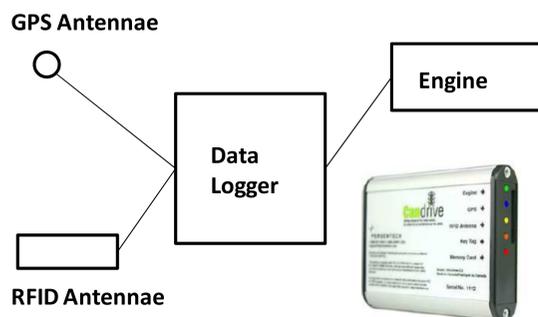


Objective

- Anonymity of participants must be maintained
 - Dataset shared with differing trust models
- Candrive: multi-year study, 1000 older drivers
 - GPS and OBDII sensor deployed in participant vehicles
- Automation required for massive (1TB) dataset

Data Set processing requirements

- Data Validation
 - GPS reception failures
 - Incomplete samples
- Data Augmentation
 - GIS map data (posted limits, road hazards)
 - Solar cycle (day/night) data
 - Weather information
- Data Anonymization challenge
 - Participants identified by number only but:
 - Location – Provides details on where a user lives/works/visits → their identity
 - k – anonymity – combinations of data elements can't be combined to re-identify participant
 - Differential Privacy – Provide differing data access based on trust models



Method

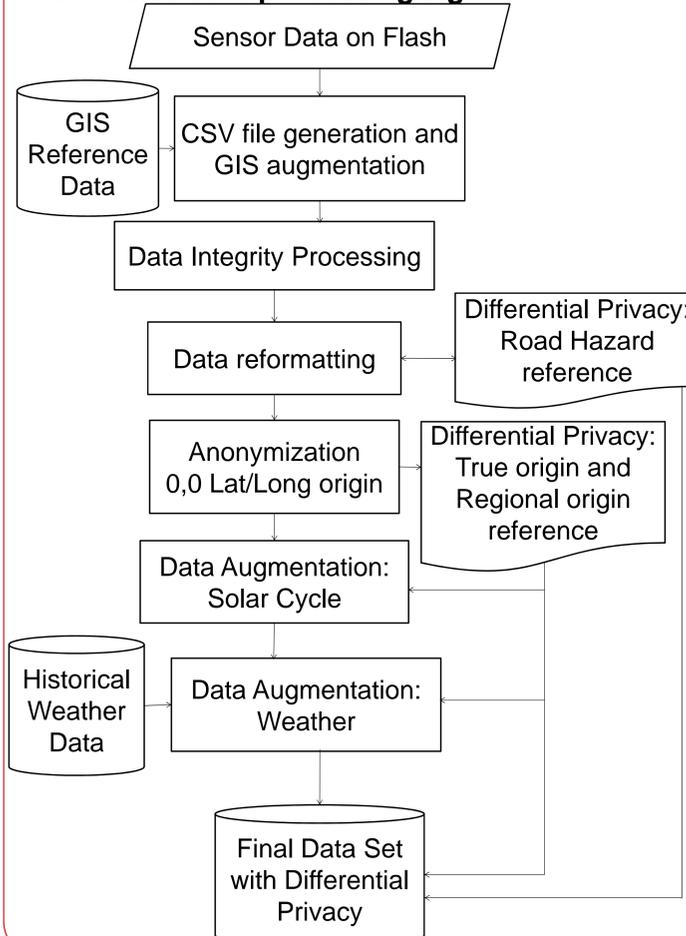
Data set statistics for participants

	Min	Mean	Max	Mean Std Dev
Total Data Size for each participant (MB)	1350	2250	4010	822

Example sensor data captured at 1Hz

Sensor	Measured Value	Units / format
GPS	Date & Time	Date string
	Longitude/Latitude	Degrees
	Velocity	km/hr
GIS	Posted Limit	km/hr
	Alerts	Text String
RFID	RFID tag number	Serial number
OBDII	Velocity	km/hr
	Throttle Position	Percentage

Automated data processing algorithm



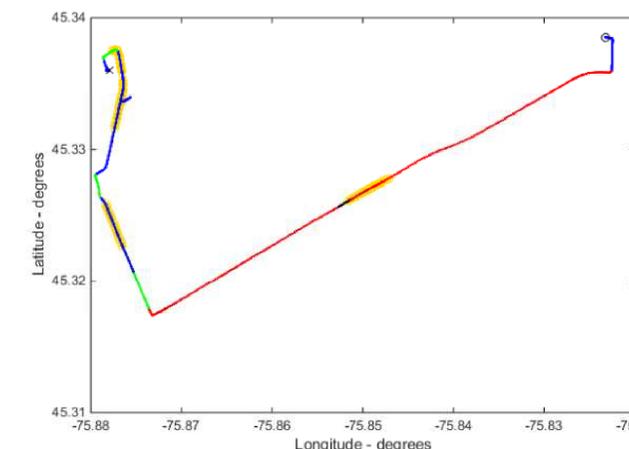
Results and Conclusions

Attribute	Value
Metropolitan Area Size	~100km East - West ~50km North - South
Variation in Sunrise/Sunset time to East/West limits	+/- 4 mins
Variation in Solar Day time to North/South limits	+/- 1 min

Preventing data relocation and participant identification through Sunrise and Sunset time

Example trip trace showing posted speed limit and location of road hazards

Legend:
 blue: 40km/hr
 green: 60km/hr
 red: 80 km/hr
 black: 100km/hr
 yellow: road hazard



	Fully Trusted User	Intermediate Trust User	Open Access User
Dataset Access	Zero referenced Longitude/Latitude trip Data		
Location of trip origin	Absolute	Regional reference only	No
Latitude Accuracy	GPS limited (+/- 10m typical)	Limit by region (Ottawa +/-25km)	+/- 100km
Longitude Accuracy	GPS limited (+/- 10m typical)	Limit by region (Ottawa +/-50km)	+/- 100km
Regional Weather Data	Yes	Yes	No
Sunrise/Sunset Data	Yes	Yes (region center)	Yes (added noise)
Road Hazards	Codes and meanings	Codes only	Codes only

Anonymity model

- 3 levels of trust
- K-anonymity
- Differential Privacy
- Ensures anonymity of participant
- Guidelines for publication

Acknowledgements

Natural Sciences and Engineering Research Council (NSERC) and industrial and government partners, through the Healthcare Support through Information Technology Enhancements (hSITE) Strategic Research Network. Candrive prospective study (www.candrive.ca) funded by the Canadian Institutes of Health Research (CIHR) Innovation Fund of the Alternative Funding Plan for the Academic Health Sciences Centres of Ontario.