

# Regional Replica FAQs

## Replica Overview

### ***What is Replica?***

Replica is a synthetically-generated representation of the movement of people and goods in a given area for a typical weekday and typical weekend during a given season. There are three ways to access the data.

1. **Reports** – Ready to use reports on specific topics, customizable with a few clicks.
2. **Explorer Projects** – A web-based interface for querying and visualizing Replica's data.
3. **Replica Data Downloads** – A variety of data files summarizing Replica data that you can download for analysis using other software.

### ***What can Replica data be used for?***

Replica can be used for both in-depth transportation planning and across many other domains. Here are a few examples of how Replica can be used across various domains of city and transportation planning:

- **City Operations:** when should we schedule water main repairs?
- **Parks, Recreation, and Facility Management:** where are visitors to our facilities coming from? How can we improve access?
- **Health and Human Services:** which neighborhoods encourage active transit?
- **Transportation Planning:** where should we build a light rail line?
- **Transit Planning and Operations:** will increased frequency attract more riders in off-peak hours?
- **City and Neighborhood Planning:** which neighborhoods lack access to grocery stores?

### ***How is Replica created?***

Replica creates a synthetic population. We use government provided census data as well as other sources that contain sociodemographic and locational information about residents living in a given area, regional housing and employment availability. We apply modeling and optimization algorithms to generate the synthetic population that is statistically similar to the census population in person-level attributes, individual households composition, as well as in aggregate. These synthetic people and households are assigned housing units and locations of workplaces and schools.

Replica then creates personas. Personas extract behavioral patterns from de-identified mobile location data collected from mobile device(s) of real people. Persona training is informed by the region's spatial and socio-demographic data as well as other auxiliary sources of human activity

in the region. They are composed of three main underlying behavioral choice models: activity scheduling, destination location, and travel mode. Each synthetic person, with its assigned persona and its travel behavior models, is “motivated to travel.”

Finally, we simulate a given day of the representative week to recreate the travel patterns for the entire metropolitan area. The output of this stage is the Replica activity table, which contains a synthetically-generated travel itinerary for each synthetic person on a given day of the season and serves as a single consistent source of all other Replica outputs. **It is not possible to re-identify any person from the sample locational data from the Replica activity table.**

At the validation stage, the ground truth data collected by the agencies within the region can be used to compare observed travel metrics with Replica outputs.

### ***What areas do Replica cover?***

Each region in Replica covers specific counties known as the core area. To see the region definitions, please see the Region Definitions document available in the Help section. In addition to the core area, Replica models certain buffer areas that have high traffic into the core area. This area is determined when the region is built. Replica accounts for movements from the buffer area to the core area, but not from buffer area to buffer area.

## Replica Data Questions

### ***How is Replica's data generated?***

The data is generated using Replica's advanced modeling tools that use a composite data set including public use population census data, proprietary locational data from telecommunications and other IT infrastructure in the region, and field observations data from customer public agencies (ground truth data). For more information, please see Replica's Methodology document available in the Help section.

### ***What data is used to represent a population in a given region?***

Replica uses samples of census demographic data, such as Public Use Microdata Samples (PUMS) and ACS data, to create a “synthetic population” that is statistically representative of the actual population in a region.

### ***How are movement activities determined in Replica's model?***

Replica creates personas to determine movement activity in the model. Personas extract behavioral patterns from de-identified mobile location data collected from mobile device(s) of real people. They are composed of three main underlying behavioral choice models: activity scheduling, destination location, and travel mode. Each synthetic person, with its assigned persona and its travel behavior models, is “motivated to travel.”

## **What is a Replica activity table?**

A Replica activity table is a data table describing all of the activities (including travel) by a synthetic population within a defined period of time, usually a typical week in a 3-month period (a season). The synthetic population is generated using the methods described above.

## **What types of mobile location does Replica use?**

Replica uses de-identified mobile location data to model individual travel behaviors, what we call personas. The below mobile location data is used at various stages of the Replica pipelines.

- Location based services (LBS) data: As people move around with their phones in the real world they use mobile apps that rely on their location. Users opt-in to sharing their location when using the apps.
- Cellular networks data: Wireless telecommunication hardware (cellular networks) records abundant technical data while maintaining connection with user's devices. The use of longitudinal de-identified data of device's locations by any third parties is described by the end-user license agreements (EULAs) of the carriers.
- Vehicles in-dash GPS data: The data on vehicles speeds and locations geo-matched to a particular road segment (telematics data) collected by GPS systems is transferred to the centralized data storage and processing systems that monitor real-time congestion. EULAs regulate the use of these telematics data by third parties.

Replica can also utilize Point of Interest aggregates as well as credit transactions data for ground truth activity data.

## **How does Replica take data privacy into consideration?**

Replica sources anonymized third party data that has been further replaced with masked identifiers, so our data sets remain intact while reducing potential re-identification concerns, to create: (a) a synthetic population that matches aggregate statistics of a region's real population, (b) estimates of aggregate movement patterns across coarse-grained locations (c) a travel behavior model that encodes a set of rules representing who moves where, when, why, and how. These components as a whole yield a "Replica" of likely transportation patterns across the region. Further, Replica has entered into contracts that govern our ability to use, share, and secure various de-identified third party data that is provided to us.

Replica:

- Uses de-identified data and/or applies de-identification measures to our data platform. These measures further strengthen the privacy of the data we source.

- Minimizes the risk of re-identifying specific trips or people. Replica's synthetic data closely matches aggregated statistics, but is not intended to match any specific underlying person in the original, ingested data.
- Does not attempt to re-identify individuals from its data sources and our Terms of Use prohibit our users from doing so as well. In particular, Replica does not join data sources on keys containing sensitive data.
- Independent models are built on different data sources in order to abstract out identifying details of any given individual before combining with other data sources.

***How do you account for underrepresented populations, like those without access to mobile phones, or those who don't participate in the census?***

Replica's synthetic population includes all people and is calibrated to the most recent Census ACS estimates. It is of course possible and sometimes the case that not all people are represented in the census, which is why our calibration to census does include small margins of error. Replica does not require seeing all mobile devices in a given geographic location, so access to mobile devices is not necessary for scaling Replica samples to represent movements of all people.

***If 7+ days of overnight stays at an identified home location is required, does Replica capture homelessness population?***

The Census does attempt to capture homelessness populations but there are known issues. People experiencing homelessness do have phones so we likely are seeing some of their movement.

***How is Replica monitoring relocation or seasonal movement for example interns or students?***

Currently the synthetic population matches most recent ACS data (as of now it is ACS 2018) and is updated annually. While there is seasonality in travel behaviors, changes in seasonal populations (such as seasonal workers) are currently not reflected with an exception of reducing school and college attendance to 20% in summer as compared to the usual period.

***What is the maximum amount of trips in which I can download data from in Reports and Explorer?***

Downloads are available in a CSV format and are limited to 1 million rows in Reports and Explorer.

***Can I upload my own data sources to Replica?***

Not currently. However, this is a functionality we are exploring.

## ***Can I use Replica to forecast?***

Replica is the first step in a virtuous cycle of planning and operational tools for cities. Replica enables customers to understand “what is” – what is happening today and in the near past? By providing current conditions and previous seasons, users have better data to make more informed decisions for the future. Seasonal updates allow users to see trends that can be projected forward.

## ***What type of ground truth data do you use to calibrate Replica's model against?***

Each region is unique. The below list includes the data points that can be used when calibrating the model.

- Bridge/Tunnel Data
- Custom Geos
- DMV Data
- GTFS
- General Sensor Count
- Non-Motorized Data
  - Pedestrian counts
  - Bike counts
  - Bikeshare
- Parking Data
- Survey/Model Data
- Traffic Data
  - Traffic counts and speeds
  - Regional O/D pairs
- Transit Data
  - Transit line counts
  - Transit Stop Boardings/Alightings
  - Railway
- TNC/Taxi Data
  - TNC data
  - Taxi data

## ***What are the most helpful types of ground truth data sources?***

GTFS, Interstates and Transit ground truth data sources are critical to starting the build of the model.

## ***How is land use data used in Replica?***

Land use and building footprints data are included in the large composite of data that Replica uses to ensure the product is an accurate reflection of the real world. Built environment data sources are utilized in the three areas outlined below.

- A combination of the land use attributes and an observed recurrent pattern of stay at most frequented locations is used to identify home and work of a device owner. Land use, popularity score of a set of venues at a given time, and the context of the stay time period within an individual's activity sequence are used to assign likely activity types to all secondary stay points.
- Land use and longitudinal employer-household dynamics (LEHD) employment data are used to redistribute commute flows between census block groups and census tracts by given industry sector. Using a similar process of locating housing units from relative occupancy weights, mobile location data are used to assign workers to the office units based on building footprints (or parcel centroids) within census block groups.
- Data services that provide auxiliary information (demographic information, land use, accessibility) for training travel behavior models are separated and deployed in isolated projects and are initialised with a separate copy of databases, isolating potentially sensitive query logs.

## ***Can you explain the resident/visitor status?***

The Resident/Visitor Status module shows the number of people who live in the core Replica region, those who live in the buffer region, and those who are visitors to the region.

All residents of the core region who make a trip on the selected day are included in the data. Only residents of the buffer region who work in the core region and make a trip on the selected day are included.

Visitors are people who do not normally live or work in the Replica region, and either stay overnight in the region, or enter and exit via a "port of entry" (usually an airport) the same day.

## ***How does Replica show domestic and international travelers both residents traveling outside the region and non-residents traveling into the region?***

Currently, Replica models movements by residents of the core region, residents of the buffer region and visitors. When downloading the population data from an Explorer query, you will see the 'resident type'. If the persona is within the core region is a resident this information will be included. If this is a visitor to the core region coming from the buffer zone, the demographic information will not be included.

We will look into tracking domestic and international travelers to show more clarity within the tool.

### ***Why can't I see demographic information for visitors?***

Replica does not include the demographic characteristics (such as age and income) of visitors. When you filter to view visitors only, you will not see any data in the demographic modules, and if you filter using a demographic characteristic you will not see any visitors.

## Defining Trips

### ***How do you define a trip? How do you define the end of a trip?***

A trip is a movement by a person between places. A trip begins when a person leaves a place, and ends when a person stops to do a non-travel activity in a place. For example: If a person walks from home to a cafe, buys a coffee, and then walks to work, two trips have occurred.

A person can use multiple modes within a single trip. For example: If a person walks to the bus stop and then takes the bus to work, this is a single trip with two trip segments. When a trip involves multiple modes the primary mode is determined using the following ranking: 1) Public transit, 2) Driving (private auto)/Auto passenger/Taxi/TNC, 3) Biking, 4) Walking.

### ***How do you determine trip purpose?***

Replica uses a trained location choice model (LCM) to determine locations choices for discretionary activities (e.g. not home/work/school) made by the device owner of the de-identified mobile location data. The model selects individual venues (businesses, shops, services) and Points of Interest (parks, places of historic interest, tourist attractions) as potential destinations. It is a model that for a given set of alternative destinations represented as attributes of particular venues in the area, ranks them based on the observed choice made by the owner of the device considering every other location ever visited by a person as an alternative. LCM includes contextual variables such as distance to home, distance to work, a score describing the deviation from a regular home-work commute, the hour of the day, day of the week, allocated travel time, and duration of the next scheduled activity. Based on the locations determined from the model outputs, the trip is assigned a purpose.

### ***What types of trips do you model?***

The purpose of a trip is determined by the type of place at the destination and other contextual information. For example: If a person is traveling to work, the purpose of the trip is 'Work'. If a person is traveling to a restaurant, the purpose is 'Eat'.

- **Home** includes all trips to a person's own home.
- **Shop** includes all trips to shops.
- **School** includes all trips to a person's school or college.

- **Work** includes all trips that ended at a person's workplace (for example: return trips from lunch), not just commute trips.
- **Eat** includes all trips to restaurants.
- **Errands** includes trips to hairdressers, auto shops, banks, and a variety of other locations.
- **Recreation** includes trips to recreational destinations such as parks and swimming pools. Replica does not include looping trips without a destination, such as walking the dog, or jogging.
- **Commercial** refers to trips by light, medium, and heavy trucks for deliveries and other commercial purposes.
- **Lodging** refers to trips by visitors to overnight accommodation such as a hotel.
- **Social** refers to visits to friends' home.
- **Region departure** refers to trips by visitors to a "port-of-exit", such as an airport, or major train station.
- **Pass-through traffic** refers to trips made by non-residents that start and end outside the region. These trips can include short stopovers within the region.

### ***If Replica doesn't model trips taken by children 5 and under, what happens to daycare / preschool trips?***

Correct, residents below the age of 5 do not receive a persona. All their travel is assumed to be represented by the travel of an accompanying adult from the household. Likely, personas include trips to daycare en route to work. For example, an accompanying adult persona would be that they start at home, drop kids off at daycare, go to work, run an errand after work, pick kids up and go home.

### ***Do you account for trip-chaining?***

Travel between two locations with a discreet purpose at the origin and destination. Some intermediate stops with short dwell time may be captured within a single trip. We do not provide definitive criteria for this dwell time.

Example: Home > Work, Work > Gym, Shop > Eat, Home > Starbucks Drive-thru > Work

### ***Who is included in the trips?***

Currently, Replica includes trips by residents of the metro region within the map outline. We also model visitors in a given area based off of the detection of a permanent home location outside of a region's boundaries along with several observed overnight stays within the region.

### ***How does Replica determine which direction bike/pedestrian trips take?***

Pedestrians take whichever side is faster based on their start/end points. The direction could be determined by origin and destination.

## Calculation

### ***How does Replica calculate vehicle miles traveled?***

Vehicle miles traveled (VMT) are computed by aggregating distances traveled by vehicles on the road network. Whenever relevant, VMT metrics are further explained in the information pop-ups in the UI.

VMT, BMT (bike miles traveled) and WMT (walk miles traveled) are available fields in the Resident reports and can be accessed in Explorer by applying the right filters. In Explorer, you can use the “Distance Summary Metrics” filters to find the total distance of different types of trips. For example: Use the Home Location filter to select an area, then use the Primary Mode filter to select 'Driving'. This will give you the total private auto VMT for residents of the area. Use the Destination filter to select all trips ending in an area, then use the Purpose filter to select 'Eat, Shop & Errands'. This will give you the total distance traveled by all modes, for trips to the area, for the selected purposes.

### ***How can I determine Single Occupant Vehicle (SOV) trips?***

Replica models both driving (drivers of auto vehicles) and auto passenger (passengers of auto vehicles) trips. To get an approximation of SOV trips, take the total number of trips for that given day that displays in Reports or Explorer based on the criteria in which you’ve filtered and subtract that number by the number of auto passenger trips shown.

### ***Is there a way to compare travel patterns?***

The best way to compare travel patterns to and from two locations is to look at the related data downloads from O/D queries run in Explorer. Download the two related queries and combine the data for comparison of underlying trips and unique people.

## Modes

### ***How does Replica determine mode of travel?***

Replica’s mode choice model consists of two distinct components, mode inference model and a mode choice model. Mode inference model takes into account geographical location of origin and destination, accessibility, and observed travel times and distances of the activities extracted from the mobile location data. The second component, mode choice model, gives the likelihood of choosing a given mode out of the available alternatives based on origin and destination with their respective start end times of travel, routes are computed for each possible mode (private auto, walking, bicycling, public transit with different access/egress types, and on demand auto) and are included into a choice set for a given trip.

For more information on mode choice model, please see Replica's Methodology document available in the Help section.

### ***How does Replica define "primary mode" and handle multi-modal trips?***

Many trips use multiple modes, such as walking to a bus stop and then riding the bus. Replica faithfully models these trips. For example, if someone takes a 2 mile bus ride to connect to a 10 mile rail trip this is a single trip with two segments (bus to rail stop, rail stop to destination) and Replica will model both of these.

### ***What kind of modes does Replica model?***

Below are definitions of modes used in Replica.

- **Driving (private auto):** Trips made by drivers in private auto vehicles. This is equivalent to the number of private auto vehicle movements.
- **Auto passenger:** Trips made by passengers in private auto vehicles. Combine this number with the number of private auto trips to get the number of people who traveled in private autos.
- **Walking:** Trips made by people walking.
- **Biking:** Biking only trips. Replica does not model scooter trips and does not separate out e-bike trips.
- **Public Transit:** Trips that primarily used public transit. For example, buses, light rail, and subways.
- **Taxi/Transportation Network Company (TNC):** Trips made by passengers in a Taxi or using a TNC such as Uber or Lyft. These are also known as For Hire Vehicles, On-demand Autos, Ride-shares, or Transportation Network Providers (TNP).
- **Commercial (freight):** Trips made by light, medium, and heavy trucks.

### ***What about other modes such as private shuttles, paratransit, scooters?***

As additional data becomes available, Replica will add new modes. Currently, these modes are not modeled by Replica.

### ***How do you model TNC?***

TNC encompasses both Taxi and On-Demand cars (Uber, Lyft, etc.) Mobile location data alone does not allow reliable identification of trips made by an on-demand auto (vs., for example, a carpool). It is preferable to use auxiliary ground truth data sources to represent travel by taxi and TNCs more accurately.

### ***What type of transit is included?***

Replica models all transit modes available in the region provided we have GTFS, this includes bus, rail, light rail, subway, ferry and even gondola.

## ***How does Replica count transit transfers and accessing transit?***

A transit to transit transfer would count as one transit trip in Replica. We use the concept of 'linked' trips.

We do not currently model access to transit, like walking to a transit station. The entire trip would have a primary mode of transit and it would count as a single trip.

## ***How do you model freight?***

Replica uses location data provided by fleet management and telematic companies to capture freight flows. These inputs are then used to train commercial vehicle personas, much like individual / resident personas are trained. Additionally, like resident movements, commercial movements are then simulated within the context of a given day and season in Replica. Where available, ground truth is used as a source of calibration for commercial travel as well.

## Reports

### ***What is a mobility report?***

This report is about the trips that ended in a particular area. It covers the trip characteristics such as mode and distance traveled. You can also learn about the demographics of the people who made the trips including the home location and household income.

### ***What is a resident report?***

This report is about the residents of a selected area. It covers their demographic characteristics such as race and ethnicity and household income. You can also learn about their travel behavior including the modes used and distances traveled.

### ***What is a worker report?***

This report is about the people who have a workplace located in a particular area. It covers their demographic characteristics such as race and ethnicity and household income. You can also learn about their commuting behavior including the modes used and distances traveled.

### ***What is a transit report?***

This report is about the movement patterns and associated information for people using selected transit routes. It covers their demographic characteristics such as race and ethnicity and household income as well as car ownership. You can also learn about their travel behavior including trip purpose and trip distance.