ADDRESSING GAPS IN THE AVAILABILITY OF TRAVEL BEHAVIOUR DATA

Marcus Wigan
Napier University
John Polak
Imperial College London
James Cooper
Napier University
Jan-Dirk Schmoecker
Imperial College London

1. INTRODUCTION

Sound transport research and policy making depends upon the availability of appropriate, high quality and up to date information. Thus, the means by which transport related data are collected, stored, processed and made available for use are issues of central concern to the transport profession.

In recent years there has been a significant expansion in the range and complexity of the issues confronting transport researchers and decision makers. Emphasis is now placed not only on the traditional concerns of capacity provision and management but also on the relationship of transport to environmental issues and health impacts and to wider considerations of land use and sustainability. Moreover, much greater emphasis is also being placed on the differential impact of transport on specific groups such as the elderly, the disabled or those groups deemed to be socially excluded. These new policy drivers have given rise to a range of new research questions calling for new and different types of modelling and analysis and hence also, underpinning data. This in turn has exposed a number of significant problems with existing data provision ranging from a basic lack of data on potentially significant aspects of behaviour (e.g., travellers' temporal constraints and preferences) to ambiguities and uncertainties in the definition of key data items. None of these problems are in themselves new, but the growing demand from the research and practitioner communities for more and more complex behavioural data has highlighted afresh their importance.

At the same time as these 'demand' factors are operating, a parallel set of data 'supply' factors are beginning to assert themselves and potentially transform the landscape of data provision:

- Many travel behaviour surveys in the UK and elsewhere are encountering increasing difficulties in obtaining the cooperation of the general public to take part.
- Despite this, there is an increasing emphasis on continuous surveys of individual travel behaviour, principally driven by the need to closely monitor the implementation and effects of policy.

- The rapid emergence of new forms of data collection technologies, especially those based on navigation and positioning technologies and web based methods.
- The increasing availability of large volumes of data from ITS and other automatic operational sources, though uncertainty remains regarding the extent to which these data will be available and configured in a manner useful for secondary research use.

This convergence between demand and supply factors creates many exciting opportunities, but also poses challenges. One of the key challenge is to identify those areas in which the opportunities offered by new and emerging technologies and methods can make the most effective contribution to relieving key data bottlenecks on research and practice.

This paper discusses some of the (sometimes unexpected) issues that have arisen in a project carried out to identify key areas in which shortcomings in the collection, processing and dissemination of behavioural data have held back needed research and, building on this, to define how best, in the light of current trends in data collection and management methods, these limitations can be addressed. The specific objectives of this work were to:

- To consult widely with the relevant professional communities (e.g., government, research, operational) to establish the existence and nature of gaps and shortcomings in existing data sources, access and management methods.
- To assess the areas where research gaps might best be addressed by a coordinated or 'model' data set collection programme.
- To report on the outcome of the review and consultation process and to make specific recommendations on data collection and integration

The project was supported by the UK Engineering and Physical Sciences Research Council. Whilst principal focus of the work was therefore on the UK data context, the review and consultation activities were international in scope.

Consultations were undertaken in Northern Ireland, Scotland and England, and at the major international conference on travel behaviour held in Switzerland.

2. CURRENT CONTEXT

Travel behaviour analysis and research draws upon an increasingly wide range of sources and perspectives. Some of these are physical and can be measured directly while others depend entirely on direct access to the individuals whose behaviour is being studied.

Traffic data such as flows are relevant as they allow models of route choice to be tested and applied, although the vehicle flows are not immediately evident as travel behaviour datasets. Microsimulation has made the detailed

behaviour of individual agents (persons, vehicles) in the traffic stream matters of direct interest in travel behaviour, and the agent approach has extended the need for different types of behavioural data such as car following behaviour and the parameters that are chosen to control the behaviour of agents in the representations of drivers, pedestrians and decision making parties (such as freight shippers). Similarly land use changes are now addressable through simulations of location choice and migration behaviour, and so any models of this process now require and make effective use of better understanding of this aspect of travel behaviour.

Some of the this extension in the types and range data sets and areas of interest is due to the increased range of analysis, data collection and modelling/representation tools now available. Some is due to a widening of the areas of interests in cross-disciplinary research, and the policies that are related to these cross cutting areas. Such areas include the study of gender effects and of social exclusion and the behavioural characteristics and responses to it in different forms. Others are required to account for the behaviour of individuals when presented with environmental, safety and other priority choices as part of extended economic evaluation developments. These extensions include the determination of choice sets considered by individuals in specific circumstances, and their views on the accessibility and mobility that they possess. In many cases the research into behaviour and modelling requires data to be collected for each project by the very nature of the behaviours and perceptions involved.

These emerging issues and approaches differ from many of the historical approaches, where a large scale cross sectional survey of a large number of household and individual characteristics would provide the basis for as whole family of modelling and other enquiries. However such surveys were not always used very well, and wider questions that could be addressed suing them were often ignored. The reasons for this in the 1980s were a combination of barriers to access – such data sets were large in the terms of the day – barriers to their use – they required scarce computing and programming skills - and barriers form the then current mind sets which were oriented overwhelmingly towards vehicle rather than personal behaviour ,and towards geographically specific summaries (especially origin-destination matrices) as the preferred basis for making use of the rich household survey data.

This emphasis widened slowly (Wigan, 1987), as personal behaviour, the impacts and responses of subgroups and gender and activity issues became more important in research. While individual choice models were put firmly on the agenda at the Transportation Research Board meeting held in Williamsberg in 1973, the use of stated preference methods grew only slowly, with the datasets generated as an intrinsic part of what was for a long time regarded as solely a research issue. Similarly the activity datasets and methods developed in the early 1980's were also research exercises undertaken to secure this new form of data rather than surveys for general or wider use by a larger research community.

Even securing the retention of the rich (and often not fully exploited) household travel surveys was still a serious question (Wigan, 1985), let alone access to these datasets.

As discrete choice modelling has developed, so too have the Stated Preference surveys, as activity analysis has grown, so too have the activity-travel diaries so as to capture the greater complexity more effectively. In most cases the state of the data capture art has grown in step with the research, and the research has usually needed data collection. This symbiotic relationship continued for perhaps two decades – throughout out the period when all types of data were scarce and usually expensive to process.

Simpler types of data were the first to be automated. Fischer-Porter pneumatic tube traffic counters were collecting nationwide data over 30 years ago, and roles for manual counts continue to the present day. However the advent of ITS systems, area traffic control sensors, and improved communications have completely revolutionised this type of data: it now can be collected economically in terabytes – but of course does not necessarily yield an equivalent increase in meaningful information.

This massive increase in the potential scale of data acquisition has now been joined by geospatial data. GPS tracking is now a reality not only for freight movement, but also for personal movements. The data qualities have risen accordingly, and the combination of GPS and survey methods on PDAs allows even more detailed data to be acquired. Again, this growth has arisen mainly in an iterative process of research projects collecting data, and research then improving how it can be used.

There are still points at which research councils and governments can profitably prime the research pump with data – especially if it is of good quality and well organised. Examples where this has been done on a large scale in the past are areas such as household travel in Baltimore in the 1990s' and an extensive transport and behaviour time series continues to be created for Puget Sound area by the FHWA.

3. DATA COLLECTION RATIONALES

Data means something when turned into information. Almost all forms of data collection are predicated to at least some extent on the purposes to which it is expected to be applied. These may be applied purposes, or to support research. The applied work is within a context where the compromises in coverage, detail and aggregation can be made with reference to the applied context for which the data set was aimed. However, for research purposes there is amore complex process at work. These processes usually fall into a small range of approaches, which can be described as:

- Here is some interesting data, let us explore it from certain points of view and see what we can learn, and what hypotheses emerge

- We have found some interesting effects in a set of data we have collected as part of a research project, and now want to see if the findings are generalisable elsewhere

- We have a research problem which demands that we go out and find out about individuals and organisational attitudes and actions, so we need to collect data as part this research and enquiry project

- We have developed some new theory or modelling techniques, and need data in a form in which it is not presently available, so we need to get some data to test our innovation

- We have a new technique or model which we wish to populate to represent an area or region, and so we need data of a wide range of types to make this model operative

- We have found an interesting or important influence in a set of data representing a specific year, and we need to find earlier/later cross sections of the same or comparable information

- We are looking at how processes develop in transport and related fields over time, so we need time series information that is not available - how can we get it? Do we need to do a prospective study?

Each of these approaches is different, some draw from, access to existing datasets, some form linkages between existing data sets, and some need fresh data collections.

In each case the research question dominates the compromises that can be made to ensure that the data processes are manageable but there are many such cases and it is a fair question to ask how much could better integrated access to sources – or even fresh integrated sources in a large special collection – address the needs of advancing both understanding and the art.

The costs of many different small – or large – scale data collections add up quickly, and those responsible for funding research may have concerns over the fractions of grant or research funds that are used for the data aspects. One approach to this is to build in a requirement to deposit the data and documentation in a cumulative archive, a solution that has proved to be very successful for many aspects of the social sciences in the form of the Social Science Data Archive (SSDA), but less successful in the transport areas.

One reason for this may be the sheer complexity of many of the transport data sets, another that the integration of several different types of data are required in may cases, and so the transport data sets have not fitted in well with the other types of data collections. Whatever the reasons, the general model offered by the SSDA framework clearly has had much to offer many areas of social science and will continue to do so.

Transport appears to need a slightly different approach. Time use data bridges the two areas of general social science and of transport) and has, through a focussed investment in data and ESRC funded personnel has over many years achieved an international exchange on a common core set of items, and excellent integration into SSDA and the equivalent organisations elsewhere. However, time use as a whole is not a complex data structure, and fits far more easily into the mainstream of social science data collections than many transport data sets.

It seems likely that the complexity of many transport data sets will require a special initiative to reach the same levels of access and understanding as is now possible with many social science data sets in the SSDA and its overseas ilk.

4. METADATA ISSUES

Mining existing transport related data sets is also fraught with a wide range of barriers, which limit the access to existing information as well as confusing and obscuring the issues of gaps in available – and indeed prospective – data resources. The only area where some progress can be pointed to is that of the historical data collections typified by SSDA of - which there are few in the transport area.

The problems of knowledge of the existence of data are now becoming more manageable - at least in principle. A brief diversion to discuss this is relevant at this point, as a number of issues recorded in the consultations bear directly on this area.

Techniques are emerging to address metadata requirements, and subsequently to enable data integration on the fly through devices such as Nesstar's Explorer tools. The others in the Nesstar (Ryssevik & Musgrave, 1999) bundle are concerned with creating formal descriptions of whatever data set one is dealing with and reformatting the description with as much or as little additional information as one chooses, and creating a machine and human readable XML file. While few, if any, will read through a 70,000 record written entirely in XML produced by such as process, it does indeed contain no more and no less information that one chooses to put into it.

While a wide range of parties in the consultations were anxious to see some sort of transport data access resource, research workers raised the issues of possible imposed standardisation or reliance on a specific software vendor. Both of these were regarded as presenting unacceptable risk to both research and to the ability to manage ones own data. Phrases like 'We don't want a Microsoft for data access and management' were typical.

The general idea of metadata has taken some time to become understood and appreciated in transport (Wigan, Grieco, & Hine, 2002), The process is still continuing, and has some way yet to go before it is generally appreciated and understood by transport research, policy and practitioner groups. Even in

the safety area, where metadata has much to offer, the ideas have as yet made little impact. The main areas where the ideas have been taken up are in Geographic Information Systems, where a number of approaches have been taken up and adopted, mainly reliant on spatial resolution and accuracy aspects as one would expect in spatial data bases. One such formal example is the FGDC (US Federal Geographic Data Committee) (Federal Geographic Data Committee, 1998), and another is the metadata specification initially adopted by the US Federal Transit Authority (FTA, 1996) – which is very precise on geospatial aspects, but almost all the transport factors are left for free text description.

Formal approaches to specifying data are now available, and Axhausen and Wigan (Axhausen & Wigan, 2003) give a detailed discussion of their application to transport surveys. Nevertheless, the best that is generally available is a set of written descriptions of different elements of data and its ownership. Even this is lacking in all too many current transport data sets.

The recurrent theme in the consultations of a need to locate and make reliable and effective use of existing data sets where possible is not due simply to a reluctance to collect data, or indeed in many cases to fund it, but more to the need to bring more and more different types of data together, often originally collected for very different purposes, as the span on information types required for research has expanded substantially – and in many cases well beyond most transport specialists range of personal knowledge.

This theme is also one that is becoming audible in discussions of community participation in transport and planning (Hine, Wigan, & Grieco, 2003), where the ability to access and make use of wide range of data types is growing as at least some of the barriers to this form of contestability in public and transport debate begin to fall .

5. THE CHANGING LANDSCAPE FOR TRANSPORT DATA COLLECTION AND RETENTION

The concerns over undue or premature standardisation are well founded fears judging from the effects on research in other fields, although researchers had a number of other agendas which should be honestly faced up to. The effort required to formulate full documentation of data sets is a perennial problem, although the simple retention of the basic data in any form at all still remains a genuine issue. Research funding bodies have moved progressively over the years from funding software production to requiring it to be an outcome from projects rather than its raison d'etre. Similarly, in more recent times data as the main outcome of a research project has fallen far from favour, although the retention and archiving of data created as part of a research project still remains in a grey area in terms of retention and documentation after the project is completed.

There has been a major shift in data collection processes over the last decade or so. While it was previously collected and used mostly via government (local

or national level) bodies, or from specific research grants, since 1990 there has been a steady migration to consulting bodies and a rise in private data collection – and retention.

This shift has been associated with a rise in privatised transport bodies, the beginnings of a private entrepreneurial data collection industry exemplified in the UK by enterprises such as TrafficMaster and iTIS, and a substantially greater awareness of the value of reducing access to data for competitive reasons. One of the best examples of this is the series of public transport franchises that did not require data provision to the relevant regulator. The Scottish bus franchises are a case in point, where much of the data is simply inaccessible, and concerns over contestability were also reported as a reason for data withholding in Northern Ireland.

These changes have made intellectual property (IP), commercial, and competitive protection issues become genuine barriers to access. In addition, data registration, data protection, privacy and constraints on data collection and repurposing through Ethics Committees have also become established over the same period (from about 1990).

In total these shifts have changed the landscape for transport data access, at the same time as certain forms of data have become available in a flood. The new richness in data is an outcrop of the maturing of ITS, and massive traffic flow data is typical of the types of data now becoming readily available. Such huge data flows present completely different problems for research – as well as new opportunities – and are at this initial stage are concerned more with the extraction of meaningful information than exploring new conceptual and behavioural ideas about transport and behaviour. The application to better understanding of behaviour will follow once new methods have been developed to make effective and appropriate use of such data streams for understanding of a wider range of issues and mechanisms. These flows have been one of the first to be commercialised, and the vendors are actively seeking to expand their product range and the added value that can be commanded (and paid for) from such sources.

This commercial motivation is one that must be recognised as it will only grow, and the intellectual property issues involved will continue to increase as licensing per project rather than for general access is a popular way of maximising the revenues form such sources.

The overall effects of these trends are to raise many new types of barriers not only to access to data sources, but also obscure what is actually available (or potentially so).

The use of data withholding as a competitive measure, and the reduction in contestability created thereby, is not new, the difference here is that so much of transport is a current and vital concern to both community and government, as well as research and commerce.

6. MODEL DATA SETS

The idea of a model data set was one mooted at the beginning of this project. It has survived - but only in a mutated form. A 'model data set' for research would be one that covered a range of transport and behavioural areas, and was integrated and accessible, well documented and of verifiable quality.

Each of these aspects must now be viewed through a very different lens than the one used at the outset. Simply assembling such a model data set from current data sources would be subject to many of the barriers already discussed. To make such an integrated data set available would require fresh data to be collected in many areas where data already existed but were not available for integration due to one or more of the following barriers:

- Privacy
- Commercial interests
- Cost
- Constraints on usage imposed by licence
- License conditions that require any additions or corrections to any data provided under agreement to be returned to the provider at no charge.

Even this latter condition is becoming problematical as regulations on privacy and repurposing become more widespread, and grandfathering of such restrictive licences become more widespread.

This emerging situation suggests that the 'model data set' issue needs two separate thrusts.

- A unified *ab initio* collection free of the burgeoning constraints implied by assembling multiple sources. This may be justified in some new situations now arising in transport policy

- A data broker or linkage role which would fulfil many of the same requirements but not address others.

Ideally both should be carefully considered, and the key aspects brought together in some way. There are many options, and an increasing range of technical approaches that can be used.

7. DATA GAPS AS ADDRESSED ELSEWHERE

Data gaps have previously been addressed in three different ways.

- Just to ensure that copies of all available materials are held
- As part of an assessment of the need for transport data
- As part of a process specifically to identify unmet research needs

An example of the first motivation: The US DoT Bureau of Transportation Statistics (BTS) was set up under the ISTEA (Intermodal Surface Transport and Environment) Act in the early 1990s, and was reauthorised under the subsequent TEA21 Act that followed five years later. It was initiated largely as a direct result of TRB initiative in the Data Committees which assembled a powerful case for the importance of data as a basis for decision making, and the need to make the data holdings of the US more effective to that end.

Another example is the long drawn out effort to make sense of the torrents of ITS data now emerging. This has taken the rather different tack of seeking to find ways of making the data accessible and specified well enough that it can be made use of. The standard developed by the ASTM E17.54 Committee as now been published (ASTM, 2003), and formal metadata issues are now being actively considered by the same group. These objectives correspond very well with some of the demands emerging from this current UK EPSRC project.

- A metadata standard specification for archiving ITS-generated data. This standard will provide the exact structure for the metadata needed in addition to those attributes required for ITS data dictionaries.

- A standard specification for archiving ITS-generated traffic monitoring data in a data dictionary for archiving traffic data, a record structure for creating data tables, and a file transfer format.

Once again, members of the TRB Data Committees (Margiotta, 1998) play a major role, and it might be useful to consider what similar groupings in Europe represent or focus similarly effective networks.

After the initial five year period of data assembly and distribution to all interested parties, BTS moved to assess if their coverage was complete. This was not a quality issue, or indeed one predicated on the research needs that would be addressed, it was aimed simply at completeness of the BTS holdings and the web based survey addressed the need for data to address gaps encountered by the respondents. The objective was primarily to pick up missing data sets and secondarily to consider the possible case to collect new data to fill gaps identified. In parallel with that initiative BTS began working on a national transportation library to make access to their holdings easier. Again, the metadata and quality issues were not addressed at this stage. However the wide distribution and the efficiency of the BTS data packaging and redistribution process has been remarkably effective in simply making US transport data available both inside the US and outside it.

An example of the second motivation, as a result of the broader transport user and behaviour insights shown to be possible to extract from household travel surveys (Wigan, 1987), the Victoria Australia Ministry of Transport undertook a project (Taylor, Young, Wigan, & Ogden, 1992a, b) to determine exactly what data was needed and what it was needed for. The emphasis was mainly on household interview or survey data, and eventually gave rise to an improved activity and household survey process in the State.

A further example is the international Delphi survey done by the Institute for Transport Studies in Sydney Australia (Wigan et al., 1995), which looked at the areas where data was needed – and specifically addressed some of the barriers to its access and use. A later paper added the modelling implications of this work (Hensher, 2000), building on the same international survey results. The questions emphasised the problems in access to data.

> "What are the most common frustrations you have faced in accessing information from… government agencies, private data agencies, universities and other sources.
> Who collects useful primary data in your country (from the same list)
> Where do you usually get your travel data (for transport planning and evaluation). list up to five sources used in the last two years.
> What in your view are the most important core urban travel data items that should be collected to service the transport planning and research community?" (Wigan et al., 1995)

In addition, this international group was asked:

"Where, in your view, does the expertise in your country lie in the following skill areas? Please rate each organisations' skills from 1 to 10 (1=very good, 10=very poor) Federal Government Universities, State/Provincial Government Subsidised Research Organisations, Local Government, Consultants"

The ratings for a range of data access, management, collection, analysis and modelling issues are shown in Fig. 1

These results make it clear that who collects and holds data makes a substantial difference on a skills quality dimension. An aspect of data gaps and barriers to access and use that is rarely explicitly addressed, but which emphasises that steps towards quality process and ratings – possibly peer review ratings of data holdings- will soon become necessary.

8. CONSULTATION COVERAGE

The results of Fig. 1 illustrate that it is important to ensure a spread of consultees on data gaps, and to identify what this spread has been. There was a major difference between the invitations issued and the attendees, with an under representation of research workers at the physical sessions. This is being addressed though a direct survey as the last stage of the work to provide and secure feedback on the outcomes of the work to date and to provide a grounded basis for research worker responses against this backdrop.

Fig. 1 Overall organisational ratings for different data and analysis functions (Wigan et al., 1995)

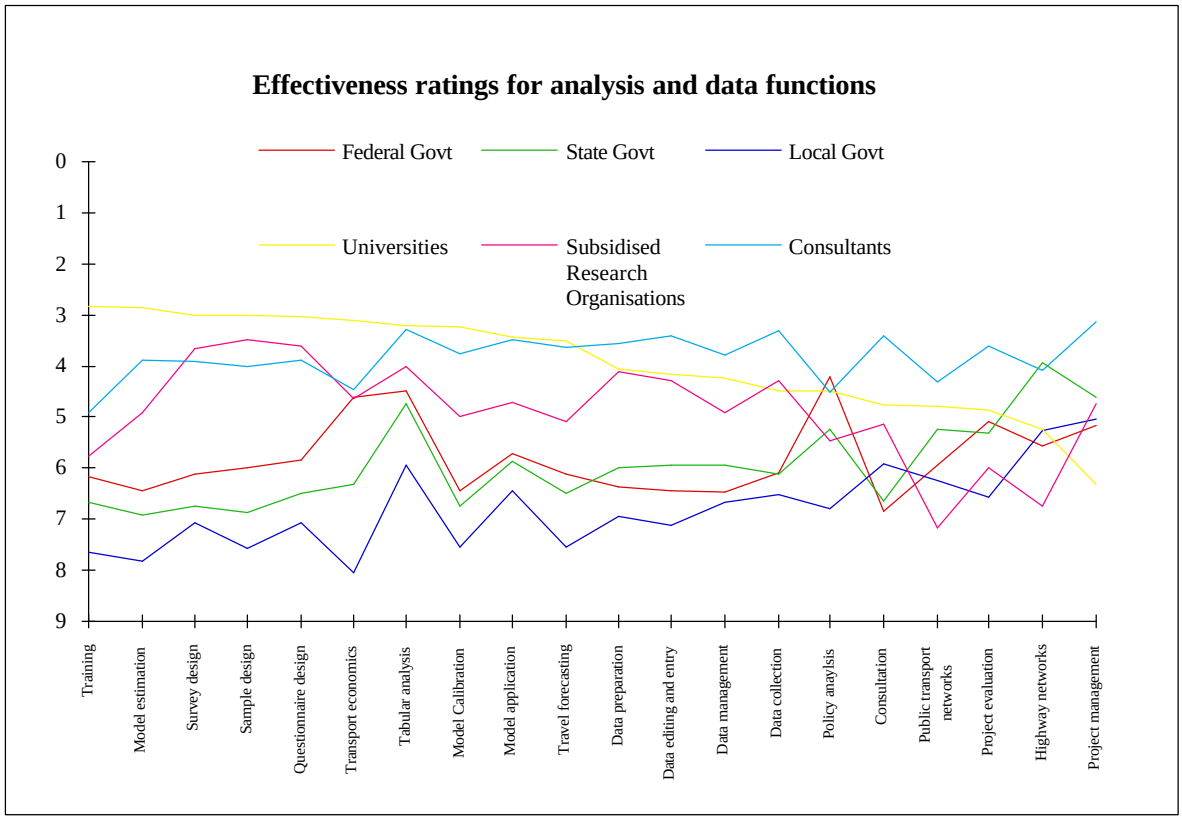


Fig. 2 A typical profile of consultation attendees.

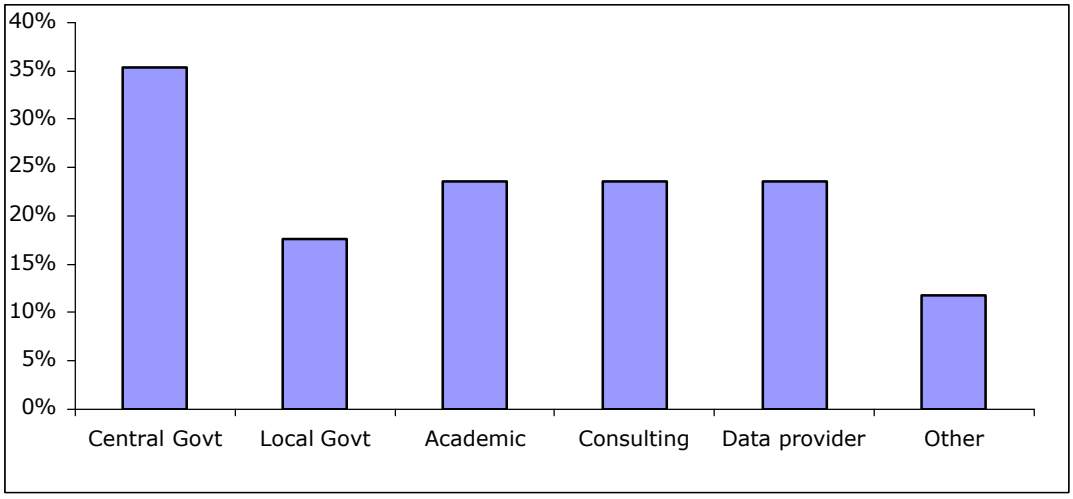
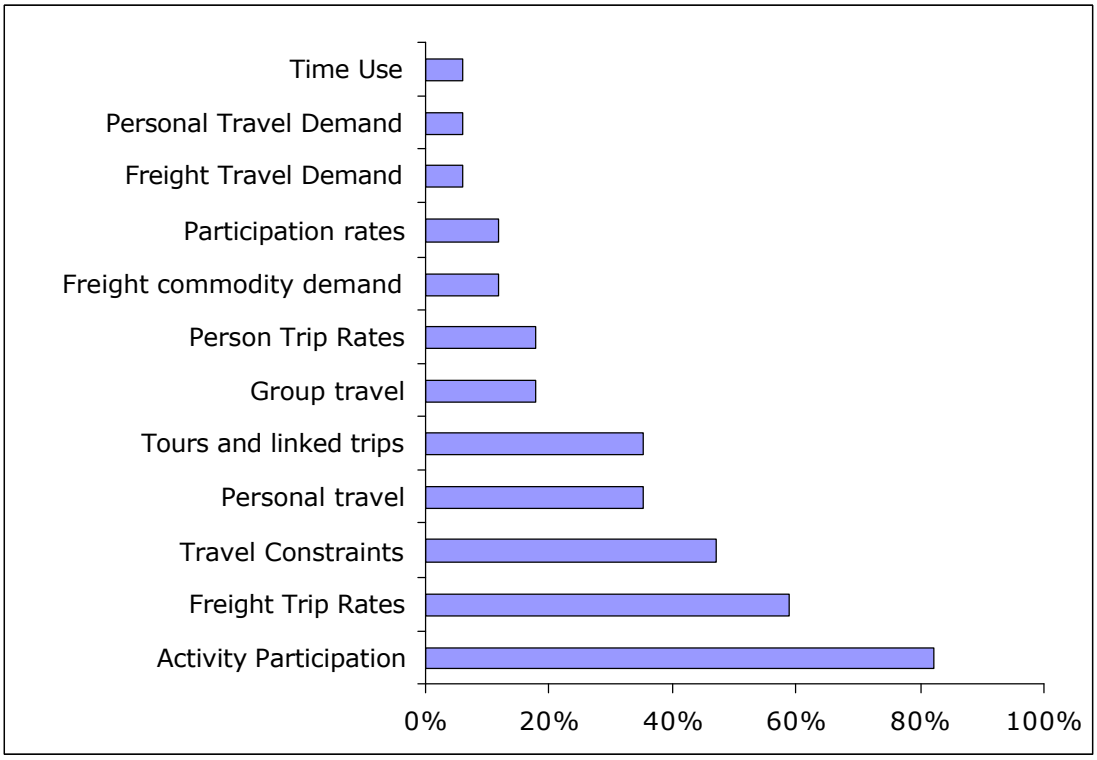


Fig 3. The types of travel data regularly used by these attendees



9. DATA GAPS AS BARRIERS

There are many different types of barriers that limit the use made of data. If the data is collected as part of a research project, then many of these simply do not arise (although subsequent prospective users may be constrained by the actions of the researcher), and so the balance between obtaining existing data and collecting ones own will often be tipped strongly towards collecting oneself if that is possible. If it is not, as is frequently the case, the form and accessibility of the data will constrain the research approach unless one is singularly fortunate. The major barrier types can be summarised as follows:

- **Availability**: In some cases the required data are simply not collected (e.g., due technical difficulty or cost) or if collected are of insufficient temporal or geographical coverage or quality to render a useful estimate of the required parameter(s).

- **Access and use**: Ownership and access rights relating to transport data are often opaque resulting in (real or imagined) barriers to access.

- **Fragmentation**: Data holdings are often fragmented, both physically and institutionally (both within and between organisations), leading to both institutional and operational difficulties in fully exploiting existing data resources.

- **Documentation**: Opportunities for the re-use and efficient archival storage of data are often limited by the lack of proper data documentation and archival processes and/or administrative provision for re-use.

- **Standards**: There exist few appropriate standards for the precise characterisation and description of transport activity. Nor are there, in general, agreed metadata conventions for the exchange of information on the precise nature of data holdings.

- **Integration**: Rarely does a single dataset contain all the information necessary to address a particular research or policy question, leading to the need to combine data from different sources – a task fraught with statistical and methodological difficulties.

- **Uncertainty**: Datasets almost invariably embody uncertainty regarding the parameters of the real-world process to which they relate, due to sampling and/or non-sampling errors. Yet data are frequently stored and transferred without taking into account this uncertainty, leading to inhibitions stemming from the perceived risk of inappropriate use.

- **Model outputs embedded in 'data'**: It is not unusual for model estimates to be sued to fill out missing data points and other lacunae in data sets prior to their operational use. This process is not always documented. A similar issue is the randomisation of values placed in small cells by census bodies – the effects are not always obvious and the process not always well documented

- **Quality**: the data coding accuracy, as well as the quality of the survey questions and design, as well as the sample response rates and design biases are all barriers. In many cases if these factors are known then effective use can still be made of the data, in other cases it precluded their effective use. The lack of information about the quality aspects is a barrier that is encountered more frequently as an issue by experiences users of secondary data…

10. CONSULTATION PERSPECTIVES

Some key findings from the consultation process are discussed in this section. The consultation process encountered the same problems as all consultations. The perspectives presented generally from the very specific contexts of each individual rather than organisational, and the response rate to the meeting invitations varied substantially as a result.

The instrumental aspect – what was likely to be achieved for the respondent by participation in the process – and the nature of the responses in terms of the perspectives on outcomes were evident. The response rates from users of third party information have in general a higher level of interest and a shorter fuse in terms of immediacy of the issues than the majority of the academically

oriented participants. This also was reflected in the response rates to the invitations as well as in the framework within which the participation contributions were expressed. The creation of data collection requirements through government regulations has had the effect of requiring certain forms of data to be collected to provide monitoring and management of different government transport objectives. A desirable result – but one that appears in some of the respondents' eyes to have displaced information collection of greater reliability and value, while not necessarily achieving the results in terms of the effective monitoring that had been the original government intent.

There is an interesting contrast between this point of view, repeatedly espoused by local government parties, and the objectives of the ISTEA Act in the early 1990s, where requirements were set both for fresh data to support all transport proposals (including the no-build options and non-motorised transport) – a very successful initiative - and for a series of management and data management processes – which were largely neglected and treated in a far less responsive manner by the States and MPOs (Metropolitan Transport Organisations). In addition the initiation of the Bureau of Transport Statics made a material difference in terms of the access to available information for all parties.

In the UK situation the management processes have been set to have data collection outcomes, but have not been matched by a BTS function funded by government. This is probably due to the healthier state of UK transport data than in the US in the early 1990s, and to a very different perspective on available data. There are few barriers to the privatisation of public data in the UK (unlike the US) and also the procedural and process guidance of the UK Department for Transport has been considerably more extensive and prescriptive than in the US.

The combined effect appears to have been a consolidation around an unspoken view that *'local data should be collected by local authorities to support local applications and funding cases'*. Nevertheless, the Nationally funded data collections – overwhelmingly the UK National Travel Survey (the NTS) – has been frequently rated the most trusted and useful data source, although unsuitable for many local applications (due to limitations in sample size).

The broad (and mandated) recent Quality Assurance Reviews of UK government transport data sources have as a result led directly to a restoration of a considerably higher sample rate for the NTS: the NTS being also the most widely and freely available data source, and subject to some of the best process and content quality procedures available.

The lack of a coordinating and open distribution organisation like BTS has two different effects on consultee perspectives:

- Concern over the low priority that data sources other than traffic counts appeared to hold in LGA budgetry outcomes, and

- Concern over access or knowledge of third party or otherwise available data sources.

At a National government level, the widespread withdrawal from outside access by privatisation of key data collected within the now mostly-franchised public transport organizations and the growth of third party dedicated data collection, broking, integration and resale organizations has also shifted the approaches towards internal data collection and availability in complex ways.

At the most comprehensive level, Transport for London has maintained the long and strong tradition of this major metropolitan region collecting and managing transport and related data as a key function. The GLC and now the GLA have continued to regard transport data as a strategic requirement – and in this substantially share national government perspectives.

The perspectives expressed by those employed in academic research displayed a remarkable degree of agreement with many of these standpoints. While some of this may have been due to a convergence of perspective due to the large scale reliance on consulting income - and thus a shift to understand and reflect specific client direct needs and standpoints) it was not clear how much was due to an implied assumption that fresh research would require fresh data as part of the specific research project.

There was however a lively appreciation of the costs of fresh data collection, and the need to locate external and existing sources where possible, Most of the barriers hypothesised at the start of the project and reviewed in the last section were confirmed and supported as real. However what was at first sight the most surprising outcome – a remarkable lack of specifics on research needing to be undertaken but debarred by the lack of available suitable data – became on analysis much easier to understand.

Research funds and grants are highly competitive, and although working conferences such as the international IATBR (International Association of Travel Behaviour Research) and Survey Methods often come up with general areas where research is needed – rarely do they go to the level of detail where individual projects can be pinpointed down to the precise data requirements involved. The long and slow process of developing 'Research Problem Statements' within the US Transportation Research Board structure should have altered us to this, but the parallels did not become evident until we had carried out this work. It is not surprising that once a TRB problem statement is published it becomes a priority target for the US funding agencies to support, and even with this direct link the process is slow and often painful, extending over several years.

A more general TRB process, the 'Millennium statements' for the year 2000 were requested by TRB from all Committees. Most Committees responded, but usually only from a subset of the committee members. The results make excellent and useful reading, and generally reflect the IATBR type of experience. This Millennium process was considerably more successful than most of the participants had expected in pointing the way. The TRB urban

data committee contribution is a very pertinent example (Limoges, Purvis, Turner, Wigan, & Wolf, 2000). The urban data committee abstract demonstrates the pertinence to the current study.

> "The exciting opportunities technology is providing researchers in the field of urban transportation data come with equal challenges. These challenges include data privacy and confidentiality, fluctuating priorities on data needs, and problems for data collection budgets given the imminent need for modernizing the transportation infrastructure. Adding to these issues is the need to train a new generation of analysts to effectively analyze varying quantity and quality of data. Certainly new technologies promise to reduce the cost while improving the accuracy of collected data, but this benefit is countered to some degree by the increasing reluctance of people to be monitored". (Limoges et al., 2000)

The collected statements of all the data committees were subsequently issued as the 'Unpublished' (sic) TRB report cited above in response to the demand.

While specific projects debarred by data lack were carefully avoided by most of the academic respondents, there was a high degree of interest in a number of themes common to all parties – access, knowledge and barriers to extant data. An entirely safe and no competitive area encountered by virtually all academic (and indeed other) participants.

The quality issues were also salient, although often in a displaced or disguised form expressed as a strong concern over a possible deadening effect from 'standardisation of surveys'. These specific responses need to be unpacked, and justify careful dissection.

11. CONCLUSIONS

The study described in this paper set out with the deceptively simple objective of establishing the nature of the data gaps affecting behavioural transport research in the UK.

The nature of these gaps have been found to be diverse and to extend far beyond the simple lack of salient data items *per se*. At least as significant an issue in addressing current data gaps is a broad set of institutional, legal and technical considerations surrounding the collection, storage, documentation, quality assurance and dissemination of behavioural data. These considerations apply both to existing data holdings and, critically, to any future attempts to institute a programme of model dataset creation.

We are strongly of the view that although individual researchers and practitioners will inevitably focus on the particularities of their specific interests and obligations, the behavioural transport research community as a whole must find a means of focusing more broadly on these structural issues. Indeed, increasingly, it does not have the luxury of ignoring them.

REFERENCES

ASTM. 2003. Standard Guide for Archiving and Retrieving ITS-Generated Data: American Society for Testing and Materials.

Axhausen, K. W., & Wigan, M. R. 2003. Public use of travel surveys: the metadata perspective. In P. Stopher, & P. Jones (Eds.), ***Travel survey quality and innovation***: 605-628. London: Elsevier.

Federal Geographic Data Committee. 1998. Content standard for digital geospatial metadata (revised June 1998). Washington D.C.: Federal Geographic Data Committee.

FTA. 1996. FTA National Transit GIS: Data standards, guidelines and recommended practices. Washington DC: Federal Transit Administration.

Hensher, D. A. 2000. Assessing data and modelling needs for urban transport: an Australian perspective. ***Journal of Transportation and Statistics*** (April): 31-50.

Hine, J., Wigan, M. R., & Grieco, M. 2003. Metagovernance, metadata and community information systems: developments in the transport sector.

Limoges, E., Purvis, C. L., Turner, S., Wigan, M. R., & Wolf, J. 2000. Future of urban data, ***Transportation data, statistics, and information technology in the new millennium: State of the art and future directions***, Vol. Unpublished Report 56: 15-22. Washington DC: Transportation Research Board.

Margiotta, R. 1998. ITS as a Data Resource - Preliminary requirements for a user service: http://www.fhwa.dot.gov/ohim/its/itspage.htm. Washington DC: Report for the Federal Highway Administration, Office of Highway Policy Information.

Ryssevik, J., & Musgrave, S. 1999. ***The Social Science Dream Machine: Resource discovery, analysis and delivery on the Web***. Paper presented at the IASSIST Conference, Toronto Canada.

Taylor, M. A. P., Young, W. F., Wigan, M. R., & Ogden, K. W. 1992a. Designing a large-scale travel demand survey: new challenges and new opportunities. ***Transportation Research A***, 26(3): 247-262.

Taylor, M. A. P., Young, W. F., Wigan, M. R., & Ogden, K. W. 1992b. Travel data: their collection and use. In E. S. Ampt, A. J. Richardson, & A. H. Meyberg (Eds.), ***Selected readings in transport survey methodology***: 351-371. Melbourne: Eucalyptus Press.

Wigan, M. R. 1985. The secondary use of transport survey data. In E. Ampt, A. J. Richardson, & W. Brog (Eds.), ***New survey techniques in transport***: 131-148. Netherlands: VNU Science Press.

Wigan, M. R. 1987. ***Australian personal travel characteristics***. Vermont, Victoria: Australian Road Research Board.

Wigan, M. R., Grieco, M., & Hine, J. 2002. Enabling and managing greater access to transport data through metadata. ***Transportation Research Record***(1804): 48-55.

Wigan, M. R., Hensher, D. A., & Raimond, T. 1995. ***Results of a Delphi study on transport research needs***. Paper presented at the World Conference on Transportation Research, Sydney, Australia.