

The two-phase life cycle model and mobile phone data

WP5 on Mobile Phone Data

E. Esteban, S. Saldaña, D. Salgado, L. Sanguiao

Statistics Spain (INE)

Madrid, 20-21 march, 2018

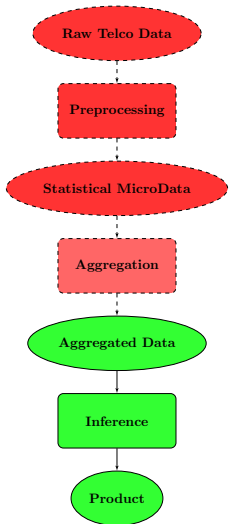


Big Data *for Official Statistics*

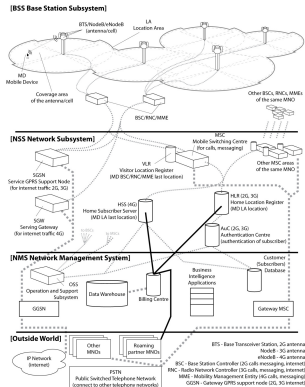
Survey	Admin	Sensors, Social Media, ...
		Remote Crowd Personal
		WWW ...

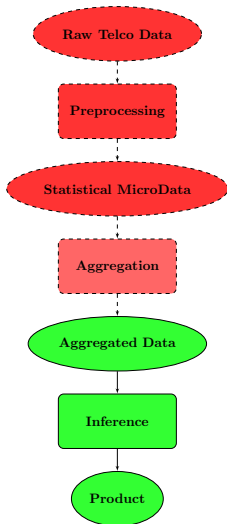
1. Data about third people, not about data provider.
2. Data at the business core of the economic activity.
3. Data generated without statistical metadata.
4. 3Vs.



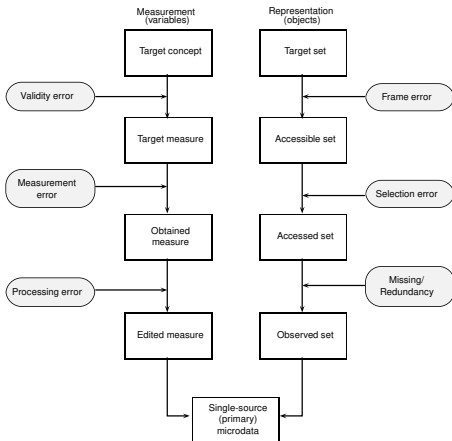


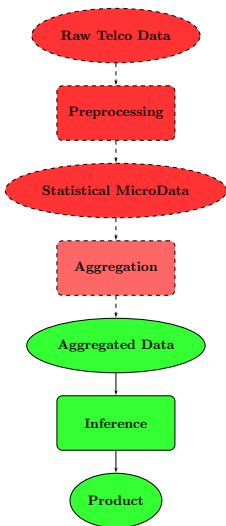
MNO Architecture



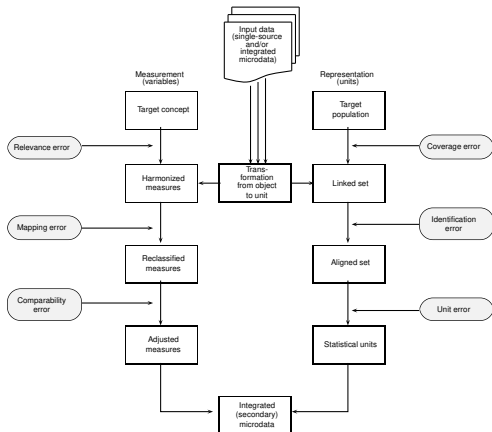


Phase 1





Phase 2



eventID	deviceID	attr1	...	attrN
1	1
2	1
⋮	1
n_1	1
$n_1 + 1$	2
$n_1 + 2$	2
⋮	2
$n_1 + n_2$	2
$n_1 + n_2 + 1$	3
⋮	3

Transformation
 object to unit

deviceID	eventID	attr1	...	attrM
1	1
	2
	⋮
	n_1
2	$n_1 + 1$
	$n_1 + 2$
	⋮
	$n_1 + n_2$
⋮				

Identification
 of individuals

individualID	deviceID	attr1	...	attrQ
1	1
	2
2	3
⋮



- Collaboration with MNOs (engineers?) needed for the primary phase assessment
 - ↪ Normalization in statistical terms of the generation of raw data.
 - ↪ Assessment of error sources for statistical purposes.
- Statistical microdata needed for the secondary phase assesment
 - ↪ Normalization of data architecture (core data model).
 - ↪ Definition and generation of variables (geolocation, ...).
- Aggregated data needed for the tertiary phase assessment
 - ↪ Normalization of file formats (SDMX?, Standards?).
 - ↪ Normalization of aggregation procedures.