Conferência Nacional dos Agentes Produtores e Usuários de Dados - CONFEST/CONFEGE

SALVADOR, BAHIA - 3 A 5 DE DEZEMBRO DE 2025

# From Logs to Policy: Integrating MPD, Surveys, and Geospatial Layers for Faster, Smaller-Area Tourism Statistics

Alfatihah Reno Maulani
Statistics Indonesia

Apoio

SENAI CIMATEC UNIVERSIDADE

Sistema FIEB
SESI | SENAI | IEL | CIEB

GOVERNO DO ESTADO BAHIA
GOVERNO PRESENTE TRABALHA PRA GENTE

Realização

IBGE
Instituto Brasileiro de Geografia e Estatística

MINISTÉRIO DO PLANEJAMENTO E ORÇAMENTO

GOVERNO DO BRASIL
DO LADO DO POVO BRASILEIRO

The Power of Integration
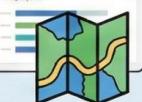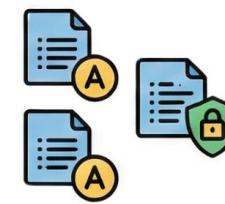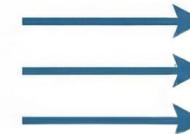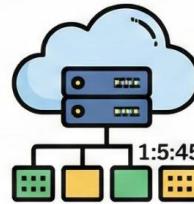
Getting to Know Mobile Positioning Data

MPD IS PASSIVE POSITIONING FROM CHARGING/NETWORK EVENTS; IT'S NOT GPS TRACKING.

WHO (hashed MSISDN) → WHEN (timestamp) → WHERE (lat-lon)

PASSIVE POSITIONING (MPD)	ACTIVE POSITIONING (GPS)
Network events, charging logs	Continuous tracking, device-based

SAMPLE CDR SCHEMA			
...	...	...	...
Hashed MSISDN	Timestamp	Cell ID	Event Type (e.g., Call, SMS)

Trust by design: privacy & access

PRIVACY AND CONFIDENTIALITY ARE SAFEGUARDED VIA MNO SANDBOX + HASHING; PROCESSING HAPPENS INSIDE TELCO ENVIRONMENT.
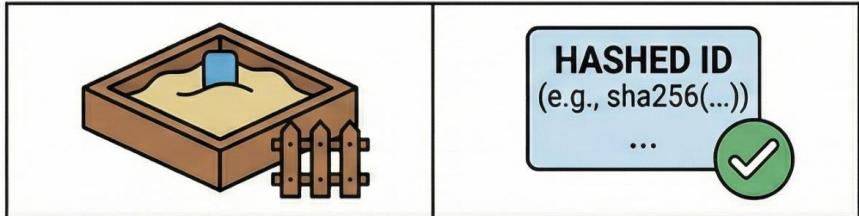
BPS
(National Statistics)

SECURE SANDBOX
(MNO Environment)

TELCO
(Mobile Network Operator)

HASHED ID
(e.g., sha256(...))
...

VOLUNTEER-BASED MODEL VERIFICATION

QA gates for MPD input
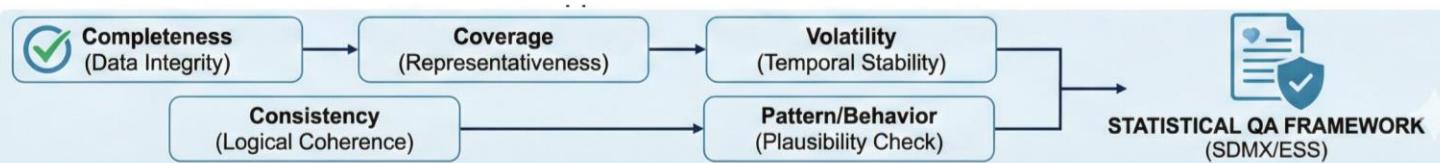
✅ **BEFORE ANY INTEGRATION, WE RUN FORMAL QA GATES (COMPLETENESS, COVERAGE, VOLATILITY, CONSISTENCY, PATTERN/BEHAVIOR).**

QA INDICATORS DASHBOARD (#1–#15)

	ID	INDICATOR NAME	DIMENSION	STATUS	VALUE
COMPLETENESS	#1	Missingness (CDR Count)	Missingness	✓ PASS	0.01%
	#2	Missingness (Subscriber ID)	Subscriber ID	✓ PASS	0.05%
	#3	Temporal Gaps (Hourly)	Temporal Gaps	✓ PASS	0
COVERAGE	#4	Unique Coordinates/LAU	Unique Coordinates/LAU	✓ PASS	98%
	#5	Spatial Coverage (Grid)	Spatial Coverage	✓ PASS	95%
	#6	Cell Tower Density	Cell Tower Density	⚠ WARNING	Low in rural
VOLATILITY	#7	Elephant-curve by hour (Activity Volume)	Elephant-curve by hour	✓ PASS	〰
	#8	Daily Subscriber Count Variance	Daily Subscriber Count	✓ PASS	〰
	#9	Weekend/Weekday Ratio	Weekend/Ratio	✓ PASS	〰
CONSISTENCY	#10	Time gaps between events	Time events	✓ PASS	<1s
	#11	Spatial gaps between events	Spatian events	✓ PASS	<500m
	#12	Speed gaps (Velocity check)	Velocity check	✓ PASS	<120km/h
PATTERN/ BEHAVIOR	#13	Home/Work location stability	Location stability	✓ PASS	High
	#14	Commuting pattern regularity	Commuting pattern	✓ PASS	High
	#15	Anomalous Travel Paths	Anomalous Travel Paths	✓ PASS	0.1%

Completeness (Data Integrity) → **Coverage** (Representativeness) → **Volatility** (Temporal Stability) →

Consistency (Logical Coherence) → **Pattern/Behavior** (Plausibility Check) →

STATISTICAL QA FRAMEWORK (SDMX/ESS)

s.id
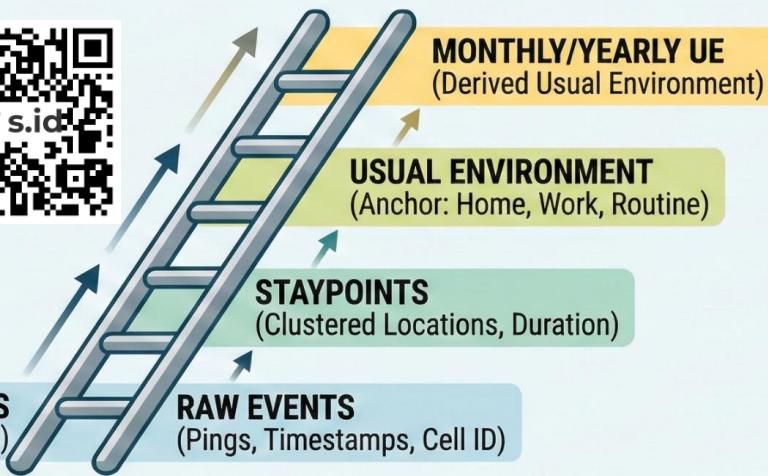
From pings to people: AMDA and concepts

WE CONVERT RAW EVENTS → STAYPOINTS → USUAL ENVIRONMENT (HOME/WORK) USING AMDA; TRIPS FOLLOW UNWTO/IRTS CONCEPTS.

MONTHLY/YEARLY UE
(Derived Usual Environment)

USUAL ENVIRONMENT
(Anchor: Home, Work, Routine)

STAYPOINTS
(Clustered Locations, Duration)

RAW EVENTS
(Pings, Timestamps, Cell ID)

RAW EVENTS
(Pings, Timestamps, Cell ID)

TRIP DEFINITION
(UNWTO/IRTS CONCEPTS)

≥6 HOURS
(Stay)

<12 MONTHS
(Duration)

NON-ROUTINE
(Not Commuting)

NOT FOR WAGES
(Purpose)

USUAL ENVIRONMENT

AMDA STEPS
(Processing Logic, Algorithms)

IRTS REFERENCES
(UNWTO Guidelines, Statistical Framework)

EXCLUDE ROUTINE COMMUTING
(Defining Tourism Activity)
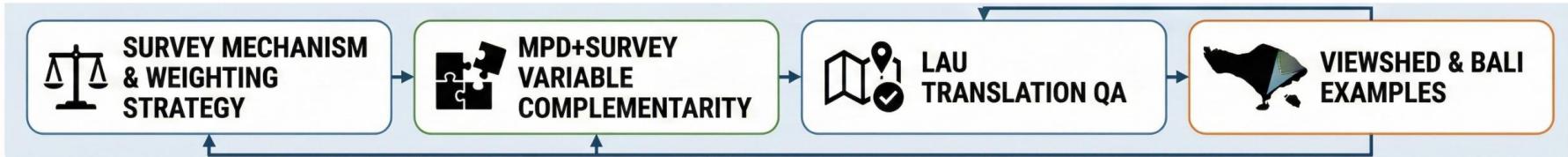
The integration: MPD × Digital Survey × Geoscience

MPD GIVES **MOVEMENT** (ORIGIN, DESTINATION, LOS); THE **DIGITAL SURVEY** ADDS **ATTRIBUTES** (PURPOSE, SPEND, DEMOGRAPHICS); **GEOSCIENCE** LAYERS ENSURE SPATIAL CORRECTNESS.

MPD
(Mobile Positioning Data)
ORIGINS / DESTINATIONS / LENGTH OF STAY
(Municipality Level, Monthly)

DIGITAL SURVEY
MODE, ACCOMMODATION, PURPOSE, EXPENDITURE
(Quota by Municipality; CAWI via SMS/WA; Incentives)

GEOSCIENCE LAYERS
TOWER LOCATIONS, LAU POLYGONS, VIEWSHED/COVERAGE

SURVEY MECHANISM & WEIGHTING STRATEGY → **MPD+SURVEY VARIABLE COMPLEMENTARITY** → **LAU TRANSLATION QA** → **VIEWSHED & BALI EXAMPLES**

Extrapolation and weights

WE SCALE MPD TRIPS TO THE POPULATION WITH TRANSPARENT FORMULAS THAT ACCOUNT FOR MULTI-SIM, OTHER MNOs, AND NON-PHONE USERS—BENCHMARKED TO SUSENAS.

FORMULA PANEL

$$Y = (T_w * MPD_{other} * NonHP * DSC * ...)$$

- **Y: TOTAL POPULATION ESTIMATE** (Scaled Trips)
- **DESIGN WEIGHT** (Initial Probability)
- **ADJUSTMENT FOR OTHER MNOs** (Market Share)
- **ADJUSTMENT FOR NON-PHONE USERS** (Demographics)
- **CALIBRATION** (Benchmark to SUSENAS)
- **OTHER FACTORS** (e.g., Multi-SIM)

✓ **DESIGN WEIGHT:** Initial scaling based on sampling probability (MNO market share).

✓ **NONRESPONSE ADJUSTMENT:** Corrects for missing data and non-participants.

✓ **CALIBRATION:** Aligns estimates with official benchmarks (e.g., SUSENAS population totals).

EXTRAPOLATION EQUATIONS (Transparent & Documented)
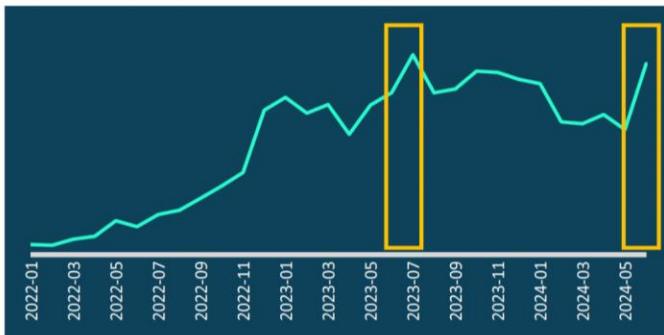
**SUSENAS WEIGHTING** (Official Benchmark & Calibration Target)

Evidence it works

The number of trips in peak season is higher than in the other months.

Using unique survey links outperforms direct links in response rate.

Apoio — SENAi CIMATEC UNIVERSIDADE · Sistema FiEB SESI/SENAI/IEL/CIEB · GOVERNO DO ESTADO BAHIA GOVERNO PRESENTE TRABALHA PRA GENTE

Realização — IBGE Instituto Brasileiro de Geografia e Estatística · MINISTÉRIO DO PLANEJAMENTO E ORÇAMENTO · GOVERNO DO BRASIL DO LADO DO POVO BRASILEIRO
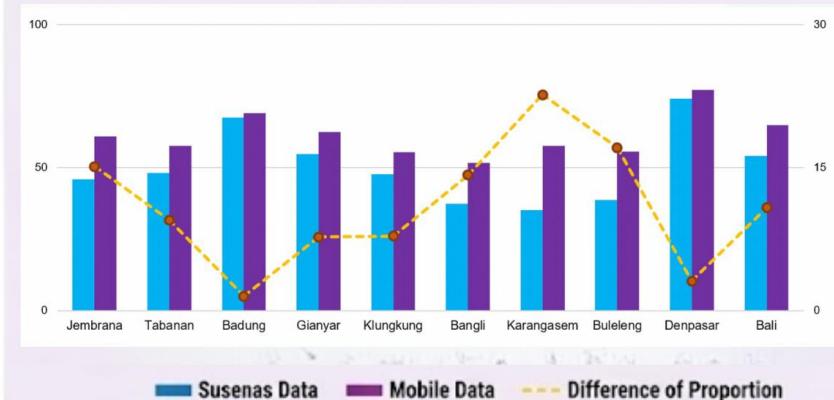
Policy-ready geospatial outputs

POPULATION COVERED BY MOBILE NETWORK IN BALI BASED ON NETWORK TYPE

2G 3G 4G

- 1 - 300
- 300- 600
- >600

VALIDATION & POLICY BENEFITS: PATTERN ALIGNMENT WITH CONVENTIONAL DATA

Susenas Data   Mobile Data   Difference of Proportion

Cost/time and sustainability

1. Granularity (Detail Level)

District/City: High detail, hyper-local data. → Province: Broad, aggregate data.

2. Periodicity (Frequency)

Monthly Cycles / Yearly Cycle

Monthly: Regular, up-to-date data. → Yearly: Annual snapshot, less frequent.

3. Surveyors (Field Presence)

No field surveyors: Automated, remote collection. → Depends on sample: Requires on-the-ground personnel.

4. Budget (Cost Estimate)

≈ IDR 15B → ≈ IDR 30−40B

≈ IDR 15B: Lower cost, efficient. ≈ IDR 30−40B: Higher cost, resource-intensive.

5. Output Frequency (Reporting)

Monthly Tourist Stats / Jan Feb Mar → Annual Only

Monthly domestic tourist stats: Continuous reporting. → Annual only: Once-a-year reporting.
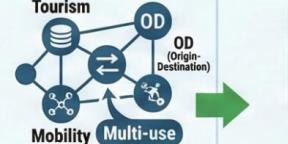
6. Pandemic Readiness (Adaptability)

No door-to-door needed: Safe, contactless. → Restricted: Vulnerable to disruptions.

7. Reusability (Data Utility)

Tourism / OD (Origin-Destination) / Mobility / Multi-use → Single-use

Multi-use (tourism, OD, mobility): Versatile datasets. → Single-use: Purpose-specific, limited application.

8. Conclusion & Summary

MPD + Digital Survey: Innovative, Cost-effective, Agile, Resilient, Versatile. Traditional Survey: Conventional, Costly, Infrequent, Vulnerable, Limited Use.
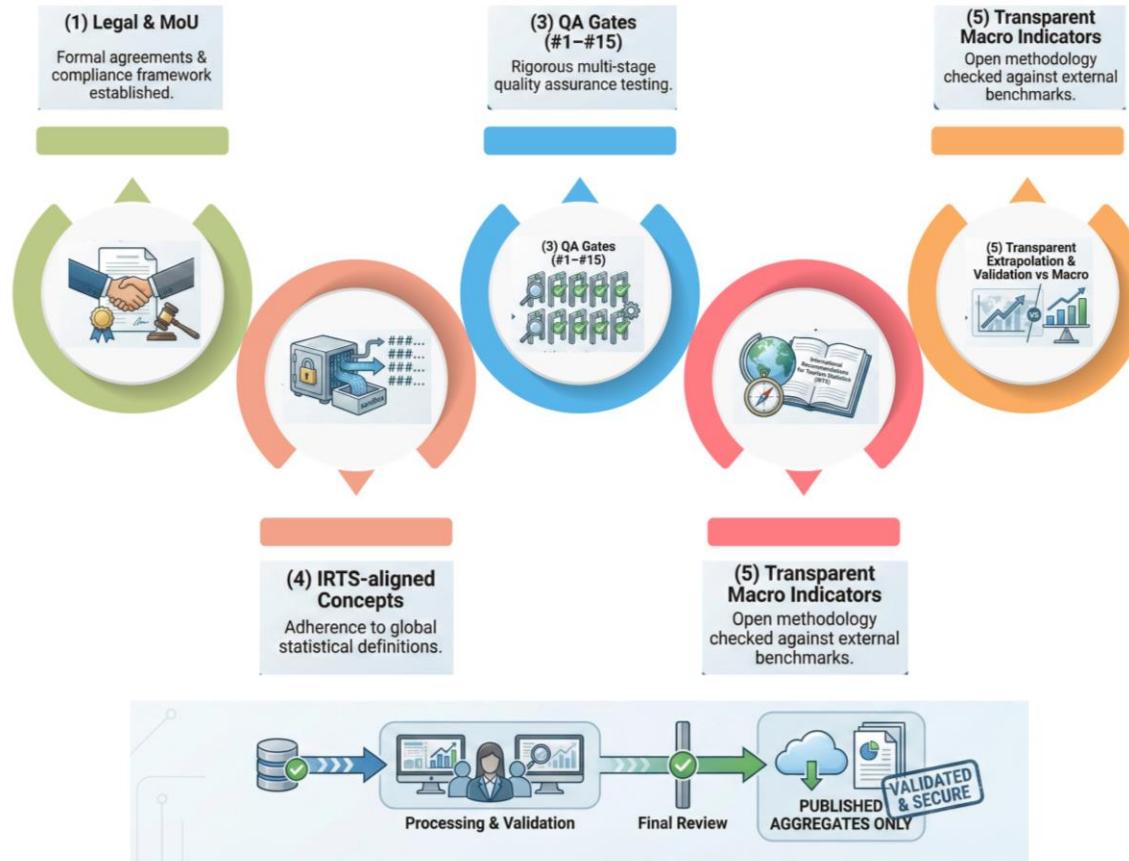
Assurance & Governance Checklist
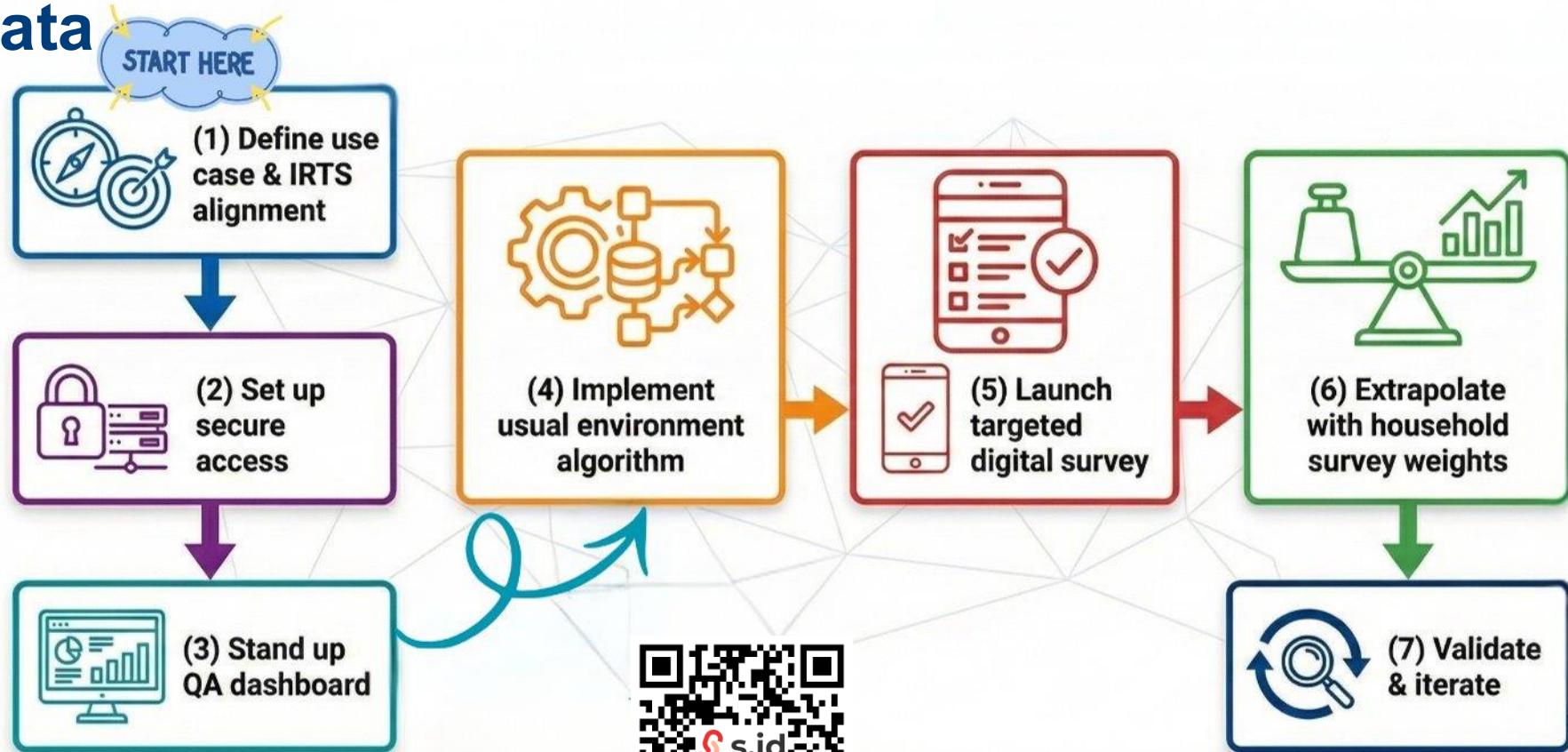
Starting a New Project using Mobile Positioning Data

Thanks!

mpd@bps.go.id

Para mais informações, acesse o QRcode

Apoio

SENAI CIMATEC UNIVERSIDADE

Sistema FIEB
SESI | SENAI | IEL | CIEB

GOVERNO DO ESTADO BAHIA | GOVERNO PRESENTE TRABALHA PRA GENTE

Realização

IBGE
Instituto Brasileiro de Geografia e Estatística

MINISTÉRIO DO PLANEJAMENTO E ORÇAMENTO

GOVERNO DO BRASIL
DO LADO DO POVO BRASILEIRO