

Additional variables to be added to the Crash Analysis System (CAS) open dataset

To: Helen Aki, Senior Manager Research & Analytics

Cc: Holly Ludlow, Manager Statistical Analysis; Rebecca Schulz, Manager Specialist Analytics; Suzanne Jones, Data Scientist; Colin Morrison, Senior Advisor Strategic Interventions; Christopher Liu, Data Scientist; Mel Smalley, Manager CAS Processors; Louise Murrell, CrashHub Project Manager

From: Carolyn Fyfe, Manager Statistical Design

Date: xx March 2018

Subject: Additional variables to be added to the Crash Analysis System (CAS) open dataset

Purpose

The purpose of this memo is to seek your approval to release the following variables in the unit record open data produced from the Crash Analysis System (CAS).

Background

We aim to make as much information as possible openly available from the Crash Analysis System (CAS), while ensuring we are maintaining our privacy obligations. We consider that CAS variables can be put into three groups:

1. **Personal identifiers**, e.g. Name, address, date of birth, licence plate, driver licence number, officer ID. These variables absolutely must not be openly released from CAS so do not form part of our open data approach.
2. **Non personal information**, e.g. Location, road characteristics, crash conditions (lighting, weather), crash severity, etc. We have made these variables [openly available](#) for each crash in the CAS since 1 January 2000.
3. **'Grey area'** e.g. Crash factor, crash cause, exact time of crash, etc. These need careful consideration before they are made openly available to ensure that we are not allowing an individual to be identified or identifiable, and that the variables are suitable for open release in terms of accuracy/reliability.

Since the release of the first variables, we have been responding to requests for additions to the dataset by adding non-personal variables as appropriate. We have also continued assessing the 'grey area' variables, including going over our assessment in an in-depth manner with the Data Futures Partnership Re-Identification work-stream lead.

The purpose of this memo is to seek approval for the release of the variables set out below in appendix 1 in the open dataset. Like the current data, this information would be provided for each individual crash from 1 January 2000. If approved, we will endeavour to make these variables openly available in time for the Hackathon on 16th March. If this is not achievable, we will make these available as soon as possible, and notify those attending the Hackathon that this is in the pipeline.

For information on the assessment of the variables currently available, please see this [memo](#).

Risks

Providing the data at unit record level allows people to observe trends in the data (which is one of the aims/benefits of open data). However, in some instances these trends may be a result of changes to coding practices by our CAS Processing team. In instances where changes to practices are known, we will include this information in the metadata. However, there is a risk that errors in coding, or coding issues (such as the current issue with alcohol suspected), could be observed and questioned. We do not think this risk should prevent release, as any issues can also be seen by existing CAS users. Moreover, being open to people questioning the data is a good feedback loop to help us constantly improve data quality. Before the release of this data the Manager of the CAS Processors will be advised of what will be released, and this risk and given the ability to advise on any data that should not yet be released due to coding concerns.

Recommendations

It is recommended that the Manager Research and Analytics:

1	Approves the release of the variables listed in appendix 1 at the individual level (ie per distinct crash), without suppression or aggregation.	YES / NO
2	Agrees that the variables listed in appendix 2 are not fit for release at the individual level (ie per distinct crash), and therefore should not be released as open data.	YES / NO
3	Notes the advice from the Privacy Officer attached in Appendix 3.	YES / NO
4	Accepts the risk that openly releasing this data may highlight any changes to coding practices/incorrect coding by the CAS Processing Team	YES / NO

Appendix 1 – Variables for release, March 2018

Format of Data

Currently the CAS open data is released in a .csv file with one row per crash, without suppression or aggregation. These additional variables will be added to the [existing dataset](#) where appropriate. In the event that the current format does not support these additional variables, we will reshape the data into a form that best supports our customers' usage of the data. (For example, to provide information about the characteristics of the people involved).

Privacy Assessment

Given the amount of information contained in this database, when assessing each individual variable for open release it is also important to consider the picture that is being provided by the data as a whole.

As crashes are often reported in detail in the media, information about the crash is likely to already be in the public domain. As such, it is reasonable to assume that someone could find the crash a particular individual has been involved in, in our dataset. We therefore need to consider whether we are exposing additional personal information about that person.

Following that assumption, if these variables are added to the dataset the following types of information could be deduced about an individual from our data (worst case scenarios, not identified instances):

- The social cost resulting from the person crashing
- Factors contributing to the crash (discussed further below)
- The driving movement the person was undertaking when they crashed
- The very broad age band of the person (very likely to have already been known at this level to identify them in the data)
- Whether or not they were wearing their seatbelt
- If a fatality would have been prevented for someone in their vehicle if they were wearing a seatbelt
- The severity of their injury (but not what the injury was)
- The severity of the injury [they caused] (but not what the injury was)
- The parties involved in their crash – eg could potentially deduce, they were over the legal limit or on drugs and crashed into a tree; they were over the legal limit or on drugs and hit a pedestrian aged under 16, resulting in a fatality (here you need to know they were the driver and fault is assumed); etc
- Their drivers licence status
- Whether they owned or rented the vehicle

To attempt to protect against absolute certainty in identification we have grouped variables that are most likely to be used as a way to link with knowledge obtained elsewhere, such as exact day and time, and the age of the person.

Factor codes

We assess the variables with the highest risk in release to be the factors contributing to a crash. However, the inclusion of these would also provide the greatest benefit, as they would enable our open data customers to understand the causes behind the crash and potentially come up with preventions for this happening again. To balance this risk and benefit, we are proposing to only openly release information about factors at the highest level. At that level, these could potentially allow someone to deduce things such as that a person was:

- over the alcohol limit or on drugs
- had an illness or disability
- doing something intentional or criminal

- driving too fast for the conditions
- failed to keep left
- conducted a forbidden movement
- in the wrong lane
- did not stop
- inexperienced, etc

For a full list please see the factor section below.

When considering factors it is vital to note that these are factors PROBABLY contributing to crashes, as recorded by the Police Officer at the time of the crash. This information has not been substantiated in a court of law. A worst case example of this could be (hypothetically) that it was believed a person was on drugs, but they were having a medical event, thereby exposing this person to gossip or rumour. This will be made as clear as possible in our documentation. In addition, we will not explicitly link crash factors to an individual. However, in some circumstances, such as single vehicle crashes, or based on movement codes, it will be reasonably easy to accurately deduce which party the factor applies to. Note also, a crash may have multiple crash factors.

Summary

While it is possible that additional information about a person could be discovered from this data, we consider the public good and potential benefit in terms of road safety initiatives that could be gained from proactively releasing this data outweigh any privacy risk in doing so.

Data for release

NB. Only new variables have been noted below. For information on the variables already available please see the [open data](#). Some variables may be renamed when released for ease of use for our customers (eg. coi_id is not immediately recognisable as 'suburb'). Descriptions will be updated to match the variable released.

The groupings below are by CAS table.

Crash

VARIABLE	DESCRIPTION	ASSESSMENT
casualty_hier	A derived flag summarizing casualties. Values: 'E' (pedestrian), 'S'(cyclist), 'O'(other non-driver), 'M' (motor cyclist), 'H' (heavy vehicle occupants, 'D' (car/van drivers), 'P' (car/van pass), 'T' (other motorised) The first in this order that applies is attached to the crash.	This is a hierarchy based on the most vulnerable person in the crash. For example, a crash in which a pedestrian was a fatality would be a 'pedestrian crash'. Release of at fault should be considered with this variable. There is limited privacy risk in releasing this hierarchy alone. Releasing this variable with at-fault would make it easier to link an individual to a fatality they caused, in the event that individual was able to be identified in the data. Note, this is most likely to occur through media reports of fatal crashes.
coi_id	Suburb name	Exact crash location is already openly available.

		This provides flexibility in the way users can understand location.
crash_month	The month (1-12) in which a crash occurred, if known. Derived from crash_date.	While year, month, location and crash descriptions may allow a crash to be identified, exact day and time will not be provided as a confounding factor.
crash_time_4hr	A field derived from crash_time. It is the time of day in 4hr blocks. Eg the value 1 is times 000-0359; the value 2 is times 0400-0759 etc. Value 0 is unknown time.	While year, month, location and crash descriptions may allow a crash to be identified, exact day and time will not be provided as a confounding factor.
multi_veh_simple	Another flag (see also multi_veh) derived from the number of vehicles which are given roles in a crash. Values 'S' (Single vehicle; multi_veh = 'S'), 'M' (multi-vehicle; multi_veh = 'M', 'C', 'O' or 'E'), '_' (Non-vehicle; multi_veh = 'F', 'G', 'P' or '_'). 'Vehicle' means non-parked vehicle. All these may also involve parked vehicles.	This is another way of presenting the data about the vehicles involved in the crash that is already available in the open data. While with the inclusion of crash cause it would be possible to identify single vehicle crashes in which alcohol was a cause, confounding factors (eg of day and time) limit the ability to identify the person involved with certainty.
mvmt_id_grp	A unique id for a movement code group. Values 'O' (overtaking), 'S' (Straight road), 'B' (bend), 'R' (rear end or obstruction), 'I' (intersection), 'P' (pedestrian), 'M' (miscellaneous), " (unknown). These are the Road Safety Report movement code groups.	The risk of consequential elements associated with this variable are considered to be negligible. Nothing indicating party at fault will be released, further mitigating any risk in releasing this variable.
mvmt_ida	The first movement code for the principal vehicle involved in the crash. Possible values range from 'A' to 'Q'. See Vehicle Movement Coding Sheet (version 1.0 April 1997) for details.	As per above. Note however terminology around 'principal vehicle' would need to be considered to ensure attribution of fault is not being identified. Suggest using 'key vehicle' to align with the messaging in the guide to coded crash reports which clearly states that 'the vehicle role does not in any way indicate driver fault.'
mvmt_idb	The second movement code for a crash. There is no direct lookup for this as its meaning depends	As per above.

	on the value of mvmt_ida as well. The descriptions are therefore found in lu_mvmt_id_ab.	
pol_att	A flag indicating whether police attended the crash. Possible values 'Y', 'N'	No privacy concerns here as just notes if Police attended the crash or if this was reported to them.
weekend	A derived flag: whether the crash was on a weekend. Values 'W' (weekend - between Fri 1800 and Mon 0559 for ordinary weekends, or 1600 on the first day to 0559 on the final day for holiday weekends); ' ' (not weekend)	Providing this derived flag rather than exact day as a confounding factor.
z2	Used to hold cal_rcs.road_class. A single character indication of the general class of road. Values 'Y' (motorway), 'S' (open road SH), 'O' (other open road), 'M' (major urban road), 'I' (minor urban road)	Not personal information, no privacy concerns.
z3	Used to hold a code for the season of the year, derived from crash_month. Possible values: 'S' (Summer; months 1,2), 'A' (Autumn; months 3-5), 'W' (Winter; months 6-8), 'P' (Spring; months 9-11)	Not personal information, no privacy concerns.
z5_obj_struck	A Y/N flag indicating whether any objects were struck during this crash	Details of objects struck are already provided in the open data. This summary indicator would be of benefit to users. There are no privacy concerns.

Person

VARIABLE	DESCRIPTION	ASSESSMENT
Age group	[new variable being derived for open data] Age group the person belongs to. Age groups will be: 0-15 years; 16 - 24 yrs; 25 - 54 years; 55 - 64 years; 65+	This variable is being derived for the purposes of the open data to provide the benefit of age information, while minimising the chance of re-identification.
ct_party_type	This is the 'party type' as used in the crosstab reports. Values 'C' (car/van/taxi), 'T' (truck) 'B' (bus), 'M' (motorcycle), 'P' (moped), 'S' (cycle/cyclist), 'E' (ped.), 'O'	There is a lot of benefit in being able to tell the role of the person. Ensuring that any identifying features (eg age) are confounding

	(other), 'U' (unknown), 'Q' (equestrian), 'K' (skateboarder), '4' (Van/Ute/4WD). Derived from pers_type,veh_type	minimises privacy concerns.
fat_prev	'Fatality preventable by wearing seatbelt'. Values 'Y' (yes, preventable), 'N' (no, not preventable), 'U' (unknown), '' (not fatal crash, or not applicable - pedestrian etc.)	No privacy concerns, but need to confirm if we should release this as it is Police opinion, not from coroner etc
inj_sev	If the person was injured, this field has a code for the severity of injury. Values 'F' (fatal) 'M' (minor) 'S' (serious) 'N' (not injured) '' (not known).	There is a lot of benefit in being able to tell the role of the person and the level of injury they sustained. Ensuring that any identifying features (eg age) are confounding minimises privacy concerns.
ltsa_role	Each party to a crash (vehicle, cycle, ped) is given an 'LTSA role' to distinguish it from other parties. The main participant is assigned role one, secondary parties role 2, and so on. Ltsa_role is also used to connect people to the vehicles they were in. In the crash_cse_code table where the cause code is environmental, ltsa_role is set to 0 (zero).	We do need a randomly generated identifier to indicate which person we are referring to. Descriptions of roles need to be written so as not to assign fault.
pers_prot	Whether seat belt etc was worn. This applies to data 1980-87. Values 'W' (worn), 'K' (unknown), 'A' (not available), 'N' (not worn). Now applies post-2001 also with values 'Y' (worn), 'N' (not worn), 'O' (not available), 'U' (Unknown) and ''. Old pre-87 data may have its values changed to the new ones.	Not personally identifying.
pers_role	The part a person played in the crash. Values: 'P' (passenger), 'D' (driver), 'W' (witness), 'O' (vehicle owner), '' (other incl. cyclist, ped, equest., skateboarder etc)	Limit to passenger, driver, cyclist; pedestrian, other. Witness and vehicle owner to be removed as witness information will not be included in the open data and whether or not the driver was the owner of the vehicle will be indicated elsewhere
pers_sex	The gender of a person involved in a crash	Combined with the other data in the dataset this will increase potential to identify an individual
pers_type	The type of person that we have	This is the road user type.

	data on. Values 'D' (driver, not including cyclist etc), 'P' (passenger), 'K' (skateboarder), 'E' (pedestrian), 'S' (cyclist), 'Q' (equestrian), 'O' (other), 'H' (wheeled ped.), 'W' (parked car owner)	Need to consider this variable along with at-fault, eg. if this and at fault are released, and a person is identified in the dataset you could establish they were the driver and at-fault. If the person was identifiable and movement codes are released, it could allow you to identify their actions (eg. distracted). Note, this is most likely to occur through media reports of fatal crashes. Within the data demographic variables will be grouped to prevent instant recognition. If releasing this, remove 'parked car owner' as we won't be providing information related to ownership/information not directly related to understanding the crash.
pers_cas_type	A derived flag used for reporting. For a person this is the same as the pers_type unless they are a 'D', 'P' or 'W' in which case it is the veh_type they were in.	Need to confirm, but understand this is the type of person that was casualty. If this is so, same considerations as casualty hierarchy apply.

Vehicle

VARIABLE	DESCRIPTION	ASSESSMENT
ltsa_role	Each party to a crash (vehicle, cycle, ped) is given an 'LTSA role' to distinguish it from other parties. The main participant is assigned role one, secondary parties role 2, and so on. Ltsa_role is also used to connect people to the vehicles they were in. In the crash_cse_code table where the cause code is environmental, ltsa_role is set to 0 (zero).	We do need a randomly generated identifier to indicate which person we are referring to. Descriptions of roles need to be written so as not to assign fault.
dvr_lic_stus	The drivers licence status. L (Learner), R (Restricted), F (Full), N (Never licenced), D (Disqualified), O (Overseas), K (unKnown)	Not considered to be identifying even when considered against other demographic variables being released. However, so as not to provide information about a person's compliance, Never Licenced, Disqualified,

		and unknown should be grouped into an 'other' category.
park_revr	A flag indicating whether the vehicle was parked or reversing. Values 'P' (parked), 'R' (reversing), 'S' (stationary), ' ' (neither or unknown).	Not personal information, and no information about fault is being included.
towing	Was the vehicle towing, and if so, what? Values 'B' (boat), 'C' (caravan), 'T' (trailer), 'S' (semitrailer), 'A' (A train), 'R' (B train), 'O' (other), ' ' (unknown)	Not personal information. Unlikely to be identifying.
veh_cc_rating	The engine capacity in cc.	Not personally identifying
veh_dam_locn	The location of the damage to a vehicle. Values : 'F' (front), 'R' (RHS), 'L' (LHS), 'B' (back), 'T' (top), 'G' (general), 'C' (concertina), 'U' (unknown), '1' (R front corner), '2' (R centre), '3' (R rear corner), '4' (L rear corner), '5' (L centre), '6' (L front corner)	Not personally identifying
veh_dam_sev	A code indicating the severity of damage to a vehicle. Values: 'N' (nil), 'M' (minor), 'E' (extensive), 'F' (fire), 'O' (overturned), 'W' (writeoff), 'U' unknown.	Not personally identifying.
veh_fault	A Y/N flag derived from the cause codes for this crash which indicates that this vehicle was faulty. Used in the cross-tab area.	Non-personal information - about the vehicle.
veh_manuf_year	The year of manufacture of the vehicle.	Essential for understanding the impact of older vehicles (safe vehicles). As make/model is not being included, this should not be identifying.
veh_own	The relationship of the driver to the vehicle. Values: 'O' (driver is owner), 'N' (vehicle is borrowed etc), 'R' (rental), " (unknown)	Not personally identifying.
veh_type	The type of the vehicle - C (car, station wagon), X (taxi), T (truck), B (bus), M (motorcycle), R (rental), V (van or ute), A (artic. truck), L (school bus), P (power cycle), U (unknown), O (other), S (bike), H (wheeled pedestrian), 4 (4WD)	Not personally identifying.
pass_front	The number of front seat passengers (note: this data is not reliable - although it should be front seat passengers the Police will sometimes record 1 meaning	Non-personal information, but need to note that the driver may sometimes counted.

	there was a driver, when they should have recorded 0)	
pass_other	Passengers in the vehicle who were not in the front or rear seat.	As per above.
pass_rear	The number of rear seat passengers.	As per above.
crash_spd	The speed of a vehicle before the crash.	This is the speed the person said they were going. This is not personal information, but it should be noted in the metadata that this is not necessarily accurately measured.

Objects

All information about objects struck is already openly available.

Factors

The following factors will be released at the highest level (ie. factor_grp_id):

All road user factors:

- Roll the following up to factor_grp_id 100 **Alcohol or drugs**, where these are proven (ie codes 103, 109, 220 and 221 below only)

Alcohol

- 101 Alcohol suspected
- 102 Alcohol test below limit
- 103 Alcohol test above limit or test refused
- 105 Impaired non-driver (pedestrian /cyclist/passenger, etc)
- 100 Other alcohol

Drugs

- 108 Drugs suspected
- 109 Drugs present
- 220 Impaired non-driver (pedestrian / cyclist / passenger, etc)
- 221 Other drugs

- Roll the following up to factor_grp_id 380 **Misjudged speed**

Misjudged speed, distance, size or position of

- 381 Another vehicle
- 383 Pedestrian
- 385 Size or position of fixed object or obstacle
- 386 Own vehicle
- 387 Misjudged intentions of another party
- 380 Other - misjudged speed, distance, size or position

- Roll the following up to factor_grp_id 500 **Illness and disability**

Released under the Official Information Act 1982

Illness

501	Sudden illness
504	Medical illness
505	Mental illness
506	Attempted suicidal
500	Other illness

Disability

502	Physical impairment
503	Defective vision
507	Impaired ability due to old age
508	Other disability

- Roll the following up to factor_grp_id 520 **Driver/Passenger**

Driver or passenger boarding, leaving or in vehicle

521	Intentionally leaving / boarding moving vehicle
523	Riding in insecure position
524	Interfered with driver
525	Opened door inadvertently
527	Child playing in parked vehicle
520	Other driver or passenger boarding, leaving or in vehicle

Driver only factors:

- Roll the following up to factor_grp_id 110 **Too fast for Conditions**

Inappropriate speed

111	Entering/on curve
112	On straight
113	Approaching a traffic control
115	When passing school bus
116	At temporary speed limit
117	At crash or emergency
118	For road conditions
119	For weather conditions
182	Travelling unreasonably slowly
110	Other inappropriate speed conditions

- Roll the following up to factor_grp_id 120 **Failed to keep left**

Position on road

121	Swung wide on bend
122	Swung wide at intersection
123	Cutting corner on bend
124	Cutting corner at intersection

125	Too far right
126	Vehicle crossed flush median
129	Too far left
120	Other positions on road

- Roll the following up to factor_grp_id 130 **Lost Control**

Lost control

131	Lost control when turning
132	Lost control under braking
133	Lost control under acceleration
134	Lost control while returning to seal from unsealed shoulder
135	Lost control - road conditions
136	Lost control - vehicle fault
137	Lost control avoiding another party
130	Other lost control

- Roll the following up to factor_grp_id 140 **Failed to Signal in Time**

Appropriate signalling

141	Failed to signal in time
145	Incorrect signal
140	Other failed to signal

- Roll the following up to factor_grp_id 150 **Overtaking**

Overtaking

151	Overtaking line of traffic or queue
152	Overtaking in the face of oncoming traffic
156	With insufficient visibility
157	Overtaking at an intersection
158	On left without due care
159	Cut in after overtaking
160	Vehicle signalling turn
150	Other overtaking

- Roll the following up to factor_grp_id 170 **Wrong Lane/Turned From Wrong Position**

Wrong lane or turned from wrong position

171	Turned from incorrect lane
173	Travelled straight from turning lane or flush median
174	Turned from incorrect position on road
176	Turned into incorrect lane
177	Weaving or cut in on multi-lane roads
179	Long vehicle tracked outside lane
184	Incorrect merging / diverging

170 Other wrong lane or position

- Roll the following up to factor_grp_id: 180 **Following too close**

Following too close

181 Following too closely
183 Motorist crowded cyclist
180 Other following too close

- Roll the following up to factor_grp_id: 190 **Sudden action**

Sudden action

191 Suddenly braked
192 Suddenly turned left/right
194 Served to avoid pedestrian
195 Served to avoid animal
196 Served to avoid crash or broken down vehicle
197 Swerved to avoid vehicle
199 Swerved avoiding emergency Vehicle
190 Other sudden action

- Roll the following up to factor_grp_id 200 **Forbidden movements**

Forbidden movements

201 Wrong way on road / motorway
202 Non-compliance with regulatory device with sign or marking
204 Driving / riding in pedestrian space
208 Motor vehicle in special purpose lane
533 Equestrian not keeping to verge
200 Other forbidden movements

- Roll the following up to factor_grp_id 320 **Did not Stop**

Did not stop

321 At stop sign
322 At full red traffic signal
324 At amber traffic signal
326 At flashing red signals (railway crossing, fire stations, etc)
327 For traffic controller
328 For school patrol / kea crossing
320 Other did not stop

- Roll the following up to factor_grp_id 300 **Failed to give way**

Failed to give way

301	At a priority traffic control
303	When turning to non-turning traffic
304	When priority defined by road markings
306	To a pedestrian
308	When entering roadway from driveway
309	To traffic approaching or crossing from the right
312	Entering roadway not from driveway or intersection
313	Failed to give way to emergency vehicle
314	Driver waved through
315	When turning right to opposing left turning traffic
316	To traffic approaching or crossing from the left
300	Other failed to give way

- Roll the following up to factor_grp_id 350 **Attention diverted**

Attention diverted by

Inside vehicle

351	Passengers
354	Animal or insect in vehicle
357	Emotionally upset / road rage
358	Food, cigarettes, beverages
359	Cell phone
361	Navigation device
362	Non cell communication device
364	Vehicle console inbuilt features: radio / heater, etc
365	Objects under driver's pedals
366	Food, cigarettes, beverages

Outside vehicle

352	Scenery or persons outside vehicle
353	Other traffic
355	Trying to find intersection, house number, destination, etc
356	Advertising or signs
363	Driver dazzled
350	Other attention diverted by

- Roll the following up to factor_grp_id 330 **Inattentive: Failed to Notice**

Failed to notice

331	Vehicle slowing, stopping or stationary in front
332	Bend in road
333	Indication of vehicle in front
334	Failed to notice control
336	Failed to notice signs
339	Failed to notice road works
340	Failed to notice markings
341	Obstructions on roadway

534 Another party wearing dark clothing
330 Other inattentive

- Roll the following up to factor_grp_id 370 **Did Not See or Look for Another Party Until Too Late**

Did not see or look for other parties until too late

371 Did not check / notice another party behind
375 Did not check / notice another party
377 When visibility obstructed by other traffic
370 Other did not see or look

- Roll the following up to factor_grp_id: 400 **Inexperience**

Lack of experience

401 In driving in fast complex or heavy traffic
402 New driver / under instruction
403 Driving unfamiliar with vehicle/towing
404 Overseas / migrant driver fails to adjust to NZ road rules and road conditions
407 Driver over-reacted
400 Other lack of experience

- Roll the following up to factor_grp_id: 410 **Fatigue (Drowsy)**

Fatigue (drowsy, tired or fell asleep)

411 Long trip
412 Lack of sleep
414 Long day (working / recreation)
415 Exceeded driving hours
410 Other fatigue

- Roll the following up to factor_grp_id: 420 **Incorrect use of vehicle controls**

Vehicle control mistakes

421 Started in gear / stalled
423 Wrong pedal / foot slipped
426 Lights not switched on
428 Parking brake not fully applied
429 Trailer coupling or safety chain not secured
420 Other vehicle controls

- Roll the following up to factor_grp_id: 440 **Parked or Stopped**

Parking

441 Parked vehicle is not visible

443	Incorrectly parked vehicle
447	Not clear of rail crossing
440	Other parking

- Roll the following up to factor_grp_id: 510 **Intentional or Criminal**

Intentional action

Showing off

431	Racing
432	Playing 'chicken'
433	Wheel spins / wheelies / doughnuts / drifting etc
434	Intimidating driving

Intentional

430	Other intentional actions
-----	---------------------------

Intentional or criminal

511	Homicide/suicide (successful)
512	Intentional collision
514	Evading enforcement
515	Object thrown (at / by / from)
518	Over the speed limit

Pedestrian factors:

- Roll the following up to factor_grp_id: 700 **Walking along Road**

Walking along road

701	Not keeping to footpath
702	Not keeping to side of road
703	Not facing oncoming traffic
704	Not on outside of blind curve
705	Wheeled pedestrian behaviour
700	Other pedestrian walking along the road

- Roll the following up to factor_grp_id: 710 **Crossing Road**

Crossing road

710	Other pedestrian crossing the road
711	Walking heedless of traffic
712	Stepping out from behind vehicles
713	Running heedless of traffic
714	Failed to use pedestrian crossing when one within 20 metres
715	Waiting on carriageway / confused by traffic
717	Stepping suddenly onto crossing
718	Not complying with traffic signals or school patrols

- 719 Misjudged speed and/or distance of vehicle
- 740 Looking the wrong way

- Roll the following up to factor_grp_id: 720 **Miscellaneous**

Miscellaneous

- 721 Pushing, working on or unloading vehicle
- 722 Playing / unnecessarily on road
- 723 Working on road
- 725 Vision obscured by umbrella or hood
- 726 Child escaped from supervision
- 727 Unsupervised child
- 729 Pedestrian from or to school bus
- 730 Pedestrian behind reversing / manoeuvring vehicle
- 731 Overseas pedestrian
- 732 Pedestrian attention diverted by cigarette, cell phone, music player
- 733 Pedestrian from or to scheduled service
- 720 Other pedestrian

Vehicle factors:

- Roll the following up to factor_grp_id: 600 **Lights and Reflectors at Fault or Dirty**

Running lights

- 601 Dazzling headlights
- 602 Headlights inadequate / no headlights or failed suddenly
- 604 Brake lights or indicators faulty or not fitted
- 605 Tail lights inadequate or no tail lights
- 606 Reflectors inadequate or no reflectors
- 607 Lights or reflectors obscured
- 608 Confusing / dazzling lights
- 609 Lights or reflectors at fault or dirty
- 600 Other lights or reflectors

- Roll the following up to factor_grp_id: 610 **brakes**

Brakes

- 611 Parking brake failed / defective
- 613 Service brake failed
- 614 Service brake defective
- 615 Jack-knifed – uneven braking
- 610 Other brakes

- Roll the following up to factor_grp_id: 620 **Steering**

Steering

621	Defective
622	Failed suddenly
620	Other steering

- Roll the following up to factor_grp_id: 630 **Tyres**

Tyres

631	Puncture or blowout
632	Worn tread on tyre
633	Incorrect tyre type
634	Mixed types (tread) / space savers
630	Other tyres

- Roll the following up to factor_grp_id: 640 **Windscreen or Mirror**

Windscreens, mirrors, visors

641	Shattered windscreen
642	Vehicle windows / helmet visors / goggles / glasses / misted / dirty / windscreen wipers
643	Rear vision mirror
640	Other windscreen / mirror

- Roll the following up to factor_grp_id: 650 **Mechanical**

Mechanical

651	Engine failure
652	Transmission failure /broken axle
653	Accelerator or throttle jammed
650	Other mechanical

- Roll the following up to factor_grp_id: 660 **Body or Chassis**

Chassis / running gear

661	Body, chassis or frame (cycle / motorcycle) failure
662	Suspension failure
665	Inadequate tow coupling
666	Inadequate or no safety chain
668	Wheel off
660	Other chassis / gear

Body / doors

667	Door / bonnet catch failed, defective or not shut
670	Inconspicuous colour
671	Blind spot
664	Other body / doors

- Roll the following up to factor_grp_id: 680 **Load**

Load

681	Load interferes with driver
682	Not well secured
683	Over-hanging
684	Load obscured vision
686	Over-dimensional vehicle or load
687	Load too heavy
688	Towed vehicle or trailer too heavy or incompatible
680	Other load

- Roll the following up to factor_grp_id: 690 **Miscellaneous Vehicle**

Miscellaneous vehicle

691	Emergency vehicle
692	Vehicle caught fire
693	Being towed
690	Other vehicle

Road factors:

- Roll the following up to factor_grp_id: 800 **Slippery**

Slippery road surface

804	Loose material on seal
807	Painted markings
808	Recently graded
809	Surface bleeding / defective
813	Deep loose metal
828	Steel / iron covers and joints
800	Other slippery road condition

Wet surfaces

801	Rain
802	Frost or ice
803	Snow or hail
805	Mud / effluent
806	Oil / fuel

- Roll the following up to factor_grp_id: 810 **Surface**

Surface condition

811	Potholed
812	Uneven
814	High crown

815	Curve not well banked
816	Edge badly defined or gave way
817	Under construction or maintenance
818	Unusually narrow
810	Other surface condition

- Roll the following up to factor_grp_id: 820 **Obstructed**

Obstructions and objects

821	Fallen tree or branch
822	Slip or subsidence
823	Flood waters, large puddles and fords
824	Road works not adequately lit/sign posted
826	Roadside object fell on vehicle
827	Object flicked by other vehicle
820	Other road obstructed

- Roll the following up to factor_grp_id: 830 **Visibility Limited**

Visibility limited

Visibility limited by road feature

831	Curve
832	Crest
837	Bank
849	Traffic signs

Visibility limited by other feature

833	Building
834	Trees
835	Hedge or fence
836	Scrub, long grass or foliage
838	Temporary obstruction, dust or smoke
839	Parked vehicle
829	Signs / billboards / hoardings
830	Other road feature limit visibility

- Roll the following up to factor_grp_id: 840 **Signs and Signals**

Signs and signals

841	Damaged, removed or malfunction
842	Badly located
843	Ineffective / inadequate / obscured
844	Necessary
845	Signals off
840	Other signs or signals

- Roll the following up to factor_grp_id: 850 **Markings**

Markings / Islands / Barriers

851	Faded
852	Difficult to see due to weather or geometry
853	Markings necessary
872	Traffic island(s) ineffective, badly located or designed
856	Barriers necessary
857	Island necessary
850	Other markings / islands / barriers

- Roll the following up to factor_grp_id: 860 **Street Lighting**

Street lighting

861	Failed
862	Inadequate for road and pedestrian crossing
860	Other street lighting factors

Environmental factors:

- Roll the following up to factor_grp_id: 900 **Weather**

Weather

901	Heavy rain
902	Dazzling sun
903	Strong wind
904	Fog or mist
905	Snow, sleet or hail
900	Other weather

- Roll the following up to factor_grp_id: 910 **Animals**

Animals

912	Household pets rushed out or playing
913	Farm animals straying
914	Farm animals attended, but inadequate warning or unexpected
915	Farm animals attended, but out of control
910	Wild Animal
911	Other animal factors

- Roll the following up to factor_grp_id: 999 **Unknown**

NO IDENTIFIABLE FACTORS

999	Unknown
-----	---------

Appendix 2 – Data assessed as NOT suitable for open release at the level of individual crash

Crash

VARIABLE	DESCRIPTION	ASSESSMENT
crash_id	The unique identifier for a crash	Unique identifiers are not to be openly released. NB. a random number may be assigned to each crash if we need to reshape the data into multiple tables. This will be randomly generated and will not enable any linkages with the source system.
crash_cause	A code for the overall cause of a crash. Values 'A' (alcohol), 'S' (speed), 'B' (alcohol & speed), 'N' (neither alcohol nor speed). Relates to cause codes 100,101,103-109 for alcohol, and 110-119 for speed.	Multiple factors may contribute to a crash. These will be provided as per the factors assessment above. As such, one 'overall' cause will not be provided.
crash_date	The date on which a crash occurred. The date or the least significant parts of the date may be unknown. Format is YYYYMMDD	Exact date will be excluded as a confounding factor to prevent re-identification by matching with external sources (eg media). An indicator of whether the crash was on a weekday/weekend will be included. Year, financial year, and whether the crash occurred over a holiday period are already available in the open data.
crash_dow	The day-of-week on which the crash occurred. Possible values range from '1' (Sunday) to '7' (Saturday). ' ' is invalid or unknown.	As per above. The weekend indicator will be included however. This will allow the identification of whether this crash was on a weekday or in the weekend.
crash_fin	Another crash severity flag, derived from crash_sev. Values 'F' (fatal, same as sev. 'F'), 'I' (injury, same as sev. 'S' and 'M'), 'N' (non-injury, same as sev. 'N' or '')	While there is no reason not to release this variable, it contains the same information as crash_sev, which is already openly available.
crash_in	Another crash severity flag, derived from crash_sev. Values 'I' (injury, same as sev. 'F', 'S', 'M'), 'N' (non-injury, same as sev 'N' or '')	As above.
crash_pol_sev	The only non-derived severity flag. Comes off the police form or INCIS and is entered. Used for cross-checking with the derived flags. Has the same values as crash_fin.	To be used internally for validating the derived flags. No value add in release.
crash_road_id	The road ID of the road segment on which the crash is located. This may not be the road ID of the road whose name is in crash_locn1 or even crash_locn2 -	As we already provide a variety of location variables, including eastings and northings, there is no benefit in including this variable. In

	eg if the crash is in the middle of a roundabout segment it will be the ID of the highest priority road which includes that segment - perhaps a SH.	addition, this variable requires expert knowledge to interpret.
crash_time	The time-of-day of the crash. 24hr clock hhmm. The exact time may be unknown.	Not being included to prevent re-identification with external sources (eg media). A derived variable providing the crash time in a 4 hour block will be released however.
crash_time_6hr	A field derived from crash_time. It is the time of day in 6hr blocks. Eg value 1 is times 0000-0559; value 2 is times 0600-1159 etc. Value 0 is unknown time.	Releasing this variable alongside crash_time_4hr would be reducing the time window to a 2hr period, which would greatly increase the potential for re-identification. The benefit in providing this slightly different time window is not considered to outway the reidentification/privacy risk in this.
locality_name	The name of the locality where the crash occurred. This is 'free' data not limited by lookup.	There is no privacy concern in releasing this variable, but there are quality considerations in terms of it being a free-text field. Given the other location information being released, there is no additional benefit in releasing this variable.
nzta_region_id	The unique identifier of an NZTA region. Values 'A' (Auckland), 'H' (Waikato/BoP), 'J' (Central), 'S' (Southern). TLAs do not map exactly onto a single NZTA region in every case, but TLAS 1-7 are in 'A', 11-27 are in 'H', 28-53 are in 'J', 54-75 are in 'S'. This attribute is blank (' ') where it is used in lu_tla and the boundaries do not match.	Not relevant for general users outside of the Agency. No additional benefit above the other geographical variables already being released.
off_att_date	The date the police attended the crash.	Exact date of crash not provided so no benefit in providing this. If we were to include it we would need to liaise with Police in terms of the information about them this is providing.
off_att_time	The time when the police attended the crash.	As per above.
pol_area_id	The unique id of a police area.	Not relevant for general users. No additional benefit above the other geographical variables already being released.
pol_dstr_id	The unique identifier of a police district. Current values CJ Auckland, DA Counties/Manukau,DC Bay of Plenty, DE Eastern, DH Wellington, DJ Tasman, DK Canterbury, DM Southern, DG Central, CX Nth Shore/Waitakere, DB Waikato, BG Northland.	Not relevant for general users. No additional benefit above the other geographical variables already being released.

z4_ak_mway	Used to indicate whether the crash falls in the area patrolled by the Police Auckland Motorway Patrol. It is used to split crashes out from the underlying Police Districts for detailed reporting to the Police. It can only be calculated after the crash is geocoded. Values: 'M' (Inside the Ak motorway patrol area), 'N' (Outside the Ak Mway patrol area), null (Yet to be calculated).	Not relevant for general users. No additional benefit above the other geographical variables already being released.
z1	Used for a Y/N flag to indicate that for this crash, the crash_locn2 field contains the name of a geocoding feature, and not the usual sideroad name.	No privacy concerns, but Colin Morrison advised no benefit in releasing this, as it is an internal systems checking variable.
qual_stus	The overall quality of the data for this crash. Has the same value as the worst individual error for this crash. Values 'F' (fatal - unusable) 'S' (serious) 'M' (minor)	No privacy concerns, but Colin Morrison advised not to release this.
pol_rptd	A Y/N flag indicating whether a crash was police reported or not (if Y, also indicates that the person who entered the crash was an LTSA user)	This variable is used to filter the data for release (we will only publish police reported crashes). This variable therefore does not need to be included in the output as they will all be 'yes'.
crash_cost	The social cost of this crash calculated by the Transit algorithm. It is dependent on TLA, urban (based on speed limit) and crash_sev.	This is calculated as an average, and is therefore not appropriate to be provided at the level of individual crash.

Person

VARIABLE	DESCRIPTION	ASSESSMENT
crash_id	The unique identifier for a crash	Unique identifiers are not to be openly released. NB. a random number may be assigned to each crash if we need to reshape the data into multiple tables. This will be randomly generated and will not enable any linkages with the source system.
pers_id	The unique id of a person in a crash. Note this is unique for the crash, not globally. In the crash_cse_code table where the cause code is either a vehicle or environmental code, the pers_id is set to 0 (zero).	Unique identifiers are not to be openly released. NB. a random number may be assigned to each person if we need to reshape the data into multiple tables. This will be randomly generated and will not enable any linkages with the source system. Only person level information cleared for release as per this document will be made available.
age_5yr	A field derived from 'pers_age'. It is the	Information about age is considered

	age in 5yr blocks: the value 20 includes ages 20-24, the value 25 is ages 25-29 etc	highly valuable, but is also potentially sensitive (for example, if there was a crash where alcohol is considered to be a cause and a young child was killed; and also in terms of assisting in identifying people involved). As such, 5 year age groups are considered too sensitive for release. However, we will derive a variable with broader age groups as per above.
pers_adr1 - pers_adr3	Address lines	Personal information.
pers_age	The age of a person involved in a crash.	Personal information – a derived variable ‘age group’ will be provided.
pers_au	The area unit in which a person involved in a crash lives.	Potentially identifying information as area units are relatively small, and when considered with the other information contained in the dataset it would be possible to further limit the potential population of interest. (Area units normally contain 3,000 – 5,000 population though this can vary due to such things as industrial areas, port areas, rural areas and so on).
pers_dob	The date of birth of a person involved in a crash. This will generally be known only for drivers. YYYYMMDD.	Personal information.
pers_hosp	A flag indicating whether a person was hospitalized. Applies to data 1980-87. Values 'Y' (yes, hospitalized), 'N' (you can guess the meaning of this one), 'U' (unknown). May also come out of INCIS.	Open data is from 1 Jan 2000. This variable is not available for that time period.
pers_hospital	The name of the hospital to which this person was sent.	Potentially personally identifying/sensitive.
pers_initials	The initials of a person involved in a crash	Personally identifying.
pers_mb	The meshblock in which a person involved in a crash lives.	Personally identifying – Meshblocks are the smallest geographic unit, and vary in size from part of a city block to large areas of urban land. Releasing this could be very close to releasing the persons address.
pers_seat_posn	The person's seating position in the vehicle. Values 'F' (front), 'R' (rear), 'O' (other), ' ' (not known). Applies to data 1980-87.	Does not apply to time period of the open data. Subject to quality, we will provide information on the number of front seat and rear seat passengers, and how many passengers were not in the front or rear seat.
pers_surname	The surname of a person involved in a crash	Personal information.

Vehicle

VARIABLE	DESCRIPTION	ASSESSMENT
crash_id	The unique identifier for a crash	Unique identifiers are not to be openly released. NB. a random number may be assigned to each crash if we need to reshape the data into multiple tables. This will be randomly generated and will not enable any linkages with the source system.
alc_blood_ref	A flag indicating the status of the blood test if one was requested. Values " (not requested), 'R' (refused - a positive result), 'T' (taken - if the actual result is known it will be in alc_blood_test).	Administrative and sensitive information.
alc_blood_test	If a blood alcohol test was taken, this is the result if we know it.	Sensitive information.
alc_evid_ref	The status of the evidential breath test if one was requested. Values " (not requested), 'R' (refused - a positive result), 'T' (taken but result maybe unknown)	Sensitive information.
alc_evid_test	Evidential breath test result if known.	Sensitive information.
alc_scrn_test	If a screening test for alcohol was used, was it positive or negative? Values: '+' (positive), '-' (negative), ' ' (not tested)	Sensitive information.
alc_susp	Whether alcohol involvement was suspected. Values: 'S' (suspected), 'N' (not susp.), 'U' (unknown), 'R' (breath level 250-400)	Sensitive and supposition.
dvr_culp	A code indicating driver culpability for this vehicle. Values '1' (1 veh, dvr at fault) '2' (1 veh, no dvr fault), '3' (>1 veh, dvr prime fault), '4' (>1 veh, dvr part fault), '5' (>1 veh, no dvr fault).	Sensitive when considered with other information (particularly demographics) being released, as it assigns fault to an individual.
dvr_frgn_cntry	The country of origin of a driver using an overseas licence.	Potentially personally identifying when considered with other information being released. Whether the driver was using an overseas

		licence is provided in the dvr_lic_stus variable.
dvr_occ	The occupation of the driver of this vehicle.	Personally identifying.
pers_id	The unique id of a person in a crash. Note this is unique for the crash, not globally. In the crash_cse_code table where the cause_code is either a vehicle or environmental code, the pers_id is set to 0 (zero).	Unique identifiers are not to be openly released. NB. a random number may be assigned to each person if we need to reshape the data into multiple tables. This will be randomly generated and will not enable any linkages with the source system. Only person level information cleared for release as per this document will be made available.
veh_make_model	The make and model of the vehicle.	Combined with the other information, could be potentially identifying, especially in the case of rare vehicles (eg. make/model for a rare vehicle, combined with location of damage to the vehicle could identify the vehicle in the crash, and therefore imply the owners involvement).
veh_spd_contrib	A flag to indicate whether this vehicle's speed contributed to the crash. Values are 'Y' for yes, 'N' (guess!) and '' for unknown.	Attributes potential fault to an individual. Should not openly release information that is a 'guess'!
veh_wof_current	A flag indicating whether the vehicle WOF or COF was current.	If a person was able to be identified, this would provide information about their compliance behaviour.
veh_wof_expy_date	The expiry date of the WOF or COF of the vehicle.	As per above.

Causes

VARIABLE	DESCRIPTION	ASSESSMENT
crash_id	The unique identifier for a crash	Unique identifiers are not to be openly released. NB. a random number may be assigned to each crash if we need to reshape the data into multiple tables.

		This will be randomly generated and will not enable any linkages with the source system.
ltsa_role	Each party to a crash (vehicle, cycle, ped) is given an 'LTSA role' to distinguish it from other parties. The main participant is assigned role one, secondary parties role 2, and so on. Ltsa_role is also used to connect people to the vehicles they were in. In the crash_cse_code table where the cause code is environmental, ltsa_role is set to 0 (zero).	As CAS data relates to factors PROBABLY contributing to crashes, we do not want to provide any information that implies a particular individual was at fault, when there is even the smallest possibility that individual could be identified in the data. Any information on factors or causes will be assigned at the crash level.
pers_id	The unique id of a person in a crash. Note this is unique for the crash, not globally. In the crash_cse_code table where the cause code is either a vehicle or environmental code, the pers_id is set to 0 (zero).	As per above.

Factors

- The following factors relating to reason for death/injury are **not to be released**

Reason for death/injury

- 531 Casualty drowned
- 532 Casualty thrown from vehicle
- 535 Electrocution
- 536 Unsecured child seat
- 537 Child restrained failure/inappropriate
- 672 Seatbelt failed / defective
- 673 Air bag failed / defective
- 530 Other reasons for death/injury