Creario – Creating MATSim models for any place on earth

Nadine Rieser
Simunto GmbH
nadine.rieser@simunto.com

Marcel Rieser
Simunto GmbH
rieser@simunto.com

Introduction

Many papers have been written about creating scenarios for MATSim. This is mostly due to the fact that in each country the datasets available for creating a synthetic population differ. MATSim itself focuses mostly on the mobility simulation. Thus, interested users must use other tools to create the initial data required for running MATSim.

In the past, tools like TASHA [1], CEMDAP [2], Albatross [3] or Eqasim [4] have been used. In addition, independent implementations for the creation of synthetic populations were written, e.g. for Austria [5], Singapore [6] or Isreal [7], just to name a few.

Over the last few years, more and more data sources have been published as open data that cover not only a single country, but the whole earth. While the level of detail in such datasets might be less compared to national or regional datasets, the worldwide availability allows to create a data processing pipeline for building MATSim models for essentially every place on earth.

Creating models on worldwide data

Creario is a new service from Simunto that allows the user to quickly create MATSim models — worldwide. It uses several datasets with worldwide coverage, but also allows users to upload local data to improve the quality of the generated models.

The initial road and public transport networks in Creario are created from OpenStreetMap [8] data, but are then enhanced with data from AWS Terrain Tiles [9] (for node elevations and link gradients) and GHSL [10] (for determining if a link is in a rural or urban context). Combining data from all these datasets allows estimating very detailed link capacities and speeds based on a link's classification, curviness, gradient, and location. The generated multimodal networks are thus not only suited for the simulation of private car traffic, but also for bicycle trips and electric vehicles where gradients and the road surface have a higher impact on route search.

Public transport data in the form of GTFS datasets can be uploaded by users. Alternatively, Crearío provides access to over 1.000 GTFS feeds from all over the world and allows users to simply select those covering their model area.

For the synthetic population, data from GHSL [10] and WorldPop [11] is used. The combination allows for a realistic spatial distribution of agents as well as reasonable age and gender attribute values. To improve the accuracy of the generated population, users can upload custom spatial data that describe the distribution of the population within the model area along with optional age and gender distributions. Demand (i.e. agent plans) is created based on data from National Household Travel Surveys. Currently, users can select between NHTS data from France and the USA, which are both available as open data. Additional surveys might be included in the future, allowing for richer and better adapted travel behavior in the generated models. Location choice for the activities in the agents' plans is done based on facilities extracted from OpenStreetMap.

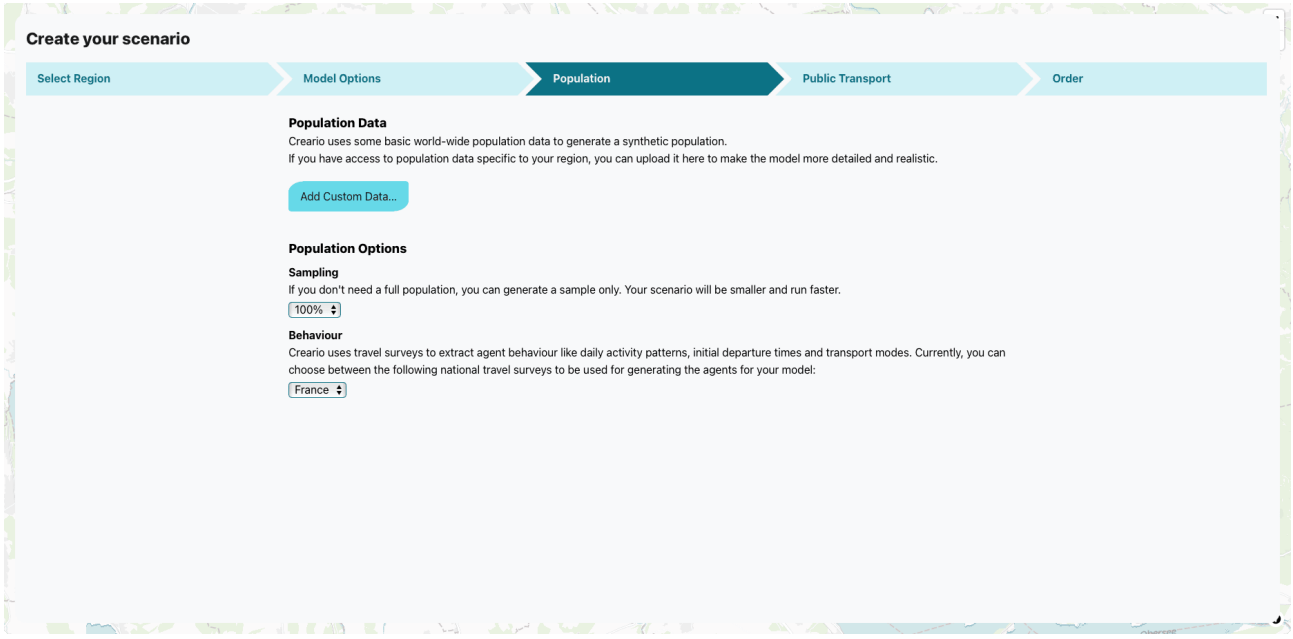

Figure 1: Crearío allows users to allow custom data describing their population.

Conclusions

The models are created by a fully automated pipeline based on datasets that provide a worldwide coverage. They will thus not be able to compete with manually created, hand-calibrated models. While the models might not be used for detailed transport planning without further refinement, they provide a good, fast and cost-effective base for such models. In addition, the created models are typically good enough for research projects that want to demonstrate certain effects applied to a specific region. To assist transport planners working with the generated models, each model comes with an extensive report that includes the results of numerous validation steps performed, highlighting potential problems or improvements that can be made to the models.

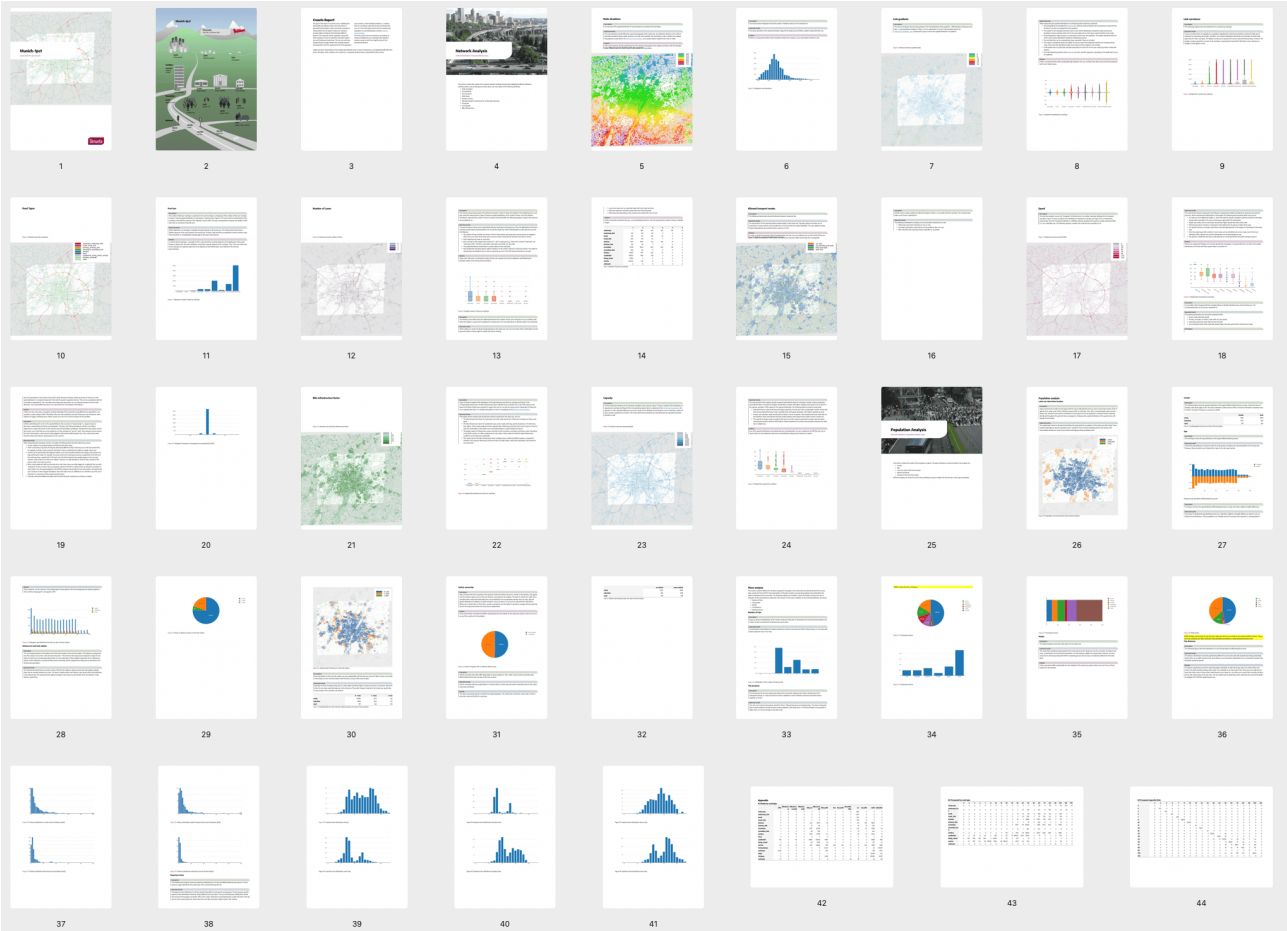

Figure 2: Overview of a generated model report.

References

1. Hao, J.Y., M. Hatzopoulou, and E.J. Miller (2010). Integrating an Activity-Based Travel Demand Model with Dynamic Traffic Assignment and Emission Models: A Case Study of the Greater Toronto Area. Presented at 89th Annual Meeting of Transportation Research Board, Washington, D.C., 2010. http://dx.doi.org/10.3141/2176-01
2. Ziemke, D., K. Nagel, and C. Bhat (2015). Integrating CEMDAP and MATSIM to increase the transferability of transport demand models. Transportation Research Record: Journal of the Transportation Research Board. 2493. 117-125. 10.3141/2493-13. http://dx.doi.org/10.3141/2493-13
3. Winter, M. K. E. (Creator), S.J. Narayan (Creator) (26 Nov 2019). Amsterdam Scenario MATSim. TU Delft - 4TU.ResearchData. https://doi.org/10.4121/UUID:6108ED85-7B24-455E-BD95-89D84E6306FA
4. Hörl, S. and M. Balac (2021) Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data, *Transportation Research Part C*, **130**, 103291.
5. Müller, J., M. Straub, G.Richter, and C. Rudloff (2022). *Integration of Different Mobility Behaviors and Intermodal Trips in MATSim*. Sustainability. 2022; 14(1):428. https://doi.org/10.3390/su14010428

6. Sun, L., A. Erath, M. Cai (2018) A hierarchical mixture modeling framework for population synthesis, Transportation Research Part B: Methodological, Volume 114, 199–212. https://doi.org/10.1016/j.trb.2018.06.002
7. Bekhor, S., C. Dobler, K.W. Axhausen (2010) Integration of activity-based with agent-based models, Arbeitsbericht Verkehrs- und Raumplanung. https://doi.org/10.3929/ethz-a-006121544
8. https://www.openstreetmap.org, accessed April 2025
9. https://registry.opendata.aws/terrain-tiles/, accessed April 2025
10. https://human-settlement.emergency.copernicus.eu/datasets.php, accessed April 2025
11. https://www.worldpop.org/datacatalog/, accessed April 2025