Methodology

Our sophisticated holistic methodology provides the most accurate data model for analysing human behaviour in space and time

The methodology used in Positium Data Mediator (PDM) to process mobile big data into meaningful and reliable statistical indicators has been developed in academic cooperation between Positium and the University of Tartu since 2004. The core data model and the statistical indicators produced by PDM are based on learnings from working with clients and academia throughout the years.

Mobile positioning data has a number of limitations and we have spent years finding solutions for each one of them. Not everyone uses a mobile phone. Some people own several. There may also be disparities due to the gender gap in mobile phone ownership, especially in low- and middle-income countries, and disparities between different socio-economic groups. Additionally, it is only phone users who are subscribed to or roaming in the mobile network which is providing the data who will be included in the dataset. The geographic, demographic, and socio-economic distribution of each mobile network operator is likely to be different and should be accounted for with appropriate methods. There is noise in the data. To address these concerns, we have developed various solutions that are included in our flagship product, Positium Data Mediator.

The purpose of the complex methodology behind PDM is to transform MPD into the most accurate data model that describes human spatio-temporal behaviour, so that the conformity of the data model to reality is only limited by the quality of MPD. Our experience working with national statistical offices and government organisations, ministries of tourism and planning, cities, municipalities and urban planners, data businesses, and international organisations (UN WTO and UN GWG) means that the methodology we use adheres to high quality assurance standards.

To ensure the best results for our clients, we apply these **11 principles to our methodology**:

1. The model is an approximation of reality and seeks to approximate to the most representative and populous entities in real life, while acknowledging the diversity of human behaviour;

2. The model cannot be static and needs to consider the dynamics of human spatiotemporal behaviour;

3. The model is not an array of snapshots of human life, but represents human life in its continuity;

4. The model cannot be better than its underlying data, but should never incorporate logical fallacies, such as:

   o people moving with unrealistic speeds,

   o people being present in multiple locations at the same time,

   o an unreasonable number of people travelling at night-time;

5. The model should describe human spatio-temporal behaviour, so sufficient effort should be made to exclude as much non-human data (machines, IoT) and biased data as possible within the limitations and possibilities of the data and domain knowledge;

6. The model needs to be based on data that has gone through several quality assurance steps so that domain-specific insights could be provided with confidence;

7. The model continuum can be interrupted with lack of data, which can be counteracted by imputation and interpolation where possible;

8. The model needs to provide comparable results across different domains without introducing controversy;

9. The model has a limited number of domains that it can be applied to;

10. The model allows retrospective view for conducting revision;

11. The model entities are classified into semantic classes that do not conflict with real life.

The core of the methodology is the **continuity model**, which is the representation of people's spatio-temporal activity with a semantic domain-specific interpretation of the

behaviour of the subscribers. The continuity model allows sophisticated queries and analysis possibilities that are incorporated into the automatic results generation of PDM.

Among others, the **methodology includes the following features**:

- universal coverage area model of the network antennae;
- probabilistic spatial interpolation using an adaptive grid and land coverage model;
- identification of stay/move sections;
- anchor point model for the recognition of regularly visited locations and semantic assumptions about those places;
- identification of long-term changes in migration, usual environment and activity spaces;
- calculation of indicators for population statistics;
- calculation of indicators for mobility statistics, and of origin-destination matrices;
- calculation of national and international statistical indicators.

Compared to other solutions, Positium Data Mediator uses a sophisticated methodology and a core data model unique to our product. From what we have seen, other solutions out there provide a simple ETL platform that aggregates raw MPD data with little or no methodology involved. From an analytics point of view, PDM will give you **better, more accurate results.**

There are several reasons why:

- reference data used during processing is temporally dynamic and customisable. For example, custom GIS layers can be used for a spatial breakdown of results, which can include custom grids, continuous or sparse polygons, etc;
- using an adaptive grid and a model of cell coverage areas as a basis for spatial interpolation gives advantages over point-in-polygon and Voronoi-tessellation approaches;
- results can be produced and an analysis can be done for several domains and a variety of indicators, breakdowns, and custom queries; our competitors mostly concentrate on a single domain and a very limited range of resulting indicators;

- PDM is automatic and systematic in the sense that it processes data from raw to results fully independently, unless there is a need to manually interrupt. Manual work is mostly necessary for human QA;

- PDM is process-flow based and controlled by configuration – different processes can be set to run, or skipped if necessary;

- There are about 400 configuration parameters in PDM which are used for technical and methodological manipulation of PDM. These parameters offer us a great range of possibilities for customising the PDM methodology to the needs of our clients.