The Challenge of Journey Data and Privacy

Regardless of who holds it, journey data poses unique privacy challenges. "Bread-crumb" data, for example, tracks a person's movement when location services are enabled on a phone. The GPS unit on the phone "pings" every few minutes or seconds, creating a detailed map of the route a person takes to walk to the grocery store or the speed at which they move. This information can be used to make informed decisions about transit service allocation or safety improvements. However, those same "bread-crumbs" could track a person to the doctor, a political rally, or a job interview. Such information is fundamentally private.

As discussed in ***Managing Mobility Data***, a guidance document co-developed by NACTO and IMLA, advances in data science and the huge increase in the volume, precision, and ubiquity of data mean that journey data is or can easily become personally identifiable information (PII). This happens in two ways:

Recognizable Travel Patterns – Even in anonymous datasets, people can be easily re-identified from their routine travel patterns – e.g., from home to work, school, stores, or religious institutions. The 2013 Scientific Report article, "Unique in the Crowd: the privacy bounds of human mobility" found that, in a dataset of 1.5 million people over 6 months, and using location points triangulated from cellphone towers, "four spatio-temporal points are enough to uniquely identify 95 percent of the individuals."[62]

Combined With Other Data – Journey data can be combined with other data points to become PII. For example, taken by itself, a single geospatial data point like a ride-hail drop-off location is not PII. But, when combined with a phonebook or reverse address look-up service, that data becomes linkable to an individual person. For example, in 2014, a researcher requested anonymized taxi geo-location data from NYC Taxi and Limousine Commission under freedom of information laws, mapped them using MapQuest, and was able to identify the home addresses of people hailing taxis in front of the Hustler Club between midnight and 6 am. Combining a home address with an address look-up website, Facebook and other sources, the researcher was able to find the "property value, ethnicity, relationship status, [and] court records" of individual patrons.[63]

Today's data management choices will impact the world we live in tomorrow. The public and private sectors alike should look to develop data practices and policies that increase the amount of information available for planning and policy making, while simultaneously increasing privacy protections and ensuring that data is protected and managed appropriately.

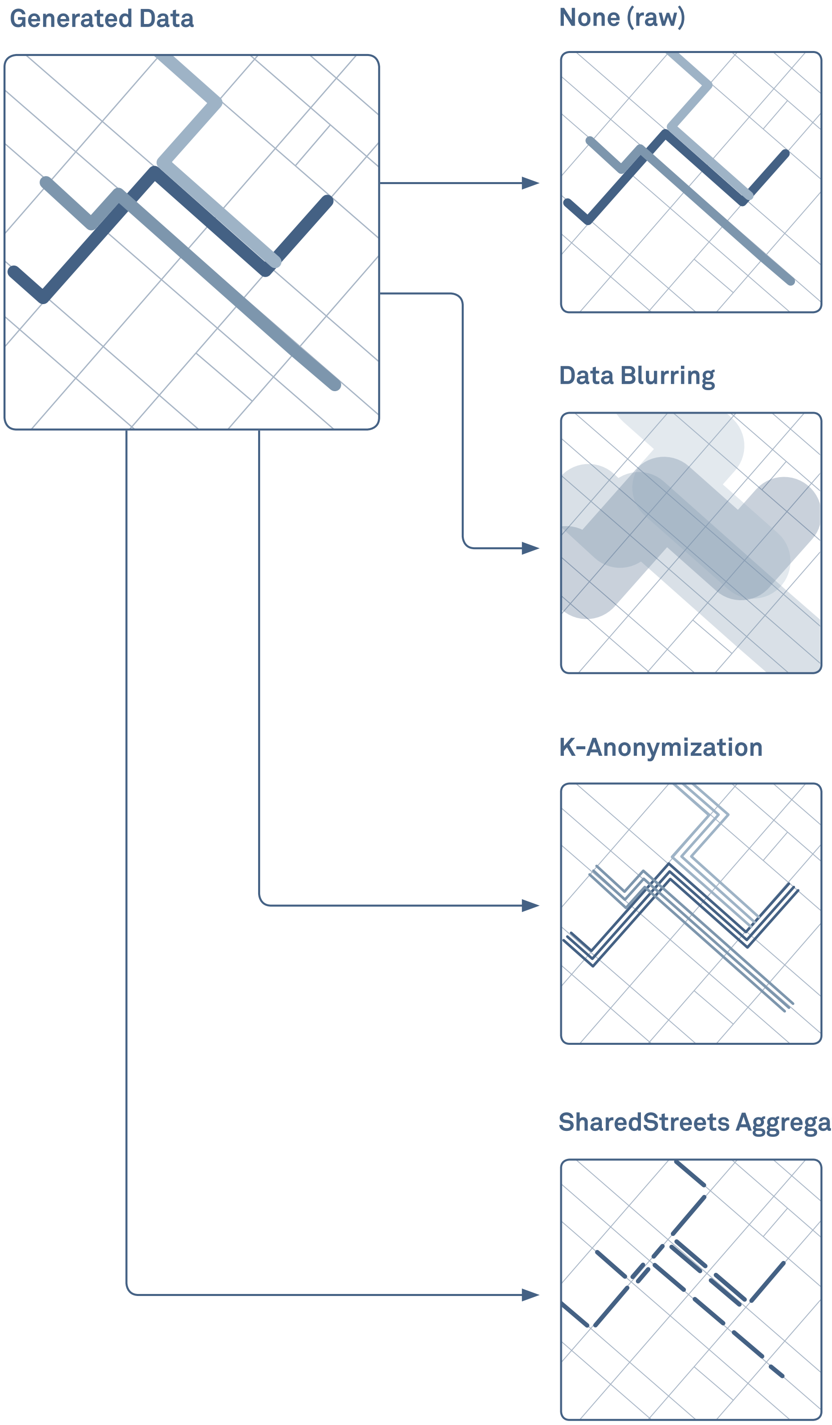
On the public sector side, cities must strengthen their data management and analysis capacity, recognizing that not all data analysis or aggregation methods are the same when it comes to protecting privacy or providing useful policy-making or planning information. Cities should also retool procurement and development processes to prioritize open standards to avoid getting locked into proprietary systems that may be unsuited to properly address privacy or planning and regulatory needs. As cities gather additional essential mobility data, they should work to educate lawmakers and attorneys on the ease with which mobility data can become PII to prevent inappropriate disclosures.

As the age of autonomous vehicles approaches, revelations about the data (over)-collection and loose handling practices of internet giants like Facebook[64], Google[65], and The Weather Channel[66], should be treated as a wake-up call. U.S. citizens lack federal-level data privacy protections, creating a state-by-state patchwork for protection. In response, calls for a "data bill of rights" are mounting.

The European Union's General Data Protection Regulation (GDPR) provides a good example of active government intervention to address privacy. First enacted in 2016, the GDPR defines basic protocols for protecting a person's privacy, including guidelines to limit the over-collection of data, rules for informed consent, and policies for anonymization, storage, and access. The GDPR is meant as a safeguard against the abuse of data by both private and public actors, who may be able to access personal information for personal use, abuse, or enforcement.

Data Anonymization Methods

Different data anonymization methods produce different results for analysis and privacy.



Generated Data

None (raw)

Data remains unprocessed. Individual trips can be tracked from start to finish creating significant privacy and liability issues for data holders.

Data Blurring

Aggregators remove decimal places from the latitude/longitude coordinates that make up each point in a GPS route (i.e. 40.6893002,-74.0444091 becomes 40.689,-74.044). Privacy issues can be reduced by decreasing the overall precision of the data itself.

K-Anonymization

Aggregators hold full unprocessed trip data until they gather enough identical trips to batch together. Data is then aggregated, and unprocessed records are deleted. Individual trip information can be accessed for the duration of time it takes to gather identical records.

SharedStreets Aggregation

Aggregators snap individual GPS points to individual street segments but divorce those points from other information about the trip in totality, such as origin or destination. Data precision remains high but an individual trip cannot be traced from start to finish. SharedStreets applies k-anonymity to data at the precision of street segments.