



Working Paper

Conceptual data model for the integrated travel survey and spatial data

Author(s):

Chalasanani, V.S.; Axhausen, Kay W.

Publication Date:

2005-08

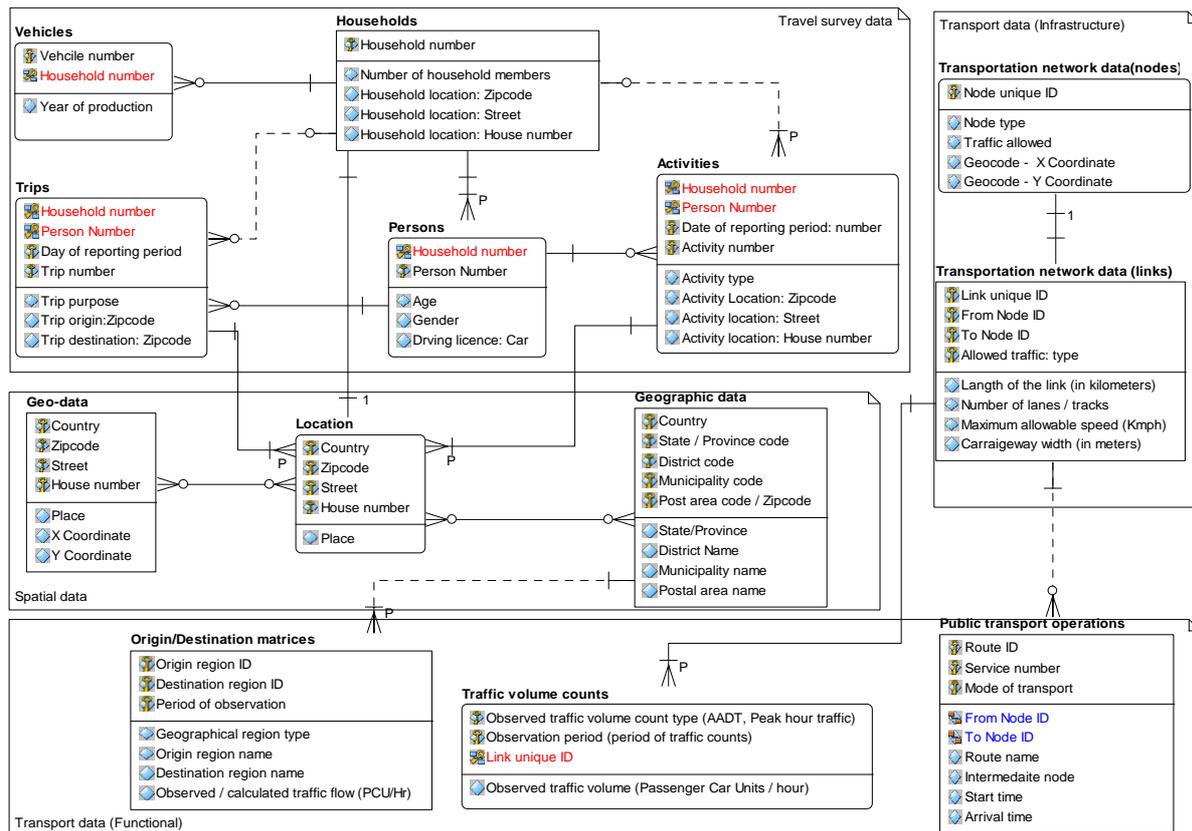
Permanent Link:

<https://doi.org/10.3929/ethz-a-006020477> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).



Submitted for the ASC conference “Maximising data value”

Conceptual data model for the integrated transport and spatial data

VS Chalasani
KW Axhausen

Conceptual data model for the integrated travel survey and spatial data

VS Chalasani
IVT, ETH Zürich
CH-8093
Switzerland

Telefon: +41 1 633 33 40
Telefax: +41 1 633 10 57
e-Mail-Adresse: chalasani@ivt.baug.ethz.ch

KW Axhausen
IVT, ETH Zürich
CH-8093
Switzerland

Telefon: +41 1 633 3943
Telefax: +41 1 633 10 57
e-Mail-Adresse: Axhausen@ivt.baug.ethz.ch

August 2005

Abstract

Everyone involved in transport and spatial planning is at some stage involved with data production or analysis. Each transport survey is conducted for a set of objectives. Data obtained from these transport surveys do not follow any specific pattern, and are thus difficult to understand. At the same time, a re-search organization conducts a wide variety of surveys ranging from simple road-side interviews to the complex travel diaries, which can be either longitudinal or cross-sectional, and differences in methodology, design, and protocols will often obscure basic similarities between them. Above all, it is almost impossible to collect complete information about the existing transportation system in a single survey. Most transport surveys collect partial information and depend on other sources for more. To understand the interactions between the datasets obtained from different surveys, a conceptual data model as a platform to integrate transport and spatial was developed. Both the transport data and spatial data were broadly classified. Transport data was classified as *travel survey data*, *transport data (infrastructure)*, and *transport data (functional)*; Spatial data were classified into *geographic data* and *geo-data*. Individual data models were developed for each classification. These data models help in streamlining the data from longitudinal surveys and standardizing the data from cross-sectional surveys. As a final step, the independent models are integrated into a single conceptual data model that represents integrated transport and spatial data. This final model facilitates easy understanding of the relationships between various data sources and allows the users to pass the information between them.

Keywords

Transport data, Geo-data, Geographic data, Transport survey, Data modelling, Entity-relationship model, Conceptual data model

1 Introduction

Improved support for the development of information systems integrating transport and geo-referenced information has been a long-term user requirement in transport planning and spatial analysis. The need to travel arises from the spatial separation of two activity locations. The continuous and mutual interaction of transport and spatial information is a central task of spatial analyses of travel patterns. Several studies have examined the spatial influences on travel patterns (Simma, 2000; Schlich, 2004), as well as the effects of various factors such as land-use (McNally and Kulkarni, 1996; Boarnet and Sarmiento, 1998), neighbourhood design (Crane and Crepeau, 1998), activity spaces (Schoenfelder, 2003), etc. This paper focuses on the issues that require integration of transport and spatial data, and proposes a solution based on entity-relationship data modelling.

Traditionally transportation professionals have collected information through various transport surveys. As technology has progressed, transport surveys have grown from simple road side interviews to complex multi-period and multi-method travel diaries. Although transport surveys are able to collect comprehensive, accurate and high quality information, they do suffer some limitations because of design and operational difficulties, as well as with respondent resistance. Transport survey data need to be enriched for the following reasons:

- Resource constraints, e.g. budget, time, etc mean that no single transport survey can collect complete information
- Outliers must be cross-checked with information from previous studies or external sources.
- Existing information, from the pre-survey process, must be integrated with the freshly ob-served data.

Considerable research has been conducted on transport data enrichments. Two important enrichments for the Microcensus 2000 (Swiss national travel survey: One day trip diary) that were carried out at IVT, ETH Zurich are - Geo-coding the households and travel end locations (Jermann, 2003) and a study on precision of geo-coded locations (Chalasanani et al., 2004). In the first enrichment, geo-data was integrated with that of observed travel

survey data to calculate the crow-fly distances. Transport network data and geographic data were integrated with the travel survey data in the second enrichment to calculate various network distances. Spatial data was also used to augment Microcensus 2000 data at ETH Zurich with stage imputations, aggregated modes of transport, accessibility indices, travel costs, and regional traffic type (Chalasani, 2005).

1.1 Background

Recognizing the growing importance of data re-usability, ETHTDA (ETHTDA, 2005), an exclusive travel data archive, was established in 2002 at IVT, ETH Zurich. Data from several surveys ranging from simple traffic counts to the travel diaries have been archived (Chalasani, 2004). Though spatial data has been used extensively in day-to-day analyses, and in enrichments of most of the surveys, no spatial datasets were archived. Furthermore, regular updating of spatial data and data added from continuous travel surveys have increased the difficulty of understanding and mining the data obtained from disparate sources. Based on the ETHTDA experience, and research at the institute, we have reached the following conclusions about the interaction between transport and spatial data:

- Transport surveys cannot independently collect all necessary information and must therefore be enriched
- A thorough understanding of existing transport and spatial information is a mandatory pre-requisite for any transport survey.
- High-end documentation, and dissemination of both transport and spatial data, maximises use and re-use of the data
- Linkages between transport and spatial data should be developed to support the integration of transport and spatial data.

In this study an attempt is made to develop a platform for integrating transport and spatial data. Linkages within and between transport and spatial data are explained by a conceptual data model using the entity-relationship approach. The main objective of this study is to explore the linkages between transport and spatial data through a set of interaction as relationships.

This paper is organized in the following way: Chapter 2 covers data modelling issues, Transport data and Spatial data are described in Chapter 3 and 4 respectively. A conceptual data model is proposed in Chapter 5 and conclusions are stated in Chapter 6.

2 Entity-relationship modelling

The entity-relationship approach to conceptual data modeling was initially developed by Peter Chen (Chen, 1978). ER/studio 6.6.1(ER/studio, 2005) is used as a tool to draw all the entity relationship diagrams in this study. Entities are the real objects and starting point of a data model. Interactions between the entities are explained through relationships. These relationships are configured for type, existence and cardinality.

The three distinct relationship types implemented in the model are:

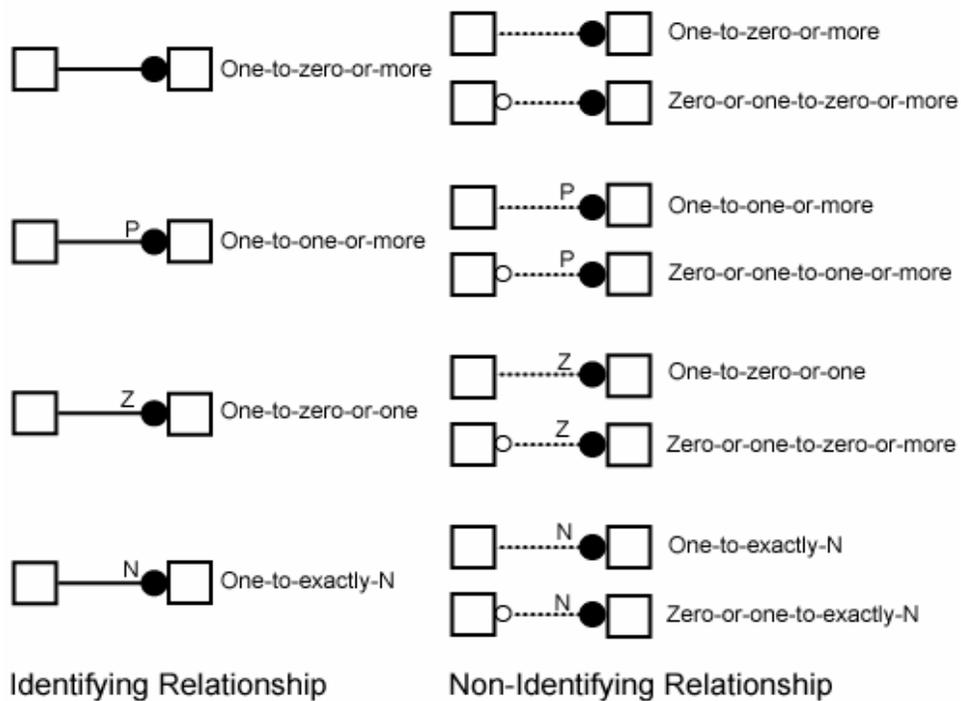
- Identifying relationships, which propagate the parent entity's primary key to the child's primary key. In the model, identifying relationships are drawn as solid lines with a solid circle terminating the child entity.
- Non-identifying relationships, which propagate the parent entity's primary key to the non-key attributes of the child. In the model, non-identifying relationships are drawn as dashed lines with a solid circle terminating the child entity. If the non-identifying relationship is optional, then a hollow diamond terminates the parent entity.
- Non-specific relationships, which denote many-to-many relationships. Because many-to-many relationships cannot be resolved, non-specific relationships do not propagate any foreign keys. In the model, non-specific relationships are drawn as solid lines with solid circles terminating both entities.

Existence describes the relationship between a pair of entities from the perspective of the child entity. Fundamentally, it asks the question, "Is a foreign key value always required in the child entity?" The possible answers are: Optional – not always, and Mandatory – always required. Existence can be enforced on the three relationship types as follows:

- Identifying Relationships, which are always mandatory.
- Non-Identifying Relationships, which can be mandatory or optional. In the model notation, optional non-identifying relationships are represented by a hollow diamond at the parent end of the relationship line.
- Non-Specific Relationships, in which existence cannot be enforced because we cannot resolve many-to-many relationships.

Cardinality describes the quantitative dimension in the relationship between a pair of entities as viewed from the perspective of the parent entity. It is read as the ratio of related parent and child entity instances. The cardinality ratio for the parent entity depends on whether the relationship is mandatory (one or more) or optional (zero or more). The model used four different cardinality ratios for the child entity: zero-or-more, one-or-more (P), zero-or-one (Z), and exactly N (N). The cardinality notation for different relationships types is illustrated in the Figure 1.

Figure 1 Notations of four cardinal ratios by relationship type



Source: ER/studio manual

Relationship existence also has implications for relationship cardinality. If a relationship is mandatory, then the cardinality must be in the form of one-to-something. If it is optional, then the cardinality would be in the form of zero or one-to-something.

Notations described in this section have been used in all the entity-relationship diagrams included in this report, but not for the spatial data hierarchy.

3 Transport data

Transport data is a generic term that covers different data types such as transportation network data, travel survey data, vehicle counts data, etc., from which comprehensive information about both the transportation system and its users can be extracted. This study classifies transport data as follows:

- Transportation system data (infrastructure)
- Transportation system data (functional)
- Transport survey data (behavioural/user reported)

The above classification is broad and general in nature, and is exclusive to this study. A detailed description of each category is covered in subsequent sections.

The following transport datasets were used in developing entity-relationship diagrams for transport data:

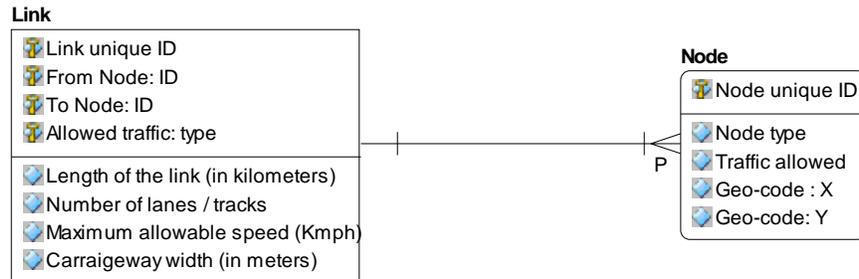
- Microcensus 2000 – One day trip diary
- 12 weeks of leisure travel – Activity based 12 weeks diary.
- IVT national road and rail network model
- Travel module of “Household income and consumption survey” 1998
- DATELINE – long distance travel survey

3.1 Transport data (infrastructure)

Transport infrastructure data contains information about the prevailing infrastructure, i.e. the static characteristics of the transportation network, represented as a set of links and nodes, important junctions, public transport stops, etc. The transport network database consists of two data files, namely links and nodes. A simple ER diagram that represents the transport network data with two entities is shown in Figure 2. This ER diagram

completely fits all the three transport network models (road network model, rail network model, and cantonal network model) available at ETH Zurich.

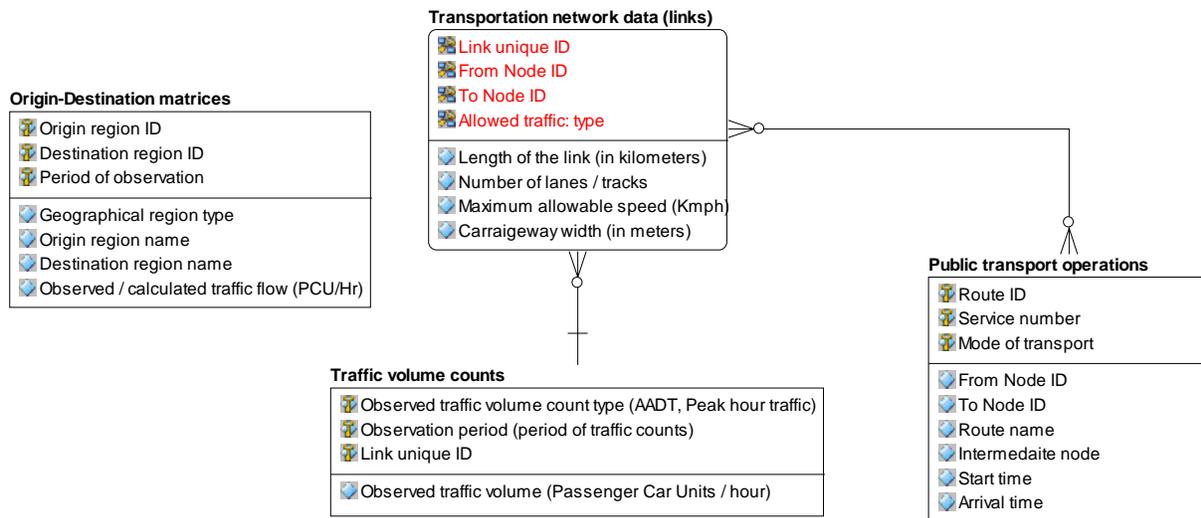
Figure 2 ER diagram for transportation network data



3.2 Transport data (functional)

Transport functional data carries information about dynamic characteristics of the prevailing transportation system. Several methods such as traffic volume counts, cordon counts, moving observer's method, etc. are used to collect the data. The functional characteristics are of two types: network operational characteristics, such as traffic movements at intersections, direction of traffic, etc., and public transport operational parameters, such as routes, schedules, frequencies, etc. A simple ER diagram for functional based transport data is shown in Figure 3. The entity "Origin-destination matrices" is not related to other entities because it is indirectly calculated from either transport survey data or traffic volume counts.

Figure 3 ER diagram for transport data (functional)

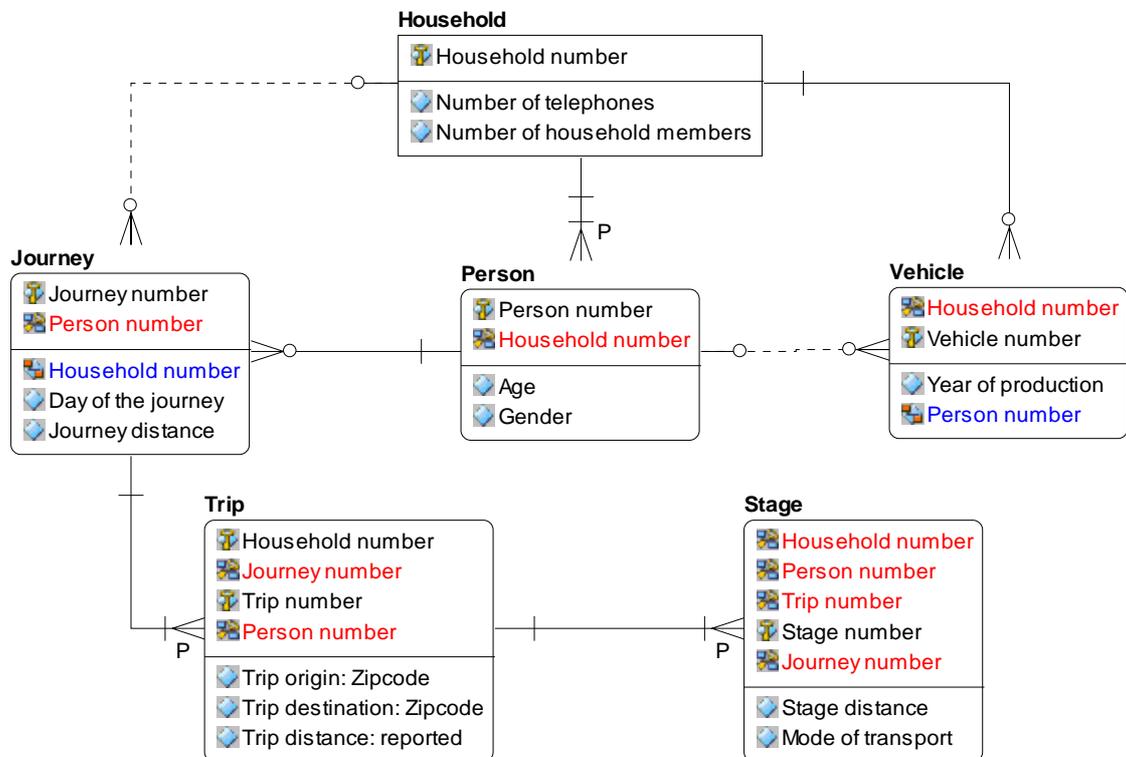


3.3 Travel survey data

Traditionally travel survey data is trip based until the activity based travel demand modelling has picked up its momentum in early 1990s. This report covers both trip-based and activity-based travel survey data. After carefully editing and error checks, a travel survey data is output to set of data files. Each data file contains information on a distinct type of object, such as households, persons, vehicles, activities, journeys, trips, stages, etc.

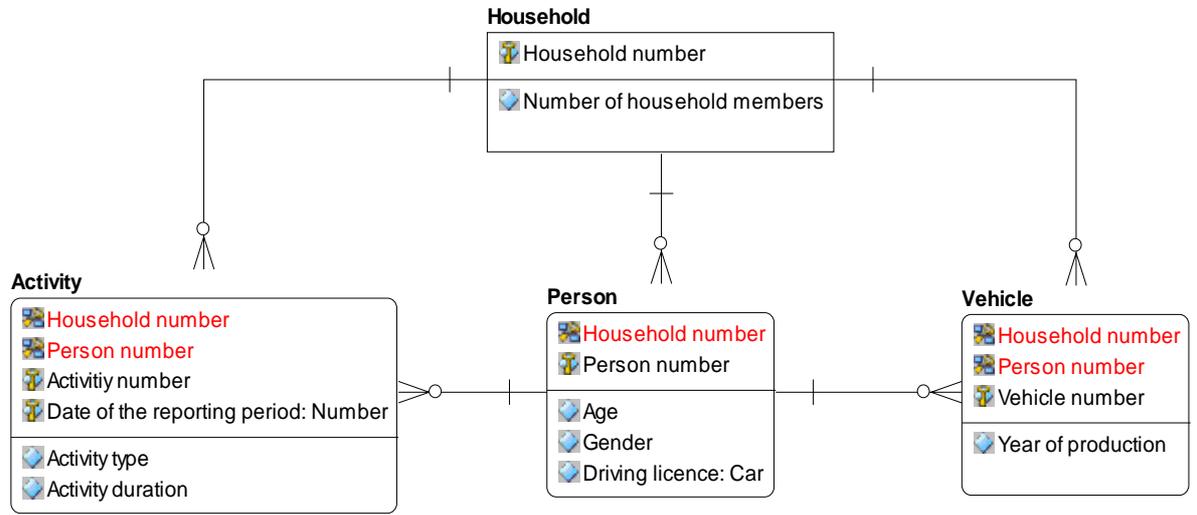
An entity-relationship diagram for a typical trip-based travel survey is shown in Figure 4. Each entity in this model is a data file. Though most of the relationships are mandatory (with solid lines), they depend on various factors such as survey context, unit of analysis, survey structure, etc. For instance, journeys can be observed at person level as well as household level, as can vehicles. As noted earlier, the structure and relationship of the ER model is survey specific. Definitions of travel terms such as ‘journey’, ‘trip’, ‘activity’ and ‘stage’, come from Axhausen (2000). This ER model represents the most used travel survey data in Switzerland, i.e. Microcensus – National household travel survey.

Figure 4 ER diagram for trip based travel survey data



Activity-based travel modelling has become more popular since the early 1990's, and is now widely used by planners all over the world. Activity-based travel survey data is much simpler in structure than trip-based travel survey data. The ER diagram shown in Figure 5 represents activity-based travel survey data from "12 Weeks of Leisure Travel", an activity survey conducted in Switzerland.

Figure 5 Entity-Relationship diagram for the activity based travel survey data



4 Spatial data

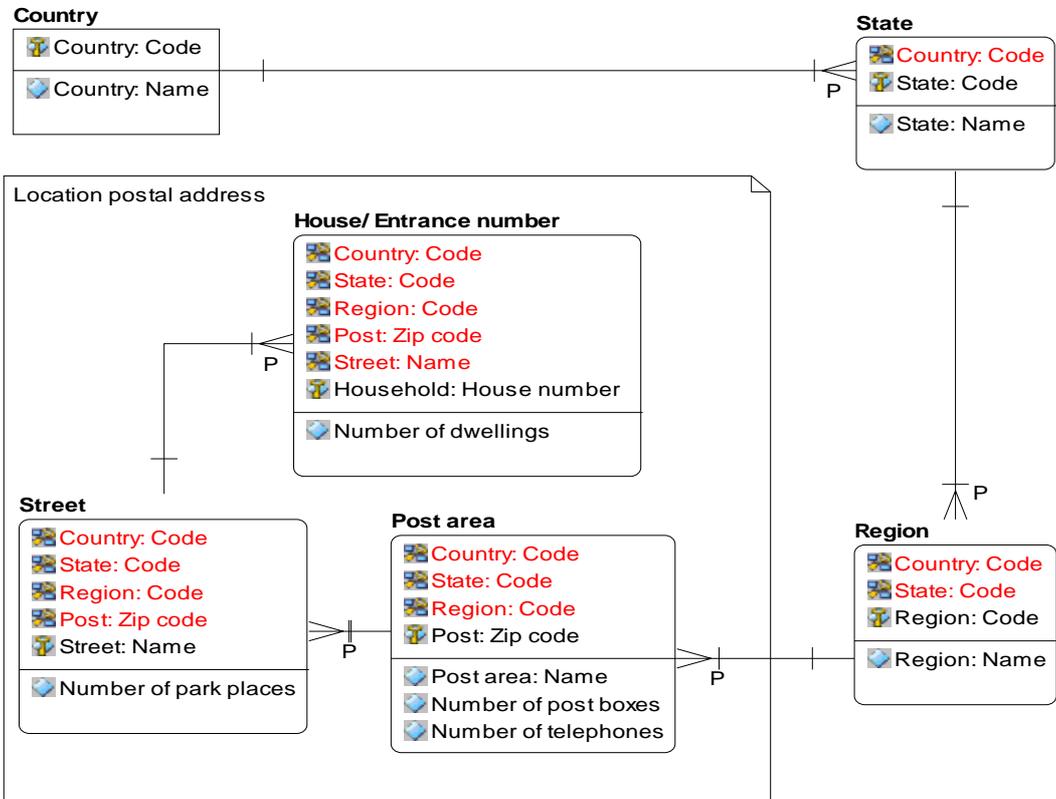
Spatial data in present context is limited to geo-referenced information i.e. geographic data and geo-data. The following spatial datasets were used in the development of entity-relationship diagrams for spatial data:

- Swiss geo-data: Geo-codes of Swiss building entrances
- Swiss geographic data (1850 – 2000): Geographic information of Switzerland

4.1 Geographic data

Geographic data is much more than electronic pictures of maps. Geographic data describes how a particular domain (continent, country, region, etc) is geographically divided according to different themes like political, administrative, transport, language, etc. Depending on its size and structure, each domain will have its own geographical hierarchy for different themes. Geographic data is less dynamic than data pertaining travel patterns. The geographical hierarchy of a country's geographic data divided for administrative purposes is shown in Figure 6.

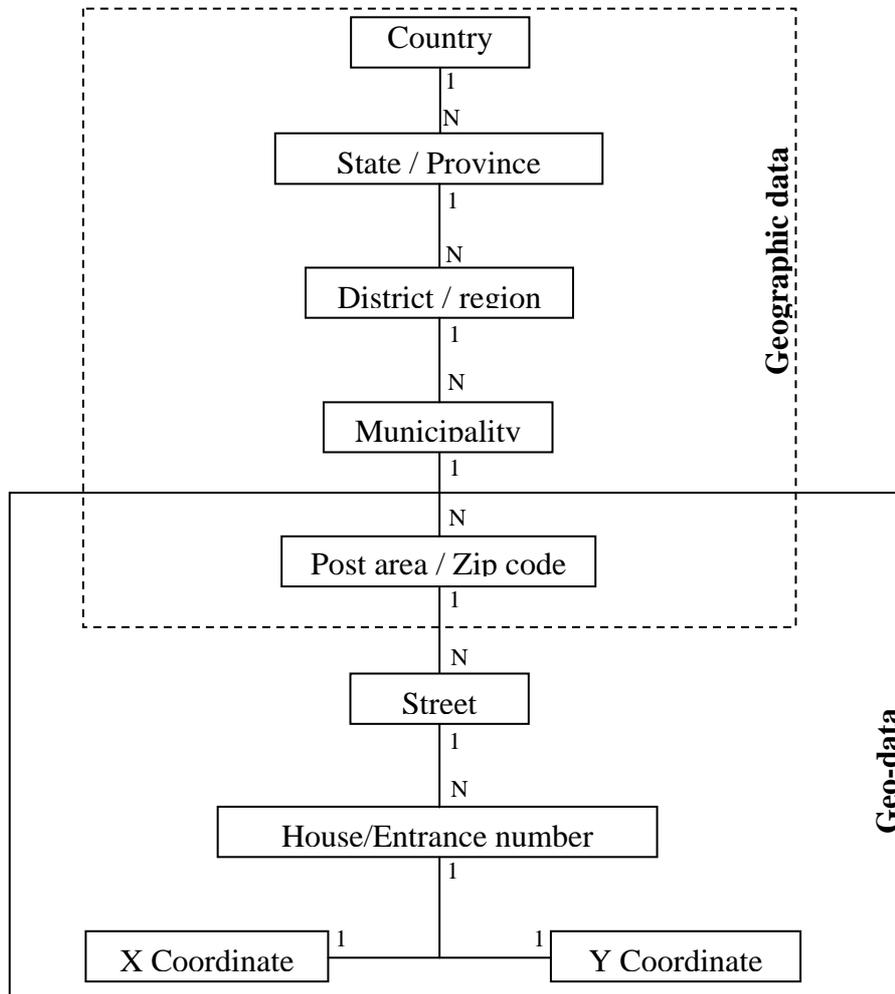
Figure 6 ER diagram for the geographic data



4.2 Geo-data

Geo-data contains geo-information that identifies distinct physical objects such as households, building entrances, post offices, railway stations, road junctions, etc., through a pair of geographical coordinates or geo-codes. Geo-data is a division of spatial data that cannot be observed or collected in most of the transport surveys, except the GPS surveys. Transport data must be enriched with geo-data to perform spatial analyses of the travel. Eventually geo-coding (process of assigning geo-codes to different locations) has become mandatory enrichment for most of the transport data. A geo-database is a database with extensions for storing, querying, and manipulating geo-data. The information hierarchy within a geo-database is similar to that of a geographic database. Figure 7 shows the spatial database hierarchy that combines geographic data and geo-data. The spatial data hierarchy differs from that of previous entity-relationship diagrams because it represents the internal structure of a single data file, while the former represents the relationships between different data files. Due to this, entity-relationship notation was not used to describe the spatial data hierarchy.

Figure 7 Spatial data hierarchy



5 Conceptual data model (CDM) for the integrated transport and spatial data

The central task of this study is to develop a conceptual data model to facilitate understanding of interactions between transport and spatial data. A model using the entity-relationship is shown in Figure 8. Each entity is a separate data file and the most important attributes are represented in the model. All the entities and relationships follow the notations explained earlier. To maintain consistency with the data classification used above, the model contains four sections: Travel survey data, Spatial data, Transport data (functional), and Transport data (infrastructure), with a note tab used to describe the grouped entities. Descriptions of each section can be found in previous sections of this report. A logical entity “location” has been added to simplify the interactions in the model. Relationships between Households and Trips, and Households and Activities entities are optional because trips and activity data are infrequently collected for households, as compared to persons. The relationship between Geographic data and Origin-Destination matrices becomes optional when the Origin-Destination matrix’s geographic region is at the lowest possible level. Relationship between Transport network data (links) and Public transport operations is optional due to the fact that all links in the network need not to be accessible to all traffic types (e.g.: Public, private). Key interactions between different entities along with the key variables are listed in Table 1.

Figure 8 Entity relationship diagram for the integrated transport and spatial data

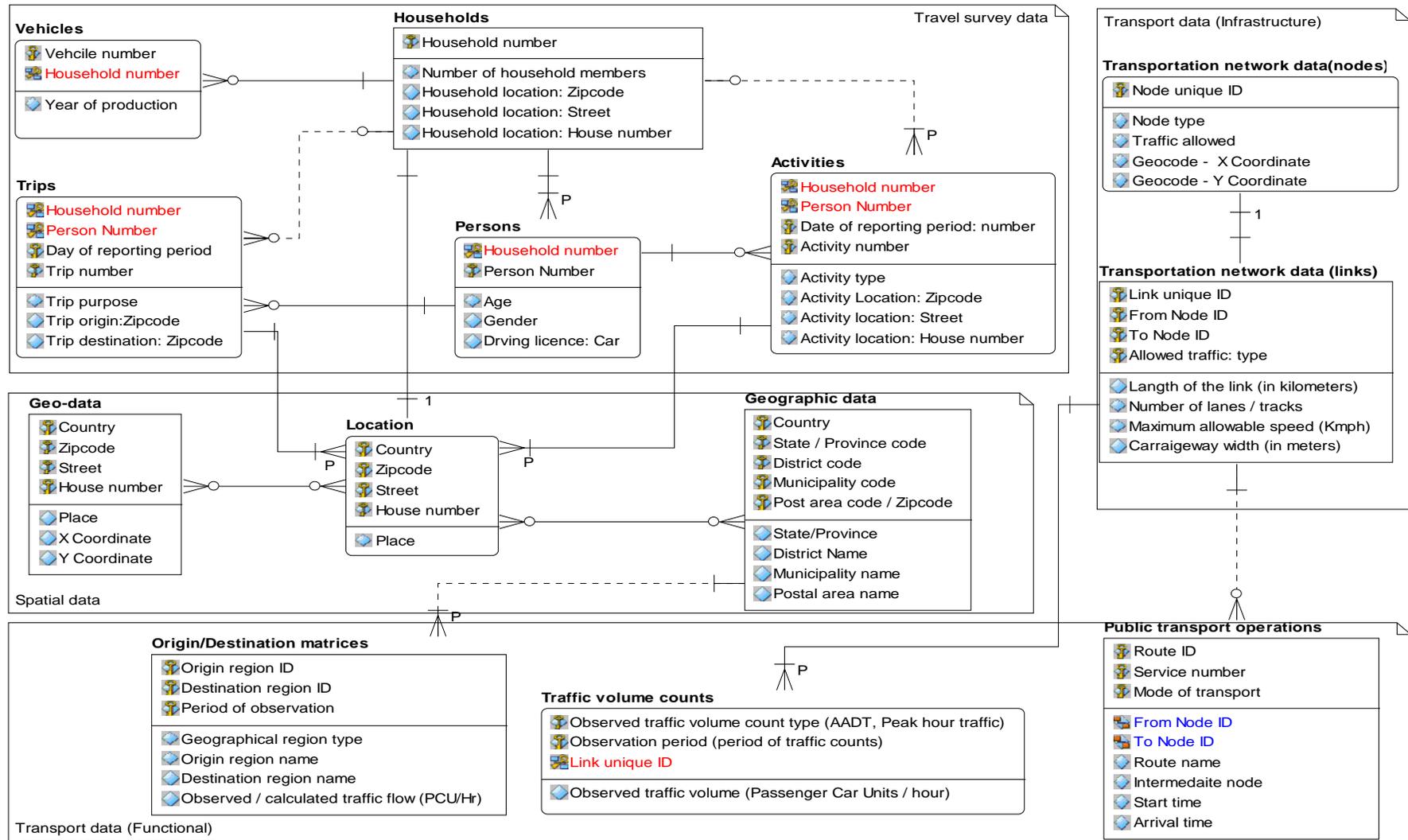


Table 1 Relationships of the conceptual data model for the integrated transport and spatial data

Parent Entity	Child Entity	Key variables	relationship
Household	Person	Household number	1 : 1 or more
Household	Vehicle	Household number	1 : zero or more
Household	Activity	Household number	1 : 1 or more
Person	Trip	Household number, Person number	1 : zero or more
Person	Activity	Household number, Activity number	1 : zero or more
Household	Geo-data	Location*	1 : 1
Household	Geographic data	Location*	1 : 1 or more
Activity	Geo-data	Location*	1 : 1
Activity	Geographic data	Location*	1 : 1 or more
Geographic data	O-D matrices	Geographic area**	1 : 1 or more
Links	Node	Node unique ID	1 : 2
Public transport operations	Links	From Node ID, To Node ID	1 : 1 or more

*: Logical entity **: Depends on the O-D matrices coarseness

6 Conclusions

An understanding of the existing information (both transport and spatial) is a basic prerequisite for any transport survey, as it not only helps in designing the targeted information i.e. the redundant information that should be targeted in the survey, but also improves the data quality. A thorough knowledge of existing transport and spatial data leads to better survey instrument design and improved post-processing of survey data. Basic enhancements to the transport survey data are highly recommended to reduce redundancy in the reported/observed transport data. When integrated with transport data, spatial data broadens the range of areas of applications, i.e. a wide range of additional problems on spatial analyses of travel patterns can be analyzed. This study employed the following steps in developing a conceptual data model for integrated transport and spatial data:

- Classify the transport and spatial data for the available set of data sources.
- Develop independent data models for each sub-section of the classification.
- Identify all the possible linkages within and between transport and spatial data.
- Build a data model using the identified linkages and individual data models.

A set of datasets were used in the conceptual data model development. The conceptual data model developed in this study will facilitate:

- Integrating geographic and geo-data with both trip-based and activity-based travel survey data
- Understanding the linkages within transport data and between transport and geo-referenced data.

This model can be extended to include geographic data with other themes, e.g. such as transport regions, (transport regions), language, etc., as well as census and social network data.

7 Literature

- Axhausen, K.W. (2000) Definition of movement and activity for transport modelling, In D.Hensher and K.Button (eds.) *Handbooks in Transport: Modelling*, Transportation paper, Elsevier, Oxford.
- Chalasanani, V. S. (2004) Travel data archiving: The art of presenting and preserving travel data, conference paper, 4th Swiss Transport Research Conference, Monte Verita, Ascona, March 2004.
- Chalasanani, V. S. (2005) Enriching the household travel survey data: The case of Microcensus 2000, conference paper, 5th Swiss Transport Research Conference, Monte Verita, Ascona, March 2005.
- Chalasanani, V.S., Ø. Engebretsen, J.M. Denstadli and K.W. Axhausen (2004) Precision of geocoded locations and network distance estimates, *Arbeitsbericht Verkehrs- und Raumplanung*, **256**, IVT, ETH Zürich, Zürich.
- Chen, P.P. (1976) The entity-relationship model – toward a unified view of data, *ACM Transactions on Database Systems*, **1**(1), 9-36.
- Crane, R. and R. Crepeau (1998) Does neighbourhood design influence travel?: behavioural analysis of travel diary and GIS data, *Working paper*, **374**, The university of California Transportation Center, UC Berkeley.
- ER/Studio 6.6.1 (2005) <http://www.embarcadero.com/products/erstudio/>, Embarcadero technologies, April 2005.
- ETHTDA (2005) <http://129.132.96.89/index.html>, April 2005.
- Jermann, J. (2003) Geokodierung Mikrozensus 2000, *Arbeitsberichte Verkehrs- und Raumplanung*, **177**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.
- McNallay, M.G. and A. Kulkarni (1996) An assessment of the influence of the land-use transportation system on travel behaviour, *Working paper*, **96-4**, Institute of transportation studies, UC Irvine, Irvine.
- Schlich, R., S. Schönfelder, S. Hanson and K.W. Axhausen (2004) Structures of leisure travel: Temporal and spatial variability, *Transport Reviews*, **24** (2) 219-238
- Schönfelder, S. and K.W. Axhausen (2003) On the variability of human activity spaces, in M. Koll-Schretzenmayr, M. Keiner und G. Nussbaumer (eds.) *The Real and Virtual Worlds of Spatial Planning*, 237-262, Springer, Heidelberg.

Simma, A. and K. W. Axhausen (2000) Mobility as a function of social and spatial factors: The case of the Upper Austria Region, Vortrag, *Land Use and Travel Behaviour*, Amsterdam, June 2000.