Where in the world is Joe Geo?
The Importance of Good Place (POI) Data

BY ALISTAIR GOODMAN, GENERAL MANAGER OF EMODO

Knowing a consumer's real-time and historical location is only half of the equation for mobile advertising. Gaining any type of intelligence about that consumer -- from building audience segments to simply using a geofence to trigger a proximity ad -- also requires good point of interest ("POI") data, the description of the physical context of the user. POI data may not be as sexy or creepy as locating a user via their phone, but it is equally if not more important for the mobile-location advertising ecosystem.

A POI data set is a hyper-correct data representation of the physical world. There are as many pitfalls associated with this type of data as there are with user location data (insert parenthetical text of first byline here). Get it wrong, and you risk sending a consumer to a place that doesn't exist, adding them into the wrong audience segment or even crediting a consumer as visiting a store for attribution purposes when they didn't. Moreover, bad POI data could lead to delivering an ad in the wrong place, irrelevant for the consumer, and charging the advertiser for doing it.

In this 2nd article in our series, we will dive into why POI data is important, the challenges of using POI data, and how this data can be effectively collected and normalized for use in mobile ad campaigns.

WHY CARE ABOUT POI?

Perhaps the best way to illustrate the importance of accurate POI data is with a specific example. Appliance company Electrolux wanted to target rich-media mobile ads to drive store visits and sales for Frigidaire Professional products--accurate POI data enabled a number of key components in this campaign.

First, highly accurate POI data ensured that all the dealer locations utilized in the campaign were correct, making it possible to create geofences around those stores and trigger ads only when a consumer was found in a geofence around the store. It also ensured that the correct address for the closest store appeared dynamically in the ad unit, enabling a consumer to easily navigate to the store. Lastly, it ensured that additional store data, such as hours and available inventory, were correct.

These same geofences around the store POI were then used to construct audience segments of likely shoppers. Other geofences around locations selling high-end appliances created additional segments of in-market shoppers. And, tight geofences around specific dealer locations were used to measure the lift in foot traffic for

consumers exposed to the advertising vs. a control group. All of these activities are built on a foundation of accurate, POI data.

In sum, accurate POI data made it possible for Electrolux to target relevant consumers at the appropriate time and place with no inaccuracies, while also allowing the company to collect valuable data about their customers behavior in the physical world and about the impact of their campaign.

example, a recent large dataset we analyzed still had Lord & Taylor stores listed in Atlanta. Lord & Taylor closed all 32 stores in the Atlanta metro area in 2003, yet it was still in the dataset more than 12 years later! These challenges mean that, despite its value, reliable POI data is often treated as a non-issue when in fact it is a critical building block of effective mobile-location advertising campaigns.

POI CHALLENGES

Like many types of specialty data, working with POI data creates its own unique challenges. To start, a single building can contain multiple business locations. Think of a large office building which may have multiple businesses on each and every floor. Figuring out which specific business a consumer has visited is not easy.

There may also be multiple semantic references to a single place. For example, 1 Bryant Park = 1111 Avenue of the Americas = The Bank of America Building = The Starbucks on 42nd and 6th. In addition, businesses are always evolving -- locations open, close, move, or change identifiers, like phone numbers or owners. And of course, you can't rule out human error, and incorrect POI data entry can lead to inaccurate or transposed POI data. In our experience, a hefty 20-40 percent of any dataset about places is wrong, even when a retailer themselves sends you a list of their store locations.

Vendors of merchant data without technology for accurately cleansing POIs contribute to the challenge by propagating out-of-date or user-sourced data. For example, a recent large dataset we analyzed still had Lord & Taylor stores listed in Atlanta. Lord & Taylor closed all 32 stores in the Atlanta metro area in 2003, yet it was still in the dataset more than 12 years later! These challenges mean that, despite its value, reliable POI data is often treated as a non-issue when in fact it is a critical building block of effective mobile-location advertising campaigns.

COLLECTING AND NORMALIZING

Unlocking the full potential of POI data requires technology that can first ingest, clean and normalize big data sets, then build context for the data depending on the particularities of a local market. Ultimately, it yields building a reference set for real-world places where each POI is as accurate as possible — what we refer to as a canonical record for a place.

On the data side, the ability to ingest and combine multiple different sources of data about places — stores, parks, transit hubs, etc. — provides the raw material. Multiple sources are preferred for several reasons; most notably because businesses are not constant, fixed entities. As mentioned above, they open, close, relocate, rename, or update phone numbers, which is why it's common to receive store listings from the merchants themselves that are riddled with mistakes. Whether data was entered incorrectly by a human or something significant about the business has changed, multiple raw data sources creates a richer clean data set.

User-generated POI data, however, create unique problems that make it hard to work with. In particular, as we mentioned above with many possible references to a place, the problem increases geometrically when many humans can change a place record. Even when "super-users" are involved in manually correcting data, multiple users with the ability to edit place data in a system creates more challenges than it solves. Instead, machines need to be able to systematically process and score place data first, leaving the exceptions to be reviewed by humans as necessary.
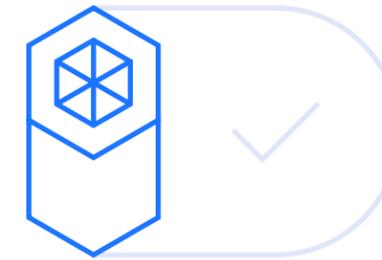
On the technology side, a system needs to be able to clean, normalize and standardize multiple different potential references to a place. These systems need the capacity to filter and score data sources and geocode locations correctly in order to create the place references. The key here is a system that can ultimately create a "canonical record for a place" — an always correct, continuously updated record. This reference can then enable other references to remain in the nomenclature of that system (i.e. 1 Bryant Park and the Starbucks on 42nd and 6th can coexist, and the system knows that they are both correct references to that place).

Included in this is the ability to correctly attach and manage both static and dynamic elements to a place. For example, Madison Square Garden is a very different "place" when the Knicks are playing than when Rihanna is performing. This is a key concept for creating place profiles — richer understandings of how a place changes over time.

Accurate place data leads to more effective mobile advertising campaigns. It's that simple. Sure there are challenges, but there are also significant rewards. POI data enables advertisers achieve the Holy Trinity of delivering the right message to the right consumer at the right time

EMODO INSTITUTE

POI Data Issues
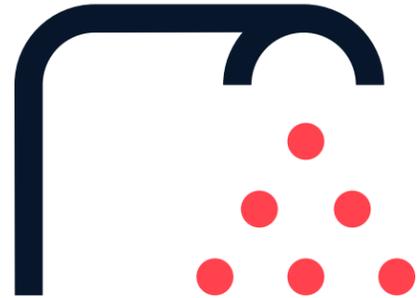
A single building can have multiple business locations, such as a large office building, which may have multiple businesses on each and every floor. Figuring out which specific business a consumer has visited is not easy.

There may be multiple semantic references to a single place. For example, 1 Bryant Park = 1111 Avenue of the Americas = The Bank of America Building = The Starbucks on 42nd and 6th.

Human error, and incorrect POI data entry can lead to inaccurate or transposed POI data.
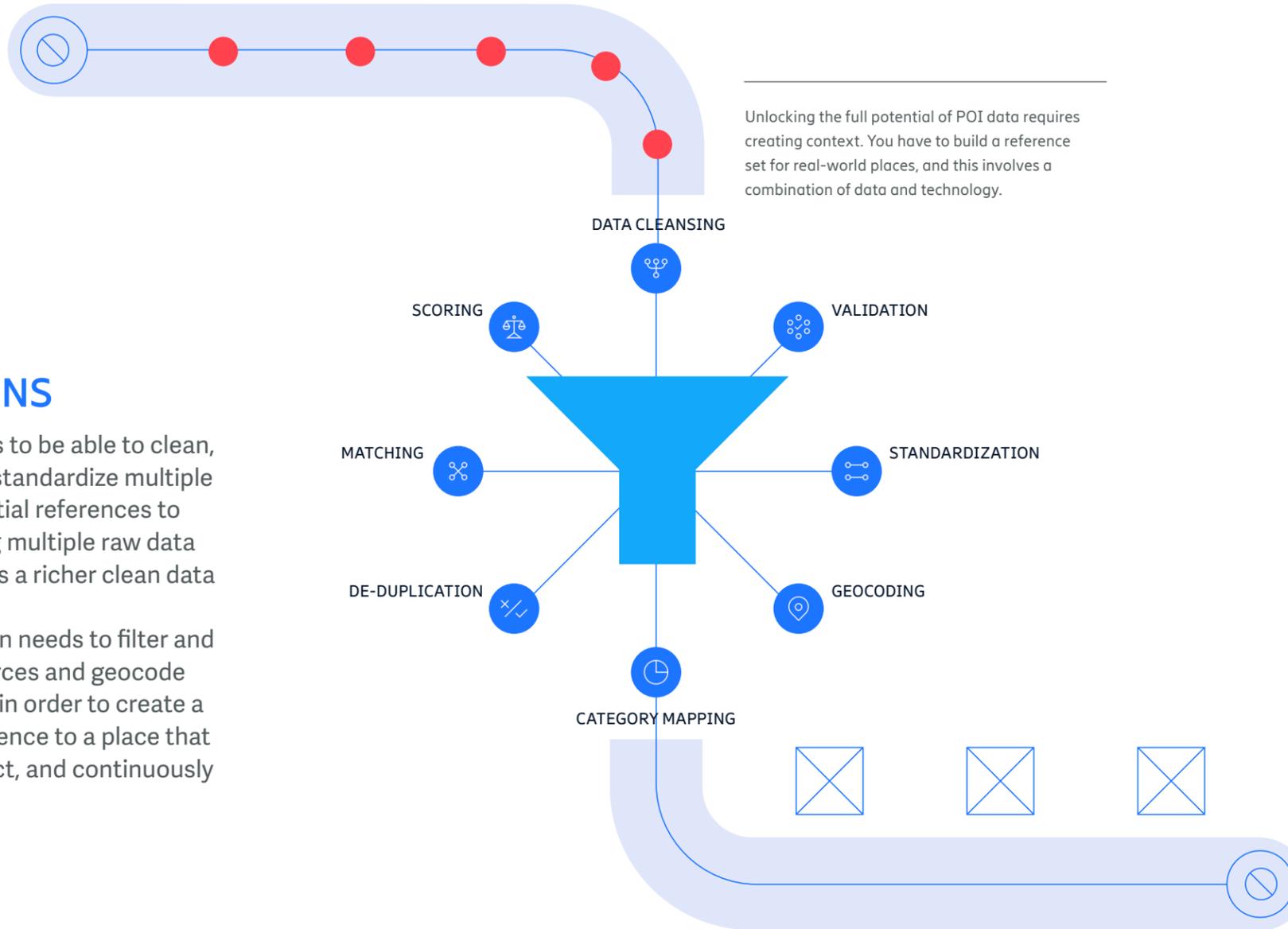
Businesses are always evolving — locations open, close, move, or change identifiers like phone numbers or owners.

Unlocking the full potential of POI data requires creating context. You have to build a reference set for real-world places, and this involves a combination of data and technology.

DATA CLEANSING

SCORING

VALIDATION

MATCHING

STANDARDIZATION

SOLUTIONS

A system needs to be able to clean, normalize and standardize multiple different potential references to a place. Getting multiple raw data sources enables a richer clean data set.

The system then needs to filter and score data sources and geocode them correctly in order to create a canonical reference to a place that is always correct, and continuously updated!

DE-DUPLICATION

GEOCODING

CATEGORY MAPPING

The result is an accurate place dataset that leads to more effective mobile advertising campaigns

EMODO INSTITUTE