THE INTERNATIONAL TRAFFIC DATABASE PROJECT

Marc Miska*
Research Fellow
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
Tel: +81-3-5452-6419, Fax: +81-3-5452-6420, mmiska@iis.u-tokyo.ac.jp

Alexandre Torday
Project Director
TSS-Transport Simulation Systems
Pg. de Gràcia 12, 08007 Barcelona, Spain
Tel: +34 93 317 16 93, Fax: +34 93 317 17 16, torday@aimsun.com

Hiroshi Warita
Assistant Professor
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
Tel: +81-3-5452-6419, Fax: +81-3-5452-6420, warita@iis.u-tokyo.ac.jp

Masao Kuwahara
Professor
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
Tel: +81-3-5452-6419, Fax: +81-3-5452-6420, kuwahara@iis.u-tokyo.ac.jp

*) corresponding author

Abstract:
For transportation studies and applications, practitioners as like as researchers spend an enormous amount of time to gather suitable data and to convert this data into a format that can be processed by their working tools or developed models. This work slows down their job and leads to tons of unused data in archives that is inaccessible for other people working in the same field. This paper presents a web based approach for an International Traffic Database to overcome this drawback and to open the possibilities for world-wide collaboration of transportation research and study.

Traffic measurements, Data warehousing, Data processing

RESEARCH MOTIVATION

Gathering real life data, for whatever type of use, is a time consuming job. A lot of data is measured and stored in several places and different formats around the world. While a lot of it is not used, other institutions gather similar data on different locations or, worse, on the same ones. In this way a lot of money and time is spend unnecessary [1]. Thus, the aim of the International Traffic Database (ITDb) project is to provide traffic data to various groups (researchers, practitioners, public entities) in a format according to their particular needs, ranging from raw measurement data to statistical analysis.

The idea of providing traffic data or data storage standards is not new and is done in several parts of the world. Among others and just to name a few, these projects and institutions provide such information:

- The National Roads Authority, Ireland [2]
- Department of Transportation (DOT)**,** Federal Highway Administration (FHWA), the Federal Motor Carrier Safety Administration (FMCSA), the National Highway Traffic Safety Administration (NHTSA) and NHTSA's National Center for Statistical Analysis (NCSA)**,** USA. [3]
- Traffic Management Data Dictionary (TMDD), Institute of Transportation Engineers, USA [4]
- Next Generation Simulation (NGSIM), USA [5]
- Region Laboratory Delft, Netherlands [6]
- ASTRA, Switzerland [7]
- Clearing house for transport data and transport models, DLR, Germany [8]

However, the traffic data available to the public is provided in changing formats, different aggregations, and with a varying density of Meta information (when existing), which allows just a limited reuse of the data or a time consuming process of formatting or understanding the data. To overcome this drawback, the United Kingdom started a project called "ITS Metadata Registry" to allow users of such data to match variables from different sources and to understand different formats more easily [9]. Instead of just trying to make the data understandable for other users, ITDb incorporates this translation or data name matching and applies it to actual data to provide a comprehensive data pool to the user.

These different formats and the many places to check for useful data are slowing down the process of data gathering and therefore the studies using them (research, model calibration, etc.). To tackle this problem, the ITDb promotes a flexible traffic data provision format based on user needs and standard habits, instead of another in-house solution for sharing data. Further, the collection of world-wide data sources and making them accessible via one single platform accelerates projects and decision making in any data sensitive field.

Reflecting all this, the ITDb is founded on three pillars: Data, platform and quality management (see Figure 1). Handling them independently in the design and development allows to focusing on technical aspects on the one hand and the fulfillment of user needs on the other hand. The ITDb is not based on existing data and tailored for its provision, but based on the final operation. The sustainable development kept in mind different user groups, transfer efficiency and a high quality standard. This is the major difference to other existing

implementations of data platforms, where the design is purely based on a single set of data types and has to be changed or adapted with every revision of the content.

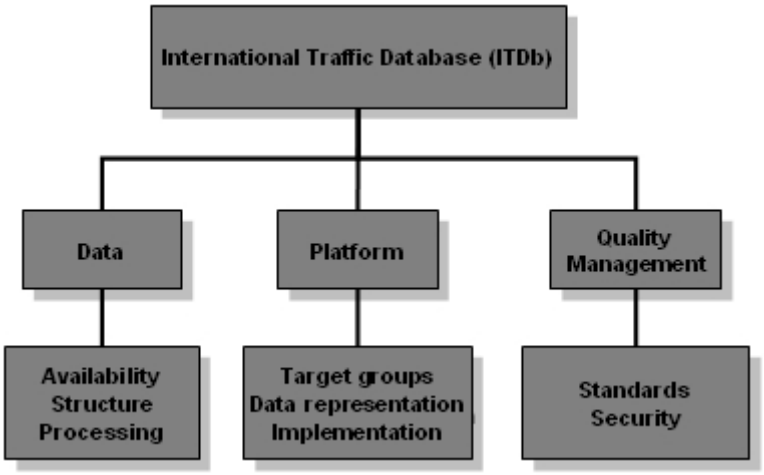

Figure 1: Three pillars of the International Traffic Database

CONTRIBUTIONS

The ITDb project is contributing to the field of traffic engineering by providing a standardized Metadata with an underlying traffic data structure and providing the whole range from raw to processed data in various levels for the usage in research, model calibration and validation, performance analysis and so on. As using traffic data has usually as final goal analyzing or assessing network performances, a wide variety of data type has to be available in order to get the full picture of the traffic behavior. For this reason, not only conventional traffic data is stored, but also accidents, traffic events, traveler information messages, parking and environmental data. This data can be collected and be downloaded in a whole package which is transformed to a requested format (based on the availability of data) or common standard used in traffic engineering. This means, that every user can format the requested data according to personal preferences. Additionally, it provides rules for sustainable data storage as guideline for quality management in the provision of traffic data for future projects.

ITDb stores traffic data, environment data and incident data from various sources in a common format. Stored Meta information which can be linked together regionally can be browsed and be used to make a collection of data of interest for the user of the database. This allows a high flexibility, since the datasets are separated from their description. Instead of downloading project specific files or browsing online databases of traffic data in various formats, ITDb allows a regional search for all available data and data type specific searches world-wide. This data will be presented to the user in various selectable formats to meet his personal needs. The ITDb database together with the storage facility for non text based data, such as maps, videos, and photograph series creates a powerful tool for gathering a comprehensive data collection of study areas.

Concerning the data storage, it is a necessity to provide standardized Metadata to enable several parties to use the available data and to ensure that this data can be shared among them.

Metadata have two main purposes, including support of improved data quality; and ensuring the longevity of data by documenting who, when, where, why, and how the data was collected. The major contribution of this work is the development of a standardized Metadata set for traffic measurement data and flexible formats for providing the measurement data, which allows the translation to commonly used existing standards [10] such as:

- The Traffic Management Data Dictionary (TMDD),
- The P1512 Incident Management Data Dictionary (P1512-IEEE),
- Traffic Model Markup Language (TMML),
- Geographic Markup Language (GML),
- Universal Traffic Data Format (UTDF),
- Digital Geospatial Metadata (FGDC-STD-001-1998),
- European traffic information exchange standard (DATEX2), and others

to ensure compatibility and easy usage.

Since traffic data is valuable information and storage as well as provision includes costs, the ITDb structure is designed to include data selling and brokerage. This means, that data provider can determine if the download of their data is free or if they would like to charge a usage fee. The costs for this data are shown on the website and are marked differently. With brokerage ITDb wants to make data exchange between parties more attractive. While companies might be more interested in the selling, research institutes might want to use the chance to exchange data for data. It allows users to offer their data in exchange for similar datasets from different location or from other sources. Additionally, providers can offer data with a price to be negotiated bilateral or data requests can be posted on the platform.

Finally, quality management has to be guaranteed as using traffic measurement data for research, calibration or validation is very sensitive to the data source reliability. Algorithms and tools can only work properly with correct inputs and quality management is therefore vital to avoid “garbage-in / garbage-out” scenarios.

To gain the user’s trust in the ITDb, all datasets available on this platform are strictly checked according to the quality management rules, which require a minimum set of Metadata information that allows the user of the data to reconstruct the exact locations of the measurements, the conditions under which the measurements took place, and the equipment used for these measurements. Next to this technical information, the International Traffic Database requires administrative Metadata including a contact institution or person for further inquiries about the data.

While a strict handling may raise the need for post processing existing data, it is supposed to be a guideline for future traffic measurement experiments and installations, to store the data in a format that allows further usage for research projects, or simulation calibration tasks.

SYSTEM DESIGN

The design of the database consists of a Meta database per country which stores the Meta data and can be browsed for network descriptions, projects, measurements, environment and incident data in certain locations and time spans. The structure of this Meta search database can be found in Figure 2.

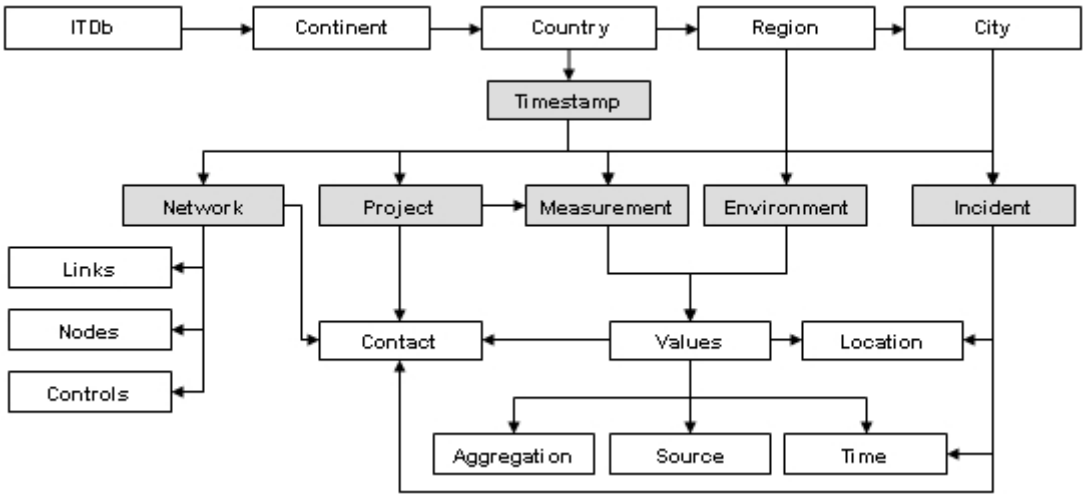

Figure 2: Structure of the Meta search database of ITDb

These Meta data is linked to the actual datasets in the data storage of ITDb. This design allows an efficient search for the user and fast access to the wanted information, because the download is not limited to certain projects, but can be accumulated from various uploaded elements. Further it allows to link data from different data sources together in the case that the user is looking for regional data provided by various institutions.

The dataflow in the ITDb system has two parts. A public front end where registered users can browse through the Meta database to search for useful datasets, and a second, hidden part, in which a scheduler will trigger the collection of data from the ITDb data storage and provides it to the user as a downloadable file via FTP or HTTP.

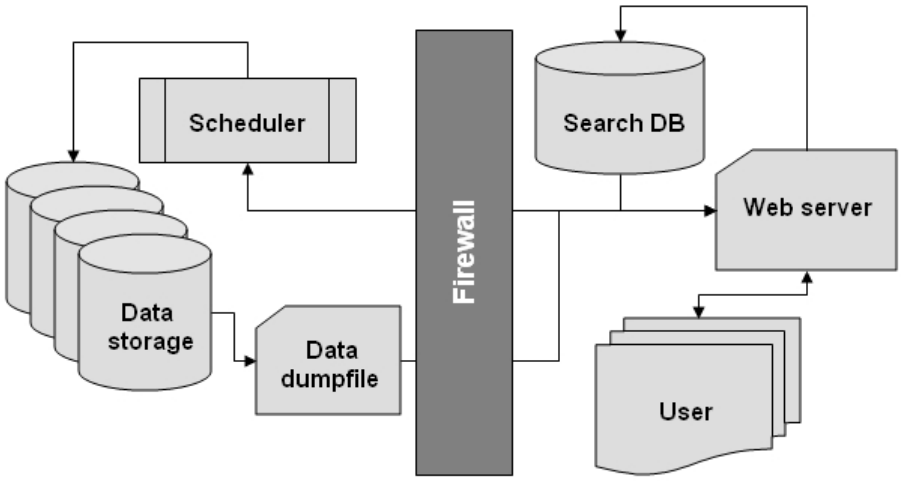

Figure 3: Dataflow between the ITDb data server and the web front end

Data protection and security is a major problem for their owners who do not want to open their data base to any usage. Thus, depending on the aggregation level, the data can be disclosed to any user, or to a subset of registered users. In order to give total flexibility on the

control of the data provision process, various option of shared responsibility between data owner and the ITDb are available as shown in Figure 4.

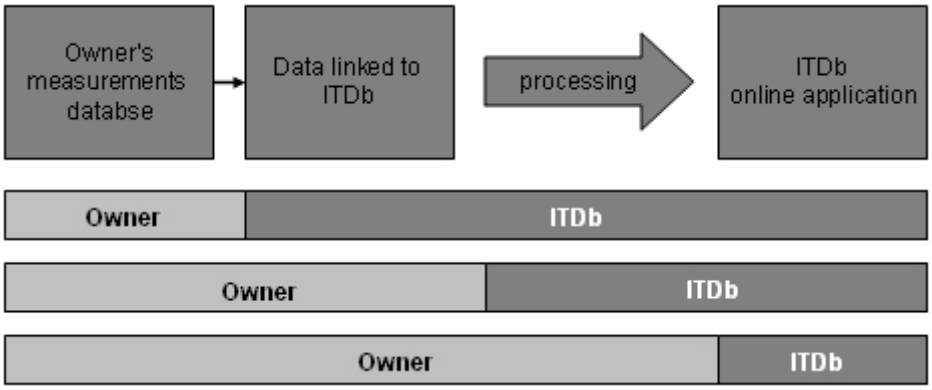

Figure 4: Possibilities of the responsibility handling for traffic data on ITDb

This concept allows the data owner to keep full control about their data, to allow ITDb to process the data, or to let the handling of the data being fully controlled by ITDb.

APPLICATION

The prototype application is a dynamic website that allows browsing all data content via keyword based searching and visually by browsing through the network and selecting the data sources of interest. The latter one is realized using the map interface provided by Google™ (see Figure 5) and allows an intuitive "shopping" for traffic data based on maps or satellite images.

Figure 5: Web front end of the International Traffic Database

The found and selected data by the user is stored in a DataCart, which acts as a shopping basket. The user can add or delete items from it and when finishing check out and the data get prepared for download. After the system collected all the data and created a downloadable package with the desired content, the user will receive an e-mail with a download link. This link will be available for 48 hours, while the data will be available until fully downloaded.

To upload data to the ITDb the user has to fill out the necessary Meta information concerning the data and choose between uploading via FTP or giving a download location for a remote upload. The data will be available on the website after the Meta information has been checked for consistency. This is necessary to keep a high standard of data quality on the website. If the information is incomplete ITDb will request further information. The prototype application can be reached via http://www.trafficdata.info.

RESULTS AND FURTHER WORKS

The database currently includes data samples from different parts of the world to give a broad coverage. Using various data source types and worldwide locations, a standardized format for storing traffic data has been defined which allows providing them based on the needs of several user groups. While researchers prefer raw, disaggregated datasets for specific analysis, practitioners are interested in processed data (e.g. speed-density relations, lane occupancy, queue discharge behavior …) to for example calibrate and validate simulation models. Public entities on the other hand might be interested in the quality of service in networks. The developed standardized format is based on existing approaches of standardization and contains next to the traffic data values a broad selection of Meta information to assure the usability of the provided data.

Registered users are able to browse through the data in different process levels. Uploading new data is restricted by the quality management guidelines provided to ensure the users of the database high quality and comprehensive data sets.

Future work includes the collection of more datasets and to use other access methods for downloading data, like the REST (representational state transfer) protocol. This would allow a direct use of traffic information in simulation models or other applications without the manual transfer of the data and would increase the convenience of using the ITDb system for research and practice.

REFERENCES

(1) Miska, MP, Lint, Hans van, (2006), A web based data source platform for modeling and simulation, Proceedings of the EWGT2006 Joint Conferences, *Advances in Traffic and Transportation Systems Analysis"*, Technical University of Bari, Faculty of Engineering, Bari Italy

(2) Online resource as of June 2007: *http://www.nra.ie/Transportation/TrafficDataCollection/*

(3) Online resource as of June 2007:
http://www.nhtsa.dot.gov/people/Crash/crashstatistics/index.htm

(4) Online resource as of June 2007: *http://www.ite.org/tmdd/*

(5) Online resource as of June 2007: *http://ngsim.camsys.com/*

(6) Online resource as of June 2007: *http://www.regiolab-delft.nl/*

(7) Online resource as of June 2007:: *http://www.dlr.de/cs*

(8) Online resource as of June 2007: *http://www.verkehrsdaten.ch*

(9) Online resource as of June 2007: *http://www.itsregistry.org.uk/*

(10) Kovvali, V, Margiotta, R, Hranac, R, Alexiadis, V, 2004, Next Generation Simulation (NGSIM) Data Format Plan, Final Report, CAMBRIDGE SYSTEMATIC, INC.