ONLINE PLATFORM FOR SUSTAINABLE TRAFFIC DATA STORAGE

Marc Miska, Project Lecturer, University of Tokyo, Institute of Industrial Science, 4-6-1 Komaba, Meguro-ku, 153-8505 Tokyo, Japan

Masao Kuwahara, Professor, University of Tokyo, Institute of Industrial Science, 4-6-1 Komaba, Meguro-ku, 153-8505 Tokyo, Japan

Shinji Tanaka, Lecturer, University of Tokyo, Institute of Industrial Science, 4-6-1 Komaba, Meguro-ku, 153-8505 Tokyo, Japan

ABSTRACT

Traffic engineers are involved in transport modelling, traffic simulation, operation optimisation and the development of methods to control and analyse traffic itself. New developments for individual traffic, public transport as well as pedestrian movements are the hope to ensure the mobility and accessibility of urban areas, to secure mobility in the less profitable countryside, to increase safety, and to limit the effects on the environment caused by transportation. Around the globe, governments declare goals in each of the mentioned fields, mostly under the umbrella of intelligent transport systems (ITS), to develop a sustainable transportation for everyone (Barcelo et. al., 2010).

To trigger the process of traffic data standardisation and to increase the availability of traffic data needed, we have developed an online platform for sustainable traffic data storage. This portal allows data providers and researchers to store there data geo-referenced, and with a minimum set of Meta information to ensure that the data will remain useful. The data, categorised per country and location, can then easily be made accessible to project partners, colleagues, or even the whole community with just a few mouse clicks. The usage of the portal is free of charge.

However, the idea behind this portal is not simply the tool, but to gather information about traffic data storage schemes, user's needs and provider's concerns. Only bringing all of them together will lead to a wide-spread acceptance and standardisation. To reach this, all public accessible data is analyzed and fed to an universal translator, who will store the data in a database in single elements and can retrieve it from there in any other known format learned before (as long as the information is available).

Keywords: traffic data, on-line, portal, standard

1. INTRODUCTION

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. 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 (Miska et. al 2007).
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.
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 fulfilment 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

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 behaviour. 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 standardised meta-information to enable several parties to use the available data and to ensure that this data can be shared among them. Meta-information 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 standardised Meta-information set for traffic measurement data and flexible formats for providing the measurement data, which allows the translation to commonly used existing standards. 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 Meta-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 Meta-information 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 (Miska et. al., 2010).

2. KNOWLEDGE MANAGEMENT (KM)

Before going into the deign of the system itself, a short introduction to the field of knowledge management (KM) which was developed to create systematic and reliable ways to manage data, information, and knowledge. In a KM context, data, information, and knowledge are distinct from each other and have a hierarchical relationship. A collection of statistics is one example of data; information is any meaningful conclusions that can be drawn from the data (anything that informs). For example, the employment rates for every year (as long as they've been collected) are data. If we look at the numbers for last year and this year we have information, such as what the figures were for each year and how they compared. Knowledge, in this context, is a broader understanding that helps us to use the information to our best advantage, and might include, for example, factors that typically lead to a decrease in the employment rate. This, translated to our field explains that stored measurement is just data, but if we store it with meta-information we get information that can be used to create knowledge about a transportation network.

An established discipline since 1991 (see Nonaka, 1991), KM includes courses taught in the fields of business administration, information systems, management, and library and information sciences (Alavi & Leidner, 1999). More recently, other fields have started contributing to KM research; these include information and media, computer science, public health, and public policy. Thus, traffic engineers should take benefit of such developments and findings. Many large companies and non-profit organisations have resources dedicated to internal KM efforts, often as a part of their 'business strategy', 'information technology', or 'human resource management' departments (Addicott et al., 2006), while road authorities have not yet established a consistent strategy. Several consulting companies also exist that provide strategy and advice regarding KM to these organisations.

KM efforts typically focus on organisational objectives such as improved performance, competitive advantage, innovation, the sharing of lessons learned, integration and continuous improvement of the organisation. KM efforts overlap with organisational learning, and may be distinguished from that by a greater focus on the management of knowledge as a strategic asset and a focus on encouraging the sharing of knowledge. KM efforts can help individuals and groups to share valuable organisational insights, to reduce redundant work, to avoid reinventing the wheel per se, to reduce training time for new employees, to retain intellectual capital as employees turnover in an organisation, and to adapt to changing environments and markets (McAdam & McCreedy, 2000; Thompson & Walsham, 2004).

KM is not new and efforts have a long history, to include on-the-job discussions, formal apprenticeship, discussion forums, corporate libraries, professional training and mentoring programs. More recently, with increased use of computers in the second half of the 20th century, specific adaptations of technologies such as knowledge bases, expert systems, knowledge repositories, group decision support systems, intranets, and computer supported cooperative work have been introduced to further enhance such efforts. In 1999, the term

personal knowledge management was introduced which refers to the management of knowledge at the individual level (Wright, 2005).

In terms of the enterprise, early collections of case studies recognised the importance of knowledge management dimensions of strategy, process, and measurement (Morey, Maybury & Thuraisingham, 2002). Key lessons learned included: people, and the cultures that influence their behaviours, are the single most critical resource for successful knowledge creation, dissemination, and application; cognitive, social, and organisational learning processes are essential to the success of a knowledge management strategy; and measurement, benchmarking, and incentives are essential to accelerate the learning process and to drive cultural change. In short, knowledge management programs can yield impressive benefits to individuals and organisations if they are purposeful, concrete, and action-oriented. More recently with the advent of the Web 2.0, the concept of knowledge management has evolved towards a vision more based on people participation and emergence. This line of evolution is termed Enterprise 2.0 (McAfee, 2006). However, there is an ongoing debate and discussions (Lakhani & McAfee, 2007) as to whether Enterprise 2.0 is just a fad that does not bring anything new or useful or whether it is, indeed, the future of knowledge management (Davenport, 2008).

KM consists of different dimensions with different frameworks for distinguishing between knowledge. One proposed framework for categorising the dimensions of knowledge distinguishes between tacit knowledge and explicit knowledge. Tacit knowledge represents internalised knowledge that an individual may not be consciously aware of, such as how he or she accomplishes particular tasks. At the opposite end of the spectrum, explicit knowledge represents knowledge that the individual holds consciously in mental focus, in a form that can easily be communicated to others. (Alavi & Leidner, 2001). Early research suggested that a successful KM effort needs to convert internalised tacit knowledge into explicit knowledge in order to share it, but the same effort must also permit individuals to internalise and make personally meaningful any codified knowledge retrieved from the KM effort. Subsequent research into KM suggested that a distinction between tacit knowledge and explicit knowledge represented an oversimplification and that the notion of explicit knowledge is self-contradictory. Specifically, for knowledge to be made explicit, it must be translated into information (i.e., symbols outside of our heads) (Serenko & Bontis, 2004). Later on, Ikujiro Nonaka proposed a model (SECI for Socialization, Externalization, Combination, Internalization) which considers a spiralling knowledge process interaction between explicit knowledge and tacit knowledge (Nonaka & Takeuchi, 1995). In this model, knowledge follows a cycle in which implicit knowledge is 'extracted' to become explicit knowledge, and explicit knowledge is 're-internalised' into implicit knowledge. A second proposed framework for categorising the dimensions of knowledge distinguishes between embedded knowledge of a system outside of a human individual (e.g., an information system may have knowledge embedded into its design) and embodied knowledge representing a learned capability of a human body's nervous and endocrine systems (Sensky, 2002). A third proposed framework for categorising the dimensions of knowledge distinguishes between the exploratory creation of "new knowledge" (i.e., innovation) vs. the transfer or exploitation of "established knowledge" within a group, organisation, or community. Collaborative environments such as communities of practice or the use of social computing tools can be used for both knowledge creation and transfer. By using these frameworks and bringing the knowledge of engineers

together with the data, we can create a valuable tool for state estimation and expert systems for traffic management.
The motivations leading organisations to undertake a KM efforts are:

- Making available increased knowledge content in the development and provision of products and services
- Achieving shorter new product development cycles
- Facilitating and managing innovation and organisational learning
- Leveraging the expertise of people across the organisation
- Increasing network connectivity between internal and external individuals
- Managing business environments and allowing employees to obtain relevant insights and ideas appropriate to their work
- Solving intractable or wicked problems
- Managing intellectual capital and intellectual assets in the workforce (such as the expertise and know-how possessed by key individuals)

These motivations are overlapping with our intentions to accelerate research, to make traffic data easily accessible and widely available, to use the expertise of research and practice to solve our mobility problems, to enhance the dialog between academia and practice and to allow deeper insight into traffic phenomena. Based on this we created our international traffic database as described below.

3. OVERALL 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.

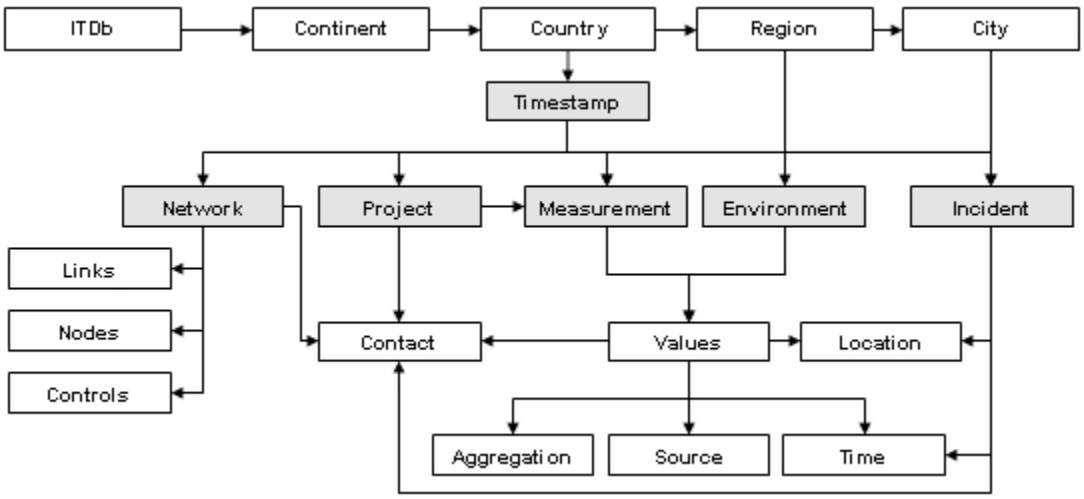

Figure 2: Structure of the Meta search database of ITDb

These Meta data is linked to the actual data sets 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 data flow in the ITDb system has two parts. A public front end where registered users can browse through the Meta database to search for useful data sets, 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.

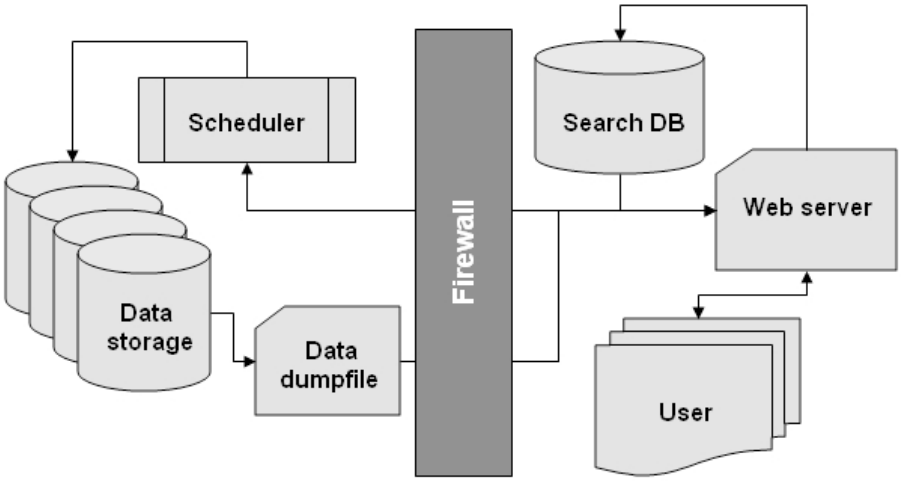

Figure 3: Data flow between the ITDb data server and the web front end

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.
To be able to search efficiently through various data sets world-wide, ITDb features a metadata search engine. Metadata for each data set is essential to give the user a complete knowledge about how to use the data and to avoid the garbage-in garbage-out phenomenon. ITDb requires the following meta information for each data set as a mandatory requirement:

- geo-referenced location of the measurement in WGS84
- time stamp of the measurement
- description of the measurement equipment
- description of the measurement installation
- information about if and how the raw measurement is processed
- what values are measured and in which form
- aggregation level of processed data
- contact information of the data owner

Further meta information will be stored, but is not required. The metadata search engine enables users to request specific data they are interested in. One could for example look for raw, disaggregated loop detector data, also known as pulse data. Then the search engine will locate all measurement spots know to the system and visualise them for the user in the commonly known environment of Google Maps, which allows an intuitive browsing through the found matches.

4. SUSTAINABLE TRAFFIC DATA STORAGE

Providing data for usage in practice and research need to be based on a good quality management to win the trust of the user in the offered datasets. It is vital the researchers and practitioners understand how the data was collected, who collected the data and how the data was processed already. For the latter it is important to have an insight of used techniques, if an automatic correction took place, how data was aggregated, and how equipment failure has been treated. Therefore, myITDb requests minimum requirements from data providers to ensure that the uploaded data can be used and, even more important, are understood by the users. We identified the most important information needed for ITDb as:

- Location
- Time span
- Measurement
- Contact

The location is important to identify the position of the measurement. Measurement spots must be easy to be identified and mapped to a network. Regardless of the measurement itself, the position in the network is needed to give any meaning to the measured values.
Time span means here the time and date of the measurement and its duration. A two hour measurement without the knowledge of the date loses valuable information. No crosschecking to special events, incidents or weather can be performed, not to mention that the day of the week has a significant influence on the traffic situation.
Of course there is the measurement itself as well. ITDb will provide its users detailed information about the measurement setup. That includes the measuring device (detector, video …), the type of installation (fixed, temporarily, experiment …), data recording (automatic by software or manually) and used correction algorithms performed on the raw data. Next to the installation issues, we demand a description of the values. This includes the aggregation, possible rounding, if the measurement is lane separated or an average, in which way it has been pre-processed and if, how much data got rejected from the sample.
No matter how much information ITDb is providing, there is always the possibility of further requests from the user to the data provider, or special questions that are beyond the usual. Therefore, ITDb will connect each single dataset to a contact person and affiliated company so that this information can be requested. This data will undergo a continuous update with an

automatic update procedure since contact persons are a subject of quick changes and stored data of a person that is not longer working at this position, is as good as no data at all.
Therefore we store meta-information with every data-set. Meta-information is information about information. For example, a if a document is considered to be information, its title, location, and subject are examples of meta-information. This term is sometimes used interchangeably with the term metadata. The ITDb meta format is (information in italics is optional):

- x-coordinate in WGS84
- y-coordinate in WGS84
- number of lanes on the road sections
- direction of road at detector location in radians
- *position along road (in m), 0 of not applicable*
- *name of road*
- *additional coordinate information (TeleAtlas, Navtec…)*
- *please feel free to add more things…*
- time (hh:mm:ss)
- date (yyyymmdd)
- measurement device
- device ID
- type of installation (fixed, temporarily, experiment)
- type of recording (manual, automatic)
- kind of measurement data
- auto correction algorithm
- aggregation (in seconds)
- rounding algorithm
- lane separated (yes/no)
- pre-processing algorithm
- percentage of remaining sample
- date registered
- last change date
- *registrar organisation name*
- *registrar phone number*
- *steward organisation name*
- *steward phone number*
- submitter organisation name
- submitter phone number
- address
- e-mail

Now to put all this into practice, we have started an online portal that consists of a public non restricted database and a secured database for limited access data and data exchange between business partners. In the following we will give a short introduction of the portal.

5. ONLINE PORTAL

The ITDb website (www.trafficdata.info) has a public part that hosts freely available traffic data from various parts of the world, and we keep our efforts to enrich the database with more and more data. The second part, the restricted area for account holders is called myITDb and is likewise free of charge.

myITDb is a data storage and sharing portal that aims to ensure a sustainable data storage for the field of traffic engineering. The data storage is divided into countries, which contain locations where the data-sets are stored. In this way, it is easy to browse the data and to find the exact datasets you are looking for without going through the trouble of keeping a good file management. myITDb can be accessed by the ITDb website. After you have logged into myITDb you will find a data browser window based on GoogleMaps. The first browser window is an country layer, containing the countries that have been added by users of myITDb.

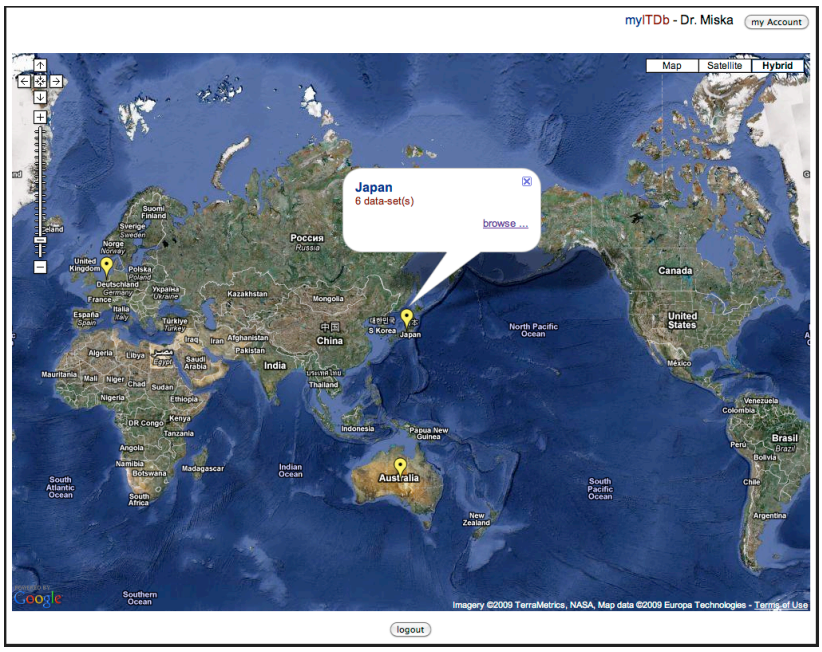

Figure 4: ITDb world-wide data browsing

If you click on “browse…” you will enter the country layer, in which the data is stored in location, also indicated by yellow markers. As with the countries, a click on the marker gives you information about the location and the amount of data visible.

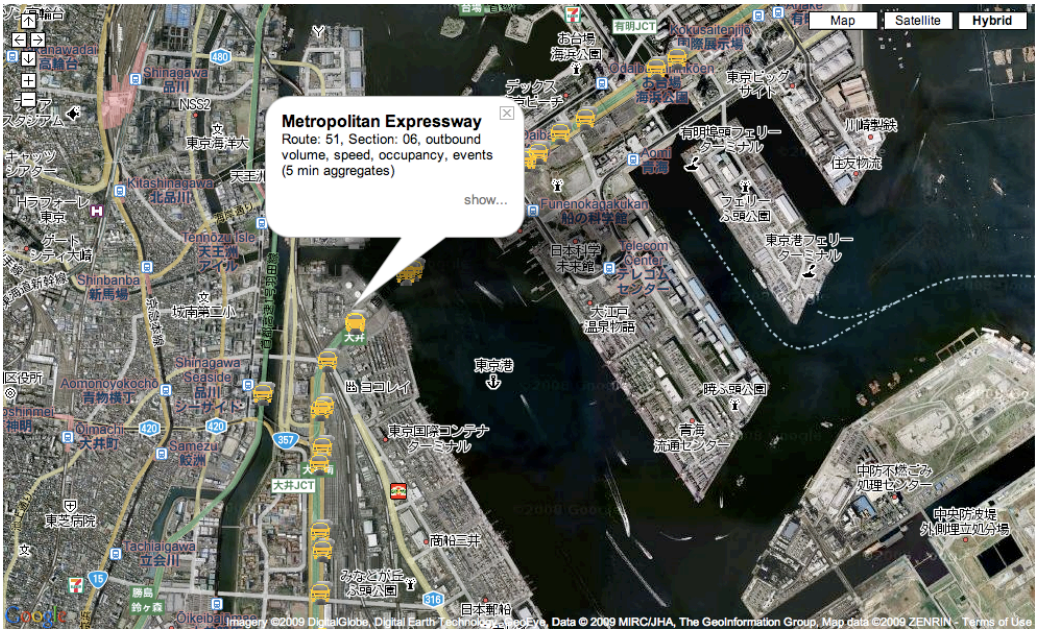

Figure 5: City scale meta search with ITDb

To store datasets geo-referenced, locations can be added for corresponding data file uploads, including the minimum set of meta information.

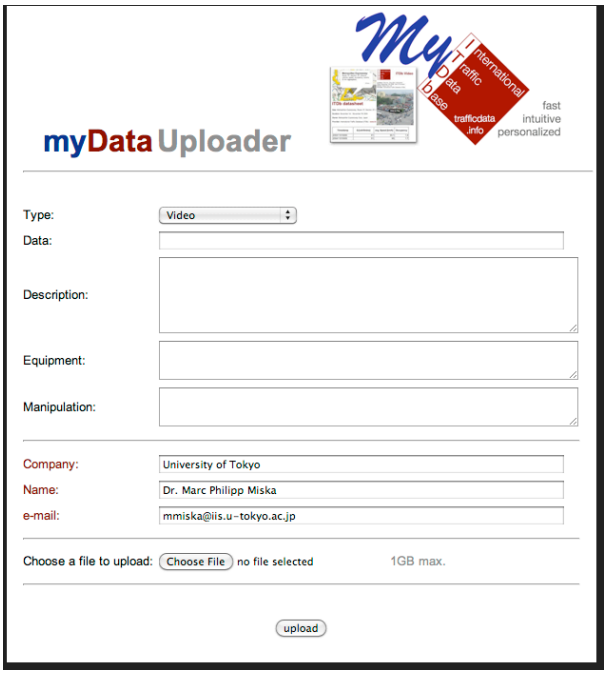

Figure 6: Data upload with minimum meta-information

To share data with others, users can create and host groups, to which other people can be invited. Uploaded data can be made visible to such groups, so that projects which use extensively shared data, or exchange data can easily manage the data flow and access rights with data repositories (see Figure 7).

Figure 7: ITDb data repository

These data repositories are location based, consist of geo referenced data and accumulate the data made available by the user and other groups the user is a member of.

Having a limited access area for restricted data download, myITDb is taking several measures to secure uploaded data from unauthorised access:

- every user has her/his own user and password combination
- during myITDb browsing, the user is continues authenticated by using cookies
- data download actions do not reveal real data location
- download links cannot be used outside the website (no e-mail of download links possible)
- every download is logged with time stamp, user information, ip address
- database containing information to link data files with descriptions, data owners or projects is not available from the outside (firewall protected - ip address not reachable)
- server ports are in stealth mode
- database access is just given from localhost
- Any other connection then HTTP is rejected by the server (no VPNs)

These measures have proven to address the concerns of data owners and disclosure agreements. However, with the portal being the result of a research project and free of charge, there is no further extend of liability.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have described an online portal for sustainable traffic data storage. Its concept is based on knowledge management, a field that is well established in the corporate world, and aims to enhance knowledge and management options for traffic networks. A set of meta-information stored with every data-set ensures that data is not simply a collection of numbers, but information that can be used in research and practice to develop new and better methods to ensure future mobility and sustainable transport. The portal is open to the public and is fully functional and in use by various parties around the world.
Future work of this project will involve next to the upload of more and more traffic data sources, the design of a toolbox for data analysis and data mining to look not only for specific data sources, but to search for traffic flow phenomena around the world. Further we are trying to extend our meta-information set for network descriptions and to provide a easy to use platform for project management.

ACKNOWLEDGEMENT

The International Traffic Database (ITDb) is part of the Energy ITS project, funded by the Ministry of Economy, Trade and Industry of the Japanese Government. Data available to the public can be accessed through www.trafficdata.info.

REFERENCES

Addicott, Rachael; McGivern, Gerry; Ferlie, Ewan (2006). "Networks, Organizational Learning and Knowledge Management: NHS Cancer Networks". Public Money & Management 26 (2): 87–94.

Alavi, Maryam; Leidner, Dorothy E. (1999). "Knowledge management systems: issues, challenges, and benefits". Communications of the AIS 1 (2).

Alavi, Maryam; Leidner, Dorothy E. (2001). "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues". MIS Quarterly 25 (1): 107–136.

Barcelo, J., Kuwahara, M., Miska, M. (2010). Traffic data availability and its standardization, Traffic Data Collection and its Standardization, Series: International Series in Operations Research & Management Science , Vol. 144, Barceló, Jaume; Kuwahara, Masao (Eds.), Springer

Davenport, Tom (2008). "Enterprise 2.0: The New, New Knowledge Management?". Harvard Business Online, Feb. 19

Lakhani, Andrew P.; McAfee (2007). "Case study on deleting "Enterprise 2.0" article". Courseware #9-607-712, Harvard Business School.

McAdam, Rodney; McCreedy, Sandra (2000). "A Critique Of Knowledge Management: Using A Social Constructionist Model". New Technology, Work and Employment 15 (2).

McAfee, Andrew P. (2006). "Enterprise 2.0: The Dawn of Emergent Collaboration". Sloan Management Review 47 (3): 21–28.

Miska, MP, Torday, A., Warita, H., Kuwahara, M. (2010). International Traffic Database - Gathering traffic data fast and intuitive, Traffic Data Collection and its Standardization, Series: International Series in Operations Research & Management Science , Vol. 144, Barceló, Jaume; Kuwahara, Masao (Eds.), Springer

Miska, MP, Torday, A., Warita, H., Kuwahara, M. (2007), The International Traffic Database Project, ITS World Congress 2007, Beijing, China

Morey, Daryl; Maybury, Mark; Thuraisingham, Bhavani (2002). Knowledge Management: Classic and Contemporary Works. Cambridge: MIT Press. p. 451.

Nonaka, Ikujiro (1991). "The knowledge creating company". Harvard Business Review 69 (6 Nov-Dec): 96–104.

Nonaka, Ikujiro; Takeuchi, Hirotaka (1995). The knowledge creating company: how Japanese companies create the dynamics of innovation. New York: Oxford University Press. p. 284.

Sensky, Tom (2002). "Knowledge Management". Advances in Psychiatric Treatment 8 (5): 387–395.

Serenko, Alexander; Bontis, Nick (2004), "Meta-review of knowledge management and intellectual capital literature: citation impact and research productivity rankings", Knowledge and Process Management 11 (3): 185–198, doi:10.1002/kpm.203

Thompson, Mark P.A.; Walsham, Geoff (2004), "Placing Knowledge Management in Context", Journal of Management Studies 41 (5): 725–747

Wright, Kirby (2005). "Personal knowledge management: supporting individual knowledge worker performance". Knowledge Management Research and Practice 3 (3): 156–165. doi:10.1057/palgrave.kmrp.8500061.