DATA FUSION FOR TRAVEL DEMAND MANAGEMENT: STATE OF THE PRACTICE & PROSPECTS

Christopher Zegras, Dept. of Urban Studies & Planning (DUSP), MIT, USA
Francisco Pereira, Cognitive and Media Systems Group (CMS), Centro de Informática e Sistemas da Universidade de Coimbra (CISUC), Portugal
Andrew Amey, DUSP, MIT, USA
Marco Veloso, CMS, CISUC, Portugal
Liang Liu, SENSEable City Laboratory, DUSP, MIT, USA
Carlos Bento, CMS, CISUC, Portugal
Assaf Biderman, SENSEable City Laboratory, DUSP, MIT, USA

ABSTRACT

This paper provides a state of the practice review of data fusion for travel demand management (TDM). Data fusion involves the seamless detection and combination of data, from multiple sources, with the goal of extracting new knowledge from the data. For understanding the challenges and possibilities for applying data fusion for TDM, we first present system architecture requirements and several data fusion models. We then provide a brief review of major relevant industry players, finding many companies now spanning across related areas such as data provision, data aggregation, and delivery to end users, with a primary focus on automobile users and roadway conditions. Examining eleven metropolitan areas in the USA, we find several characteristics apparently associated with more advanced data fusion adoption, including degree of automobile dependence and presence of "high tech" industry. We conclude by identifying some prospects for data fusion for TDM, as revealed through the analyses.

INTRODUCTION

The rapidly evolving area of Information and Communication Technologies (ICTs) has clear implications for a range of travel demand management measures, spanning across the relevant user decision points, including the decisions to undertake activities, where to perform activities, and the transportation modes and routes to choose. Metropolitan areas today have quite literally become saturated with various types and sources of real-time data that can, in theory, be utilized to improve mobility services by influencing demand, allowing users to make better-informed decisions about their mobility via seamless integration and delivery of the necessary information (including prices and externalities), when and where it is needed. These data sources include "traditional" transportation sources, often associated with Intelligent Transportation Systems (ITS), such as toll road operators, public transportation services, road sensors, image capturing and commercial fleet tracking.

At the same time, the increasing ubiquity of a range of different mobile devices and other ICT-related technologies introduces new information sources, including distributed mobile sensor networks, mobile devices, direct citizen engagement, and web-based platforms which provide close-to-real-time information (e.g. on city events).

A principal practical challenge to capitalizing on the travel demand management (TDM) potential offered by these data sources relates to the need for integration or *data fusion* – compiling and aggregating the data into an augmented and added-value whole in such a form that applications can access specifically relevant information, otherwise inaccessible from individual sources, with appropriate representation and level of detail. Data fusion poses both technical challenges, related to gathering the data in a timely and consistent fashion and computationally manipulating it for different user groups; and, institutional challenge, related to the numerous public and private agencies and companies potentially involved and issues such as financing, data ownership, privacy concerns, ownership of the computational platform for data collection and fusion, etc.

This paper aims to provide a state of the practice review of data fusion by focusing on technical and institutional aspects. Technically, we present system architecture requirements and several models for data fusion from a TDM perspective. Institutionally, we briefly review some of the predominant industry players in relevant application fields (primarily in N. America and Europe) and then examine case studies of several metropolitan areas in the United States, exploring various factors which might lead to advanced data fusion adoption for TDM applications. We conclude by identifying some prospects for data fusion for TDM, as revealed through the technical and institutional analyses.

DATA FUSION AND TDM: OVERVIEW

Data fusion (DF) involves the seamless detection and combination of data, from multiple sources, with the goal of extracting new knowledge from the data and generating improved information (including estimations, predictions) that can be transmitted to relevant users for better decision making. More specifically, we consider that a system uses *data fusion* whenever:

- More than one source of data is being fed simultaneously;
- Each data source has distinct inherent properties (i.e. specific technology, type of data, etc.); and,
- Data sources are integrated to create at least one sort of unified information.

Considerable work exists on the topic of multi-sensor data fusion, the integration of distinct low-level signals into a unified result (e.g., estimating a precise position from a GPS receiver and an accelerometer); at the level of information fusion (i.e. the integration of two or more signal-level processed sources), however, much more work remains. When aiming to fuse data into higher level information that people can perceive and use to control complex tasks, we face an increase in the number and variety of types of sensors[1] that can be combined. Data fusion of dramatically different types and levels of representation becomes increasingly complex, also increasing the quantity of information the system must handle. For a range of end use sectors (e.g., transportation), the employment of more than one sensor can bring increased robustness and reliability, larger coverage, increased dimensionality of measurement, confidence in and reduction of measurement time, but, often, at higher costs (Thomopoulos, 1989).

In transportation applications, we can envision three basic "classes" of DF user groups: transportation system users (e.g., passengers), service providers (e.g., public transport, supply change management, disaster response), and system planners (e.g., government planning agencies). In the most general sense, these user classes operate at the operational, tactical, and strategic levels, with approximately immediate-, short-/medium-, and longer-term time frames, respectively. For example, a traveller could use DF applications to assist in an immediate, mode choice (operational) decision; a public transport company could use DF applications to modify certain routes (tactical); a planning agency could use DF applications to integrate long term transportation and land development plans. No formal barrier exists between these user classes and time-frames, as, for example, planning agencies may make operational decisions.

In this paper, we focus primarily on DF implications for operational decisions by system users, arguably the heart of TDM. In theory, the range of sensors present in a metropolitan area provides the raw information necessary for a potential traveller to: determine the most appropriate time and place to carry out an activity (e.g., knowing the state of the queue at, say, a Bank), what mode or combination of modes to take (and the time-cost-externalities trade-offs that might be relevant), how to take those modes (e.g., when to leave, reliability estimates), and what route to take. The information exists today to develop real-time, nearly omniscient personal travel planners for individuals; yet the DF challenge remains of effectively capturing this information and delivering it to the users.

Computational Architectures for DF

Work in the fields of Physics, Computer Science and Mathematics has tackled important challenges of sensor fusion. For example, estimators such Kalman Filters, Neural Networks, Fuzzy Sets or Bayesian Networks already allow for the aggregation of information from different sources. However, these are ideal for *signal* level detail (e.g. aggregating GPS positioning with accelerometer information) as opposed to tasks that demand *information* level detail (e.g. inferring that a car is at a traffic light rather than in congestion by adding GIS map and speed information). We thus need to consider a broader system, one able to cope with several levels and kinds of information, integrate it, and add value to it.

Data fusion technology targets the problem of aggregating data, recorded from multiple *sensors*, together with knowledge in order to more accurately estimate conditions in the environment and allow for a variety of applications (Wang, 2001). The heterogeneous nature of the data sources demands a robust model that embodies different levels of integration and

[1] From now on, we use the term *sensor* to represent data sources in general

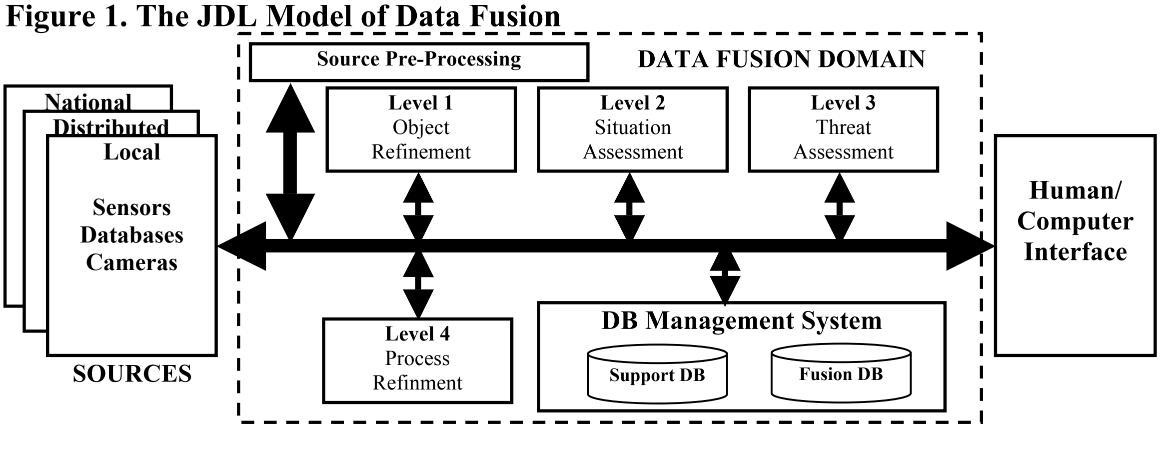
some specific semantics or protocol to communicate between all system-components. Esteban et al. (2005) synthesize the architectural issues that must be taken into account to develop a platform for multi-sensor data fusion:

- Sensor distribution for network formation: Should sensors be organized in a parallel or a serial (iterative) bus, or combination of both? A parallel sensor configuration is more adapted to identical and to physically and/or distinct sensors, whilst a serial configuration is appropriate to a system where one sensor delivers information to another, augmenting the knowledge available in a hierarchical form.
- Level of data representation: A multi-level architecture can enrich the information available, fusing data and knowledge from different sources through different treatments, and providing data with different degrees of representation according to need.
- Architecture type: Centralized (using raw data) or decentralized (using a pre-processed data)? The former requires less computational capabilities in the sensors and a central hardware capable of dealing with a greater quantity of data, while the latter distributes computational power through the system nodes, adding complexity to the DF process.
- System feedback: Allows for control of the system via recommendations provided by the architecture's different nodes and levels, implying, of course, more complex architecture.

Several models developed thus far face some or all of these issues. We now describe three of the most representative ones currently in use. JDL (Llinas et al, 2004), first proposed in 1986 as a result of a sub panel from the US Department of Defense to aid the development of military applications (Esteban et al, 2005), presents 4 levels (see Figure 1).

- Level 1, object refinement attempts to locate and identify objects (can be further divided into four processes: data alignment, data association, object estimation, object identity).
- Level 2, situation assessment attempts to construct a picture from incomplete information provided by Level 1, that is, to relate the reconstructed entity with an observed event.
- Level 3, threat assessment interprets Level 2 results in terms of possible operational opportunities, analyzing relative advantages/disadvantages of different courses of action.
- Level 4, process refinement loops around these three levels to monitor performance, identify potential sources of information enhancement, and optimize allocation of sensors.

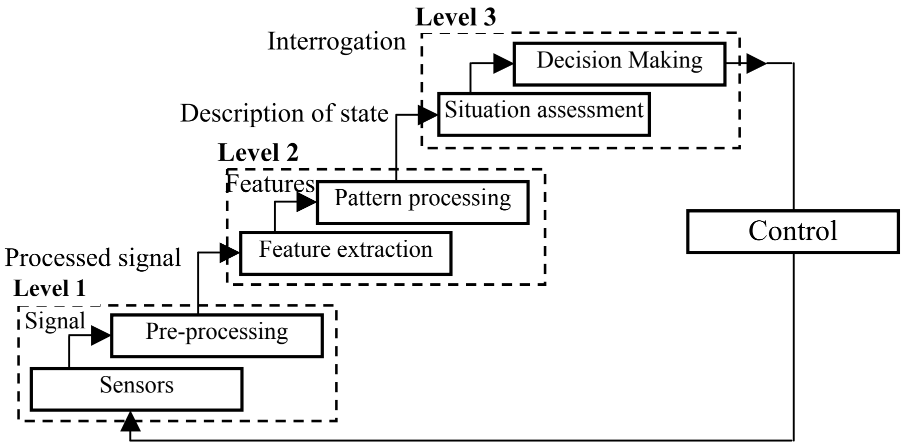
Figure 1. The JDL Model of Data Fusion



JDL assumes a parallel organization of input data (all information fed into the pipeline), although a serial process could be acceptable. It has several internal levels of information representation, not implying a specific one for input. A centralized architecture, it does all the "pre-processing" itself. Finally, the system has a feedback mechanism.

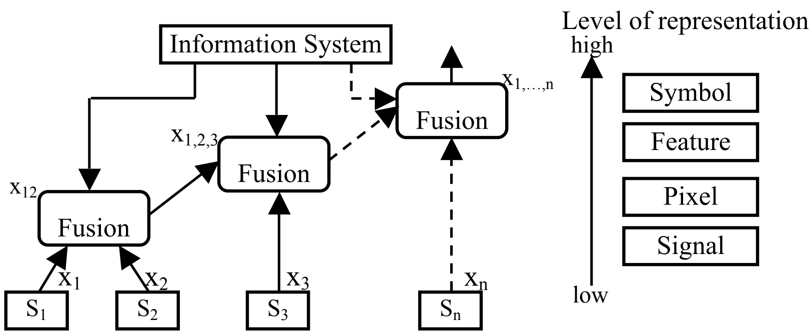
Harris et al. (1998) proposed the Waterfall model, hierarchical in nature, with the information from one module provided to the next (Figure 2). The last module (Decision Making) delivers enough information to the control module to calibrate and configure the sensors. Each of the architecture's three levels has two modules, with a closed loop acting in the system. The first level gathers and transforms data from the environment, delivering the processed data *and* information about the sensors to the next level. The second level extracts and fuses the main features of the data from the first level, thus reducing the quantity of data transmitted and increasing information richness. Building on the previous levels' processing, the third level creates a scenario of events and assembles possible routes of action.

Figure 2. The Waterfall Model of Data Fusion



The Waterfall model does not clearly state that the sources should be parallel or serial (though processing is serial). It assumes centralized control, allows for several levels of representation (similar, in this aspect, to JDL), and proposes a feedback mechanism.

Figure 3. Luo and Kay's (1988) Model of Data Fusion



Luo and Kay (1988) presented a hierarchical model, different from the Waterfall model. While in the Waterfall model all data gathered are processed in a sequential way for all modules, in the Luo and Kay model data from the sensors are incrementally added on different fusion centres (multi-sensor fusion), thus increasing the level of representation from the raw data or signal level to more abstract symbolic representations at the symbol level.

This model explicitly proposes the parallel input and processing of data sources, which may enter the system at different stages and levels of representation (as depicted in Figure 3). It is decentralized and does not assume a feedback control.

For TDM, the choice of which architecture to implement in each case depends highly on physical, economic, and institutional constraints. Multiple sources of information (e.g. different private data providers and types of data) or centralized institutional relationships (e.g. several private data providers send information to a single public institution) require parallel sensor distribution. Serial distribution is mostly possible with a consortium of closely related providers (that enables the sequential provision of information through a virtual pipeline). The data representation level is strongly linked to the degree of involvement and mutual confidence of institutions, particularly when considering the value of data detail. Higher detail (low level of representation) allows for better accuracy, signifying higher value added; higher detail also introduces important privacy concerns which may pose barriers. Higher levels of representation permit information abstraction that can often be useful (e.g. traffic managers can focus on movement patterns, not individuals). If these choices cannot be clearly decided in the beginning, a flexible model (e.g. Luo and Kay's) will be a good option.

Due to their nature, TDM applications can become extremely complex (a broad geographic distribution of many different data sources, end users, control mechanisms); on the other hand, their control tends to be centralized, suggesting a centralized organization of the architecture. However, this model is typically less reliable since it depends on a single entity and communications with it. Although the design of distributed architectures is considerably more complex, this complexity makes it more flexible to new additions. This factor should be taken into account, particularly for rapidly growing metropolitan areas.

Finally, feedback raises a number of challenges to DF-based TDM applications, relating to decisions on how to act in the environment to make the system more efficient and how to continuously tune the sensor levels to adapt to the intentions of the control levels. This can become important both for TDM and for quality improvement in the DF system (e.g. changing sensors' parameters to improve estimates). The ideal system will use the sensor information to control the actuators (traffic lights, variable message signs, etc.) with the users following every suggestion made by the system. It will also improve the sensor performance with attention to the dynamics of the system. Due to individual users' behavior and/or the quality of available technologies, the feedback loop design must contribute more to the efficiency of the system than to its entropy.

Industry Players Relevant to DF for TDM

While we cannot easily know the underlying computational architectures and models employed in relevant transportation DF activities, we can identify the scope of activities undertaken by established industry players.[2] To better understand the industry landscape, we reviewed major players primarily in North America and Europe, aiming to be comprehensive, rather than exhaustive, in our coverage. We found the largest and most recognized industry

[2] For our purposes, "established industry player" means that the company is currently offering a product or service to the marketplace and has some form of revenue model in place.

players are mostly operating in the Advanced Traveler Information Systems (ATIS) marketplace. One of the main reasons behind this level of activity in the ATIS market appears to be consumer demand. Non-ATIS applications do exist but have been overshadowed by developments in the ATIS market. Table 1 summarizes the industry players identified.

Generally, the largest number of players appear to be in the areas of data provision, whether GPS-based, cellular phone-based, and/or traditional sensor-based. Some of the largest industry players in <u>data provision</u> generate data from all three sources, such as the UK-based iTIS Holdings, which gathers traffic data from stationary sources and GPS-enabled fleet vehicles and has been experimenting with cellular-based floating vehicle data. This and similar data provision firms (e.g., INRIX, TrafficMaster) have a range of end users, including governments (iTIS, then TrafficMaster, had the contracts to supply the UK Department for Transport with traffic flow and congestion data) and insurance companies (TrafficMaster provides the sensors and data processing for Norwich Union's (UK) Pay-as-you-Go variable rate insurance). A growing number of data-providing firms operate in the cellular phone-based realm, utilizing a range of data capture techniques (e.g., pattern matching of signal channels, tracking signal "handoffs" from one cell to another) to infer travel speeds and times; these firms operate as data wholesalers (e.g., selling traffic data for broadcast on Internet sites), provide fleet tracking applications, etc. Of the various firms focusing on <u>end user devices</u>, perhaps TomTom Navigation has the broadest scope. Originally focused on in-vehicle navigation devices (GPS-enabled), they have expanded into a variety of related real-time travel information (including mobile phone-based traffic delay inference technology), software applications for mobile phones and PDAs.

In public transport-focused services, only a few firms operate. NextBus uses GPS-based, automatic vehicle location (AVL) systems already installed on buses to gather service data and provide arrival time predictions for particular routes, via the Internet, Cellphones and PDAs, and at stations. Currently operating in a number of public transport markets across the US, NextBus' most completely covered metro area appears to be San Francisco, with fairly extensive coverage for the Muni and AC Transit Bus systems. HopStop provides schedule-based (i.e., not real-time) public transport departure/arrival times and estimated travel times for combinations of bus, rail and walking (or each mode individually), allowing users to indicate preferences regarding number of transfers or use of private shuttle services. Google has also recently entered the fray with Google Transit, which provides travel time estimates for trips on both public transport and private automobile, thereby allowing users to compare different driving modes (using fixed schedules and roadway travel times and published fares and estimated driving costs).

Overall, as one might expect, this is a dynamic high tech field, with a number of mergers and various collaborations. For example, TomTom Navigation recently acquired TeleAtlas, one of only two large digital map providers worldwide, and has recently established agreements to use mobile phone data (via Vodafone or its partners) in five European countries to develop "high definition" traffic information; Westwood One (the largest radio network in the USA) which owns SmartRoute Systems has agreements with TrafficCast to provide predictive traffic information to its customers and with Yahoo! Maps to broadcast traffic information online; and, many of the end user devices (e.g., Dash, Garmin) have agreements with various data providers to deliver real-time information to their consumers.

Table 1. Established ITS Industry Businesses – Categorized by Type of Business Activity

Industry Players		Deployment Locations		Data Provider - GPS-based	Data Provider - Cellular-based	Data Provider - Traditional Sensors	Data Aggregators / Distributors	Digital Mapping	End User Devices	Public Transport	Parking	Other
INRIX		US, UK		☑		☑	☑					
iTIS Holdings		UK, EU, US, Israel		☑	☑	☑	☑					
TrafficMaster		UK		☑		☑	☑					
Skymeter		CAN		☑								
CellInt		US, Israel			☑							
IntelliOne		US			☑							
AirSage		US			☑							
DeCell		EU, Israel			☑							
TrafficCast		US, China			☑	☑	☑					
Trisent		Scotland			☑							
SpeedInfo		US				☑						
Sensys Networks		US				☑						
GlobisData		CAN		☑		☑						
TrafficGauge		US		☑		☑	☑	☑	☑			
Clear Channel		US					☑					
Westwood One		US				☑	☑					
AA UK		UK		☑			☑					
TrafficLand		US				☑	☑					
Navteq/Traffic.com		US, UK, Worldwide			☑	☑	☑	☑	☑			
TeleAtlas		Worldwide						☑				
OpenStreetMap		Worldwide						☑				
EveryTrail		US						☑				
TomTom Navigation		EU, UK, US			☑		☑		☑			
Garmin		US, EU, UK							☑			
DASH Navigation		US			☑		☑		☑			
GM OnStar		US		☑					☑			
NextBus		US		☑						☑		
HopStop		US								☑		
Google Transit		US, CAN, EU								☑		
ParkingCarma		US									☑	
SpotScout		US									☑	
Vodafone Plc		EU, UK			☑							☑
Delcan		US, CAN, EU					☑					☑

Notably (and perhaps evidence of the "market' influence in heavily auto-dependent US where most of our identified firms operate), public transportation services receive relatively little attention and no firm currently provides integrated multi-modal information enabling users to make fully informed travel decisions based on real-time data and related estimates.

From a data fusion perspective, many of these commercial applications (e.g. TomTom, TrafficCast, TrafficMaster, INRIX) already possess sophisticated capabilities. Besides the signal level (through Kalman Filters, for example), they integrate GIS maps with localization, historical data with predictive models, and cellular data with road sensors and network. Some of them (e.g. TrafficCast) include very complete DF engines that aggregate disparate information, such as weather prediction, archival flow and real-time sensing (of Cell-phones and GPS probes) into prediction information of speed and travel time. The majority also provide end-user services (e.g. on mobile devices, web pages) and therefore cover the whole spectrum from sensors to vehicle drivers. While it is difficult to know what DF model and architectures are followed precisely, we can expect some general characteristics: in general they should be centralized (single company, many external sources); inputs can be either parallel or serial depending on the heterogeneity of sensors (the more different the sensors, the greater the need for parallelization); feedback loops are only possible when there is strong integration with the infrastructure (via contracts with the public sector).

METROPOLITAN CASES FROM USA

A number of national governments – such as Germany, Japan, Korea, Singapore, the UK, and the USA –have initiatives supporting ITS applications, with clear DF implications, including: national ITS plans with architectures and standards, guidance on relevant public-private partnerships, financial support to other levels of government for demonstrations, and data provision. As seen above, private industry has also been involved in the design of system architectures, developing applications & algorithms and deploying ITS systems. Despite the importance of these two players, the most relevant DF applications for TDM purposes play out in metropolitan areas, where the majority of people carry out their daily activities. We now turn to an examination of ITS implementation in metropolitan areas, looking specifically at the degree to which DF-based multi-modal ITS applications have taken hold. Through this analysis, we aim to identify, preliminarily, some of the factors apparently associated with varying degrees of implementation. Here we focus on select metropolitan cases in the USA; the findings may apply elsewhere.

Degree of ITS Adoption

We distinguish the degree of ITS/DF adoption in the metropolitan areas by proposing a 2 dimensional spectrum according to real-time and multi-modal deployment (see Table 2). We can consider several degrees of orthogonal relationships according to increasing phases of deployment, ranging from the lowest, which consists of single-mode, static services (e.g. a web-based route planner for private vehicles), to the highest, which consists of an integrated, real-time (and predictive), multi-modal travel information service. The complexity of underlying architectures and models is necessarily proportional to the dimensions described: single-mode, static services demand little or no data fusion, real-time, isolated systems demand signal level fusion, while integrated, predictive and multi-modal systems need architectures with many components, several layers of representation and parallel inputs.

To some degree, these service spectra outline the different problems that metropolitan areas will face in developing DF-based ITS applications. At the low end of the spectrum, data availability and technological problems will be the major issues that need to be overcome. At the high end, institutional cooperation and ITS systems integration likely present the major challenges. Rarely, if ever, will a single agency or firm within a metropolitan area be the generator and provider of both real time traffic and public transport information, indicating a clear need to integrate across stakeholders to achieve DF for TDM.

Table 2. DF "Sophistication" Dimensions

		Modality		
		Single mode	*Multi-modal, separate systems*	*Multi-modal, integrated system*
Time	*Static*	Table-based system, no sensors	Table-based systems, no sensors	Table-based system, many tables, no sensors, synchronization and communication between subsystems needed
	Real time	Real Time Traffic Conditions (RTT), sensor fusion needed	RTT, sensor fusion needed	RTT, sensors, tables fusion and synchronization needed; complex communication
	Predictive	RTT, sensors and historical data fusion needed	RTT, sensors and historical data fusion needed	RTT, sensors, tables and historical data fusion and synchronization needed; complex communication

Metropolitan Area Cases

We now explore characteristics which we hypothesize may influence the degree of DF adoption within specific metropolitan areas. For the purpose of comparing relevant characteristics across different metro areas in a consistent way, we focus on a specific national setting (USA), as this allows for relative control over national-level factors of potential influence and also enables fairly consistent and easy data collection, due to various centralized data sources. We hypothesize that several different factors may contribute to variations in the degree of DF adoption for ITS: transportation system performance, presence of advanced technology industries in the region, degree of national government support for ITS deployments in the region, and the institutional arrangements governing transportation planning and investment, with a particular focus on the metropolitan-level transportation planning organizations (MPOs).

Table 3 presents the variables and our expectations regarding their influence on DF adoption. To begin to examine these hypotheses, we utilize case studies, a relevant approach when posing "how" or "why" questions, regarding "contemporary" phenomena over which investigators have little control, in situations with unclear "boundaries between phenomenon and context" and "many more variables of interest than data points" (Yin, 2003; p. 13). We selected 11 different metropolitan area cases for this analysis, aiming to represent a range of data fusion outcomes (the outcome of interest) as well as a manageable range of variation in the hypothesized influencing factors.

While Table 3 also provides detail on the variable definitions and sources, we need to clarify a few points regarding the validity and comparability of the metrics. The TTI-based metrics at least have the benefit of fairly consistent measures in time and space. Estimating the relative importance of a metro region's technology industry was difficult given the varying definitions and methodologies used by different groups; we settled upon the Metropolitan New Economy Index from 2001. We chose the number (i.e., count) of federal government grants to represent ITS support[3]; the sources analyzed do not account for all federal government grants provided to metro regions, however these were the ones most likely to involve investment in relevant applications and technologies (although the funds are discretionary and may not necessarily have been applied to ITS/DF). Including a measure of state level support would improve the analysis, but this information was not readily available. Finally, the data are not available at fully comparable geographic areas of aggregation, although the MPO and UZA are similar enough, nor necessarily for precisely the same years (see Table 3 for details).

Table 3. Variables and Expected Influence on DF Adoption in US Metropolitan Areas

Variable[1]	Hypothesis	Source[2]	Year	Spatial Scale[3]
Population	Small cities might have little demand for DF based applications; very large cities may present management challenges to DF	TTI	2005	UZA
Congestion Levels	Cities with higher congestion levels might have greater impetus for DF	TTI	2005	UZA
Congestion Increase	Cities experiencing more rapid increases in travel delay might have more demand for DF	TTI	1996-2005	UZA
Auto Dependence	Cities with a higher dependence on auto travel relative to public transport might have less demand for integrated DF	TTI	2005	UZA
"High Tech" Industry	Cities with a higher relative share of technology or knowledge-based industry might have a stronger "local lobby" for deploying advanced DF	PPI	2001	n.a.
Federal ITS Support	Cities receiving a greater share of Federal government support for ITS should have more advanced DF	FHWA	1998-2007	MPO
MPO Tax Authority	MPOs with some fiscal independence might have more flexibility for DF implementation	AMPO, MPOs	2005, 2008	MPO
MPO Representation	MPOs that have more elected representation on their Boards might be more empowered to implement DF	AMPO, MPOs	2005, 2008	MPO
MPO Jurisdictions	MPOs that represent a larger number of jurisdictions might face greater challenges in DF implementation	AMPO, MPOs	2005, 2008	MPO
Local Funding Share	MPOs with a greater share of non-Federal, non-State funding for operations and administration might have more flexibility for DF deployment	AMPO, MPOs	2005, 2008	MPO

Notes: (1) Congestion levels measured as Annual Hours of Delay per Peak Hour Traveler in 2005; change in congestion measured as net change in Annual Hours of Delay per Peak Hour Traveler from 1996 to 2005; relative auto dependence measured as ratio of total daily freeway+arterial vehicle miles traveled ('000s) to annual public transport passenger miles traveled (mns); high tech industry presence based on Metropolitan New Economy Index; Federal ITS support measured as a count of metro regions receiving funding through the FHWA Integrated Corridor Management "Pioneer" sites program, USDOT "Urban Partnerships" Congestion Initiative, or through the FHWA ITIP/TTID (511 Implementation Grants) program; text contains more description on Metro Planning and MPO accountability measurement.
(2) TTI (2007); PPI (2001); FHWA (2006a, 2006b, 1998); AMPO (2005); MPOs are: MUMPO (2008), OKI (2008), DRCOG (2008), Metroplan Orlando (2008), SPC (2008), SABCMPO (2008), SANDAG (2008), PSRC (2008), MMC (2008), MTC (2008), Metro (2008).
(3) UZA: urbanized area; n.a.: information not available on specific spatial scale; MPO: UZA + expected developed lands over next 20 years.

[3] Using a count rather than dollar figure should suffice for showing the relative impact of federal funding on ITS adoption at the metro level.

Table 4. Selected US Metropolitan Areas and Performance on DF Variables of Interest

Metro Area	Multi – Modal, RTT	Population (000s)[1]	Congestion Levels[1]	Congestion Increase[1]	Auto Dependence[1]	Federal Support	High Tech Industry	MPO Characteristics			
								Taxation Authority	Represent-ation[2]	Jurisdictions (county, local)[3]	Local Fund Share (%)
San Diego CA	Yes 1998	2,896 (3,2nd)	57 (2,1st)	19 (1,1st)	114 (6,2nd)	2	5	Yes	LE, n.a.	1, 18	43
Seattle WA	Yes 1998	3,009 (2,2nd)	45 (6,2nd)	-6 (11,3rd)	54 (10,3rd)	3	3	No	LE, E	4, 70	20
Portland OR	Yes 2002	1,729 (7,2nd)	38 (9,3rd)	2 (7,2nd)	59 (9,3rd)	0	15	Yes	RE	3, 25	93
Denver CO	Yes 2001	2,088 (5,2nd)	50 (4,1st)	10 (4,1st)	98 (7,3rd)	0	7	No	LE, Appt.	9, 56	9
San Francisco CA	Train 2008	4,156 (1,1st)	60 (1,1st)	3 (6,2nd)	36 (11,3rd)	3	1	Yes	LE, Appt.	9, 101	77
Minneapolis MN	Partial 2008	2,520 (4,2nd)	43 (7,2nd)	9 (5,1st)	132 (5,2nd)	2	10	Yes	Appt.	7, 189	52
Charlotte NC	No	860 (11,3rd)	45 (5,2nd)	19 (2,1st)	222 (1,1st)	0	30	No	n.a.	2, ~17	0
Cincinnati OH	No	1,620 (8,2nd)	27 (10, 3rd)	1 (8,3rd)	188 (2,2nd)	1	34	No	LE, Appt.	8, 198	29
Orlando FL	No	1,360 (10,3rd)	54 (3,1st)	-3 (9,3rd)	183 (3,2nd)	0	25	No	LE, Appt.	3, n.a.	19
Pittsburgh PA	No	1,838 (6,2nd)	16 (11,3rd)	-3 (9,3rd)	95 (8,3rd)	0	37	No	Appt.	10, 1(city)	n.a.
San Antonio TX	No	1,362 (9, 3rd)	39 (8,2nd)	17 (3,1st)	156 (4,2nd)	1	49	No	LE, Appt.	1, +25	0

Notes: Data sources and variables correspond to those presented in Table 3, except for Multi-Modal, Real Time Traffic provision (via integrated or separate systems), which corresponds to the Table 2.
(1) For the numbers in parentheses: the first number represents the city's rank among the 11 cities, when ranked from highest to lowest measure on the variable; the second number represents the tercile of the city when 39 major US cities are ranked from largest to smallest measure on the variable.
(2) LE, Appt. = locally elected officials appointed to MPO board; LE, E = locally elected officials elected to MPO board; RE = regionally elected MPO members; Appt. = MPO members appointed; n.a. = not available.
(3) The first number represents the number of counties represented in the MPO, the second number represents the number of towns/cities represented in the MPO.

Table 4 presents the 11 cities analyzed, including their status regarding fulfilment of "Multi-Modal, Real Time Traffic (MM-RTT)" Data Fusion and their measures on the variables of interest. Six of the 11 metropolitan areas have partial or full "MM-RTT" DF implemented. These metropolitan areas tend to be mid-sized (note the largest US metro areas, New York City and Los Angeles, were not included in our analysis); San Francisco, the largest of the cases examined, has been slower than some of the medium size areas in implementing advanced ITS. In terms of transportation system performance, a mixed pattern emerges with respect to congestion. While several of the most congested areas have advanced ITS, other relatively highly congested areas (Orlando, Charlotte) do not. Similarly, when examining the relative increase in congestion, a mixed pattern also emerges – San Diego and San Antonio had the highest growth rates in congestion; the former has achieved MM-RTT, the latter has not. Perhaps not surprisingly, the relative auto dependence of a region does seem to be associated with more advanced ITS applications; for the most part, those regions with advanced ITS applications have lower automobile dependence, which makes sense as higher public transport use would translate into demand for multi-modal ITS applications.

Interestingly, several measures not directly related to transportation system performance seem to be associated with advanced ITS applications:

- Federal grant support does seem to play a partial role: two early adopters (San Diego and Seattle) received a relatively large number of Federal grants, as did two relatively late adopters (San Francisco and Minneapolis); Denver and Portland, adopting early this decade, received none.
- The presence of "high tech" industry does seem to be associated with ITS adoption; the six MM-RTT adopters are in the top 15 "high tech" metropolitan economies; the non-adopters among the cases do not break the top twenty, supporting the hypothesis that the presence of high tech industry will advance the ITS cause in a metro area.
- No clear pattern emerges with respect to MPO jurisdiction or form of representation.
- The two financial-related MPO characteristics reflect some apparent relationship to ITS adoption; four of the six MM-RTT adopters have MPOs with some local taxation authority and (perhaps relatedly) five of the six fund transportation with 20% or more from local revenue sources, suggesting that local direct financial responsibility in the sector might also help advance ITS adoption, even in the presence of Federal support.

One must keep in mind the limitations of the case study analysis; it provides suggestions regarding some of the influencing factors rather than definitive answers. The number of cases limits the generalizability of the results. Furthermore, even where some fairly clear association appears evident, we must be careful not to confuse correlation with causation. For example, while it seems that high tech industry presence might encourage ITS adoption, the possibility exists that the high tech industry is drawn to places that have adopted advanced ITS (evidence of the region's broader support and interest in high tech industry).

Overview Comparison to EU Data Fusion Adopters

While the USA has been an early and active promoter of ITS at the national level, this has not necessarily translated into the full-fledged multi-modal, integrated applications that promise to significantly impact TDM applications, especially at the metropolitan level. Across the Atlantic, except for a few examples, the scenario seems somewhat similar. Nonetheless, a number of somewhat advanced examples exist, worth a brief comparison with the USA cases. Berlin, Germany, offers possibly the most prominent example in terms of achieving a multi-modal, predictive and integrated model. The Berlin case also provides an interesting public-private partnership example. In 2000 the city began a 10-year public-private partnership with a Daimler Chrysler/Siemens-led consortium to provide new detection devices, a state-of-the-art Traffic Management Center (TMC), and a number of value-added user services (Siemens, 2008; PTV-Berlin, 2003; Rupert et al., 2003). At the same time, the city is developing a Traffic Control Center (TCC). The overall approach integrates both "hard" and "soft" measures. Hard measures, to be controlled by the public sector, include bus lanes, bike lanes, access control, and parking restrictions. Soft measures, to be controlled by the private sector, include traveller information and new user services and require coordinated information exchange. The city provides all its data from the TCC, which covers the hard measures and allows for better management and control of road traffic, to the private partner at the TMC., The TMC and TCC are connected and have a common datapool (Rupert et al., 2003). Data come directly from a wide number of sources (infrared road sensors, inductive loops, cameras, Floating Car Data), in an open and expandable system. In terms of DF, the Berlin system functions at several different levels, including integration of data sources with other information (e.g. police information), although the degree of real integration as opposed to separate availability within the TMC remains unclear.

Other EU examples include Munich, Cologne, Stockholm, Helsinki, London and Rome. All of these cases represent on-going projects and, with the exception of Munich's MOBINET which has a fully integrated system, not fully integrated multi-modal systems. Most have several independent subsystems that achieve some degree of Data Fusion (e.g. Name Plate Recognition, FCD, RTT). They also have table-based multi-modal information with partial real time information (e.g. Helsinki provides real time positioning of buses and trams) and some predictive capabilities (e.g. Cologne).

OUTLOOK

Overall, although much of the necessary technology exists, the elaborate use of DF for TDM remains far from its potential. Both technical and institutional challenges remain, challenges which likely correlate directly and positively to the size and complexity of the tasks at hand. A number of different computer architecture models exist, with the best architecture for any particular case dependent upon numerous context-specific constraints relating to the number of different information sources, the relationships among relevant institutions, data detail (level of representation) necessary and possible, and so on. The ultimate architecture underlying a TDM-oriented data fusion application will likely need to: be flexible enough to enable a high degree of accuracy while ensuring respect for privacy *and* ease of abstraction (e.g., to higher level traffic patterns); accommodate a (likely) broad geography and number of jurisdictions and agencies; incorporate a diverse range of sensor types; enable various potential applications and delivery media to users; and, allow for some degree of feedback to improve both the efficiency of applications (e.g., TDM) and the DF system itself (e.g. modifying sensors). Overarching these general specifications come questions regarding the degree of centralization: a centralized architecture allows for clearer control over the varying dimensions of complexity; on the other hand, a less centralized system may prove more robust and likely more flexible to new additions.

The private sector seems to be undertaking a good share of the necessary activities, with many companies now spanning across related areas such as data provision (from various sensor types), data aggregation, and delivery to end users. In reviewing relevant business activity (primarily in North America), we perceive at least two relevant trends. First, the most advanced applications appear for the private vehicle-based (automobile) users, providing real time traffic conditions, route choice suggestions, and so forth, including via the increasingly prevalent in-vehicle devices. The most advanced services tend to be subscription-based and, for the moment, information available seems to be confined to major highways and arterials. Second, private sector activities on the public transportation side seem much more limited, with only a few companies active in the area. The one company providing real-time information with predictive capabilities, NextBus, has service providers (as opposed to travelers) as its direct clients. The current tilt in activity towards car-based applications may be simply due to perceived (or actual) market potential (nationally, public transport accounts for just 5% of all trips in the USA), a more difficult revenue model to implement for public transport, and./or some combination of these and other factors.

We suspect that the greatest eventual societal value for TDM or related transportation applications will come when data fusion can be utilized to introduce the suite of needed information at the necessary moment(s) in time, allowing, users to answer questions such as "should I travel now for that purpose? Should I take this mode and if, so, what time should I leave? What are the time-money-reliability-emissions trade-offs of my various options?" and so on. Market forces alone may well not provide enough incentive to develop a fully operable, integrated multi modal real-time (with predictive capabilities) application that would be necessary to answer such questions. The public sector will play a key role in bringing such an application to realization, yet the institutional challenges remain non-trivial and may indeed exceed the technical challenges. In some cases, existing agreements (contracts) with parties responsible for system elements (e.g., with a company to operate and maintain traffic signal control systems) may significantly hamper data fusion by, for example, prohibiting data sharing. This raises important issues related to data "ownership" and privacy concerns, issues which require adequate attention for DF applications to reach their potential.

Our brief examination of several metropolitan area cases from the USA suggests factors which might accelerate multi-modal, real-time DF adoption. Market potential plays some apparent role, as more auto-dependent places do not yet have such DF applications in place and high tech industry presence does seem associated with adoption of more advanced systems. Governance structures may also play a role, as relevant local authorities with more financial independence are also associated with more advanced DF. Ultimately, advanced DF will require public-private partnerships within metropolitan areas, perhaps following the Berlin approach. Such partnerships will not only have to create the right incentive structure to ensure maximization of the public good, but will also have to work to create the right standards, etc. Fully integrated DF systems require interoperability protocols and efforts are already underway to standardize transport systems communication, normally based on XML (e.g. DATEX in EU, TIH in the UK, NTCIP in the US, to name a few). When we reach widespread communication of Car-to-Infrastructure (C2I), Car-to-Car (C2C) and among different infrastructure subsystems, the opportunities for, and challenges to, DF for TDM will grow exponentially.

Finally, important questions remain relating to how users will actually respond to the information generated and made available through such systems. In other words, will users utilize the information to make "better" travel decisions, those consistent with the goals of TDM? And, will such information further blur the lines between users, service providers, and planners? We can fairly characterize the current state of DF in transportation as analogous to "transport1.0," where data providers (public or private) collect, process and publish the data. However, pervasive computing environments and the Internet make possible a new model of "transport2.0," where end users can contribute information to describe travel conditions and more.[4] This may encourage citizens to increase their participation in the planning and operation of the transportation system, introducing stronger bottom-up elements. Such developments would mirror the "open source" software model and, more generally, the new communication methods, applications and usage patterns appearing almost daily (e.g., Blogs,

[4] A number of "transport2.0" projects already exist, in which citizens send traffic information (via internet or phone) or even map corrections (e.g. TomTom MapShare).

Wikis, etc.). Using unstructured data such as a web page containing a transport-related news story, pictures (e.g. Flickr), audio and video (e.g. YouTube) for TDM applications present a challenge in terms of DF, requiring either the structuring of data via transformation (involving technologies such as natural-language processing and semantic mapping) or creating specialized data mining tools. While certainly a challenge, such applications may provide large scalable benefits, ultimately reducing the need to deploy physical hardware throughout the transportation infrastructure.

ACKNOWLEDGMENTS

This publication was made possible by the generous support of the Government of Portugal through the Portuguese Foundation for International Cooperation in Science, Technology and Higher Education and was undertaken in the MIT-Portugal Program.

REFERENCES

Association of Metropolitan Planning Organizations (AMPO). (2005). 2005 Profiles of Metropolitan Planning Organizations. AMPO Publications.

Denver Regional Council of Governments (DRCOG). (2008). DRCOG Website – About DRCOG: http://www.drcog.org/index.cfm?page=AboutDRCOG.

Esteban, J.; Starr, A.; Willetts, R.; Hannah, P. and Bryanston-Cross, P. (2005). A Review of data fusion models and architectures: towards engineering guidelines. *Neural Computing & Applications*, 14, 273–281.

Federal Highway Administration (FHWA). (2006a). Integrated Corridor Management (ICM) Pioneer Sites - Leaders, Innovators in Congestion Management. USDOT, FHWA, Research and Innovative Technology Administration, Intelligent Transportation Systems Division: http://www.its.dot.gov/icms/pioneer.htm.

Federal Highway Administration (FHWA). (2006b). Urban Partnerships Congestion Initiative. National Strategy to Reduce Congestion on America's Transportation Network Initiative: http://www.upa.dot.gov/.

Federal Highway Administration (FHWA). (1998). Transportation Technology Innovation and Demonstration (TTID) Program. USDOT, FHWA, Office of Operations, Real-Time Traveler Information Program: http://ops.fhwa.dot.gov/travelinfo/ttidprogram/ttidprogram.htm.

Harris C.J., Bailey A., Dodd T.J. (1998). Multi-sensor data fusion in defence and aerospace. *Aeronaut J*, 102(1015), 229–244.

Llinas, J., Bowman, C., Rogova, G., Steinberg, A., Waltz, E., White, F. (2004). Revisiting the JDL Data Fusion Model II: http://www.fusion2004.foi.se/papers/IF04-1218.pdf.

Luo R., Kay M. (1988). Multisensor integration and fusion: issues and approaches. *SPIE Sens Fusn 931*, 42–49.

Marchal, F., Hackney J. and Axhausen K.W., 2004, Efficient map-matching of large GPS data sets - Tests on a speed monitoring experiment in Zurich. Arbeitsbericht Verkehrs- und Raumplanung, 244, IVT, ETH Zürich,

Mecklenburg-Union Metropolitan Planning Organization (MUMPO). (2008). MUMPO Website – About MUMPO: http://www.mumpo.org/About_Us.htm.

Metroplan Orlando. (2008). Metroplan Website – About Us: http://www.metroplanorlando.com/site/about/default.asp

Minneapolis Metropolitan Council (MMC). (2008). Metropolitan Council Website – About the Council: http://www.metrocouncil.org/about/index.htm.

Ohio, Kentucky, Indiana Regional Council of Governments (OKI). (2008). OKI Website – OKI Overview: http://www.oki.org/overview/index.html.

Portland Metro (Metro). (2008). Portland Metro – About Metro: http://www.oregonmetro.gov/index.cfm/go/by.web/id=24201.

Puget Sound Regional Council (PSRC). (2008). PSRC Website – About PSRC: http://psrc.org/about/index.htm.

Progressive Policy Institute (PPI). (2001). The Metropolitan New Economy Index: Benchmarking Economic Transformation in the Nation's Metropolitan Areas. PPI Technology and New Economy Project: http://www.neweconomyindex.org/metro/index.html.

PTV-Berlin (2003) Traffic Management Centre (VMZ) Berlin: VMZ keeps Berlin moving: www.ptv-ag.com/download/logistics/anwenderberichte/Ref_VMZBerlin_e.pdf.

Rupert, B., Wright, J., Pretorius, P., Cook, G., Hutchinson, K., Kell, W., Lister, H., Nevarez, M., Sanders, L., Schuman, R., Taylor, R.,Almborg, J. (2003). Traveler Information Systems in Europe. Federal Highway Administration. FHWA-PL-03-005.

San Antonio-Bexar County Metropolitan Planning Organization (SABCMPO). (2008). San Antonio-Bexar County MPO Website – What We Do: http://www.sametroplan.org/pages/About_MPO/whatwedo.html.

San Diego Association of Governments (SANDAG). (2008). SANDAG Website – About SANDAG: http://www.sandag.org/index.asp?fuseaction=about.home.

San Francisco Metropolitan Transportation Commission (MTC). (2008). MTC Website – About MTC: http://www.mtc.ca.gov/about_mtc/about.htm.

Siemens (2008). Traffic Management Systems - Solutions from Siemens, AG: http://www.siemens.nl/its/.

Southwestern Pennsylvania Commission (SPC). (2008). SPC Website – About SPC: http://www.spcregion.org/about.shtml.

Texas Transportation Institute (TTI). (2007). 2007 Annual Urban Mobility Report. Texas A&M University System: http://mobility.tamu.edu/ums/.

Thomopoulos, S. C. (1989). Sensor integration and data fusion. *Proceedings of SPIE 1198, Sensor fusion II: Human and machine strategies*, 178–191.

Wang, X.; Qi, H.; Iyengar, S. S. and Chakrabarty, K. (2004). Multisensor Data Fusion in Distributed Sensor Networks Using Mobile Agents. *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 6, 740-753.

Yin, R.K. (2003). *Case Study Research: Design and Methods*. Third Edition, Sage, Beverly Hills.