Data Integration to Create Large-Scale Spatially Detailed Synthetic Populations

Yi Zhu[1] and Joseph Jr. Ferreira[2]

Abstract

Many planning support systems and, indeed, some 'smart city' initiatives begin with time consuming efforts to integrate cross-agency data describing current conditions in sufficient detail to support 'what-if' exploration of urban development options. Integrating data from different sources has become increasingly challenged as available datasets, and the relevant urban modeling efforts, become more disaggregated and spatial-temporally detailed. Open data initiatives, with unprecedented amounts of embedded georeferenced information, have made web services and crowdsourcing attractive. However, data from such sources are typically imperfect and their integration is complicated by syntactic and semantic differences. We develop an ontology-based data integration mechanism to fuse data from different sources in generic ways that can utilize semantic information to minimize the labor involved and facilitate updating as new data are acquired. As a test application, we evaluate, filter, adjust, and integrate building information from heterogeneous data sources for use in an agent-based microsimulation model of transportation and land-use dynamics in Singapore. Third-party data about building size, age and use added substantial value to the official datasets generally available from government agencies.

[1] Y. Zhu (Corresponding author)
Singapore-MIT Alliance for Research and Technology
1 CREATE Way, #09-02 CREATE Tower, Singapore 138602
e-mail: zhuyi@mit.edu

[2] J. Ferreira
Department of Urban Studies and Planning, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA
e-mail: jf@mit.edu