Title :

A Regional GTFS Research Agenda for ASEAN

Full name of Speaker :

Sye Loong Keoh (UGS)
Yung-Wey Chong (USM)
Chee Kiat Seow (UGS)
Qi Cao (NUS)
Achmad Basuki (UB)
Sharul Kamal Abdul Rahim (UTM)

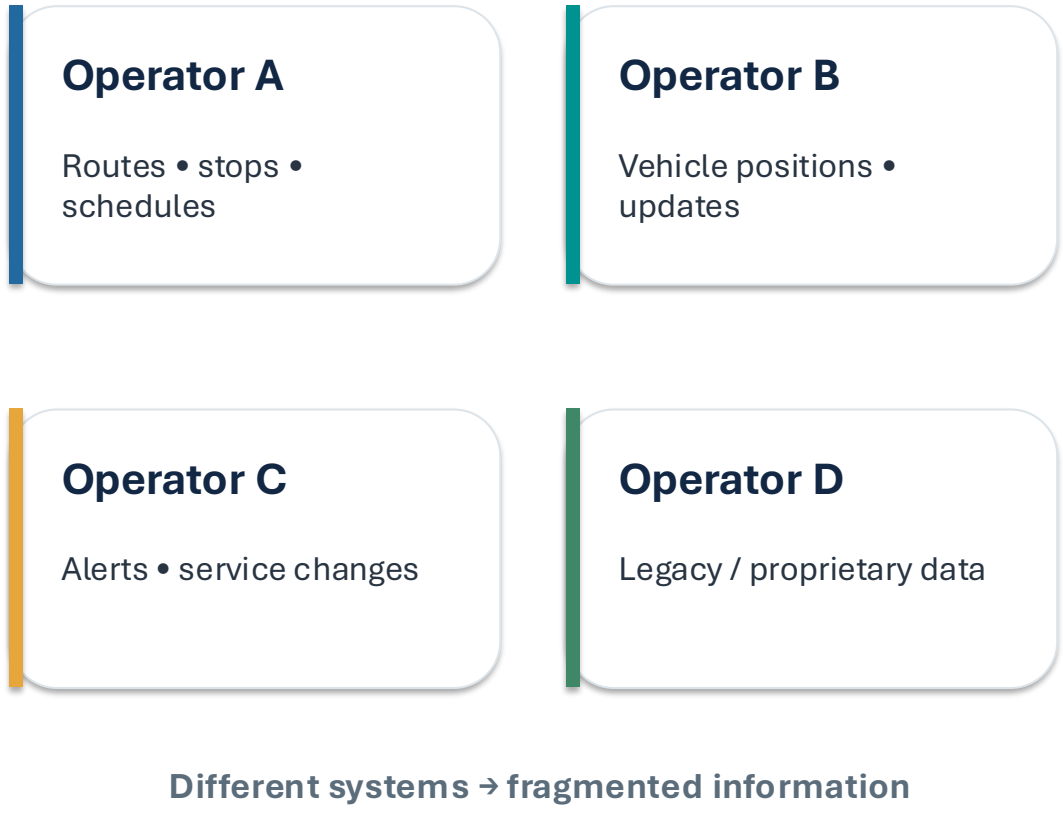
Institution :

University of Glasgow (Singapore)



Public transport data is often fragmented.

- Operators use different systems, data structures and update mechanisms.
- Static schedules and live operational information may sit in separate platforms.
- Integrating services across operators can require costly, one-off interfaces.
- Passengers experience fragmented information when journeys span networks.



Two problems hold back ASEAN bus services

Passengers can't trust arrival times

Most deployed systems rely on rule-based historical averages. They don't adapt to live congestion and don't carry over from one year to the next.

Transit data is fragmented

Unlike Europe, North America and parts of East Asia, most ASEAN cities publish no standardised, interoperable transit data. Cross-city research, regional apps and multi-country model training are practically impossible.

Better predictions need shared data, and shared data needs a common standard.

A Regional GTFS Research Agenda for ASEAN

GTFS Static

The planned network

- Routes and route shapes
- Stop locations
- Timetables and service calendars

anchors

GTFS Realtime

What is happening now

- Live vehicle positions
- Trip updates: delays, cancellations
- Service alerts

Static data is the reference that real-time feeds point to. Together they form the data layer every intelligent transit system is built on.

Together, they provide a practical foundation for interoperable passenger information and transport services.

COMMON DATA MODEL | OPEN INTERFACES | REUSABLE INTEGRATION

ASEAN's open transit data today

Singapore	**Real-time**	LTA publishes real-time feeds through its open data portal
Malaysia	**Partial real-time**	Static and real-time for government rail and stage buses, with vehicle positions only
Thailand	**Partial static** **Partial real-time**	Several cities have partial GTFS static data
Indonesia, Vietnam, Philippines, Cambodia	**No public feed**	GPS fleet tracking is active, but no publicly available feeds

The GPS data already exists. What's missing is a common, open format, and IVO is well placed to coordinate it.

Project vision and objectives

Build and validate a reusable approach for integrating public transport data across ASEAN operators using GTFS and GTFS-Realtime to provide ASEAN GTFS Repository

1 Assess

Understand participating operators' data sources, formats and operational needs.

2 Map & transform

Map heterogeneous source data to GTFS / GTFS-RT and document conversion rules.

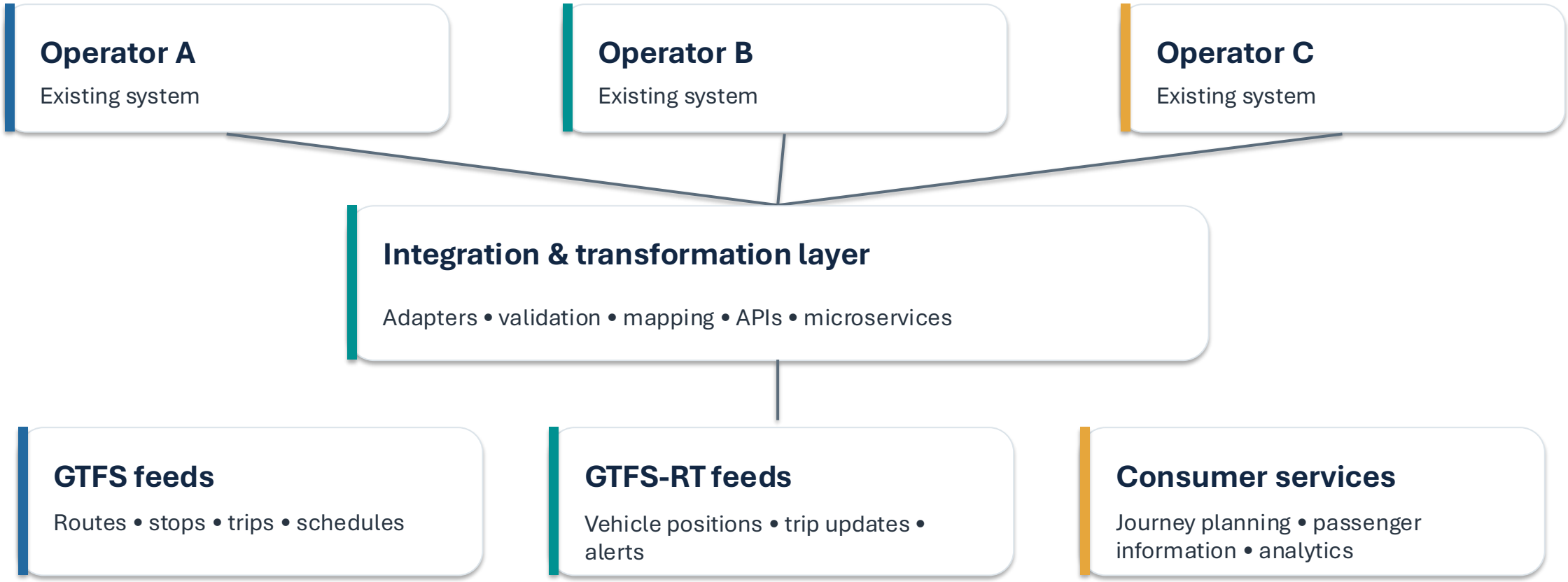
3 Integrate

Develop a lightweight integration layer using APIs and microservices.

4 Validate

Evaluate data quality, timeliness, interoperability and practical deployment needs.

Proposed technical approach



Proof of concept: Johor Bahru, Malaysia

P211 — 22.3 km
Taman Universiti ⇔ Larkin Sentral

P411 — 31 km
Kulai ⇔ Larkin Sentral

Two free public bus services

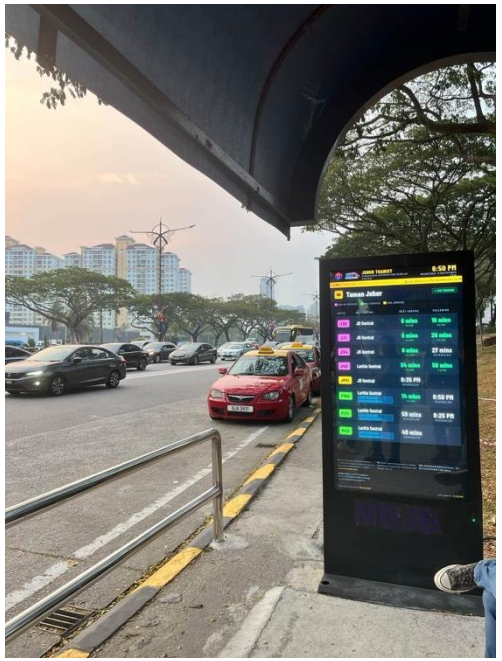
BLE proximity sensing at stops
IoT detection of bus arrivals [4]

GPS fleet tracking
Continuous vehicle positions

Passenger information displays
Now operational

Partners
Johor Transport Corporation (PAJ) and three city councils

Fleet data collected January 2025 to June 2026



Gradient boosting gives accurate bus ETAs

Seven models trained on GPS and BLE fleet data. Gradient boosting beat neural networks and the historical mean on every route.

15–26 s

RMSE per stop segment

LightGBM, across routes

0.87–0.95

R^2

Variance explained

7–24%

Better than today's system

Improvement over the operational baseline

Accurate, data-driven arrival prediction is feasible with the fleet data ASEAN operators already collect.

Closed-loop ETA: one pipeline for every city



Fleet IoT & GPS

BLE/GPS feed, vehicle positions

GTFS-RT feed

Anchored to GTFS static

ML pipeline

Rolling recent training window

ETA to passengers

Displays and journey apps

Observed arrivals flow back to retrain the model, keeping predictions current

With regional GTFS-RT feeds, this architecture deploys in any member city without bespoke data integration.

Every city can participate in this initiative easily!

1

Publish GTFS static

Routes, stops and timetables. No real-time systems needed.

2

Stream vehicle positions

Turn existing GPS fleet tracking into GTFS-RT feeds.

3

Close the loop

Trip updates, alerts and ML-based ETA to passengers.

Three work packages

WP1

ASEAN GTFS + GTFS-RT Repository

Coordinate IVO member cities to publish interoperable static and real-time feeds as a shared, open regional dataset

Who: All IVO member countries

Output: GTFS-Validated Open ASEAN Transit Data Open Data Repository

WP2

Multi-city cross-year ML validation

Validate the ML training accuracy, and training window across ASEAN cities using GTFS-RT data from WP1

Who: Any ASEAN country

Output: Regional ETA training pipeline, guidelines and benchmarking

WP3

Deployment and Integration with the cities

Deployment of IoT/GPS/Passenger Information Display/Mobile Apps by developers or the operators

Who: Any ASEAN country

Output: Deployment specification and cost-benefit model for developing cities

A Regional GTFS Research Agenda for ASEAN

A ready ML pipeline

Transferable to any city with fleet-tracking data

Open regional data

Research without one-off data agreements

Regional benchmarks

ETA benchmarks and training-window guidelines

Deployment guidance

Deployment architecture, guideline and cost-benefit model

Join us

Talk to us during the Forum. Tell us:

- your city
- the data you hold
- the work package that fits
- how can we work together!

From a single-city proof of concept to a regional public transport intelligence platform.