

CPDP Pipeline Output Documentation - II

See also [Old Pipeline README](#) (still being updated for changes with new data)

See also [CPDP document types and entity relation model - II duplicate](#) for details on the database schema that this pipeline feeds into

See also: <http://invisible.institute/download-the-data>

SchemaSpy relational map: <http://users.eecs.northwestern.edu/~jennie/courses/cs339/lab1/cpdb-schema/relationships.html>

Database dump here: <http://users.eecs.northwestern.edu/~jennie/courses/cs339/lab1/cpdb.sql>

Officers

Deduplicated list of officers from all given files. Starting from a 2017 roster, each new file that contains police entities is successively merged in. For the most part, except with explicit exceptions, requires first name, last name and appointed date to consider two entities to be the same.

File: `officers/final_profiles.csv`

CPDP Table: `data_officer`

- **UID** - universal id, i.e., unique officer id. used to link officer to other files
- **first_name** - First name of officer
- **last_name** - Last name of officer
- **middle_initial** - First character of middle name
- **middle_initial2** - First character of second middle name if any
- **suffix_name** - Suffix name, e.g., I, JR, SR
- **birth_year** - Birth year of officer
- **race** - Race of officer
- **gender** - Gender of officer
- **appointed_date** - The date that officer is appointed
- **resignation_date** - The date that officer resigned (NOTE: not always given, so you'll see officers who are not active without a resignation date. requested another FOIA to hopefully supplement this)
- **current_status** - Last known status of officer, 0 if inactive, 1 if active
- **current_star** - Last known badge number of officer
- **current_rank** - Last known rank of officer
- **current_unit** - Last known unit number of officer
- **foia_names** - Name of foias this officer appeared on
- **match** - For each file, columns that were to determine match between entities. Format is {foia_name}: {dash separated list of columns} E.g., "unit_history__2016-03: first_name_NS-last_name_NS-appointed_date" means that this officer matched from the first file it appeared on (say, roster_1936-2017) to this unit_history file because they both had the same first name, last name and appointed date.
- **profile_count** - Number of profiles this officer appeared in

Complaints

A combination of 9 different files given in various formats from 3 different agencies. There will sometimes be columns that seem useful but were only included in certain releases. Attempted to note most of them below, but can always just look at null value counts grouped by filenames to confirm.

File: complaints/complaints-complaints.csv

CPDP Table: data_allegation

Complaints/allegations file. Where did the complaint take place and when.

NOTE: provided format for address changed between files, the site will display the address string from "old_complaint_address" if it exists, and if not, construct one by combining add1, add2 and city.

- **cr_id** - complaint register id, unique id for each complaint. NOTE: format changed/reset sometime in 2019 to be prefixed by the year.
- **complaint_date** - Date complaint was made
- **incident_date** - Date when incident occurred
- **incident_time** - Time that incident occurred
- **add1** - Street number of the address
- **add2** - Street name
- **city** - City name
- **state** - State code
- **zip** - Zipcode
- **old_complaint_address** - Older format of address which has a full address string rather than separated out fields
- **location** - type of location incident took place
- **location_code** - Can likely ignore, often note given, code supposed to correspond to location field
- **last_filename** - last complaint file that this complaint appeared on
- **filenames** - all filenames this complaint appeared on
- **latitude** - geocoded latitude from address string
- **longitude** - geocoded longitude from address string
- **beat** - Police beat number

File: complaints/complaints-accused.csv

Table: data_officer_allegation

Maps a complaint (cr_id) to an officer (UID), shows category and findings.

- **cr_id** - complaint register id, unique id for each complaint. NOTE: format changed/reset sometime in 2019 to be prefixed by the year.
- **UID** - unique officer id, used to link to final-profiles to get officer info
- **complaint_category** - category of complaint, note that this changes between files and this is currently showing the category of the most recent file. Looking into possibly changing this preference before release.
- **complaint_code** - code representing complaint categories. More stable between files than category, but still changes file to file

- **final_finding** - final status of complaint, most likely the one to use for analysis
- **days** - days for penalty, only given in last 4 files so often null
- **final_outcome** - final outcome of accusation, sometimes not given and not given at all in 2 files
- **final_penalty** - only given in last 4 files, combination of a code and number of days for each penalty
- **recc_finding** - recommended finding of accusation
- **recc_outcome** - recommended outcome of accusation
- **recc_penalty** - recommended penalty of accusation
- **last_filename** - last filename this complaint appeared in
- **filenames** - all filenames this complaint appeared in

File: complaints/complainant.csv

Table: data_complainant

- **cr_id** - complaint register id, unique complaint id
- **gender** - Gender of complainant
- **race** - Race of complainant
- **age** - Age of complainant at time of incident
- **birth_year** - Birth year of complainant
- **last_filename** - Last filename this complainant appeared in

File: complaints/civilian_witnesses.csv

Table: data_complainant

- **cr_id** - complaint register id, unique complaint id
- **gender** - Gender of civilian witness
- **race** - Race of civilian witness
- **current_age** - Age of civilian witness at time of incident
- **birth_year** - Birth year of civilian witness
- **last_filename** - Last filename this civilian witness appeared in

File: complaints/investigators.csv

Table: Investigator

- **cr_id** - complaint register id, unique complaint id
- **first_name** - First name of investigator
- **last_name** - Last name of investigator
- **middle_initial** - First character of middle name, in upper case
- **suffix_name** - suffix name, e.g., JR or SR
- **appointed_date** - Date when investigator was appointed
- **current_rank** - Rank of investigator at time of investigation
- **current_unit** - Unit number of investigator at time of investigation
- **last_filename** - Last filename this investigator appeared in

File: complaints/victims.csv

Table: victims

- **cr_id** - complaint register id, unique complaint id
- **age** - age of victim
- **birth_year** - birth year of victim
- **gender** - gender of victim
- **race** - race of victim
- **injury_condition** - injury condition of victim, only given in last 5 files so often null
- **last_filename** - Last filename this victim appeared in

TRRs

Tactical response reports, or use of force reports. A combination of 5 different FOIA requests, the first 3 are in a format consistent with what's on the existing production site, while the last 2 are in a new format. Tried to note where fields are unique to the old or new format. If there are fields in either format that seem useful for some analysis but not present here, reach out and we can include them.

Old TRR format, actual document: <https://igchicago.org/wp-content/uploads/2019/11/CPD-11.377.pdf> New format actual document: <http://directives.chicagopolice.org/forms/CPD-11.377.pdf>

File: trrs/trr_main.csv

Table: trr_trr

Main file, maps each tactical response report to an officer. Gives location data, some flags, some flags derived from other files, additional officer data and then some data on the subject of the force.

- **UID** - unique id for officer, maps to final_profiles
- **tr_id** - TRR id
- **rd_no** - unique record number of event, NOTE: will be replaced before being uploaded to site or database
- **event_no** - event number, not uploaded to site
- **beat** - Beat where the TRR incident occurred
- **block** - Block where the TRR incident occurred
- **direction** - Direction of street where the TRR incident occurred.
- **street** - Street where TRR incident occurred
- **location** - Location where TRR incident occurred
- **trr_datetime** - Date and time of incident
- **indoor_or_outdoor** - Indicates whether incident happened outdoor or indoor.
- **lighting_condition** - Lighting condition at time of incident
- **weather_condition** - Weather condition at time of incident
- **notify_OEMC** - Whether or not Office of Emergency Management and Communications was notified (missing = False).
- **notify_district_sergeant** - Whether or not district sergeant was notified (missing = False)
- **notify_OP_command** - Whether or not OP command was notified (missing = False)
- **notify_DET_division** - Whether or not detective division was notified
- **number_of_weapons_discharged** - Number of weapons discharged in TRR
- **party_fired_first** - Which party fired first, member or officer
- **location_recode** - Recode of location for simplification
- **taser** - Indicates whether taser is used

- **total_number_of_shots** - Total number of shots fired in TRR
- **firearm_used** - Whether or not TRR involved firearm
- **number_of_officers_using_firearm** - Number of officers using firearms for this same RD_NO, aggregate value
- **officer_assigned_beat** - Beat number member was assigned to during TRR incident
- **officer_on_duty** - Whether or not member was on duty during TRR incident.
- **officer_in_uniform** - Whether or not member was in uniform during TRR incident.
- **officer_injured** - Whether or not member was injured during TRR incident.
- **officer_rank** - Rank of officer at time of incident
- **officer_unit_id** - unit the officer was in at the time.
- **officer_unit_detail** - Foreign key to another data_policeunit that the officer was in temporarily
- **latitude** - geocoded latitude for address string
- **longitude** - geocoded longitude for address string
- **subject_id** - Ignore this field, not making any attempt at deduplicating subjects based on limited demographic data
- **subject_armed** - Whether or not the subject was armed.
- **subject_injured** - Whether or not the subject was injured.
- **subject_alleged_injury** - Whether or not the subject alleged an injury.
- **subject_age** - Age of subject at time of incident
- **subject_birth_year** - Birth year of subject
- **subject_gender** - Gender of subject.
- **subject_race** - Race of subject.
- **last_filename** - last filename this trr appeared on

File: trrs/trr_actions_responses.csv

Table: trr_actions_responses

- **trr_id** - unique id for trr, maps to trr_main
- **rd_no** - unique event number, won't be uploaded to site
- **person** - Type of person committing the action (member/officer or subject).
- **resistance_type** - Type of subject resistance, active or passive
- **action** - Action taken by officer/member/subject
- **other_description** - Description of member action for other and other force member actions
- **member_action** - More specific description of force type
- **force_type** - Description of member force (corresponds to action_sub_category, other and other force -> 2)
- **action_sub_category** - Same as action_category, except some levels are broken into sublevels, i.e. physical holding (3.3) and taser displayed (3.1) are the same level, but different severity within the level
- **action_category** - Action level 0 (member presence) to 6 (firearm), only applicable for officer/member actions
- **resistance_level** - Level of subject resistance, generated from member action.
- **last_filename** - last filename this trr action response appeared on

File: trrs/trr_statuses.csv

Table: trr_trrstatus Note: does not exist in new TRR format, so not in the last 2 files.

- **trr_id** - unique id for TRR, maps to trr_main
- **rank** - Rank at time of TRR status
- **star** - Star at time of TRR status
- **status** - Status of TRR at time for an officer in the chain of command.
- **status_date** - Date when status was updated
- **status_time** - Time when status was updated
- **age** - Age at time of TRR status
- **first_name** - First name of officer
- **last_name** - Last name of officer
- **middle_initial** - First character of middle name
- **middle_initial2** - First character of second middle name if any
- **suffix_name** - Suffix name, e.g., I, JR, SR
- **birth_year** - Birth year of officer
- **race** - Race of officer
- **gender** - Gender of officer
- **appointed_date** - The date that officer is appointed

File: trrs/trr_subject_weapons

Table: trr_subjectweapon

- **trr_id** - unique id for TRR, maps to trr_main
- **weapon_type** - Type of weapon
- **firearm_caliber** - Firearm caliber
- **weapon_description** - Description of weapon (usually when weapon_type is other)

File: trrs/trr_weapon_discharges

Table: trr_weapondischarge

- **trr_id** - unique id for TRR, maps to trr_main
- **rd_no** -
- **weapon_type** - Type of used weapon
- **weapon_type_description** - Description of weapon
- **firearm_make** - Firearm make
- **firearm_model** - Firearm model
- **firearm_barrel_length** - Firearm barrel length
- **firearm_caliber** - Firearm caliber
- **total_number_of_shots** - Total number of shots fired
- **firearm_reloaded****d** - Whether or not firearm was reloaded
- **number_of_catridge_reloaded** - Number of catridge reloaded
- **handgun_worn_type** - Handgun worn type.
- **handgun_drawn_type** - Handgun drawn type.
- **method_used_to_reload** - Method used to reload
- **sight_used** - Whether or not a sight was used
- **protective_cover_used** - What type of protective cover was used
- **discharge_distance** - Range of distance of discharge
- **object_struck_of_discharge** - What was struck by discharge.

- **discharge_position** - Position of discharge.
- **accidental_discharge** - Whether discharge was accidental or not. New to last 2 files.
- **discharge_to_destruct_animal** - Whether discharge was to "destruct" animal. New to last 2 files.
- **fadischrg_fired_at_veh_i** - Indicator of whether firearm was discharged at vehicle. New to last 2 files.
- **fadischrg_first_shot_cd** - Whether firearm discharge was first shot. New to last 2 files.
- **fadischrg_make_cd** - Discharged firearm make code. New to last 2 files.
- **fadischrg_model_cd** - Discharged firearm model code. New to last 2 files.
- **fadischrg_reloaded_i** - Discharged firearm reloaded indicator. New to last 2 files.
- **objstrck_disch_memweap_new_trr** - Whether discharged member weapon struck an object. New to last 2 files.
- **taser_arc_cycle_cd** - Taser arc cycle code. New to last 2 files.
- **taser_arc_oth_desc** - Taser arc other description. New to last 2 files.
- **taser_contact_stun_cd** - Taser contact stun code. New to last 2 files.
- **taser_dartid_no** - Taser dartid number. New to last 2 files.
- **taser_probe_dischrg_cd** - Taser probe discharge code. New to last 2 files.
- **taser_prop_inventory_no** - Taser prop inventory number. New to last 2 files.
- **taser_spark_display_cd** - Taser spark display code. New to last 2 files.
- **taser_trigger_cd** - Taser trigger code. New to last 2 files.
- **weaponcontribute_subinjury** - Whether weapon contributed to subject injured. New to last 2 files.
- **weapon_dischrg_count_new_trr** - Weapon discharge count (not sure why separate, but given with suffix 'new_trr' in original FOIA file). New to last 2 files.
- **weapon_dischrg_does_not_apply** - Whether weaopn discharge applies or not. New to last 2 files.
- **weapon_dischrg_selfinjur_mem_i** - Whether weaopn discharge led to member self injury. New to last 2 files.
- **weapon_dischrg_selfinjur_sub_i** - Whether weapon discharge led to subject self injury. New to last 2 files.
- **weapon_type_new_trr** - Weapon type (not sure why separate, given with suffix 'new_trr' in original FOIA file). New to last 2 files.

File: awards/awards.csv

Table: data_award

- **UID** - unique id of awarded officer, maps to final_profiles
- **award_id** - unique id of the award
- **award_type** - Type of the award
- **award_start_date** - Start date of the award
- **award_end_date** - End date of the award
- **current_award_status** - status of the award.
- **award_request_date** - Date when the award is requested
- **requester_full_name** - Full name of the award requester
- **ceremony_date** - Date when the ceremony was held

Salaries

Comes from the deparatment of human resources, i.e., a completely differnet agency than the other files. Has the fewest fields in common and requires the most work in entity resolution as this is the data that's

most likely to have officers with small but meaningful differences in names, and different appointed dates.

File: salaries/salaries.csv

Table: data_salary

- **UID** - unique id of officer, maps to final_profiles
- **pay_grade** - Pay grade
- **year** - Year
- **rank** - Rank (multiple ranks can correspond to same pay grade)
- **salary** - Salary for that year (likely computed from pay grade and tenure, if you quit in that year the salary is not impacted in the data)
- **employee_status** - Employee type.
- **org_hire_date** - Most common organization hire date (by any policing organization)
- **start_date** - Most common start date as a police officer in the CPD
- **spp_date** - Date of most recent promotion/rank change
- **age_at_hire** - Age at time of hiring

Settlements

File: settlements/settlements.csv

- **UID** - unique id of officer, maps to final_profiles
- **case_id** - unique id for lawsuit case
- **complaint** - settlement complaint
- **incident_date** - date when incident occurred
- **location** - location where incident occurred
- **narrative** - text summary of lawsuit
- **settlement** - settlement amount

Unit Histories

File: unit_histories/unit-history.csv

Table: data_officerhistory

- **UID** - unique id of officer, maps to final_profiles
- **unit_id** - unit id
- **unit_start_date** - First date when officer joined the unit
- **unit_end_date** - Last date of officer at the unit