**A data mashup for understanding fine scale human dynamics**

*A position paper for 2015 Summer Specialist Meeting on Representing Human Dynamics with Big Data, Social Media, and Social Networks in Hyperlocal Contexts*

Author: Atsushi Nara, Department of Geography, San Diego State University

A Data mashup is the integration of heterogeneous data from multiple sources as a way of producing new insights and services, a mashup is often done in a web platform. In the era of *Big Data*, a wide variety of data is available from public and private sector organizations as well as citizens via internet and social media, this data can explain human mobility in much more detail than ever before. This position paper addresses challenges and potentials of data mashups as a tool for moving knowledge toward the understanding of fine-scale human dynamics.

Understanding and modeling complex human dynamics is an important issue for many applications, ranging from urban planning, transportation management, and computational animation to physical, social, behavioral, psychological, medical, and geographical studies. The comprehension of dynamic characteristics and patterns of how individuals make decisions on their trip behaviors, how they physically move, and how they interact each other and their environment across scales facilitates the development of a realistic simulation model. Such a model is a useful tool for exploring simulation scenarios and investigating decision-making alternatives, for instance, destructive disaster events and emergency evacuations, disease outbreaks, epidemiological dynamics, and health studies and promotions.

In classical approaches, modeling efforts commonly look at human dynamics in a way that is relatively coarse, aggregate, static, and inflexible across spatial and temporal scales due to limited data availability. In the last decade, on the other hand, fine-resolution individual-scale data about human dynamics are becoming more available largely due to the advancement in location aware technologies and services (LATS), ubiquitously distributed geo-sensor networks, and mobile computing. These include but are not limited to Global Positioning System (GPS), WiFi positioning system, Radio-Frequency Identification Device (RFID), surveillance camera, and various kinds of portable smart devices with LATs, which enables us to collect a large amount of location data associated with time. Spatially and temporally fine-granular timestamped location data can reconstruct individual trajectories and describe dynamic movement behaviors in detail.

A data mashup based on fine-granular individual-scale trajectory data acquired from multiple sources provides new opportunities to explain and ultimately help modeling of complex human dynamics often in the form of model construction, calibration, and validation. Individual-scale data can avoid conventional data scaling problems such as ecological fallacy and modifiable areal unit problems by aggregating data from bottom-up. Fine-granular data can also explain spatially and temporally detailed movement behaviors, while multiple data sources collected from different populations can describe various movement behaviors. However, a data mashup is not a trivial task because multiple data sources are often heterogeneous in many ways including scale, format, structure and quality. Furthermore, to take the advantage of data heterogeneity for explaining various movement behaviors, the understanding of underlying data structure and characteristics as well as the awareness of data correlation and variation are required. Thus, a data mashup is challenging and involves complex data management and analytical procedures.

As a demonstrative case, this position paper presents a data mashup example for studying human mobility based on geotagged posts collected from three social media sources: Twitter©, Instagram©, and Foursquare©. The data consists of the total of 6,318,917 geotagged posts provided by 471,153 users in San Diego County, CA, between 2014/12/07 and 2015/05/17. This paper discusses data collection, data filtering, and analytical processes for revealing underlying social media data structures and characteristics related to human dynamics. Lastly, it facilitates discussions on how heterogeneous data from multiple sources could describe spatially and temporally fine-scale human dynamics in an integrated way by demonstrating a comparative study between the human dynamics extracted from three social media data and the commuting dynamics estimated based on U.S. Census Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES).