THE ROADIDEA DATA MEDIATION –

DATA SUPPORT FOR ITS DEPLOYMENT

Rene Kelpin
German Aerospace Centre (DLR)
Institute for Transport Research
Rutherfordstrasse 2
12489 Berlin/Germany
Fon: +49-30-67055217, Fax: +49-30-670558217, E-mail: rene.kelpin@dlr.de

ABSTRACT

Data – or information – is the main requirement for ITS. Intelligent transport services (ITS), applications and systems strongly rely on reliable data sources, which have to be given with a sufficient degree of coverage and penetration. Moreover, available data has to be well described in order i.) to allow a best possible knowledge about it, ii.) to find it in an huge data archive, and so, iii.) to utilize it in a best possible way. Finally, data has to be accessible without any borders. In an optimal environment necessary data sets are available free of charge; network access is publicly available without bureaucratic and technical restrictions. Reality tells a different story. European data availability of sources relevant for ITS purposes is very heterogeneous.

In the European research project ROADIDEA given availability of data sources for the transportation sector was analysed thoroughly. The project focussed on a definition of a road map for new and radical innovation for the European transport sector valid for the next decades. Special attention was focussed on the investigation of available data sources from a European perspective. Long haul road transportation relations - for instance from South of Italy to Northern Finland – were addressed, in order to provide similar ITS services along the entire route with special respect to specific climatic conditions. It is clear, that for such an approach a best possible knowledge about data availability in a continental scale is absolutely indispensable.

This paper describes the work and the results of the data related work package of the project ROADIDEA with regard to the project objectives. The major outcome of the ROADIDEA data research was the implementation of a ‘data mediation’, which targets at a combination of technologies applicable in a – from a data perspective – optimal transportation world.

INTRODUCTION

The main objective of the EU FP7/INFSO research project ROADIDEA is the development of new and innovative services for European transport systems. It aims at improvements of road traffic conditions and forecasts with special respect to certain weather conditions in different parts of the European continent.

The main prerequisite for achieving this goal is the knowledge about and the availability of valuable and high quality road traffic and weather data sources, which have not been used systematically so far. The base for the targeted approach to be applicable all over Europe is twofold. It consists of the following elements:

- a comprehensive investigation of available data sources in Europe and
- a reliable data mediation between data/service providers and the project data archive.

In a dedicated data work package (WP2) of the project ROADIDEA general data availability in Europe was analysed. After investigating information about available data sources in the project member states and storing corresponding data set descriptions in a preliminary data archive in the first work task 2.1 of WP2, available data was analysed in terms of its usability (task 2.2) and interoperability with other data sets (task2.3). Taking main findings of those analyses into consideration, a common data model and a corresponding data format was developed (task 2.4). Finally (task 2.5), both the archive and the format were implemented into the ***data mediation architecture*** as a base for project related data input and output relations. Right from the start all WP2 efforts were aiming at the preparation and the definition of such a data mediation system.

The main goal of the data mediation approach is to mediate between data providers and data users, on the one hand, and a data handling architecture, on the other – as secure, reliable, and generic as possible. Direct access to the project data archives and repositories is not desired. New data sets need to be well described and analysed before being stored in a common data format. Data availability requests have to be made and responded in a highly standardised form. If these constraints are given, implementation of the data mediation architecture is the best choice. It has to be considered as the middleman between the data archive and the data applications, such as transport services in the given project context.

The given paper first briefly describes the project data source investigation and its results, as well as the main findings of the work tasks 2.2 – 2.4. The focus of this paper is on the data mediation idea and the description of the ROADIDEA data mediation architecture.

In a project extension, called ROADIDEA International Cooperation Aspects (INCO), a similar investigation of available data sources and its utilisation for ITS services is currently being carried out in North America (USA and Canada) until September 2010. As a comparative study it may be considered as an attempt to derive lessons learnt from European and North American transportation systems and apply them on the other side of the ocean. Assuming a more open data policy in North America, existing ITS services are to be compared with those given in Europe. Thus, this paper also briefly describes first results and findings of the ROADIDEA INCO project extension.

METHODOLOGY

METHODOLOGY OF DATA SOURCE INVESTIGATION AND MAJOR RESULTS

In order to perform a best fitting data source investigation a data survey questionnaire was designed using the experience of the DLR clearing house for transport data. This questionnaire was used to identify and to describe data sources available in the project member states or ascertained and utilized by ROADIDEA project partners. This data source identification covered the following initial main data classes (see (8)):

- data from vehicles,
- data from infrastructure and
- weather monitoring.

During the data collection period, about 60 data sets were identified and described. Figure 1 shows the assignment of the investigated data sources to the mentioned survey data classes (some data sets were assigned to more than one class).

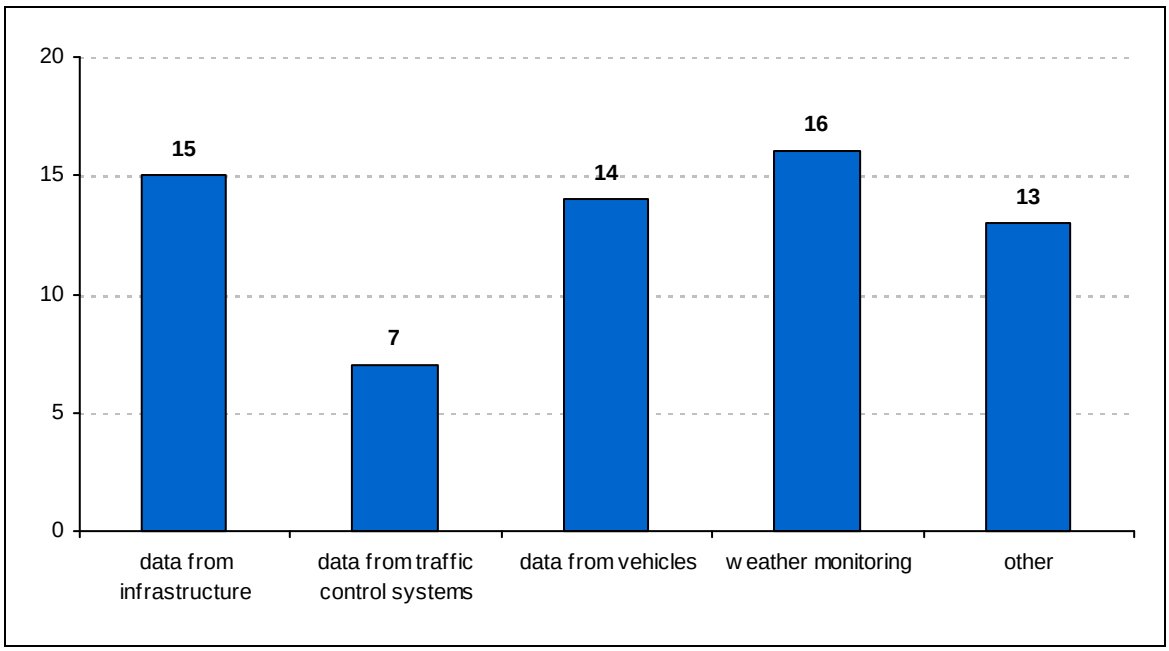

Figure 1. Assignment to preliminary data classification

An upcoming discussion showed that this classification was not likely to allow a proper distinction between specific data sources and its classification. So, a more detailed data classification was proposed and applied to the ROADIDEA data survey. Based on this classification the assignment, shown in Figure 2, gives a more detailed overview of investigated data sources. It allows a fairly good assessment of the accomplished survey as such, as well as of its results. With 24 resp. 49 data sets the main data classes 'traffic monitoring' and 'weather conditions' are well filled with respect to the main project purpose.

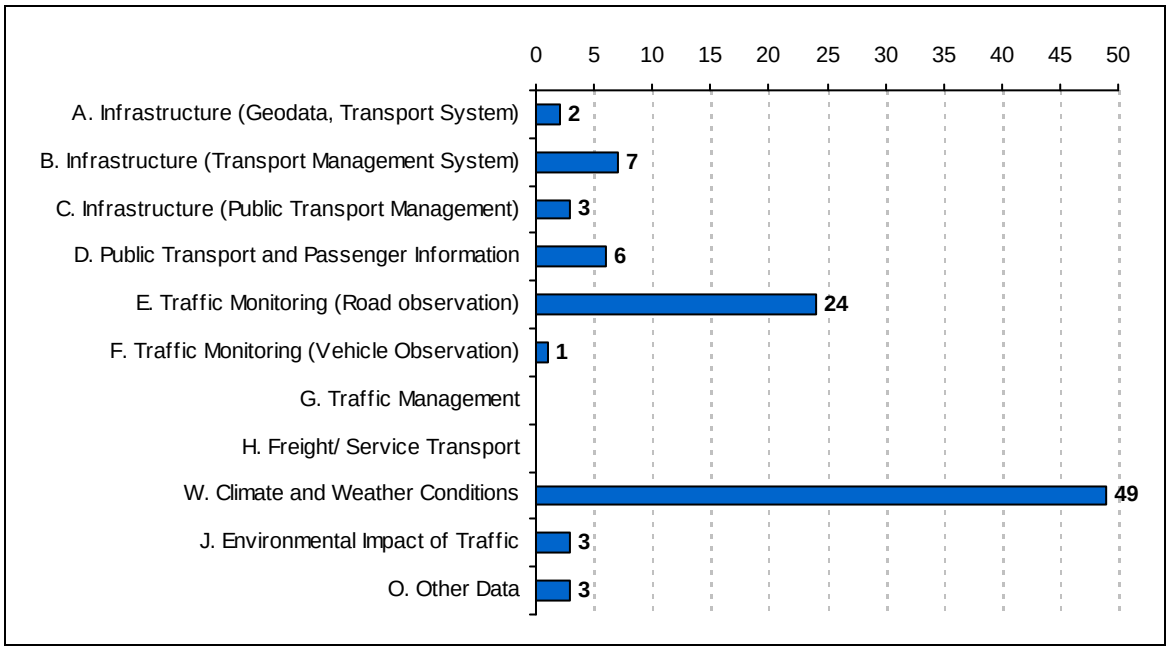

Figure 2. Assignment to final data classification

More detailed data source analyses took place in Germany, Finland, Croatia and Italy. The focus of these analyses was the comparison of general data availability in different European countries, with special respect to the degree of development in terms of data policies.

In addition, general availability of several data sources relevant for the ROADIDEA project purpose all over Europe has been investigated. Initial contacts with providers of these data sets were established; willingness to contribute to the project was discussed and partly agreed. Among more general elements (e.g. location, time, etc.), data access possibilities, licence fees and data quality matters were also considered thoroughly with the data source descriptions. This was very important in order to support other ROADIDEA work packages dealing with business models and service evaluation. Thus, while describing, comparing and analysing the data, a very good understanding of its contexts and its contents was obtained.

Participating project countries can - more or less - be covered with the same or similar kinds of data sources. However, a general data availability for ITS is not yet guaranteed all over Europe. Some well developed countries (e.g. Sweden, Finland, Germany, and Italy) are providing huge amounts of reliable and well maintained data of high quality, while in other developing countries in South- and Eastern-Europe, data responsibilities and availabilities are unclear or under development. Hence, during the investigation of data availability some partners came in contact with various obstacles and problems. These problems were rather bureaucratic than technical. In order to convince road authorities, public and private data providers, and governments to collaborate with, and generally support research projects, a lot of work is still to be done. The most sustainable way to find new collaborators and data providers is to show the success of e.g. innovative ROADIDEA services which might be achieved, with data the companies are providing.

More information, charts and data examples can be found in (1).

INTEROPERABILITY OF INVESTIGATED DATA SOURCES

The main result of the analysis of interoperability of available data sources was the recommendation of a minimal data catalogue for European ITS.

Within corresponding work tasks of WP2, available data sets have been analysed in terms of comparability and interoperability. In part, it is possible to evaluate accuracy of data sources by using data from similar sources. Data gaps can be closed; systematical or technical bias can be identified and eliminated. Moreover, with the combination of – at first sight – independent data sources, new information can be obtained. Hidden content can be derived

from useless data sources. For instance, (3) describes the attempt to identify a functional relation between permanent road sight surveys with for example, Floating Car Data (FCD) and road weather situations. The attempt finally failed: The road network penetration rate with FCD was not high enough to allow a derivation of deterministic rules and reliable statements describing such a functional relation. Nevertheless, the general question still exists: Is it possible to derive information about weather conditions from the traffic status on well observed road networks and vice versa?

But, the success of such analyses mainly depends on data: (freely) available across the board, precise and accurate, and reliable. In order to obtain a Europe-wide availability of data relevant for ITS purposes, corresponding data classes/sources have to be identified and classified. The definition of those data classes was recommended as the major conclusion of the consideration of data interoperability in WP2. In order to create ITS services valid and applicable all over Europe, most necessary data sources have to be identified and publicly provided for the transportation sector. Bureaucratic barriers are to be overcome; license and fee problems have to be negotiated at a European level. That means that an EC program is needed, which has to be applied in all EC member states as a common data policy.

More information can be found in (2).

THE ROADIDEA COMMON DATA FORMAT AND THE PROJECT DATA PLATFORM

In order to make different data sets comparable and able to be combined, a common base had to be found, with which data sets can be described in the greatest detail needed. Such a common description was to be considered as a main prerequisite for a project data platform - and for the desired ROADIDEA data mediation architecture WP2 of the project ROADIDEA was aiming at.

Taken into account all known and conceivable data sets for the specific project background, a set of necessary descriptors was designed, which was given a structured shape and functionality with the well developed XML standard. With the ROADIDEA XML data schema, a descriptive data format was designed, which suits best the underlying project demands. An assortment of existing data formats, architectures, schemes, and standards (e.g. TransXML, DATEX, GTF, etc.) was analyzed in terms of applicability for the ROADIDEA purpose. No existing format, however, was suitable for the ROADIDEA approach without applying major changes and adaptations. Nevertheless, parts of some schemes, such as Dublin Core (DC) and Geography Markup Language (GML), were used.
However, it was decided to create ‘another’ proprietary and problem oriented XML schema.

This XML-based data format is suitable for the description of all project related data sources; it is as slim as possible for the given attempt, and free of unnecessary format overhead. Figure 3 shows the ROADIDEA XML scheme data model.

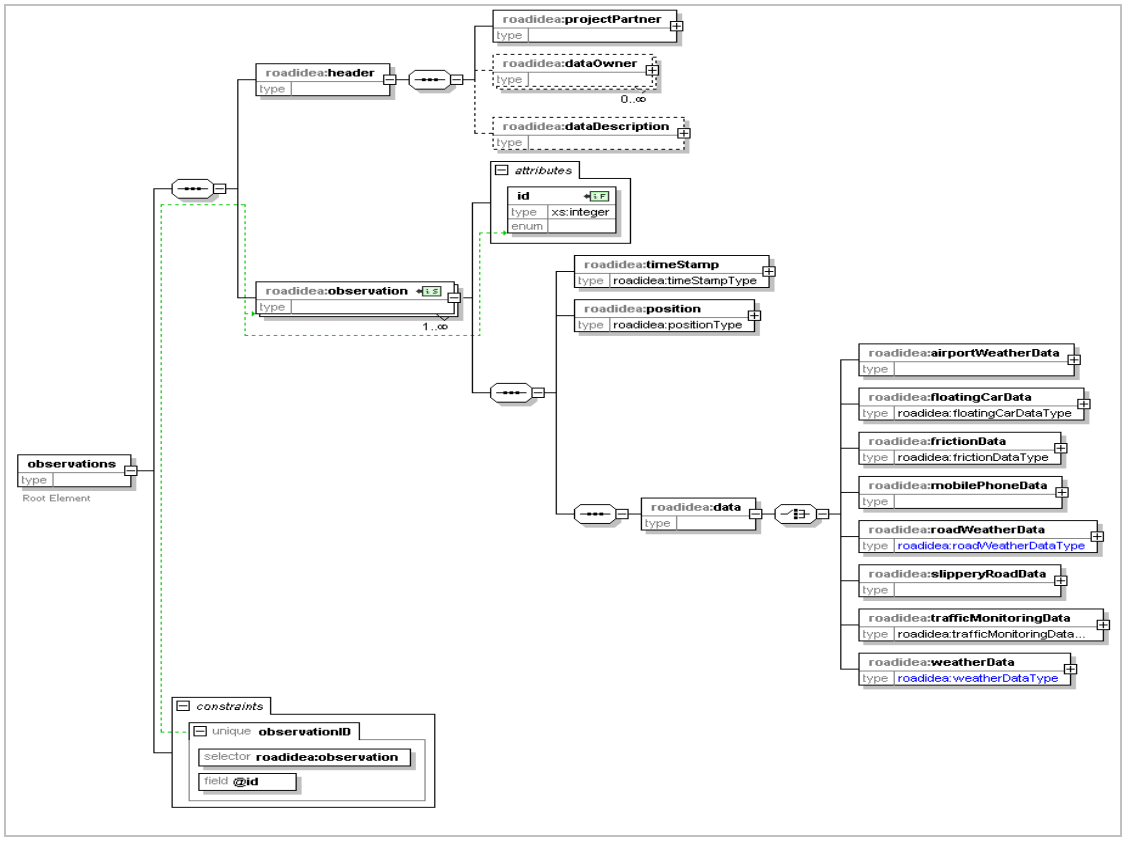

Figure 3. ROADIDEA data model

More information can be found in (4). The ROADIDEA data archive is set up with the open sources software Eprints. Find more information about the software at (9).

THE ROADIDEA DATA MEDIATION ARCHITECTURE

As already said above, the general idea of data mediation is the creation of a “middle-man”-system, which is responsible for data relations into and out of a data handling system. Due to licence, privacy and data security policies, direct access to the data is not allowed. Specifically, data availability requests to the data archive have to be processed and answered in a well defined manner. Interfaces for both human interactions (e.g. via a web service) and application triggered (‘machine-made’) requests (e.g. via an API) are to be provided with the same system. Predefined and generic dictionaries have to be implemented, which translate requests into a language the entire system is able to understand and to work with. The internal communication is highly standardised; it is based on XML-schemes. As for the ROADIDEA project context, the system has to be able to deal with very different kinds of

data - from road traffic and weather observations, to radar satellite images and road friction measurements. Corresponding data descriptors and contents have to be stored in the same archive, have to be merged and compared with respect to very specific application and utilisation purposes. The data mediation architecture may be considered as a nut-shell around the ROADIDEA data archive with well defined input and output relations.

The achievements of the preceding work tasks of WP2 were taken as preparations for creating the backbone, on which the final data mediation system was based. It took into account not only the project data sets, but also the results from various data surveys and discussions in all related project work packages, and from the ROADIDEA project innovation seminars. With the two innovation seminars a lot of new ideas with a lot of – so far – unknown data needs and application aspects were created (see also (6) and (7)). All these information, expectations and specifications have been used while designing the XML data format, the data model, and finally the ROADIDEA data mediation system.

The ROADIDEA data mediation distinguishes between Input and Output schemes.

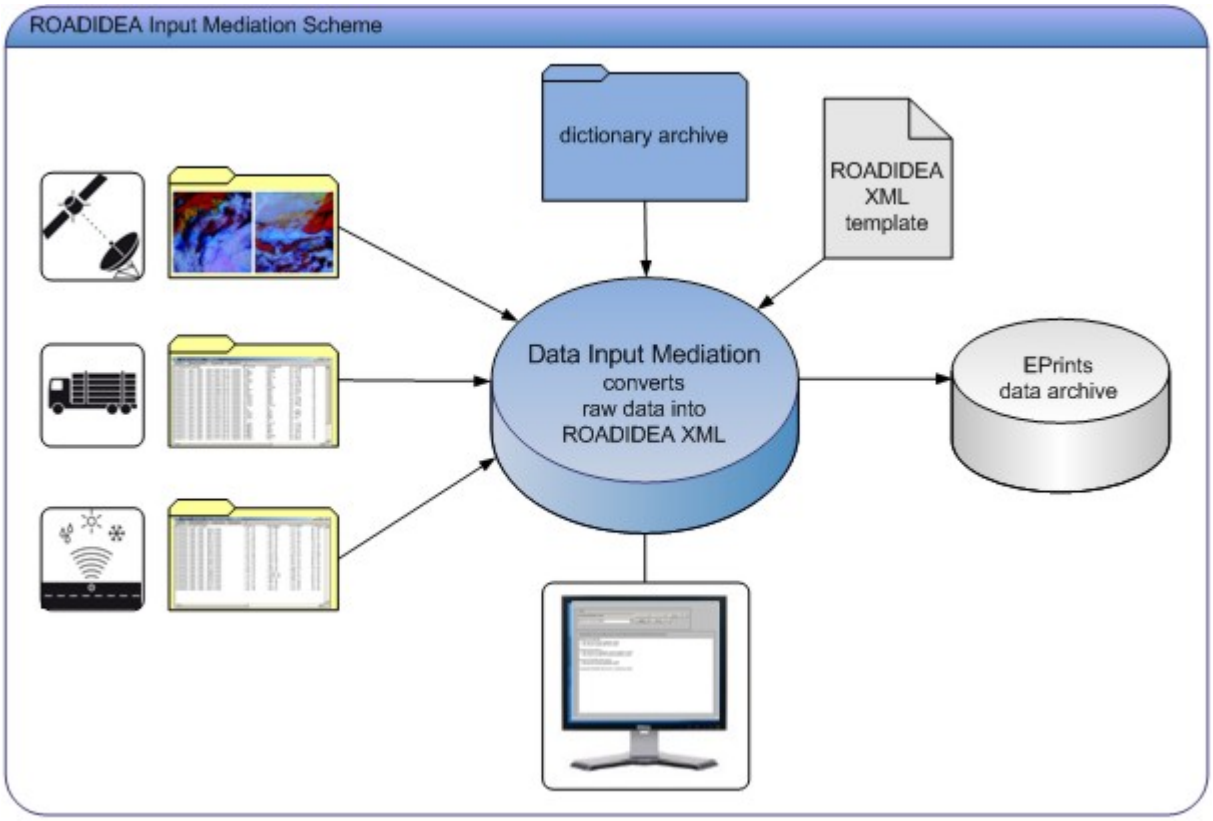

Figure 4. Input Mediation Scheme

Figure 4 shows the data mediation output scheme. The main element of both the input and the output scheme is the ‘Mediation Core’. It is accessing the data archive, the dictionary archive and several XML template directories. On the left side three examples illustrating the

diversity of ROADIDEA data sources are shown. To every data source a corresponding dictionary was defined, which describes the data source information implicitly and explicitly. This means that, beside the raw data content (i.e. codebook), all descriptive information is given with the dictionary. Using these dictionaries, input data sources are transferred into the ROADIDEA XML format and stored into the internal data archive.

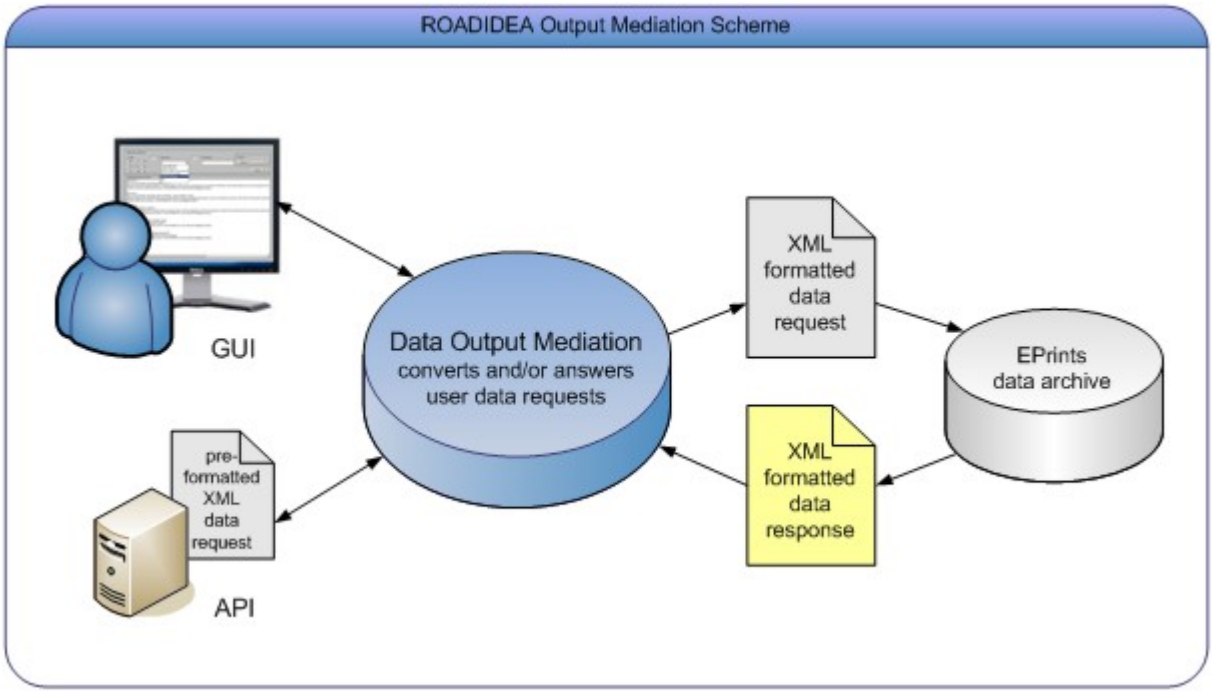

Figure 5. Output Mediation scheme

Figure 5 gives an overview about the data mediation output scheme. It shows that the system is able to deal with both ‘man-made’ and ‘machine-made’ data requests. After processing (communication and translation) a specific data request, a corresponding XML-request is sent to the data archive. If data fitting to the selected choices and keywords is found in the data archive, the XML data request will be responded to with a list of data sets. The list comprises the most important data set information (ID, title, abstract and internal contact information). ‘Machine-made’ requests will be answered with a data set list stored and submitted in a specific XML format.

More information and a detailed description of the mediation application can be found in (5).

FIRST ROADIDEA INCO FINDINGS

The major goal of the ROADIDEA INCO (International Cooperation Aspects) was the comparison of available transport related i.) Data, ii.) Applied models and methods, and iii.) the utilisation of both data and methods for innovative ITS services. While this paper is

dealing with data aspects of the ROADIDEA ‘mother’ project, this sub-chapter only describes main data related findings of the INCO project extension.

First of all, it has to be pointed out, that the transportation systems of the U.S. and Europe as a whole can hardly be compared in a general manner. While U.S. federal states can easily create cooperation networks and apply ITS initiatives and programs under the ‘umbrella’ of different public managing, planning and funding authorities – such as U.S. Department of Transportation (USDOT) or Federal Highway Agency (FHWA) – without cultural and regulation triggered barriers, European Union still exists of 27 separated member states with more than 20 languages and – even more important – a huge amount of different cultural, historical and political backgrounds, and – when it comes to transport related the data aspect - regulative policies. Nevertheless, the ROADIDEA INCO data consideration tried to compare given trans-regional ITS systems, taking U.S. mainland states and EU member states as two comparable entities in terms of transportation. Special respect has been laid on cross border ITS services applying and merging data from different sources with innovative methods and sophisticated models in order to improve road safety and driving conditions – and so, decreasing road traffic fatalities, congestions and delays.

The following listing of conclusions may be taken as a final INCO report preview.

- A lot of local solutions for local problems are given in both the U.S and Europe. Conversely, this means that an operational trans-regional approach to tackle similar issues globally is mostly missing on both sides of the Atlantic Ocean, even though lessons learned in other regions are widely taking into account when setting up new local solutions and research communities are well networked.

  However, corresponding activities and initiatives are under consideration.
- Data acquisitions focuses differ from state to state – following national or federal strategic decisions. In Europe no minimal availability of data necessary for elementary ITS services is given in a continental scale. So, cross border implementation of services along long distance routes throughout Europe are barely applicable.

  The same only partly applies to the U.S. Nowadays a more strategic direction towards common data coverage is given by regulative authorities (USDOT, FHWA, etc.). While corresponding decisions are made by only a few authorities, these decisions are more reliable but also less flexible. Once pursued strategic directions can not be easily corrected. Due to different national strategies in Europe, development paths are more diverse and competitive.
- The issue of data related regulations are important in the U.S. and Europe similarly. As for the U.S., there is definitely no “free data policy”, as it is rumoured from time to time outside the U.S. With the USDOT CLARUS initiative on road weather data one aspect of transport related data can be considered as “free” available. But, this advantage is lightly used up by regulations, public and private road traffic data

providers saddle their data with.

The European Commission, as the roofing European body, is working on corresponding recommendations, which could clarify data access regulations for EC member states. But, to date it is unknown how to implement those recommendations and how to convince member states to make those recommendations to national law.

- The “Networks-of-Networks” idea has to be matured and applied generally and globally. A variety of problem oriented networks, which providing quiet good local solutions for different local problems (traffic and weather), have to be combined globally and interdisciplinary. Corresponding networking funding schemes in the U.S. and Europe have to be developed and synchronised – not least from a data perspective.

CONCLUSION AND FUTURE WORK

With the described data mediation architecture, a powerful tool for project-related data access and exchange is given. This architecture has to be considered as the basis on which data related project applications may be premised. While developing the data mediation architecture, it was the main goal of WP2 to provide an easy-to-use option for implementing new applications and utilisations (from a data point of view). The system is able to establish connections between data providers, on one side, and service providers (data users), on the other. The applied data sources of the project pilot ‘Gothenburg traffic monitoring’ are implemented as a demonstration case for showing the feasibility and functionality of the system. This implementation can be found and tested at http://www.roadidea.eu/packages/wp2/default.aspx . Example data sets (for the Gothenburg pilot) can be transformed into the ROADIDEA XML-format and uploaded into the data archive; data requests can be created, adapted and sent to the system, which will be responded with corresponding data set listings.

However, the underlying data archive determines the value of applicability and reliability of the system. That means that a system like the described ROADIDEA data mediation architecture strongly depends on available (investigated, described and stored) data sets and their descriptions. Thus, the data source investigation may never be considered as concluded. In the remaining run-time of the project, ROADIDEA WP2 continued with the data source investigation in countries other than those participating. General information about data availability for ITS in Europe is desired as one major outcome of the ROADIDEA data work package. This went hand in hand with an intensive discussion and deployment of the idea of a data catalogue (minimal set of data) for ITS. On a European level it has to be agreed upon a minimal set of data sources, which is most necessary for the development and implementation

of elementary ITS applications. In the 21st century, applicability of ITS can not simply end at the frontiers. But, such an approach can only be followed if similar data availability is given by *all* member states in the European Union. This - not completely new - idea was raised with one of the early WP2 tasks, and it got a lot of attention within different working groups in the European Commission, as well as in some collaborating research projects such as EASYWAY. If forces can be joined in the near future, valuable and reliable results can be expected, which might influence the deployment of services for the transportation sector deeply.

Once again: In order to create a win-situation for all contributors, which are

- public and private service providers,
- public and private data providers and
- research institutions,

in a well developed European ITS system, data providers have, first of all, to be persuaded to collaborate. They have to be persuaded to allow cheap or – for specific deployment and research purposes – even free access to their data bases. This has to be supported by European rules and policies, but also by local governments and authorities. Data has to get a stronger focus than it already has - right from the beginning of the basic research up to the implementation of the commercial services.

ACKNOWLEDGEMENTS

Commission of the European Communities Information Society and Media Directorate-General has provided major financial support for the ROADIDEA project. All partners of ROADIDEA - VTT Technical Research Centre of Finland, Finnish Meteorological Institute, Destia Ltd, Logica Suomi Ltd, Klimator AB, Semcon Caran AB, Demis BV, Pöyry Infra Traffic GmbH, ARPAV, Centro Meteorologico Teolo, Road Safety Engineering Bureau Hungary, Meteo-Info d.o.o. Croatia, and Amanova d.o.o. Slovenia -
are thanked for their enthusiastic work and contributions to the success of the project.

REFERENCES

(1) “ROADIDEA D2.1 Present availability of data for transport systems and services”. Available at www.roadidea.eu.

(2) “ROADIDEA D2.2 New types of data for transport systems and services”. Available at www.roadidea.eu.

(3) “ROADIDEA D2.3 Integration of data useful for ROADIDEA system”. Available at www.roadidea.eu.

(4) “ROADIDEA D2.4 Model for data storage in a defined uniform archive”. Available at www.roadidea.eu.

(5) “ROADIDEA D2.5 Integration of data useful for ROADIDEA system”. Available at www.roadidea.eu.

(6) “ROADIDEA D5.2 Results of the first innovation seminar”. Available at www.roadidea.eu.

(7) “ROADIDEA D5.3 Results of the second innovation seminar”. Available at www.roadidea.eu.

(8) Dalgleish, M. and Hoose, N, *Highway Traffic Monitoring and Data Quality* (London, 2008), Artech House.

(9) “EPrints - Open Access to Institutional Repositories”, Available at http://www.eprints.org