

---

## Abstract

**Author:** MSc Michał Zawodny

**Title:** Method of aggregating data from large volumes, diverse origins, and structures in order to improve and create transport models

**Keywords:** transport modelling, simulation, traffic engineering, big data, combining, O-D matrix

**Abstract:**

The topic of this PhD dissertation is the use of a large number of data sources that fit the definition of Big Data for implementation in transport models. The dissertation reviews the state-of-the-art literature on travel modeling, Big Data, and simulation. Based on the review, a research gap was identified in the form of low integration with an increase in the number of data sources. Two methods were created to achieve the research goal. The first method involves combining travel matrices from Big Data sources despite their incompatibility and diverse origins. A second method was carried out to conduct the research as independently as possible from the biases of existing and known agglomerations. It involves simulating a realistic urban center without non-standard properties of existing centers that may negatively affect the results. For this purpose, an urban center was simulated, with trip generation and spatial distribution among transportation zones. To visualize and evaluate the results, an illustrative macroscopic model of this agglomeration was created using the PTV Visum program. Due to the research in the thesis being limited to spatial distribution and based on O-D matrices, the illustrative model does not include modal division. The transportation network was simplified and limited to the main arteries. The results of the main method were analyzed on the network created in the Visum scenario manager. The method involved taking into account the O-D matrix with appropriate weights and excluding unrealistic values on specific relationships in the matrix. To evaluate whether combining the selected nine data sources is justified, another eight "competitive" variants were created. The reason for the variants was to obtain better results. The variant using the full method with all available data sources provided the best results in the total number of trips and in the distribution of traffic flows over the simplified transport network of the simulated center. Six data sources were obtained from the literature, and the remaining three were simulated using literature descriptions and heat maps representing O-D matrices from these data sources. The advantages of the method were indicated. The direction of further research was planned.