Minimum datasets for pedestrian safety

2025 Australasian Road Safety Conference
20 - 23 OCT - PERTH, WESTERN AUSTRALIA

Dr Glen Koorey
Director, Principal Transportation Engineer
ViaStrada Ltd, Christchurch, New Zealand

VIASTRADA
TRANSPORT PLANNING AND DESIGN

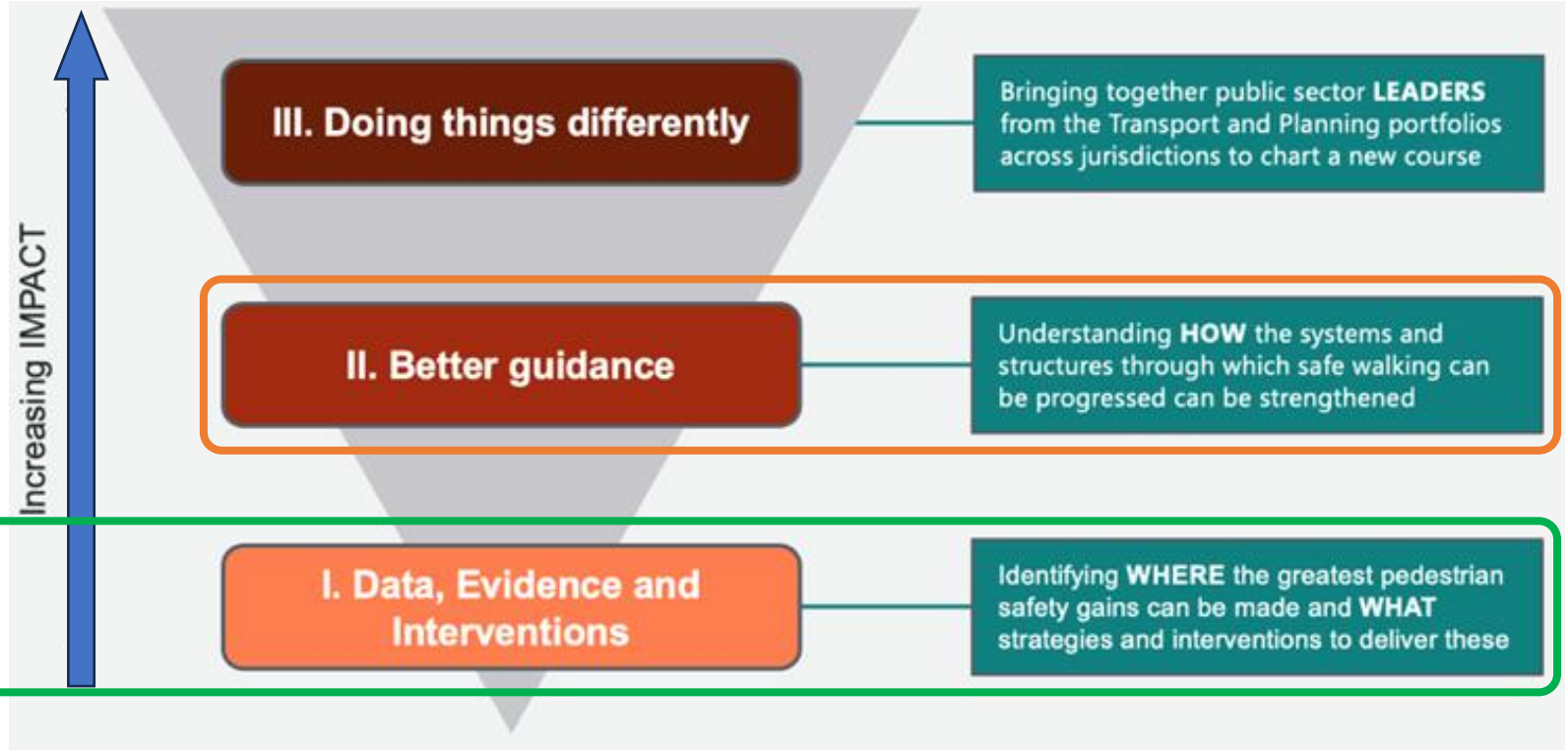
Photo: Syauqi Muhammad

Keeping People Safe When Walking

- Ongoing Austroads research project
  - **Stream 1** already completed, **Stream 2** now underway

Austroads

*Purpose: To help chart a path to **zero pedestrian serious trauma** by establishing a greater level of understanding and insights into pedestrian safety risk, trauma, and cost-effective countermeasures.*



abley HEA CONSULTING

Stream 2: Work Package 1

Our objectives

*Define the **minimum pedestrian data requirements** to progress pedestrian safety*

- That data could include:
  - Trip usage/purpose data
  - Demographic data
  - Safety/Injury data
  - Asset/Facility data
  - Path user type data

Key outputs

- **Research Report**
  - Identify available **ped'n safety data sources**
  - Including **literature review**
  - **Survey & interview** results
- **Data Framework** tool
  - **Spreadsheet** of data sets & relevant factors
  - Sources of data **most relevant** to ped'n safety
- **Practice Guideline**
  - **Practical and easy-to-read** document
  - Guidance on **types** of pedestrian safety data

Literature review

Report 2:
Literature review

Report 3:
Data recommendations

STREAM 1:
(Alavi et al 2025)

68 references: pedestrian safety (general)

9 references: pedestrian safety data (theoretical)

1 practical application (not in reference list)

STREAM 2:
Min pedestrian data requirements research report

Expanded theoretical pedestrian safety literature review:
30 references

Sources from
- *Australia*
- *New Zealand*
- *International*

Expanded strategic / practical applications literature review:
61 references

Industry Survey

Built off previous Stream 1 survey – digging deeper…

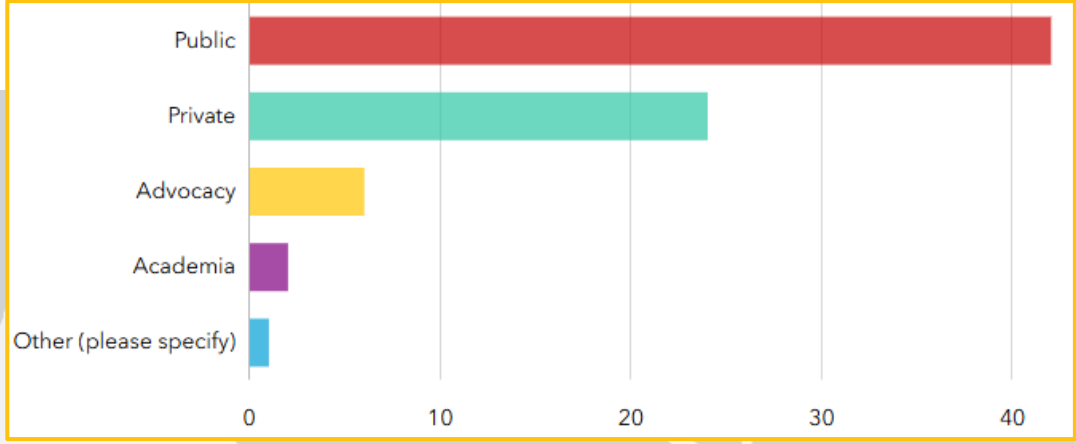
- Questions about professionals' data use in 4 key categories:
  - Vehicle Volumes
  - Pedestrian Volumes
  - Injuries and Crashes
  - Road/Street and Path Infrastructure & Environment
- Is the data:
  - existing and easily accessible? Who owns it?
  - of high quality, accuracy, and spatial coverage?
  - missing anything important for pedestrian safety?

Survey

Who responded?

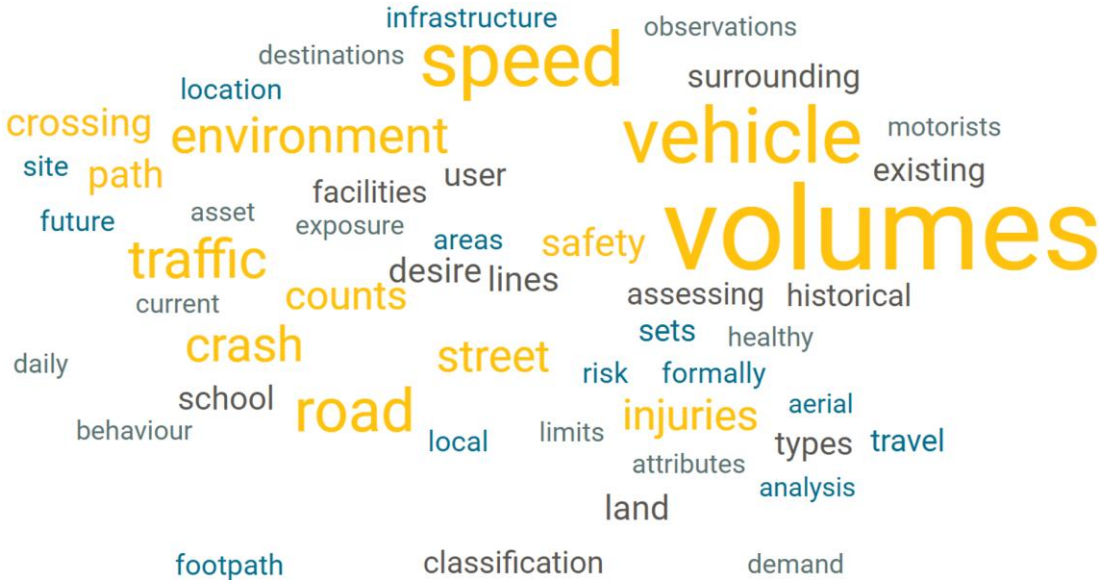
Online survey tool

- **70** total respondents
  - 42 from NZ
  - 28 from AUS
- All states (except NT)
- All cities over 500k
- Some medium cities

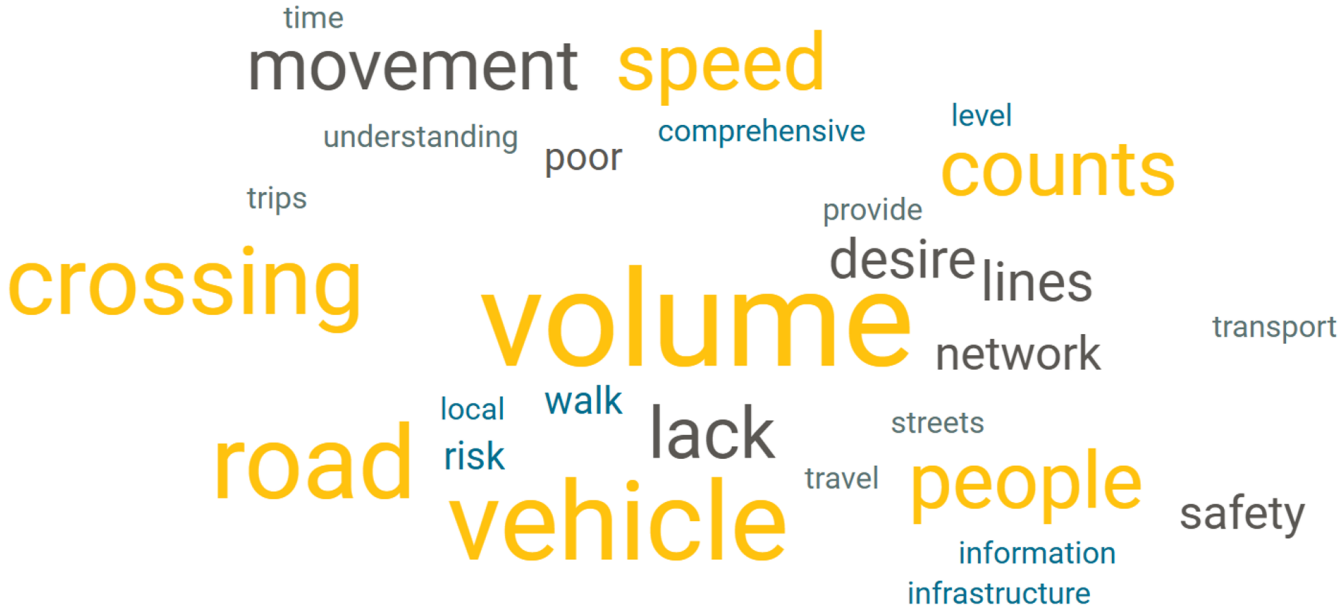


Survey findings: Important datasets

*what are the THREE **most important** transport data sets you use or collect?*



*What is the most critical **gap** in the pedestrian-related transport data you interact with…?*

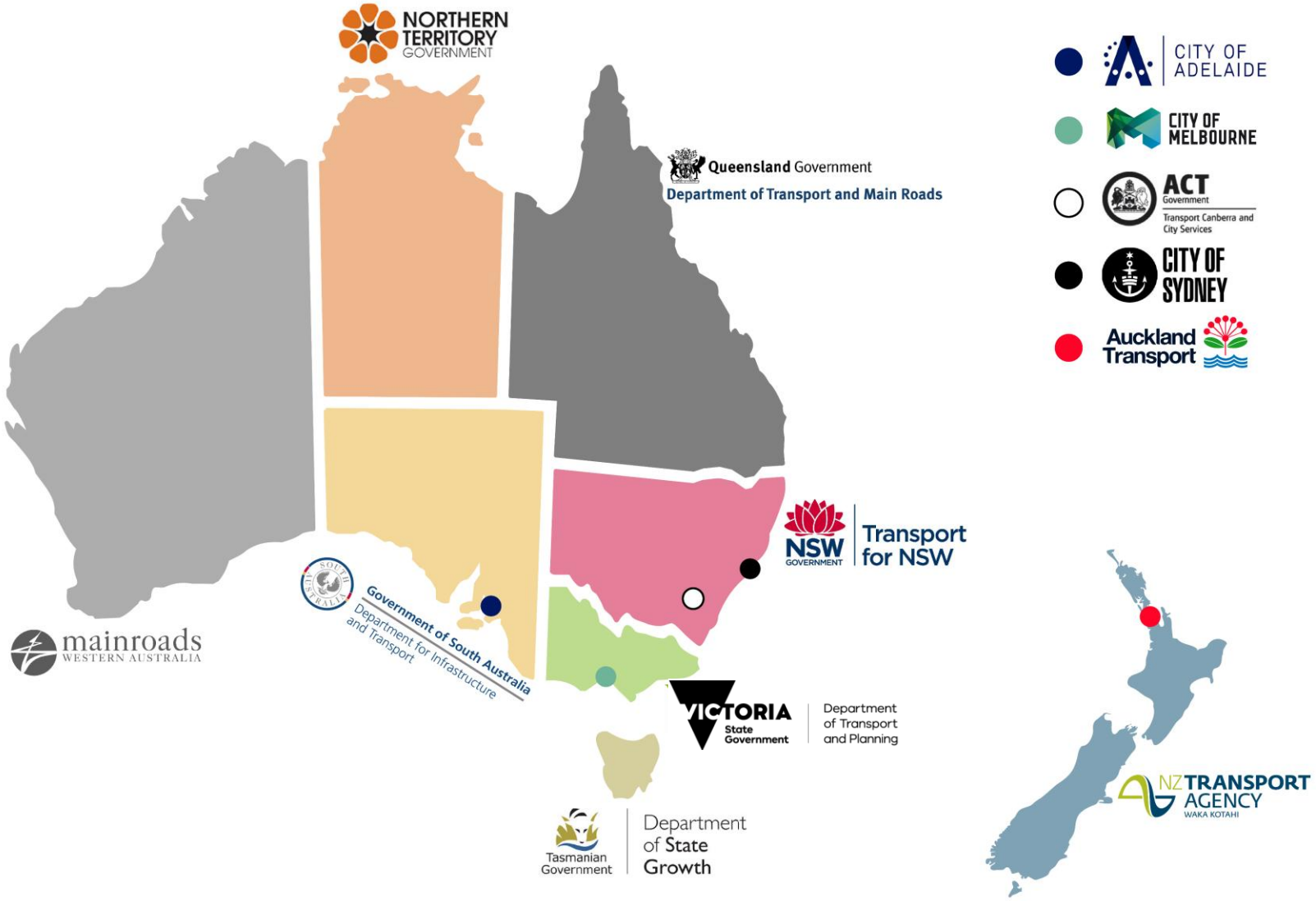


Other key thoughts and comments:

- ***All** data types are important for pedestrian safety and most are good*
- ***Pedestrian data is worse** than for other modes for all data types*
- ***Keep things simple** – too much guidance and data can be more hindrance than help*
- ***Just do something** – we know what works, we don't need more data, just **get on with it***

Jurisdiction Interviews – Who?

- **13** State, Territory or City organisations interviewed
- *Designed to be more **conversational***
  - *Allowing in-depth discussion/detail about specific topics and free expression on complex issues*



Interview findings
Common themes

- Site-specific data
  - Mostly managed by **local councils**
  - Important for identifying **high-risk pedestrian sites** though

- Crash data consistency
  - Widely utilised, consistent and easy to access – **reactive data** though
  - **Under-reporting** is a challenge – integration of hospital data could help

- Pedestrian crossing data and trip patterns
  - **Limited datasets** with path/crossing data – usually for specific projects
  - **Spatial coverage** is a challenge – use proxy measures instead?

Interview findings
Common themes cont'd

- Pedestrian asset and environmental information
  - Usually **locally held** but can be somewhat limited in detail
  - Much more detailed data sets exist for **road infrastructure**

- Technology
  - **AI, Crowd-sourcing, Mobile phones** could provide useful data
  - Currently some issues with **Cost**, **Coverage**, and **Privacy** concerns

- Proxy measures
  - Potential proxies could include: Hospital casualty data, Crossing types, Traffic volumes & speeds, Use of public transport stops

Determining min. data required: Criteria

Score	Availability / Accessibility	Quantity / Coverage	Database format	Data Quality	Contribution to addressing ped'n safety	Data Collection Difficulty/Cost
1	**Very low** - proprietary or restricted	**Very low** - Single location or aggregated to entire country	**Very low** - unable to be used in analysis	**Very low** - not relevant, some inaccuracies	**Very low** - little to no influence on ped'n safety investment & decisions	**Very high** - high financial or resource intensive cost to collect the data
2	**Low** - behind paywall or licensed	**Low** - neighbourhood or pilot studies	**Low** - needs conversion to be used in any analysis	**Low** - slightly relevant, possible inaccuracies	**Low** - some influence on ped'n safety investment & decisions	**High** – some financial or resource intensive cost to collect the data
3	**Acceptable** - some work required to access it	**Acceptable** - local area	**Acceptable** - can be used in analysis, no location information	**Acceptable** - relevant and accurate	**Moderate** - a fair amount of influence on ped'n safety investment & decisions	**Medium** - moderate financial or resource intensive cost to collect the data
4	**Medium** - public access	**Medium** - region or state	**Medium** - can be used in analysis, some location information	**Medium** - reliable, relevant and accurate	**High** - strong influence on ped'n safety investment & decisions	**Low** - minor financial or resource intensive cost to collect the data
5	**High** - public access and download	**High** – national	**High** - allows easy analysis (geospatial or similar)	**High** - detailed, reliable, relevant and accurate	**Very high** - very strong influence on ped'n safety investment & decisions	**Very low** - Data is gathered by existing sources

Final Research Report

Sections: *(100 pages)*

- Summary / Glossary
1. Introduction
2. Literature Review
3. Surveys and interviews
4. Discussion
5. Conclusions & Recommendations
- References
- Appendices (Survey/Interview questions)

Data Framework: Categories & Use levels

Human factors

Vehicle and technology

Physical environment

Injuries or crashes

Exposure or volumes

Use level	Strategic policy	Tactical	Intervention at local project / engineering
	Vision, objectives, policies and targets. E.g population data is a metric for exposure.	*Intervention type region or area-wide. E.g., specific travel data to assess risk factors.*	*Project planning & design. E.g., pedestrian crossing data to identify priority locations or design details.*

Data Framework tool: Factors/Databases

From modified Haddon Matrix, combined with Injury Causation Chain	*(Refer also to accompanying research report)*		*Some will be population or fleet level datasets, others will be location, incident or cross tabulations with other factors (e.g. crashes)*	*How accessible data is to practitioners & public, incl. whether data is openly available or requires specific access permissions*	*Extent of data coverage, whether it encompasses local, regional, or national scope and captures a wide range of relevant variables*	*How data is presented, with focus on usability & compatibility of the data for analysis purposes. Geospatial formats facilitate easier integration & analysis*	*Depth and reliability of the info provided, considering the accuracy, currency, and relevance of the data to ped'n safety concerns*	*Dataset or KPI / measure effectiveness in reducing DSIs*	*Level of financial or resource-intensive cost to collect the data*	*Overall score (1-5)*	*(all 6 factors)*	*Overall score (1-5)*	*(only 3 factors)*	*This column is preserved for now in order to move dataset examples if it is determined that it isn't really influential or relevant enough to the ped traffic safety needs*	*These data sets provide relatively limited information for the particular factor of interest, often only for a few locations or with just a few relevant measures*	*These data sets provide reasonably useful information for the particular factor of interest, probably covering either some parts of the area being investigated and/or providing at least some relevant measures*	*These data sets provid high-quality inform particular factor of i covering a full area a measur*
UID	**Factor Level 1 / Category**	**Factor Level 2**	**Factor Level 3 (if applicable)**	**Attributes**	**Availability / Accessibility**	**Quantity / Coverage**	**Database format**	**Data Quality**	**Contribution to addressing ped'n safety**	**Data Collection Difficulty/Cost**	**Assessment** *(fill all in to give accurate score)*	**Recommendations**	**Targeted Assessment**	**Recommendations**	**Not relevant / not minimum: examples of datasets**	**Acceptable: examples of datasets**	**Good: examples of datasets**	**Ideal: example**
H1	Human	Socioeconomic		population (numbers, density, mix) and projections	High - public access and download	High - national	High - allows easy analysis (geospatial or similar)	High - detailed, reliable, relevant and accurate	High - strong influence on ped'n safety investment & decisions	Very low - Data is gathered by existing sources	4.8	Use	4.7	Use		AU: Census (DB26), NZ: Census (DB08)		
H2	Human	Socioeconomic		demographics (e.g., age, gender, ethnicity and indigenous status, language)	High - public access and download	High - national	High - allows easy analysis (geospatial or similar)	High - detailed, reliable, relevant and accurate	Moderate - a fair amount of influence on ped'n safety investment &	Very low - Data is gathered by existing sources	4.7	Use	4.3	Use			AU: Census (DB26), NZ: Census	
H3	Human	Socioeconomic		socioeconomic attributes (education level, income level, deprivation index)	High - public access and download	High - national	High - allows easy analysis (geospatial or similar)	High - detailed, reliable, relevant and accurate	Moderate - a fair amount of influence on ped'n safety investment &	Very low - Data is gathered by existing sources	4.7	Use	4.3	Use				
H4	Human	Socioeconomic		social capital and community engagement	Medium - public access	Medium - region or state	Medium - can be used in analysis, some location information	Medium - reliable, relevant and accurate	Low - some influence on ped'n safety investment & decisions	Very low - Data is gathered by existing sources	3.8	Consider	3.7	Consider				
H5	Human	Health and wellbeing		health and wellness indicators (e.g., obesity rates, physical activity)	Medium - public access	Medium - region or state	Medium - can be used in analysis, some location information	Medium - reliable, relevant and accurate	Low - some influence on ped'n safety investment & decisions	Very low - Data is gathered by existing sources	3.8	Consider	3.7	Consider				
H6	Human	Health and wellbeing		cognitive and physical capabilities (e.g., vision, reaction time)	Medium - public access	High - national	Acceptable - can be used in analysis, no location information	Medium - reliable, relevant and accurate	Low - some influence on ped'n safety investment & decisions	Very low - Data is gathered by existing sources	3.8	Consider	3.7	Consider		AU: Census Health & Disability (DB14), NZ: Census (DB08)		
H7	Human	Travel behaviour		safety awareness and education enforcement (e.g., speed limits, red light cameras)	Medium - public access	Medium - region or state	Acceptable - can be used in analysis, no location information	Acceptable - relevant and accurate	Moderate - a fair amount of influence on ped'n safety investment &	Low - minor financial o resource intensive cos to collect the data	3.5	Consider	3.3	Consider			AU: Digital Atua of Australia - median heavy vehicle speed limits (DB58)	AU: Locations of spe cameras - Queenslar
H8	Human	Travel behaviour		road user perception and attitudes towards road safety	High - public access and download	Medium - region or state	Acceptable - can be used in analysis, no location information	Acceptable - relevant and accurate	High - strong influence on ped'n safety investment & decisions	Low - minor financial o resource intensive cos to collect the data	3.8	Consider	3.7	Consider		NZ: Road User Perception and Attitudes - Waka Kotahi - focusing on vehicular modes (DB09)	AU: ESRA survey data for all modes, including pedestrians - but only aggregated to national scale (DB89)	AU: Customer satisfa modes (NSW) (DB69)
H9	Human	Travel be		(e.g., risky be aviour,	High - public access and download	Very low - Single location or aggregated to entire country	Acceptable - can be used in analysis, no location information	Acceptable - relevant and accurate	Moderate - a fair amount of influence on ped'n safety investment & decisions	Medium - moderate financial or resource intensive cost to collect the data	3.0	Consider	3.0	Consider			AU: ESRA survey data for all modes, including pedestrians - but only aggregated to national scale (DB89)	
V1	Vehicle and technology			vehicle fleet composition, age /mass, and	Acceptable - some work required to access it	High - national	Medium - can be used in analysis, some location information	Medium - reliable, relevant and accurate	Moderate - a fair amount of influence on ped'n safety investment &	Low - minor financial o resource intensive cos to collect the data.	3.2	Consider	2.3	Don't use		AU: Registration data from NEVDIS (DB81) matched to ANCAP (DB79), NZ Motor vehicle registrations (DB38)		
V2	Vehicle and technology			technology penetration and standardisation	Acceptable - some work required to access it	High - national	Acceptable - can be used in analysis, no location information	Medium - reliable, relevant and accurate	Low - some influence on ped'n safety investment & decisions	Low - minor financial o resource intensive cos to collect the data.	2.8	Don't use	2.0	Don't use		Rightcar vehicle safety ratings (DB81)	AU: Registration data from NEVDIS (DB81) matched to AU: ANCAP (DB79)	
V3	Vehicle and technology			vehicle technology (e.g., Advanced Driver Assistance Systems)	Acceptable - some work required to access it	High - national	Acceptable - can be used in analysis, no location information	Medium - reliable, relevant and accurate	Moderate - a fair amount of influence on ped'n safety investment &	Low - minor financial o resource intensive cos to collect the data.	3.0	Consider	2.3	Don't use			AU: Registration data from NEVDIS (DB81) matched to AU: ANCAP (DB79)	
P01	Physical environment	Urban planning		density of roads and street connectivity	High - public access and download	High - national	High - allows easy analysis (geospatial or similar)	Medium - reliable, relevant and accurate	Very low - little to no influence on ped'n safety investment & decisions	Low - minor financial o resource intensive cos to collect the data	4.0	Use	3.0	Consider		International: Open Street Map (DB33)		
P02	Physical environment	Urban planning		motorisation level and availability of public transport	Medium - public access	High - national	High - allows easy analysis (geospatial or similar)	Acceptable - relevant and accurate	Moderate - a fair amount of influence on ped'n safety investment & decisions	Low - minor financial o resource intensive cos to collect the data	4.0	Use	3.3	Consider		International: Open Street Map (DB33) AU: Census vehicle ownership data SEIFA (DB14)	AU: Locations of public transport stops - railways (DB61) International: bus stops Open street map (DB33) or International: GTFS (DB82)	NZ: Public transport p datasets (DB83) pair ThinkProject bus stop (formerly RAMM, DB5

51 separate factors identified

128 separate databases identified

Data Framework tool: Top Datasets v2

Based on the **3** most relevant criteria:

- Population *(numbers, density, mix)*
- Posted and travelling speed *(including probe speed data)*
- Mechanism and severity of injury *(e.g. ICD code, AIS and MAIS levels)*
- Demographics *(age, gender, ethnicity and indigenous status, language)*
- Socioeconomic attributes *(education level, income level, deprivation index)*
- Road hierarchy and functional classification
- Vehicle traffic volumes and mix
- Drug or alcohol use in crashes
- Facilities for pedestrians (geospatially mapped) - crossing roads
- Crash location (road or path) and type *(DCA/CAS codes)*
- Road user characteristics *(age, gender, ethnicity)*

Final Pedestrian Data Framework

(building off Stream 1 work)

Category			Most relevant factors or attributes
Human factors	Behaviour		• Behavioural factors (e.g. risky behaviour, distraction) • Safety awareness and education enforcement (e.g., speed limits, red light cameras) • Road user perception and attitudes towards road safety
	Health and wellbeing		• Health and wellness indicators (e.g. obesity rates, physical activity) * • Cognitive and physical capabilities (e.g. vision, reaction time)
	Socioeconomic		• **Population (numbers, density, mix) *** • **Demographics (e.g. age, gender, ethnicity and indigenous status, language) *** • **Employment and income level *** • Education level and access to information
Vehicle and technology			• Vehicle fleet composition, age, and type • Vehicle regulation and vehicle safety technology penetration and standardisation • Technology (e.g. Advanced Driver Assistance Systems)
Physical environment	Urban planning		• Density of roads and street connectivity * • Motorisation level and availability of public transport *
	Land use		• Land use mix and density * • Space and capacity of land uses * • Proximity to destinations and trip generators *
	Road transport systems	Roads, streets and paths	• **Road hierarchy and functional classification (e.g. Movement & Place) *** • Road geometry (e.g. intersections, roundabouts) • Geometric design variables (e.g. lane width, curvature) • **Posted and travelling speed (including probe speed data)** • Facilities for pedestrians and bicycles (along roads or in open spaces) *4 • **Facilities for pedestrians and bicycles (crossing roads) *** • Emergency medical services and response time • Intervention effectiveness (Crash Modification Factors)
		Public transport	• Public transport supply, e.g., stops & routes * • Public transport access and safety *
	Environment		• Light & atmospheric conditions

Category		Most relevant factors or attributes
Injuries/Crashes	Injury severity	• **Mechanism and severity of injury (e.g. road crash or hospital trauma codes)**
	Human factors	• **Drug or alcohol use** • Distraction and inattention • Fatigue and physical capability
	Vehicle factors	• **Vehicle type**, weight, and age • Crashworthiness and pedestrian harmfulness • Design and technology
	Physical environment	• Public requests or complaints regarding pedestrian infrastructure • Road geometry and design (e.g., curvature, superelevation) • Traffic control devices and systems (e.g., signals, signs) • Impact speed and severity • Temporal and atmospheric conditions (e.g., rain, fog)
	Socioeconomic	• **Road user characteristics (e.g., age, gender)**
	Crash	• **Crash location (road or path) and type (DCA or CAS codes, for example) *** • Crash causation and contributing factors (e.g. speeding, driving under influence) • Impact speed • Anecdotal stories of crash victims or families
	Near-miss	• Near miss data or conflict studies from video analytics
	Post-crash	• Emergency medical services and response time • Medical care
Exposure	Micro	• Pedestrian volume and mix (along streets & paths) • Pedestrian volume and mix (across streets) • **Vehicle traffic volumes and mix**
	Macro	• Trips (duration and distance from travel demand surveys & network count programmes) • Population-based measures (e.g., exposure rate, mode share) *

Minimum pedestrian safety data sets

** indicates data sets that may have good geospatial location information*

Practice Guideline

Sections: *(33 pages)*

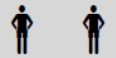
- Summary

1. Introduction
2. Practical Considerations
3. Types of pedestrian safety data
4. Recommended minimum data requirements

*Aim is to be a **concise & practical** guide for practitioners to use*

Practice Guideline: Types of Data

- Tables provided→ summarising key information about each data type
  - Ped'n travel data
  - Ped'n casualty data
  - Ped'n facility data
  - Motor vehicle data
  - Other potential metrics

Name of data type	Census data
Method of data collection	This is usually undertaken at a national, state or territory scale, where the entire population is asked to manually or electronically report their main means of travel to work (and sometimes other trip purposes such as study or education).
Relevance to pedestrian safety	 Walking journeys of all trip purposes collected usually provide an indication of kms or hours travelled on foot, or numbers of trips made. This provides indicative usage measures that can be used to determine relative casualty risk per unit of travel.
Cost of data collection	$ $ $ Widespread data collection and analysis can be a very costly survey effort. The AU Census is every five years, and the NZ Census has been discontinued.
Scale	Strategic, Tactical (Programme)
Advantages of data set	Census data provides a very comprehensive picture of pedestrian trip numbers, which can provide a useful input into land use planning and modelling at both localised and wide levels.
Disadvantages of data set	Census data is often only focused on specific trip purposes such as commuting, which underplays how much walking is occurring – especially amongst non-working age people. It may also focus on the "main" mode of travel, overlooking any walking trips made as part of the overall journey. Given the long interval between censuses, data can be out-of-date.
Other notes	While Census surveys with larger samples may provide good information about the relative "depth" of walking numbers for specific trip types, household travel survey data is useful for getting an idea of the "breadth" of walking travel patterns.
Case study	The Socio-Economic Indexes for Areas (SEIFA) ranks areas according to their relative socio-economic advantage and disadvantage using Census data. When cross-tabulated with other census data such as travel mode, a fuller picture of pedestrian exposure can be developed (DB-10).

Conclusions – minimum data

Recommended minimum pedestrian safety data:

- **Population** data (numbers, demographics)
- **Crash** data (location, type, severity, users, factors)
- **Road & Traffic** data (classification, volumes, speeds)
- **Pedestrian** facility data (crossings, paths)

Various ways to collect these data sources:

- **Data Framework** tool lists several potential data sources
- Further guidance is outlined in **Practice Guideline**

Use of pedestrian safety data

Useful to distinguish between measures that record:

- **Monitoring** of progress towards better pedestrian safety
  e.g. no. of pedestrian casualties (+ associated descriptive info)
- **Implementation** of better pedestrian safety environments
  e.g. % of low-speed streets, no. of raised pedestrian crossings

*Monitoring on an ongoing basis provides important understanding of **how well** a jurisdiction is meeting its pedestrian safety objectives, **but** you need changes in other measures to **produce** that safer pedestrian environment…*

Thank you!

Questions?

VIASTRADA

TRANSPORT PLANNING AND DESIGN

glen@viastrada.nz

/glenkoorey/

www.viastrada.nz

2025 Australasian Road Safety Conference

20 - 23 OCT - PERTH, WESTERN AUSTRALIA