')

points_gdf = raw_gdf.loc[raw_gdf.type == 'Point']
points_gdf = points_gdf.set_crs(epsg=4326)

polygons_gdf = raw_gdf.loc[raw_gdf.type == 'Polygon']
polygons_gdf = polygons_gdf.set_crs(epsg=4326)

polygons_gdf = polygons_gdf[['geometry']].copy()

gdf = gpd.sjoin(polygons_gdf, points_gdf, op='contains')
gdf.drop(columns=['index_right'], inplace=True)

gdf = gdf.to_crs(epsg=32636)
gdf['buffer'] = 0
# km
buffers = [5, 10, 15, 20, 25, 35]

bgdf_list = [gdf]
for b in buffers:
    bgdf = gdf.copy()
    bgdf['geometry'] = bgdf['geometry'].buffer(b*1e3)
    bgdf['buffer'] = b
```

Geospatial Units of Analysis



Defining a Unit of Analysis

- In many non-spatial studies, the unit of analysis for an intervention is well defined:
  - ✓ Giving a pill to individuals for a disease - the individual person is the unit of observation.
  - ✓ Applying fertilizer to a number of fields - the individual field is the unit of observation.
- In geospatial analyses the unit of analysis usually needs to be theoretically defended.
  - ✓ If you build a canal, over how many meters or kilometers from the canal do you expect that you should be able to observe an impact?
  - ✓ If you conduct a training on sustainable farming techniques in a city center, what is the size of the 'catchment area' where you expect the benefits to accrue?
- Before any statistical analysis can be performed, the unit of analysis must be defined.

Defining a Unit of Analysis

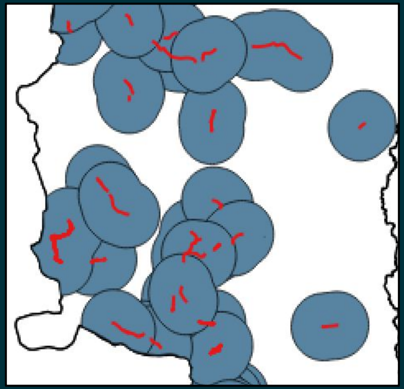
Examples of project data and corresponding units of analysis:

Type of Project	**Example Project**	**Unit of Analysis**	**Alternative (less precise spatial analysis)**
Field Based	Field level irrigation improvements	Polygon outlines of fields	Point location of field
Household Based	Cash transfer provided to household to improve resilience)	GPS of household survey respondents	Village in which household is located
Area Based	Distributing improved seed or providing irrigation for entire village)	GPS of specific village households / fields benefitting from intervention or spatial feature defining precise extent of village	Buffer around approximate GPS centroid of village
Linear Features	Network of roads which were improved	Grid cells within buffer around improved roads	villages/districts in which roads were improved
Broad Scope Program	Agricultural subsidies allocated based on administrative unit	Administrative boundaries for district (or other units) at which interventions were allocated	-

Units of Analysis: Buffers

- Assign a specific distance (i.e., 1km, 5km, 10km) from the known intervention point, and test for impacts within that area.

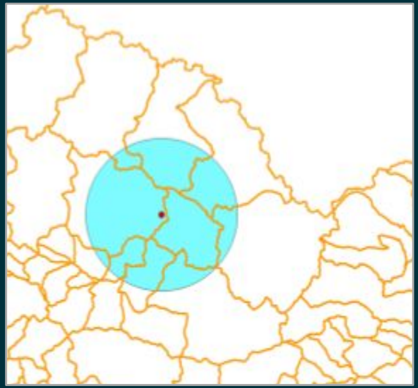
- Can be applied to point-based interventions (i.e., a training) or line-based interventions (i.e., a canal or road), or polygon (i.e., protected area).



Road projects (lines) buffered to 5km and clipped to administrative zones

Canal network (lines) buffered

Irrigated fields (polygons) buffered to examine nearby impacts



Village coordinates (point) buffer, unclipped

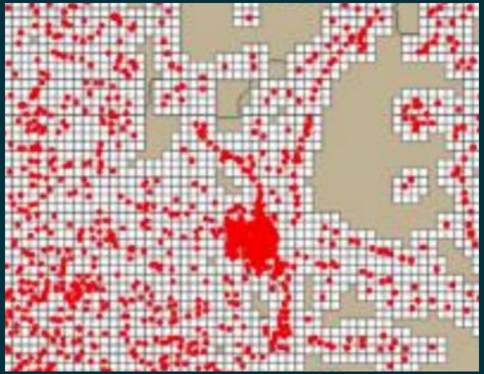
Units of Analysis: Grid Cells

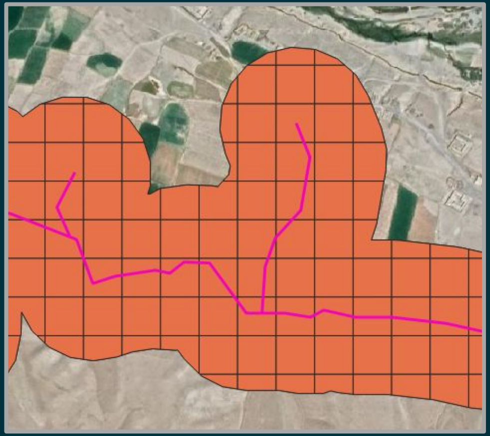
Advantages

- ✓ Many units of observation
- ✓ Can identify precise changes over study region
- ✓ Can mitigate challenge of not knowing the exact buffer radius you need to select

Disadvantages

- x Both study region and grid cell size must be defended theoretically.
- x Underlying data must support chosen resolutions.
- x Statistical challenges are larger (independence of units can be scale dependent).





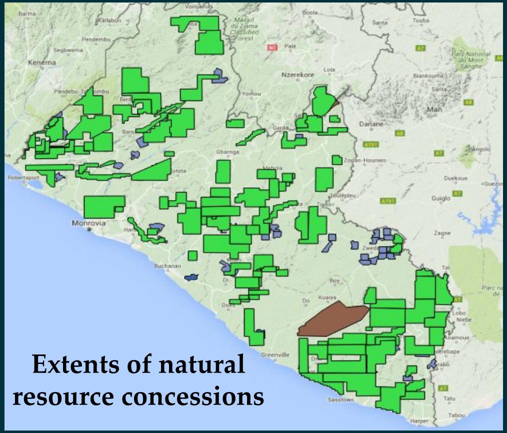
Units of Analysis: Precise Features

Advantages

- ✓ Works very well when the exact area an intervention would have impacted is known
- ✓ No or little spatial measurement imprecision

Disadvantages

- x Many programs do not have a precisely defined spatial scope
- x The costs to generate a precisely defined spatial scope can be significant



Extents of natural resource concessions

Catchment areas served by irrigation projects

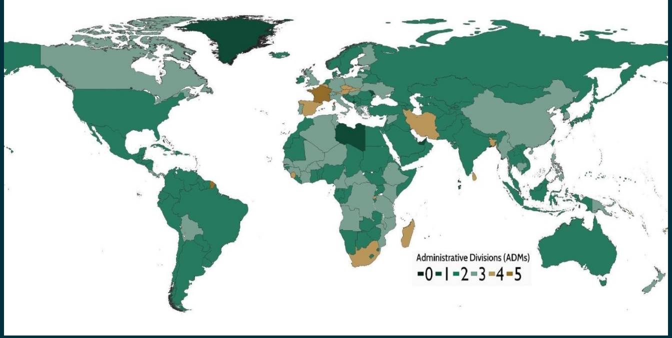
Units of Analysis: Administrative Boundaries

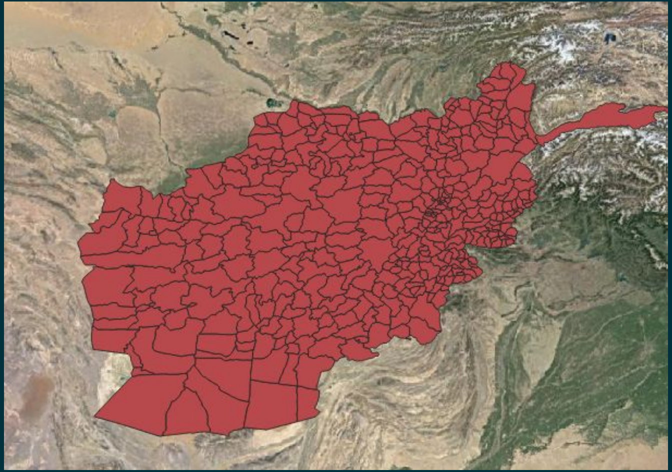
Advantages

- ✓ Helpful when an intervention is anticipated to affect an entire decision-making unit.
- ✓ Frequently the same units used for census activities.

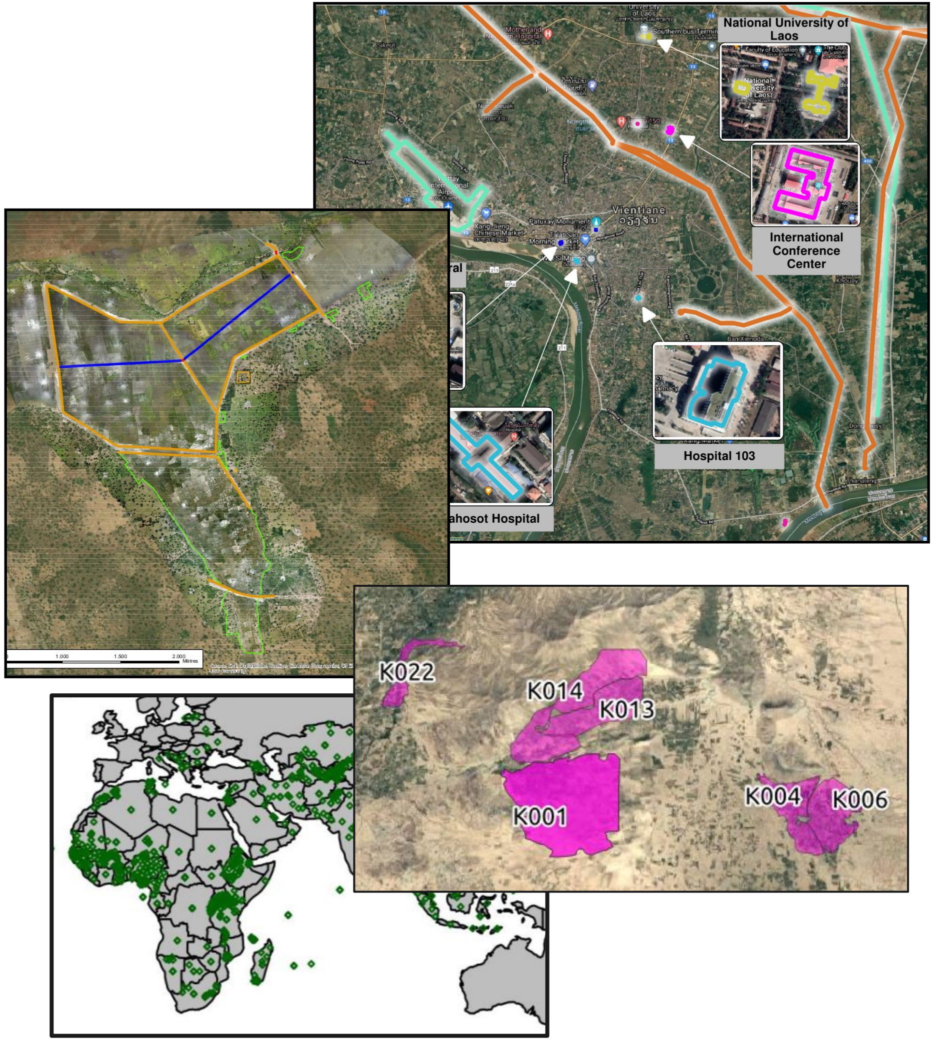
Disadvantages

- x Can change in unmeasured ways over time.
- x Can be of variable size and have variable underlying measurement qualities.





Preparing Project Geospatial Data

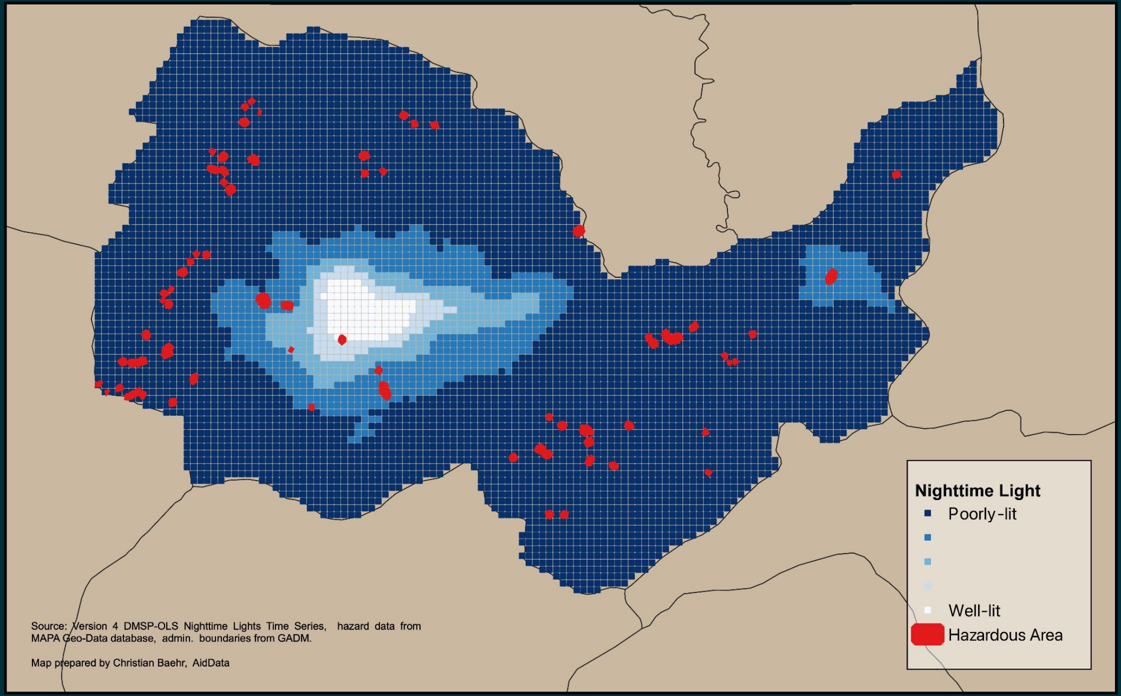


Project Data: Existing Geospatial Data

Polygons to measure spatio-temporal rollout of indigenous land demarcation program in Brazil (1997-2008)

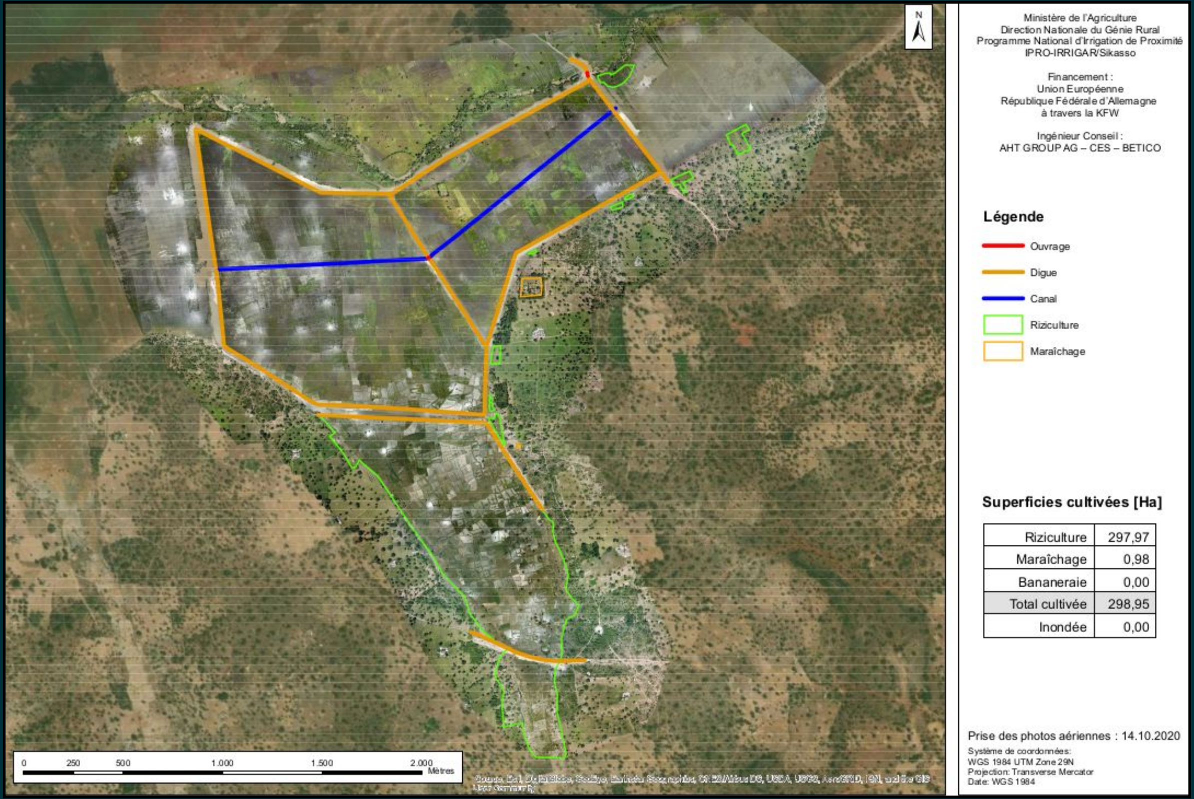
Project Data: Existing Geospatial Data

Impacts of Mine Clearance in Afghanistan



Project Data: Georeferencing Maps

Irrigation Projects in Mali



Project Data: Geocoding Project Records

Retrospective coding to cover records where explicit geospatial data was not collected

Source Material:

- Project documents
- Local government Aid Information Management Systems
- Donor information systems
- Other official, media, and third-party sources

Includes “geoparsing” and “geocoding”:

- **Geoparsing** - the initial process of identifying place names or other location information in documents (project reports, PDF files, etc.)
- **Geocoding** - translating location information into standardized, machine readable geospatial information (coordinates, administrative boundaries, etc.)

Project Data: Geocoding Project Records

Multiple possible approaches to geocoding, with varying levels of precision and amount of time required

Point Based: single point with codes to describe precision level of the location

- Ex: Coordinates in center of district and code which represents “district level”

Precise: uses lines or polygons to represent the true (or best approximation of) actual spatial features defining where investments went

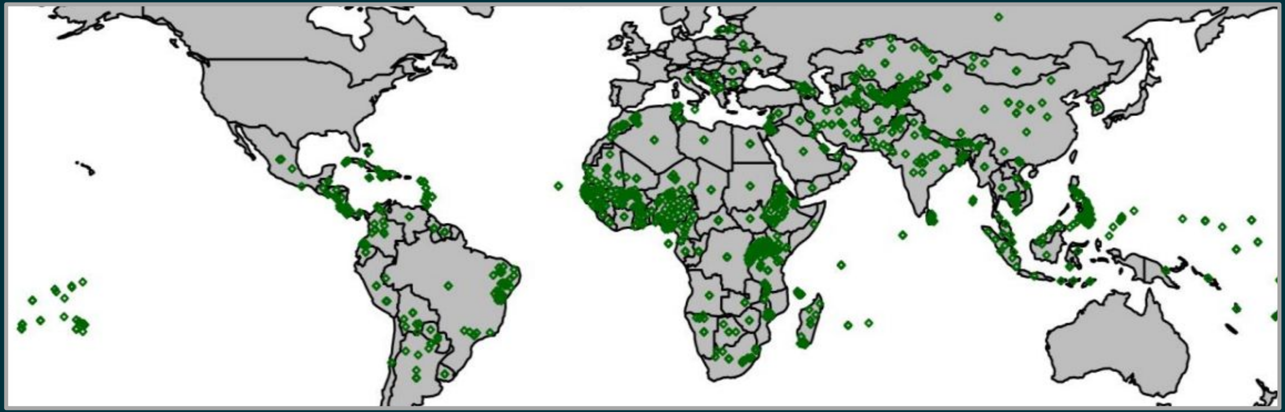
- Lines defining infrastructure development such as roads, power lines, irrigation networks
- Polygons for land tenure, protected areas, etc.

Project Data: Geocoding Project Records

Point based may be preferable at large scales where less precise location information is likely to be available in project records.

point base coding schemes have inherently limited spatial detail (e.g., project in town recorded as coordinates of town center and code which identified it as a populated place)

Example: GEF historic portfolio of land degradation projects (202 projects, 1704 locations around world)



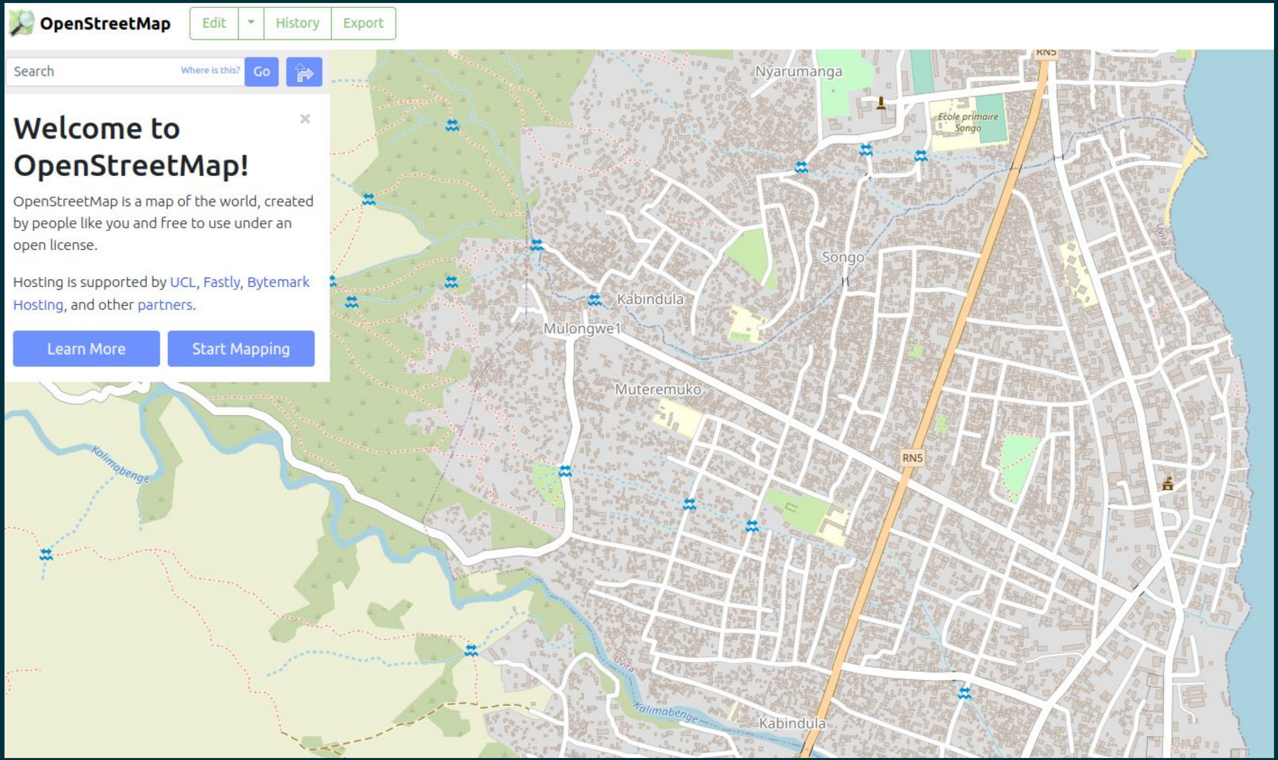
Project Data: Geocoding Project Records

Precise geocoding requires more time and underlying information, but enabled far more types of spatial analysis and understanding of local impacts

Example geocoding roads: point based vs precise. Left: recording centroids of districts road runs through vs true road path. Right: recording start and end point of road vs true road path.

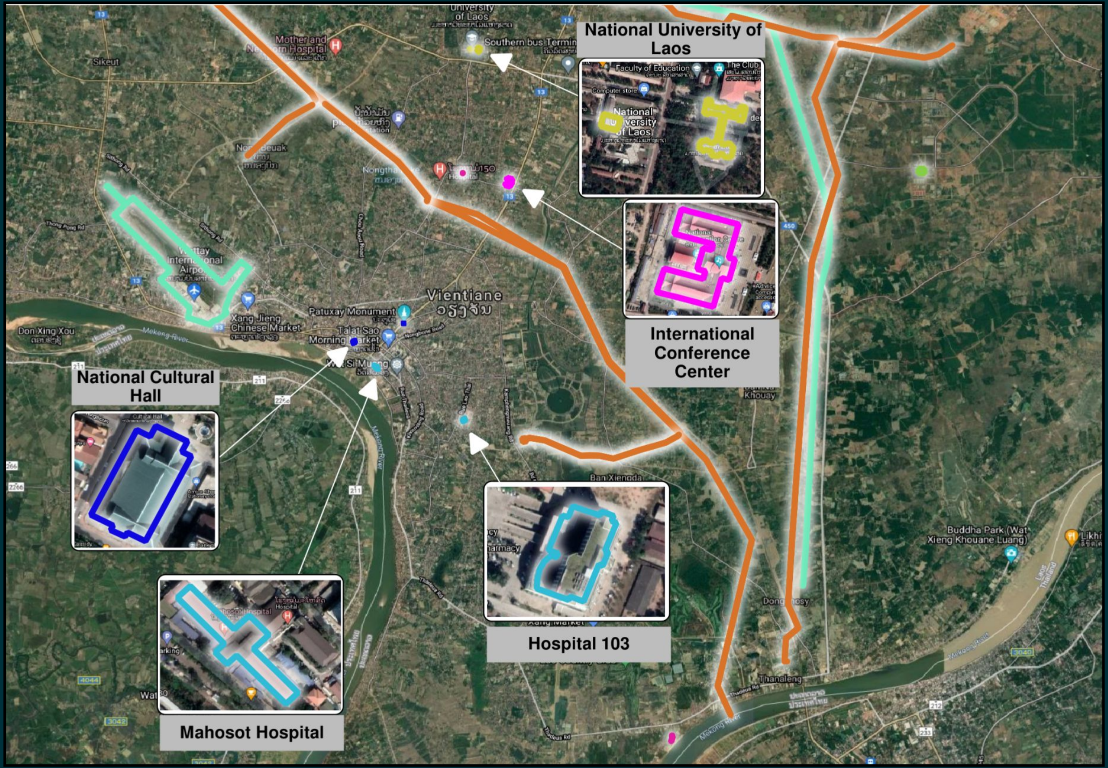
Project Data: Geocoding Project Records

Alternative to generating precise features is using data from OpenStreetMap (OSM)



Project Data: Geocoding Project Records

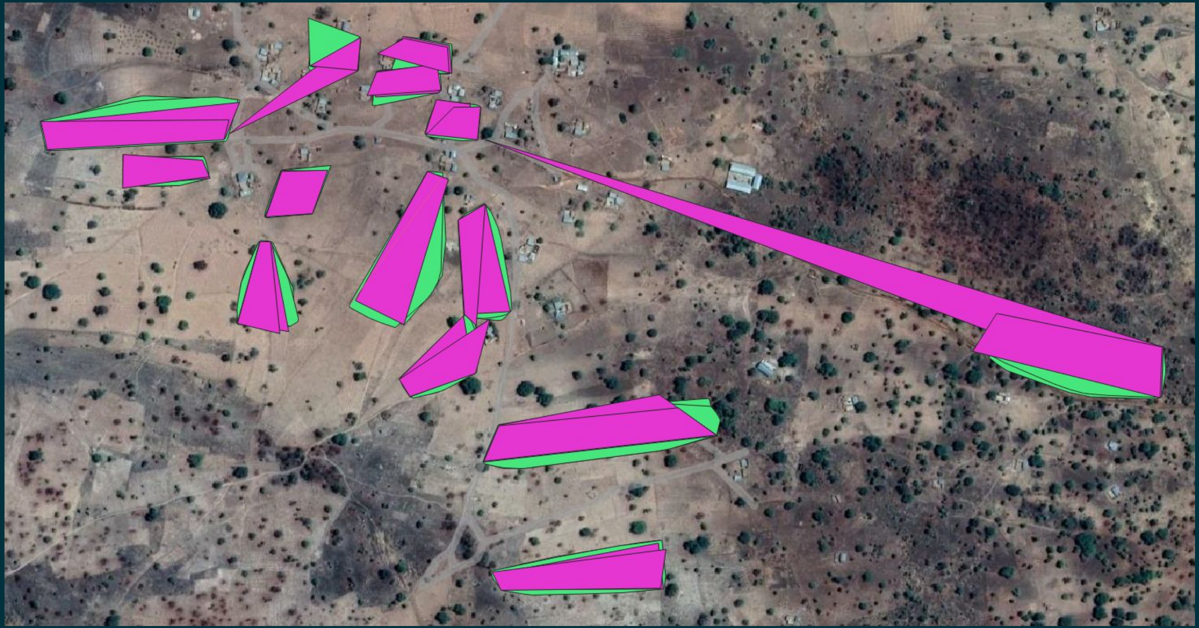
Currently being used to geocode AidData's Global Chinese Development Finance Dataset



Features from OSM representing Chinese financed projects

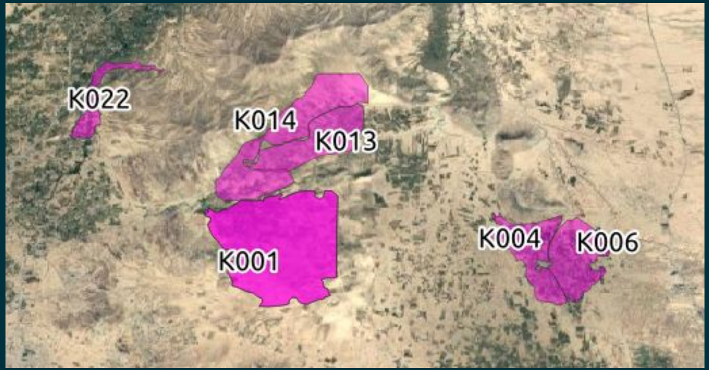
Project Data: Collecting New Data

Generating new geospatial data in the field using GPS to record borders, etc.



Example of farm plot data collected using two methods in Northern Ghana

Project Data: Unique Identifiers



All geospatial must have some unique ID for each feature that can be used to reference the feature during analysis and when joining with other data

Project Metadata

Project Metadata: Overview

Includes any additional information describing key characteristics of the project that can be tied to the geospatial unit of analysis. Examples:

- construction or project start/end dates, cost of the project / site, status of the project at a location,

The type of metadata collected depends on the particular project and evaluation needs, but in general you should address the three following steps.

1. Identify all relevant metadata which is available from project documentation and needs to be extracted and/or formatted.
2. Determine a methodology for standardizing and extracting the metadata consistently across all project records.
3. Include a unique identifier associated with the project so that metadata can be linked to geospatial data.

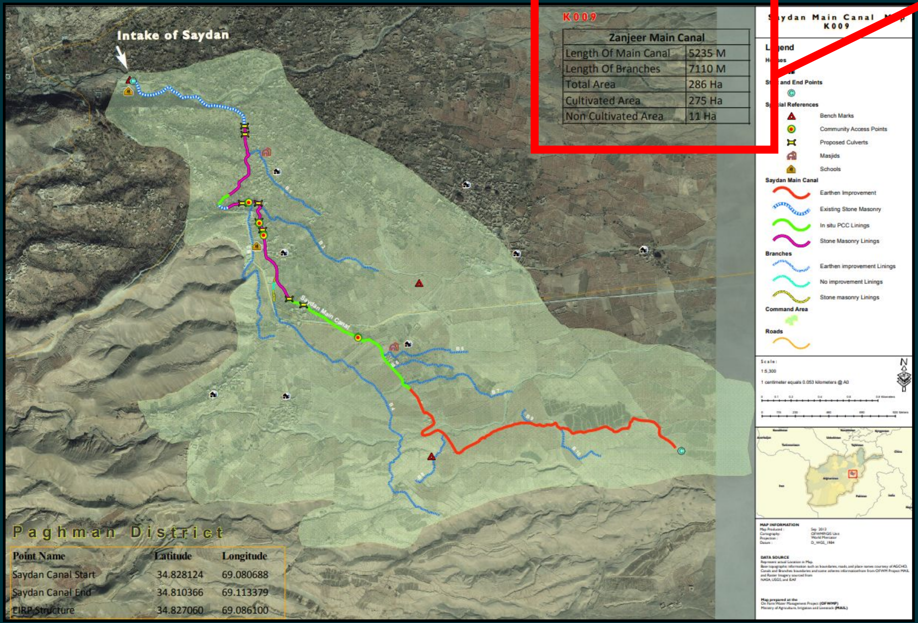
Project Metadata: Overview

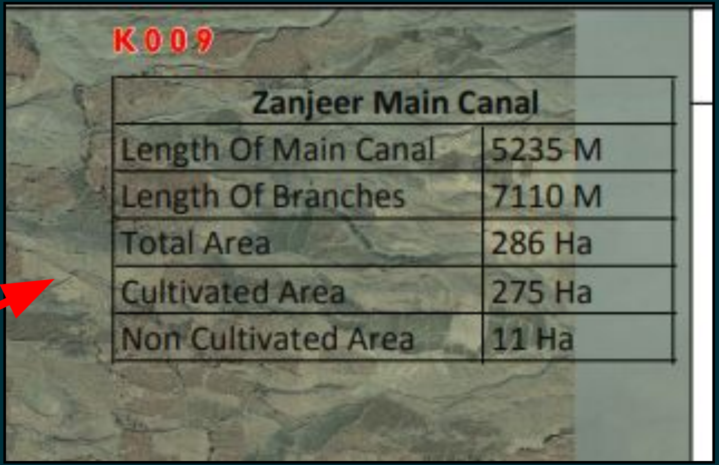
Joining existing data

fid	TYP	SPEZ2	NOM	NOM2	CDC	SECTEUR	ANNEE	HA2	METHODE
142	MARE	Consolidatio...	Gambatane	Gambatane	2354	DIRE	2010	300	Google Earth...
141	MARE	Secteur Mares	Sarfountan	Sarfountan	2403	DIRE	2011	30	Google Earth...
140	MARE	Consolidatio...	Doukou 4	Doukou 4	2353	DIRE	2010	150	Google Earth...
139	MARE	Secteur Mares	Sora	Sora	2347	DIRE	2009	150	Google Earth...
138	MARE	Secteur Mares	Tinam	Tinam	2401	DIRE	2010	150	Google Earth...
137	MARE	Secteur Mares	Bankore	Bankore	2343	DIRE	2008	70	Google Earth...
24	MARE	Secteur Mares	Fawla	Fawla	2609	GOUNDAM-K...	2001	200	K1
23	MARE	Secteur Mares	Baifandou I	Baifandou	2607	GOUNDAM-K...	2001	60	K1
22	MARE	Secteur Mares	Bora	Bora	2303	DIRE	2000	80	Mericol/K1
21	MARE	Secteur Mares	Fonkoura	Foukoura - K...	2305	DIRE	2000	250	Mericol
20	MARE	Secteur Mares	Maharafasseré	Maharafassere	2312	DIRE	2001	45	Mericol/K1/K2
19	MARE	Secteur Mares	Bondeye	Bondeye	2306	DIRE	2001	55	Mericol
18	MARE	Secteur Mares	Mayoro	Mayoro	2619	GOUNDAM-K...	2004	60	Mericol/V1.1
17	MARE	Secteur Mares	Fara-fara	Fara-fara	2608	GOUNDAM-K...	2001	40	Mericol/V1.1
32	MARE	Secteur Mares	Akar Kara	Akar Kara	2613	GOUNDAM-K...	2002	50	K1
31	MARE	Secteur Mares	Bellah Kaka	Bellah Kaka	2612	GOUNDAM-K...	2002	60	K1

Project Metadata: Overview

Extracting data from records





Questions?

Module Contact Points:

- Seth Goodman (sgoodman@aiddata.org)
- Rachel Sayers (rsayers@aiddata.org)
- Ariel BenYishay (abenyishay@aiddata.org)

AIDDATA
A Research Lab at William & Mary