

GeoPandas ITS V2.0 Pipeline

Updates and Logic Decisions

This document outlines the enhancements and decisions in the ITS V2.0 GeoPandas-based implementation.

1. Revised Task Force Region Boundaries - [here](#).
 - 1.1. Processed using revised Task Force Region Boundaries
2. GeoPandas Implementation of the ITS V2.0 Pipeline
 - 2.1. Key Modules Implemented
 - Data Enrichment Module - The enrichment process is tailored for each agency. It transforms agency fuel treatment datasets into the Task Force standardized schema and enriches each record with vegetation type, ownership, county, WUI designation, Task Force region, and treatment year. Agencies included - USFS, CalTrans, CNRA, BLM, NFPORS, NPS, Industry Spatial, Industry Non-Spatial, PFIRS, IFPRS.
 - Data Append Module - Unified process for consolidating multi-agency datasets.
 - Activities Report Generation.
 - Footprint Report Generation.
 - 2.2. Performance & Efficiency Gains
 - Data Processing Time: Reduced from 7 days → **8 hours**.
 - Data Processing + Review (Human in the loop) Cycle: Reduced from 45 days → **2 weeks** (Assuming 10 days for data review by data steward or data provider agency point of contact).
 - Efficiency Gains in ITS GeoPandas Pipeline enables rapid updates and timely decision-making. ITS GeoPandas cut data processing from 7 days to 8 hours (95% faster) and reduced review cycles from 45 to 14 days (69% faster).
 - 2.3. Debugging & Explainability: Improved through removal of black-box functions, enabling reproducibility of both results and errors.

Metric	ArcPy Implementation	GeoPandas Implementation	Speed Improvement	Efficiency Gain (%)	Impact
Data Processing Time	7 days (1 week)	8 hours (1/3 day)	~21× faster	≈ 95% faster	Rapid turnaround for data updates and releases, stakeholder requests.
Processing + Review Cycle	6 weeks (45 days)	2 weeks (14 days)	~3× faster	≈ 69% faster	

3. Main Policy (Treatment/ Activities COUNT to MAS) Decisions

BLM (DOI)

- Keep only polygons that fall inside California and whose **Year** is between **start_year** and **end_year**.
 - Set **ACTIVITY_QUANTITY** to **BLM_ACRES**. If **BLM_ACRES** is 0 or missing, use **GIS_ACRES** instead. Set **ACTIVITY_UOM** = "AC" (acres).
 - Mark every record as **ACTIVITY_STATUS** = "COMPLETE". Set **Year** from **ACTIVITY_END.year**.
 - Set **Crosswalk** using multiple passes based on treatment type/subtype, keywords in the name, and comments.
 - Set **COUNTS_TO_MAS** = "NO" for any record where **ACTIVITY_END** ≥ **2024-10-01** (these do not count toward MAS after that date).
-

IFPRS (DOI)

- Keep records where **State** == "California" and **Completion** ≥ **2023-10-01**, then filter so **Year** is between **start_year** and **end_year**.
 - Set **Acres** from a calculated value if it's non-zero; otherwise compute it from **polygon(Shape_Area)**. Units (UOM) = "AC" (acres).
 - Convert **Initiation** and **Completion** to dates with function **safe_date_convert** (handles nulls and bad values). Use these for **ACTIVITY_START** and **ACTIVITY_END**.
 - Map the original IFPRS status codes into standard status: **ACTIVE** / **PLANNED** / **COMPLETE** / **CANCELLED**.
 - Determine **Crosswalk** (activity category) using rules on Type, SubType, Name, and Notes (e.g., "broadcast", "pile", "thinning").
 - **TRMT_GEOM** = "POLYGON". Build **TRMTID_USER** from project, county, primary_ownership_group, WUI, and primary_objective.
-

NFPORS – Polygons (BIA/FWS)

- Keep polygon records for agency **BIA** and **FWS**, keep only **st_abbr** == "CA", and keep records where **act_comp_dt** ≥ **1995-01-01** or the completion date is null.
- Set **ACTIVITY_QUANTITY** = **gis_acres**, **ACTIVITY_UOM** = "AC". Overwrite the quantity with **TREATMENT_AREA** to fix over-counting in multipart polygons.

- **ACTIVITY_END** comes from **act_comp_dt**; if that's missing, use **modifiedon**. **ACTIVITY_STATUS** comes from **trt_statnm**.
- Set **PROJECTID_USER** = "NFPORS" + **trt_id**. Set **Crosswalk** from **type_name**. Set **AGENCY** = "DOI".
- Set **COUNTS_TO_MAS** = "NO" when **ACTIVITY_END** ≥ 2023-10-01.

NFPORS – Points (BIA/FWS)

- Select BIA + FWS points and keep only records where **statename** = "California".
- Set **ACTIVITY_QUANTITY** to **totalaccomplishment** if it's non-zero; otherwise use **plannedaccomplishment**. Only set **TREATMENT_AREA** when the units are acres (**UOM** = "AC").
- Choose **ACTIVITY_START** from actual vs planned initiation dates (ignore obviously invalid years like < 1901). **ACTIVITY_END** comes from the actual completion date.
- Set **ACTIVITY_STATUS** from initiation date plus the **iscompleted** flag.
- Map **iswui** to **IN_WUI**. Set **TRMT_GEOM** = "POINT", **Crosswalk** = **typename**, **AGENCY** = "DOI".
- Set **COUNTS_TO_MAS** = "NO" when **ACTIVITY_END** ≥ 2023-10-01.

NPS (National Park Service)

- Keep records where **ActualCompletionDate** > 1995-01-01 or null. Dissolve features by treatment/admin fields (to remove duplicates).
- Set **ACTIVITY_QUANTITY** = **TreatmentAcres**, **ACTIVITY_UOM** = "AC".
- **ACTIVITY_END** from completion date when present or if missing, use October 1 of the completion fiscal year.
- Map **Activity_Status** from **TreatmentStatus**:
 - "Completed" → **ACTIVITY_STATUS** = "COMPLETE"
 - "Initiated" → **ACTIVITY_STATUS** = "ACTIVE"
- Set **Crosswalk** from **TreatmentType**. If missing, fall back to rules based on **TreatmentCategory** (e.g., Fire/ Mechanical).
- Set **COUNTS_TO_MAS** = "NO" when **ACTIVITY_END** ≥ 2023-10-01.

USFS (US Forest Service – FACTS)

- Keep only USFS records where **STATE_ABBR** == "CA" and **ACTIVITY_CODE** is in a curated list.
- Apply special filters for codes **1117**, **1119**, **2510**, **2341** based on **FUELS_KEYPOINT_AREA** and dates:

- Keep **2341** and **2510** only if **FUELS_KEYPOINT_AREA** \in {**3**, **6**}.
 - Keep **1117** and **1119** only if **FUELS_KEYPOINT_AREA** = **6**.
 - Require at least one of **DATE_COMPLETED**, **DATE_AWARDED**, **NEPA_SIGNED_DATE** to be present. And **DATE_COMPLETED** \geq **start_year-01-01**.
- Drops previously included ~46,000 acres of **2341** records that had no keypoint value (so they no longer count toward MAS).
 - Set **ACTIVITY_QUANTITY** = **NBR_UNITS_ACCOMPLISHED**, if that's missing, use **NBR_UNITS_PLANNED**. Units **ACTIVITY_UOM** = **UOM**.
 - **ACTIVITY_END** comes from **DATE_COMPLETED** or, if missing, **NEPA_SIGNED_DATE**.
 - Compute **ACTIVITY_STATUS** from date: **COMPLETE**, **ACTIVE**, **OUTYEAR**, **PLANNED**, **CANCELLED** depending on date(**DATE_COMPLETED**/**DATE_AWARDED**/**NEPA_SIGNED_DATE**).
 - Set **PROJECTID_USER** = "USFS-" + **NEPA_DOC_NBR**, **TRMTID_USER** = **SUID**.
 - Map **WUI** from **ISWUI**. Set **TRMT_GEOM** from the geometry code.
 - Build the **Crosswalk** from **ACTIVITY**, cleanup for special cases.
 - Filter records so **Year** is within [**start_year**, **end_year**] before enrichment and domain assignment.

CNRA (Polygon / Line / Point)

- Keep only CNRA records where features, activities, and projects can be linked by valid IDs (GlobalID and treatment ID);
 - If **AGENCY** is missing, set it to "CNRA". If **ACTIVITY_STATUS** is missing, default to "COMPLETE".
 - When **ACTIVITY_END** is missing, compute it based on the status and start date.
 - Set **Crosswalk** = **ACTIVITY_DESCRIPTION**, **Source** = "CNRA".
 - Set **TRMTID_USER** = **GlobalID** + "-CNRA", **PROJECTID_USER** = **PROJECTID** + "-CNRA".
 - If the input feature type is "Point" but the dataset also has non-point geometries, force the geometry type to point by taking the centroid.
 - Fix **ADMINISTERING_ORG** by changing "MRCA" to "SMMC".
 - Enrich using geometry type, trim to the template schema, and assign domains.
-

CALTRANS

- Inner-join activity and treatment tables on **Highway_ID** and **Calendar_Year**. If any invalid geometries remain after repair, stop the process (fail).
- Set **ACTIVITY_QUANTITY** = **Production_Quantity**.
- Set **ACTIVITY_UOM** from **UOM**, converting any "ACRE" variants to "AC". Only set **TREATMENT_AREA** when the units are acres.
- Set **ACTIVITY_STATUS** = "COMPLETE". **ACTIVITY_START** Date comes from **WorkBeginDate** and **ACTIVITY_END** date from **WorkEndDate**.
- Set **PROJECTID_USER** = **Highway_ID**.
- **TRMTID_USER** is built from **Highway_ID** / **From_PM_C** / **To_PM_C**, county, region, and WUI to make it more descriptive.
- Set **Crosswalk** = **Activity_Description**, **TRMT_GEOM** = "LINE".
- After enrichment, apply domains and set **ownership** to **STATE** for all records.

PFIRS

- Drop PFIRS records for agencies handled in other pipelines (Cal Fire, USFS, FWS, BLM, NPS). Apply burn-type filters so most timber company broadcast burns are excluded, except for FWS Forestry LLC.
- Set **ACTIVITY_QUANTITY** = **Acres_Burned**, **ACTIVITY_UOM** = "AC".
- Set **ACTIVITY_END** = **Burn_Date**, **ACTIVITY_STATUS** = "COMPLETE".
- Map **Agency** using lookup table. Industrial timber normalized to "Timber Companies". Set **Crosswalk** from **Burn_Type** (broadcast vs pile, etc.). Set **TRMTID_USER** = "PFIRS-" + **index** (unique ID per record).
- Spatial de-duplication - Using appended treatment polygon dataset, the code selects only polygons where **ACTIVITY_DESCRIPTION** is "BROADCAST_BURN" or "PILE_BURN", then, for each year, it does a spatial join with PFIRS points and removes any PFIRS points that fall inside those polygons, so the same burn is not double counted.
- Apply domains and then reset **AGENCY** from **ORG_ADMIN_p** as the final value.

Timber Spatial

- Validate schema, reproject to **EPSG:3310**, normalize Status ("Exists" → "Active"); clip to California; filter Year within [start_year, end_year].
- **ACTIVITY_QUANTITY** = **GISACRES** as float; **ACTIVITY_UOM** = "AC".
- Dates from Year only (**ACTIVITY_START** = Jan 1, **ACTIVITY_END** = Dec 31); status fields = uppercased Status.
- Set **PROJECTID_USER** = "TI-" + **OBJECTID**; org/funding = TIMBER/PRIVATE; **ACTIVITY_DESCRIPTION** = Objective; **Crosswalk** = **ACTIVITY_DESCRIPTION**;

TRMT_GEOM = "POLYGON"; TRMTID_USER = PROJECTID_USER.

- Enrich polygons, trim to template, apply domains.

Timber Nonspatial (Representative Grid Points)

- Validate Excel schema; rename human-readable fields (e.g., “ACTIVITY CATEGORY”) to internal _ columns; convert to TF schema and then to GeoDataFrame.
- ACTIVITY_QUANTITY from ACTIVITY_QUANTITY_; ACTIVITY_UOM from ACTIVITY_UNITS; later dissolved by key attributes and summed.
- ACTIVITY_START / ACTIVITY_END copied from _ versions; descriptions cleaned (e.g., “Group Selection” → “Group Selection Harvest”).
- Funding set to PRIVATE/PRIVATE_INDUSTRY; AGENCY and all admin orgs copied from ADMIN_ORG_NAME; Source = "Industrial Timber"; TRMT_GEOM = "POINT";
- Generate fake grid of coordinates clustered by ACTIVITY_CAT
- Dissolve by description/status/veg/county/dates/admin org to get an aggregated sum of activity quantity
- Set IDs PROJECTID_USER = "TI-" + index, TRMTID_USER = PROJECTID_USER.
- After enrich_points, update_enriched_data adjusts ACTIVITY_CAT / PRIMARY_OBJECTIVE based on activity and crosswalk, sets COUNTS_TO_MAS = "YES", force Taskforce Region to be “NON_SPATIAL”, and aligns ADMINISTERING_ORG = AGENCY; domains applied.

Integration & Reporting Pipelines

Append (Build Point / Line / Polygon Layers)

- Find all individual agency enriched GDBs and classify each layer as point/line/polygon. Optionally, use --geom_type to process only one geometry type at a time.
- For each selected layer (before appending): Run **categorize_activity** → **standardize_domains** → **counts_to_mas(start_year, end_year)** again and overwrite the layer so logic is up to date.
- During concatenating, drop empty geometries only if **all** those rows have **COUNTS_TO_MAS == "NO"**. If any empty geometry has **COUNTS_TO_MAS == "YES"**, treat it as an error and stop.
- Fix invalid geometries. Add **VALID_GEOM** to indicate:
 - **"VALID"** – geometry was valid
 - **"FIXED"** – geometry was invalid but successfully repaired
 - **"INVALID"** – geometry could not be repaired and was reduced to a centroid point

- Add **CORE_CRITERIA**: a count of how many of these core fields are valid/present:
 - Administering org
 - Activity description
 - Activity end (if status is COMPLETE)
 - Activity status
 - Activity quantity
 - Activity UOM
- Concatenate by geometry type, reproject to **EPSG:3310**, and clip to the California boundary using Dask **sjoin**. drop added boundary fields.
- For points, add the **Timber Nonspatial** layer back in (since those points are deliberately placed outside California's boundary).
- Save as **appended_poly**, **appended_line**, **appended_point**.

Activity Report

- Read **appended_poly**, **appended_line**, **appended_point**, then combine them into a single dataset.
- Keep only records with **COUNTS_TO_MAS == "YES"** (activities that count toward MAS).
- Convert all geometries to representative points of original geometry (e.g., polygon centroid, line midpoint).
- Keep fields - AGENCY, ADMINISTERING_ORG, Ownership group, county, region, Description, activity category, vegetation, Status, quantity, units, end date, Year_txt, VALID_GEOM (original geometry validity), CORE_CRITERIA (core field completeness count)
- Derive **ENTITY_TYPE** from **AGENCY** (State / Federal / Timber Companies / None)
- The activity report GeoDataFrame is saved as a feature layer into a geodatabase file.

Footprint Report

- **Create Buffers**
 - **Points**- Buffer each point into a circle using the reported acres (AC) to set the radius. Keep only records where **COUNTS_TO_MAS == "YES"** and the radius is positive.
 - **Lines** - Buffer each line based on reported acres and segment length to create polygons. Drop lines whose resulting buffer is unreasonably wide (extreme CalTrans cases).
 - **Polygons** - Keep only those where **COUNTS_TO_MAS == "YES"** and **TREATMENT_AREA < 100000** acres.
- For each **Year** -

- Separate **TIMBER** points (**AGENCY** == "**TIMBER**") and treat them at “face value” (acres).
 - Union polygons + non-timber points,
 - split lines into intersecting vs non-intersecting with this polygon+point union
 - Dissolve by **TRMTID_USER** for both line and polygon+point groups.
- For intersecting lines, scale **ACTIVITY_QUANTITY** by the fraction of buffered line area **outside** the polygon+point union.
- For polygons+points, create:
 - full dissolved polygons, exploded “spaghetti” geometry, and centroid “meatball” points, plus a combined line/timber table,

Save these per year to a temp GDB. Then:

- Use ArcPy **FeatureToPolygon** to create non-overlapping polygons (NOPs) from spaghetti.
- Spatially join meatball points to assign each NOP to the treatment with the largest **ACTIVITY_QUANTITY**.
- Restore the final polygon geometry from dissolved polygons, then append the line and timber pieces.
- Save the result as **Footprint_Report_<year>**.