NII Shonan Meeting Report

No. 2012-12

Social Issues in Computational Transportation Science

Monika Sester Ouri Wolfson Glenn Geers Masaaki Tanizaki Stephan Winter

December 17-20, 2012



National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Social Issues in Computational Transportation Science

Organizers:

Monika Sester (University of Hannover, Germany) Ouri Wolfson (University of Illinois at Chicago, USA) Glenn Geers (NICTA, Australia) Masaaki Tanizaki (Hitachi Ltd., Japan) Stephan Winter (The University of Melbourne, Australia)

December 17–20, 2012

Acknowledgements

The organizers wish to thank Hiroki Ohashi, whose notes of the discussions went into this report.

1 Overview of the Meeting

This meeting was bringing together researchers working in areas contributing to Computational Transportation Science (CTS). CTS is an emerging discipline that combines computer science and engineering with the modeling, planning, social, and economic aspects of transportation. The discipline studies how to improve the safety, mobility, and sustainability of transportation systems by taking advantage of information technologies and ubiquitous computing. In this seminar we focused on the social computing aspect of CTS.

Focusing on social computing was influenced by the many recent developments in social networks and crowdsourcing for transportation as well as the integration of persuasive technologies, behavioral economics in social computing. The researchers and practitioners from industry reviewed the development in this direction, discussed issues and solutions, and planned joint publications on CTS from the social computing perspective.

Social computing and information processing taps into the wisdom of crowds, and relies on the (ubiquitous) connectedness and communication ability of the members forming the society. Provided such infrastructure, cooperation in terms of computing and information processing becomes feasible and forms new research questions. While social computing is making its way into many disciplines, it is not obvious how real-time social interaction between mobile and stationary individuals (people, vehicles, goods, and infrastructure) can improve transportation. From a state where every individual is acting autonomously in isolation, or with minimal (visual) interaction with their environment, it is quite a paradigm shift to think of transportation as an interconnected, communicating and cooperating complex system. Such a paradigm shift brings up research questions in multiple dimensions including, but not limited to:

Managing Competition and Collaboration among Travelers. Travelers in a traffic network often need to make decisions about activities such as routing and parking. With the advent of location based services, wireless communication devices, and car navigation systems with real-time traffic, travelers now have the information to help them make these decisions. However, the models and tools needed to take advantage of this information are lacking.

For example, the route guidance methods that are used by car navigation systems are based on choosing a shortest path. But the shortest path could be sub-optimal if a large number of vehicles choose the same path, leading to a "herding" effect. This in turn will backfire, i.e. lead to a longer travel, because as the number of vehicles on a route increases, so will the travel time on the route. Similarly, a bus, a train, or an available parking slot that seem attractive to a traveler may become much less desirable if many travelers make the same choice. In other words, the pervasiveness of real-time travel information renders the existing tools inadequate, and the decision-making unsupported.

The inappropriateness of existing methods results from competition among travelers with real-time information. But the travelers also cooperate since much of the travel information is provided via some sort of crowdsourcing, a phenomenon that should be encouraged and incentivized because it increases the efficiency of the overall system.

Crowdsourcing. In a crowdsourcing system, services are provided by the users themselves rather than by a business or organization. The transportation information that may be crowdsourced includes real-time traffic information of road segments, information about car accidents and available parking slots, ride sharing opportunities, and so on. Crowdsourcing may be implemented using either the client/server model, or the peer-to-peer model, or a combination of the two. In the transportation environment, peers can be highly mobile and the peer-to-peer communication is subjected to disconnections when a short-range wireless technology is used. In this case, the peer-to-peer crowdsourcing introduces special challenges to incentive mechanisms. Many issues remain to be solved, among them transactional/atomicity issues in pricing schemes, rep-utation management in non-pricing schemes, game theoretic schemes, privacy issues, and exploiting social networks like Twitter of Facebook, where it is assumed that friends share trustworthy information.

Behavioral Economics and Persuasive Technologies. Behavioral economics uses social, cognitive and emotional factors in understanding the economic decisions of individuals and institutions performing economic functions. Behavioral economics has made inroads in transportation in the areas of survey design, prospect theory, and attitudinal variables. Further infusion into transportation will lead to significant benefits in terms of increased ability to both predict and influence behavior. Computational transportation science can contribute to more sustainable transport by transferring findings in behavioral economics to transportation, with a focus on personalized information and social influences to change behavior.

2 Participants

Caitlin Cottrill, Singapore-MIT Alliance for Research and Technology, Singapore Thierry Delot, INRIA Lille and University of Valenciennes, France Stefan Funke, Universität Stuttgart, Germany Glenn Geers, NICTA, Australia Shunsuke Kamijo, University of Tokyo, Japan Steve Liang, University of Calgary, Canada Dirk Christian Mattfeld, Universität Braunschweig, Germany Harvey Miller, University of Utah, USA Hiroki Ohashi, Central Research Laboratory, Hitachi, Ltd. Japan Dennis Schieferdecker, Karlsruher Institut für Technologie, Germany Monika Sester, Institut für Kartographie und Geoinformatik, Leibniz Universität Hannover, Germany Masaaki Tanizaki, Hitachi, Ltd., Kyoto University, Japan Piyushimita (Vonu) Thakuriah, University of Illinois at Chicago, USA S. Travis Waller, University of New South Wales, Australia Stephan Winter, The University of Melbourne, Australia Ouri Wolfson, The University of Illinois at Chicago, USA



Figure 1: The participants.

3 Overview of Talks

Note: presentation slides are available from http://www.nii.ac.jp/shonan/seminar022/.

Collaborative Mobility: Using Geographic Information Science to Cultivate Cooperative Transportation

Harvey Miller, University of Utah

Transportation systems are facing unprecedented challenges in the 21st century. Increasing the efficiency of transportation systems alone will not solve these problems and may exacerbate them. Instead, we must extract new transportation capabilities related to more cooperative decision-making across a wide range of time horizons, spatial scales and decision contexts. This talk discusses the role of sensed transportation, geographic information science and social media to cultivate transportation systems where participants share, cooperate and act collectively to solve operational, tactical and strategic mobility and accessibility problems. This talk also provides a vision of the future by imaging a seamless multimodal transportation system combined with a virtual environment where data streams are fused, interpreted and made available with tools for human engagement and shared decision making. This talk concludes by outlining a research agenda.

Discussion:

- As people have begun to realize that the data are invaluable, they started to conceal them. Access to data is a critical problem.
- We need some open data initiative to accelerate our research.
- Is mechanism design a way to design persuasive services moving people from pursuing individuals' goals to pursuing systems' goals? Vickrey auction is a good example, ending typically an epsilon above system optimum.

Current Issues of Urban Japanese Transport Systems

Masaaki Tanizaki, Hitachi Ltd., Japan

Regarding the current transportation situation in Japan, there are two typical different situations and issues in urban/local area individually. In order to apply IT technology for two different areas, it is important to develop technologies based on standardized specification.

Firstly, regarding the transportation situation in the urban city area, it has been equipped with public transportation networks such as train, subway and bus. These services provide congested train schedules, for example, train comes at five minutes interval by major train company operation. Additionally, some train and subway companies have shared their tracks with each other for passenger convenience by connecting their train services. On the other hand, weakness for accident and emergency is an issue, because influence of an emergent event may spread among railway companies which share the tracks. In such case, it is important for passengers to get information quickly, however, broadcasting information by train company is too slow. Therefore, we propose to develop a gathering and broadcasting information system which utilizes two kinds of resources: One is SNS information such as twitter, which is quick but has low reliability, and the other is public information by railway companies, which has high reliability but slow. How to control the tradeoff relationship between two different resources is the key.

Secondly, regarding transportation situations in local areas, public transportation facility has not been supported. Consequently, the current issue is that many people extremely depend on their cars. The other issue is that local government needs high maintenance cost for social infrastructure such as road/water/gas/electricity because of city area expansion and low population density. In order to solve these issues, local government in Toyama has started to build public transport systems by expanding an extensive and user-friendly light rail transit network, and also to intensify the urban functions such as private residences and commercial facilities. As an expected goal with these activities, the distances citizens have to travel are shortened and CO2 emission decrease because of modal shift from their cars to public transportation services. In order to achieve this goal, there are future subject as follows; monitoring method of people movement (car, bus, LRT, and so on), incentive design for activation of modal shift.

Finally, there is another discussion for interoperability of transportation technique. In order to promote utilization of technique among urban/local area, it is important to consider mutual utilization of techniques based on standardized specification of data model. ISO, CEN, OGC have published data model standardized specification based on each target field as follows:

- Road networks:
 - ISO 14825:2011 ITS Geographic Data Files(GDF5.0)
 - ISO17267:2009 ITS Navigation Systems API
 - ISO 24099:2011 Navigation data delivery structures and protocols
- Public transportation networks:
 - CEN/TC278 IFOPT (Identification of Fixed Objects in Public Transport)
- Indoor networks:
 - OGC Indoor GML/CityGML

These data model should be a basis of consideration in application, simulation and system design.

Discussion:

- Data models and standardized specifications are needed to integrate various kinds of information.
- Marginal cost pricing, congestion pricing...
- Utilizing user generated content is another good approach, for example, Twitter data.
- Reliability of the information is key.

Cooperation versus competition in vehicular networks

Thierry Delot, Université de Valenciennes, France

In the last decade, a number of wireless and small-sized devices (e.g., PDAs, smartphones, sensors, etc.) with increasing computing capabilities have appeared in the market at very affordable costs. Some of these equipments have started to be embedded in modern cars in the form of on-board computers, navigation devices or even multimedia centers. Today, thanks to wireless networks, two vehicles nearby (within communication range of each other) can share relevant information about various events (e.g., an emergency braking, a traffic congestion, a driver exhibiting risky behaviour, etc.) or resources (e.g., a parking space released, an available charging station for electric vehicles, etc.). All these trends are motivating a great amount of research to try to develop suitable data management strategies for vehicular networks. For instance, exchanging dynamic data (i.e., data whose relevance can change very quickly) in a vehicular network has been widely studied these last years. This constitutes the building blocks for new data gathering techniques, assuming that data sources/producers are numerous (e.g., sensors embedded in the car or deployed along the roads, crowdsourcers, or even classical data sources such as web services accessible through mobile telephony networks. In this talk, we will focus on the data management techniques to facilitate cooperation or manage competition between drivers.

Discussion:

- The duration of events (or the value / validity of information over time) is an issue. How long does the influence of an accident last?
- Can V2X communication be integrated?
- Issues with penetration rate of vehicles, reliability of information created by users, and involving those for whom the information is not relevant (incentives?).

Integration of information and optimization models for routing in city logistics

Dirk C. Mattfeld, with contributions from Jan F. Ehmke, Technical University Braunschweig, Germany

As urban congestion continues to be an ever increasing problem, routing in these settings has become an important area of operations research. This monograph provides cutting-edge research, utilizing the recent advances in technology, to quantify the value of dynamic, time-dependent information for advanced vehicle routing in city logistics. The methodology of traffic data collection is enhanced by GPS based data collection, resulting in a comprehensive number of travel time records. Data Mining is also applied to derive dynamic information models as required by time-dependent optimization. Finally, well-known approaches of vehicle routing are adapted in order to handle dynamic information models.

Competition in Intelligent Transportation

Ouri Wolfson, University of Illinois, Chicago, USA

The proliferation of mobile devices, location-based services and embedded wireless sensors has given rise to applications that seek to improve the efficiency of the transportation system. In this talk we show that many of these applications have to consider the competition among travelers, and demonstrate how to use game theory in order to do so. Parking is used as an example of a large class of competitive applications in Intelligent Transportation.

IT Challenges toward the Solution of Traffic Problems in Emerging Countries

Hiroki Ohashi, Hitachi Ltd, Japan

In emerging countries, traffic congestion has become one of the social problems as their rapid urbanization. The first step for solving this problem is to understand the current traffic situation. In order to understand the current traffic situation, developed countries use road side sensors or taxi probe system. However, it is difficult to use these methods in emerging countries due to the following two reasons. Firstly, the road side sensors require too much initial investment to widely spread in emerging countries. Secondly, since the major mode of transportation in emerging countries is motorbike, the data collected by taxi are not enough. To overcome these difficulties, we promote a probe system using smartphones. Since smartphones are usually equipped with GPS receivers, we can understand the position of the smartphones using the GPS data. The technical challenge to realize the smartphone probe is to automatically understand the modes of transportation of the smartphone holders, because they usually use multiple modes of transportation in their daily life. This study proposes the novel method of classifying the modes of transportation of the smartphone users based on a generative model of the vibration during drive. We have conducted an experiment to evaluate our method in an emerging country. As a result, we achieved over 80% precision and over 50% recall, which are accurate enough to realize the probe system using smartphones.

Discussion:

- Incentive to let people to install the data collection application: payment models are not sustainable, and privacy problem is an issue.
- Accuracy of the GPS data can be problem, even if it doesn't show in the test data. The required accuracy depends on the application; some applications may also expect map-matching.

Innovative Methodologies for Large-scale Stochastic Dynamic Transportation Network Systems

S. Travis Waller, University of New South Wales, Australia

The history and current state of transport network modelling for planning is discussed. In particular, major theoretical concepts and deployed applications are covered. Further, research contributions covering recent advances in stochastic adaptive routing algorithms and transport network equilibrium with recourse are presented (which have potential future applications for ITS planning).

Discussion:

- Thinking about dynamic equilibrium is interesting.
- Behavioral issues: Some people always look for the best way, other people prefer a priori fixed cost.

The Algorithmics View on Route Planning

Stefan Funke, University of Stuttgart, and Dennis Schieferdecker, KIT, Germany

We give an overview of recent speed-up techniques for the classical shortest path problem that have been developed in the algorithmics community over the last few years. These techniques allow, after a preprocessing phase, the answering of shortest path queries several magnitudes faster than straightforward Dijkstra (including its tuned A^* or bidirectional variants) and can also be extended to more complicated settings such as time-dependent edge costs, constrained shortest path etc.

Contraction Hierarchies are the basis of many advanced route planning techniques. This includes, among others, time-dependent shortest path computation, finding meaningful alternative routes to the shortest path and batched shortest path computation. Time-dependent Contraction Hierarchies allow to answer both, earliest-arrival queries and profile queries, when link costs are variable in time. Computation takes in the order of milliseconds and additional memory overhead is tiny. Recent results even make it possible to use combinations of time-dependent and time-invariant costs.

Finding good alternative routes to a shortest path quickly was an open problem until recently. Now, Contraction Hierarchies based methods identify multiple alternative routes in less than a millisecond. These routes are modelled as a concatenation of two shortest paths, i.e. the route is described by a source, target and via node. To ensure the meaningfulness of the route, three criteria have to be fulfilled: stretch, overlap and local optimality.

Batched shortest-path computation encompasses all problems in which a naive solution has to compute subproblems multiple times. By identifying these tasks, computing them only once and storing this preprocessed data in a cacheefficient manner, queries can be speed up by orders of magnitude. Contraction Hierarchies are useful in this context as their small search spaces are directly reflected in a reduced of data that has to be preprocessed and stored. Some of the problems that fall into this category are many-to-many searches (i.e. distance matrices), point-of-interest searches and advanced ride-sharing methods.

Following, there is a small selection of literature with regards to advanced route planning techniques. Note that we restricted the list to publications that were shown during the presentation:

Time-dependent Algorithms

• Batz et al.: Time-dependent Contraction Hierarchies (ALENEX 2009)

- Batz et al.: Time-dependent Contraction Hierarchies and Approximation (SEA 2010)
- Batz, Sanders: Time-dependent Route Planning with Generalized Objective Functions (ESA 2012)

Alternative Routes

- Abraham et al.: Alternative Routes in Road Networks (SEA 2010)
- Bader et al.: Alternative Route Graphs in Road Networks (TAPAS 2011)
- Luxen, Schieferdecker: Candidate Sets for Alternative Routes in Road Networks (SEA 2012)

Batched Shortest Path Algorithms

- Geisberger *et al.*: Faster Detour Computation for Ride Sharing (ATMOS 2010)
- Geisberger, Sanders: Engineering Time-Dependent Many-to-Many Shortest Paths Computation (ATMOS 2010)
- Geisberger: Advanced Route Planning in Transportation Networks (PhD Thesis 2011)

Discussion:

- Dijkstra on large (nationwide) networks cannot be solved on today's smartphones.
- For ride sharing systems can the driver be informed about the cost to pick someone up at some location, and about the benefit he/she can get by doing so?
- Can POI (Points Of Interest) be integrated in the routing algorithm?
- Adding multiple preferences / criteria: theoretically NP hard problem, but practically solvable in reasonable time.

Transportation in a Connected and Ubiquitous Sensing Environment: Challenges and Opportunities

Piyushimita (Vonu) Thakuriah, University of Illinois at Chicago, USA (currently University of Glasgow, UK)

The presentation focused on challenges and opportunities for the transportation sector from Big Data and Open Data initiatives. Mobility intelligence may be considerably enhanced by connecting transportation information horizontally to environment, energy, health and other sectors by means of ongoing smart city-type initiatives; however, connecting dynamic streams of information pose significant challenges. Real-time transportation information can also be linked vertically to data within the transportation sector, such as synthetic outputs from transportation planning and operational models, as well as administrative, survey and census data, therefore opening up opportunities for for a broad range of near-term decision-making; these strategies are being attempted by the emerging area of urban informatics.

An example of challenges that result with more information integration was given by considering the case of linking real-time transportation (speed) information to real-time weather information and by highlighting the spatial and temporal inaccuracies that can occur and resulting propagation of errors. Under bad weather conditions, the percent improvement in speed forecasts obtained by machine learning due to real-time weather information varies significantly across different weather conditions, time periods or traffic congestion levels and locations. Results from simulations also showed that the accuracy of the weather forecasts can matter considerably in making speed forecasts, such that if weather forecasts (particularly precipitation forecasts) are degraded to below a threshold, the resultant speed forecasts are worse off than not having the forecasted weather information at all. The presentation drew implications for sensor fusion (where to measure, how to temporally integrate, which sensors to integrate at one point and so on).

An example of opportunities that arise was given by linking information gleaned by real-time bus arrival information effects to a range of small-area urban indicators (constructed from models and other data) that describe the social, economic, built environment, housing and economic characteristics of communities within cities. Near-term decision-making capabilities enabled by this type of vertical linkage of information—and subsequent spatial clustering include spatially-targeted market identification strategies for Location-Based Services and transit operations and management, among others capabilities.

GeoSensor Web Lab: Connect, share and use world-wide sensors

Steve Liang, University of Calgary, Canada

In recent years, large-scale sensor arrays and the vast data sets they produce world-wide are being utilized, shared and published by a rising number of researchers on an ever-increasing frequency. The increased amount of available data is being driven by sensor motes which are monitoring changes in everything from climate to water to transportation systems. With the rapidly increasing number of large-scale sensor network deployments, the vision of a World-Wide Sensor Web (WSW) is becoming a reality. Similar to the World-Wide Web (WWW), which acts essentially as a "World-Wide Computer", the Sensor Web can be considered as a "World-Wide Sensor" or a "cyberinfrastructure" that instruments and monitors the physical world at temporal and spatial scales that are currently impossible. This short talk presented the vision, the potential impact, and the state-of-the-art of the Sensor Web. It also presented *TrafficPulse*, a participatory mobile urban sensor web that harnesses voluntary use of smart phones as a cost-effective, ready-made and pervasive sensor web that allows us to query, model, understand, and visualize the city's mobility information in real-time.



Figure 2: The excursion to Kamakura.

4 Summary of Discussions

Besides the talks of participants, a series of sessions was reserved for group discussions.

4.1 Session I

The first section focused on topic collection and topic selection. We identified these themes (Figure 3), in this order as being of most common interest:

- 1. Mechanism design: A way to address cooperative versus competitive transport?
- 2. Crowdsourcing
- 3. Benchmarking datasets and information infrastructure for computational transportation science
- 4. Non-real time decision support systems
- 5. Role of computational transportation science for different stakeholders (allocated to concluding session)

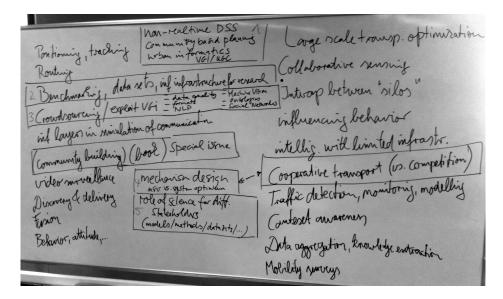


Figure 3: The collected topics of the first session.

4.2 Session II

Two breakout groups were formed to search for proposals on Topic #1: Mechanism design. The following plenum shared and discussed the breakout group results.

Group 1 considered road pricing as one of the solutions: changing the price dynamically according to the traffic situation. However, how to quantify the congestion or other factors for pricing remains an important problem. Easier to realize would be to show people some options for their travel, including multiple modes, and indicating the different costs, emissions and other factors. Known choices may help to balance traffic. It had been acknowledged that car drivers might not care on emissions. The minimum solution would be a seamless multimodal transportation service.

Group 2 added to these thoughts. People behave habitually because of lack of information about transport options, and because of low reliability of such information where available. Furthermore, the incentives to change behavior may be insignificant (see drivers' lack of care about the emissions). So, why not giving up all behavioral modelling and simple take tolls for every road? Pricing can be adapted over time to reach system optimum of traffic load.

4.3 Session III

Two breakout groups were formed to search for proposals on Topic #2: Crowd-sourcing. The following plenum shared and discussed the breakout group results.

Group 1 reported three discussed topics: applications of crowdsourcing data, incentivizing crowdsourcing, and types of crowdsourcing data, especially via DSRC or Google data. Incentivizing was seen possible by mechanisms such as (a) receiving in return aggregated data (information)—note that people do submit data to services if the services are good enough (e.g., Google Maps), (b) receiving virtual/real currency (such like airmiles), (c) earning reputation in a social community, (d) transportation safety (transmitting data contributes to your safety), (e) games and fun. Group 1 identified a need for research on reliability of crowdsourced information.

Group 2 identified a number of research questions (but no single proposal). How many users are needed for sampling particular phenomena? Are cyclists a target group that is more willing to contribute crowdsourced data: more likely tech-savvy, and suffering from lack of information? If, in a world of ubiquitous sensing (Internet of Things), everything will be sensed anyway, isn't the true challenge getting access to this data and integrate this data rather than motivating the crowd for sensing? Since reliability is an issue, could people (the crowd) be presented sensor readings for verification, in the style of captcha? Could crowdsourced sensing be integrated in participatory planning and crowd engagement, to motivate people? If data mining and knowledge discovery is applied on natural language sources (social media) then placenames need to be resolved that may not be known in databases, and that may not be known by people living elsewhere (no 'captcha' style verification)—human intelligence still required, but local knowledge is involved.

4.4 Session IV

This session was reserved for matchmaking between participants for discussing ideas in areas of common interest.

4.5 Session V: Next steps

The concluding session was dedicated to identify tangible outcomes, by that way addressing Topics #3 and #5 from above in passing. The session also reviewed the objectives and expected results set at the beginning:

- Discovering social computing in CTS;
- Growing a community in CTS;
- Seeking interaction and collaboration with scientists and practitioners from ITS.

We observe a transformative force on transportation science by Big Data, the Internet of Things, and social computing. The challenges are of computational nature, and thus, many disciplines push into transportation science such as computing, artificial intelligence, economics, or geographic information science. One challenge for a growing CTS community is bridging between these disciplines and transportation science, planning and engineering. Harvey Miller and Vonu Thakuriah have founded a TRB subcommittee on Computational Transportation and Society (reflecting in the name an emphasis on the applications of a CTS), which is an important step in this direction given the nature of the TRB. They call for broader support from the CTS community.

A Sixth International Workshop on Computational Transportation Science (2013) is on the horizon. We discussed the advantages of continuity (attaching it again to ACM SIGSPATIAL) and of flexibility (showing presence at other conferences, e.g., INFORMS). Depending on enthusiastic organizers there could be

over Vonu

Figure 4: The to-do list of the concluding session.

multiple such workshops. However, we also discussed how to attract people and papers in an environment of competitive conferences and research performance frameworks. Linking the calls to selected issues of journals should help.

There was, however, no support of an own journal or an own conference due to lack of critical mass. Instead there was a sense that growing the community is more a long-term goal that should be pursued with patience, supported by a multitude of small efforts such as having workshops, promoting everywhere the term 'computational transportation science' (keyword lists, CVs, etc.), selected issues in related journals, or an edited book shaping the term. Travis Waller offered to organize a special issue in Computer-Aided Civil and Infrastructure Engineering. A book proposal will be developed by Harvey Miller and Vonu Thakuriah in January or February 2013. The book should be a solid review of the different contributing areas to CTS, such that it can be used as a textbook for a graduate seminar on CTS. It may contain some perspective papers as well.

The community could also consider to define it through a (long-term) grand challenge or some (short-term) benchmarking or application challenges. In this context Monika Sester and Dirk Mattfeld pointed to a new transportation research infrastructure that could be deployed to collect benchmarking datasets: the Application Platform for Intelligent Mobility (AIM), http://www.dlr.de/ fs/en/desktopdefault.aspx/tabid-6422/10597_read-23684/. With this meeting starting out to discuss the role of social computing within CTS, a follow-up meeting in Dagstuhl, December 2013, could focus on such a benchmarking exercise, either data collection or data integration and knowledge discovery.