DEVELOPMENT OF A KNOWLEDGE-BASE FOR A PAN-EUROPEAN RAIL FREIGHT CORRIDOR - A CASE STUDY

ISLAM, Dewan Md Zahurul, Freight Logistics Research Group, NewRail - Centre for Railway Research, University of Newcastle Upon Tyne, NE1 7RU, UK, *Email for correspondence: dewan.islam@newcastle.ac.uk and telephone: +44 1912223972;*

ZUNDER, Thomas Hagen, Freight Logistics Research Group, NewRail - Centre for Railway Research, University of Newcastle Upon Tyne, NE1 7RU, UK.

Grashoff, Poul Sander, Demis bv, Rotterdamse weg 183c, 2629HD Delft, The Netherlands.

ABSTRACT

An effective and informative pan-European transport corridor contributes towards the European Commission's policy objective of establishing an integrated trans-European transport network (TEN-T). RETRACK is the "REorganisation of Transport networks by advanced RAil freight Concepts" project, funded under the European Commission (EC) FP6 Programme, and aims to contribute to the Commission's aspirations of a commercially viable modal shift of freight traffic from road to rail to help achieving sustainable mobility.

This paper presents a European rail freight corridor knowledge base, using a case study - RETRACK KB. The RETRACK KB aims to facilitate stakeholders with data and information for example, on infrastructure condition; supply and demand of rail freight, corridor connectivity with alternative hubs and spokes; the market conditions; institutional conditions; and environmental conditions. The information is intended to be for generic operational issues as well as requirements for a particular railway corridor.

The paper describes specification of the rail freight knowledge base, followed by a discussion on the design (including content and structure) and main software elements (including table data software, geo data software, web portal and content management and document repository) and construction of RETRACK KB.

Keywords: knowledge-base, rail freight, corridor, Europe

INTRODUCTION

Since 1991 the European Commission has adopted many directives and reform packages to transform the national rail freight transport market in Europe into a pan-European market where both incumbents and private operators can operate freely and commercially. Another important aspect of these initiatives is to achieve an integrated trans-European transport network, where a pan-European corridor approach (rather than a traditional national boundary focused approach) is an important step forward. To explore the effectiveness of this market, the EC funded two projects: CREAM for incumbent operators (CREAM, 2012) and RETRACK for private operators (RETRACK, 2012) under the FP6 programme. This research paper reports a part of the research work conducted under the RETRACK project, where private railway undertakings, market-oriented rail freight operators, experienced IT and training specialists and leading European research and development organisations have worked together to design, develop and implement a new and innovative pan-European rail freight service concept. Within this work, corridor assessment was conducted, through literature reviews and interviews with different direct and indirect stakeholders. This pan-European rail freight concept began with a planned service setup on a West-East corridor linking major ports of the North Sea (e.g. Rotterdam) with major industrial areas in Germany, Austria, Hungary and the Black Sea (e.g. Constanza). It also explored, at a later stage, the potential for extending this corridor further, to Russia and China across a number of promising East European and Central Asian countries (e.g. Kazakhstan), applying the service model developed in the project. The ultimate objectives of the RETRACK project were to explore the potential for a shift of cargo from road to rail and to create an effective and scalable freight corridor between economic core areas in Western and Eastern Europe. The lessons learned in implementing this corridor based project, and the invaluable research outputs and information collected are found to be useful for other corridors. Rail freight operators considering expansion of their services to another corridor, for example, have a need for information on the infrastructure condition on the corridor and for information on the sort of problems that may arise in designing, developing and operating a pan-European service in another corridor in Europe. The RETRACK project concluded that it is necessary to create a corridor based European rail freight knowledge base to support research for - and development of rail freight corridors.

What is a corridor based rail freight knowledge base?

The idea of a knowledge base (KB) system is to enable users to have ready access to the documented base of facts, sources of information, and solutions. There is a large volume of literature on knowledge management and knowledge based systems. A KB is developed in response to knowledge management objectives for a large variety of country, domain or organization applications. The KB applications range from collection, organization and retrieval of special information in document management systems up to expert systems providing artificial intelligence based on user inputs.

Synonyms for knowledge are information; understanding; discernment; comprehension; erudition; scholarship (Dictionary.com, 2011). A KB system refers to a system for managing knowledge for supporting the creation, capture, storage and dissemination of information (Wikipedia, 2011). In terms of information technology (IT), a KB system is a machine-readable resource for the dissemination of information, generally online or with the capacity to be put online. A KB system is considered to be an integral component of knowledge management and is used to optimize information collection, organization, and retrieval (SearchCRM.Com, 2011).

In the case of the RETRACK project there is a need for a rail freight KB, based on a corridor approach. If we look at literature for rail freight KB most references concern train scheduling and train control systems (e.g. Hui et al., 2012; Peng He, 2011; Castorini et al., 2010; Cheng et al, 2009; Liu et al, 2008; Fay, 2000; Chiang et al, 1998). These applications show the use of formal ontologies and typically use expert system approaches. Rail accidents and incidents are another domain for KB system application (e.g. Maalel et al, 2010).

For rail freight policy questions only a limited number of literature references can be found (e.g. ETIS, 2011b and c; Wigan et al, 2007; Scholz, 2002) Scholz (2002) uses case studies to illustrate eleven different methods for integration of qualitative and quantitative knowledge in the case of complex problems. Also the RETRACK problem field of policy information in the case of rail freight corridor studies is highly complex and needs both qualitative and quantitative information. The RETRCK KB should allow stakeholders to analyse and find information on infrastructure conditions, rail freight supply and demand; the market conditions; institutional conditions; corridor connectivity and environmental conditions. It should include information on generic operational issues and requirements of particular railway corridors. This demands a wide range of new and existing data and documents to be brought together in an immediately usable, downloadable and searchable form; also it needs data exploration and thematic map forms of examination, to bring the costs and time required to access and use such information, into the operational domain.

In light of this the RETRACK rail freight corridor knowledge base system requirements are defined as follows:

- Have an organization in a general section and separate sections for each corridor
- Contain information on infrastructure conditions, rail freight supply and demand; the market conditions; institutional conditions; corridor connectivity and environmental conditions.
- Relevant (qualitative) documents and (quantitative) data by rail freight corridor must be easily stored and made searchable and downloadable
- Regional data must be accessible in the form of thematic maps, tables and charts
- Network / corridor infrastructure elements and data must be accessible in an on-line geographic information system (GIS).

Methodology

The current research applies qualitative and quantitative research methods in three consecutive stages: forming of philosophy, design stage, and construction.

Applying qualitative research, the first stage is forming the *philosophy* along the following perspectives:

- **Setting generic nature** of information needs: to ensure the system architecture and the technical platform for supporting RETRACK-KB are built with a generic nature, allowing them to be applicable to the study of European rail freight corridors in general;
- **Setting uniform procedures:** conducting surveys, queries and standardised, structured questionnaires so that they can be used for the study of any other European rail freight corridors;
- **Dialogue with TENTec** Information System (DG Mobility and Transport, 2012) at the European Commission to interchange relevant data into each other's systems, to harmonise interfaces between systems, and to explore the possibility of merging into one structure.

Applying quantitative methods, the second stage is the *design* of RETRACK-KB, in two steps: (a) the design of technical platform - addressing what types of data of a rail freight corridor will be included, and how to host and present these data; (b) the system architecture, which deals with the actual building of this technical platform, based on its design. The generic nature of RETRACK-KB is revealed and embedded in this design stage.

The final stage - *construct* - refers to the work of feeding data into the RETRACK-KB platform. This contains two steps: (a) defining procedures and requirements for data feeding (data receiving and submitting); (b) data feeding.

The paper continues by describing specifications of the rail freight knowledge base, followed by a discussion on the design (including content and structure) and main software elements (including table data software, geo data software, web portal and content management and document repository) and then construction of RETRACK KB. A final section summarises the research work.

SPECIFICATION OF A RAIL-FREIGHT KNOWLEDGE BASE

The RETRACK KB is a rail freight knowledge base. It includes (knowledge on) generic operational issues and requirements of a particular railway corridor. This demands a wide range of new and existing data and documents to be brought together in an immediately usable, downloadable and searchable form; also data exploration and thematic map forms of examination, to bring the costs and time required to access and use such information, into the operational domain.

The RETRACK KB has to make complex transport datasets easily usable. This is illustrated by the range of analysis, access, time series diagrams, graphs, thematic mapping origin-destination (OD) matrix displays, desire-line mappings and many other tools, normally only accessible after time consuming, heavy lifting of securing a data set, setting it up, understanding it, loading it into a specialised analysis system and producing special maps and diagrams to communicate the results.

The RETRACK KB needs all of these tools built in, so that the process of data set search, identification, exploration and visual and tabular and regression analyses are all part of the same access session. It needs to be presented, not only as a multidimensional repository, but also as a high level productivity tool, designed for globally distributed researchers, and supporting them with tailored tools, as well as visualisation and data exploration and documentation.

Rail freight transport has long needed such a system; not only to make effective use of existing data and information, but also to raise productivity of rail freight industry and speed of progress of the research and projects undertaken in this increasingly complex domain. In this respect the REORIENT Knowledge base at www.reorient.org.uk provides a base to build upon in the RETRACK project (Wigan et al, 2007). Another key project is ETISplus (ETIS, 2011a-f) within the 7th Framework Programme of the EU. The lessons learned, and software tools built, under these projects, are useful in the design and implementation of the RETRACK KB.

Furthermore, on-line Graphical Information System (GIS) tools needed to be used to manage and standardise the inputs; therefore an appropriate network editor and data collection support system was implemented. A multi layered security and data screening system was also a requirement. Considering these aspects, the rail freight RETRACK KB approach requires the followings:

1) A specification;
2) Tools to support it (and assurances of solid third party support for the key server engines);
3) Special glue to craft user interface and integration;
4) A clear vision of how it will make research more productive and technology transfer more effective and economic.

The specification of this generic rail freight knowledge base consists of 1) web portal 2) document repository for handling (qualitative) information, 3) (quantitative) table data viewer and editor 4) On-line GIS for network elements and data

DESIGN OF THE RETRACK KB

RETRACK KB Content

Kolesa (2007, p. 191) suggests a 'tool for building knowledge bases from textual resources. It is based on two assumptions. First, if the specific purpose of the knowledge base system is well understood in advance, then during knowledge acquisition, the knowledge base can be constructed in a simple (but special-purpose) manner. Second, each text has to be processed manually by a user'. Chklovski and GIL (2005 P. 35) outlines the following five key design features of a knowledge base:

1) Create and fine tune templates to acquire specific types of semantic relations;
2) Provide guidance and feedback on the form and type of the answer sought;
3) Acquire knowledge incrementally, breaking up collection of complex statements into several acquisition steps;
4) Automatically post-process the knowledge to repair or discard entries; and
5) Direct multiple contributors to validate and evaluate previously entered statements.

Considering the necessity of the RETRACK KB, the themes and structure of the proposed content are prepared on five themes:

- Infrastructure supply and demand;
- Market condition;
- Institutional conditions;
- Connectivity of the corridor;
- Environmental conditions.

For each of these themes typical policy questions were derived and then analysed for the indicator and supporting parameter data needs to answer these questions. The resulting list of questions, indicators and parameters were defined content of the RETRACK KB.

RETRACK KB Structure

Three main storage types are defined for the information / data / knowledge:

- Table: data that has no specific geographical dimension attached in the form of network topology elements (i.e. links or nodes);
- Geo-data: data that is linked to a specific geographical dimension in the form of network topology elements (i.e. links or nodes);
- Document: information and knowledge captured in the form of readable documents or blobs (e.g. video).

CONSTRUCTION OF THE RETRACK KB

For each topic a list was created of questions and indicators needed to answer the question were defined, as well as supporting parameters to derive the indicators. The complete list is can be found in RETRACK (2012). The parameters for the construction of the RETRACK KB were classified as one of three forms: table data and/or geo-data and/or document. These forms are implemented in a generic way. The main elements of the RETRACK KB are:

1. A Web Portal and content management system to encompass all RETRACK KB software elements;
2. A public and a restricted Document Repository and a full text document search engine.
3. Table data: consists of a data viewer to look at stored data in the form of tables, thematic topographic maps and charts and a data editor to update stored data;
4. Geo-data: Network data viewer and editor to handle the road, rail and water network infrastructure data and topology;

Web portal and content management system

Knowledge discovery from a knowledge base is an important issue, due to its potential to offer 'tools to effectively tackle the information overloading problem' (Rajaraman and Tan, 2002, p.669). They suggested knowledge discovery can be broadly classified as document browsing and content mining. The browsing methods usually work at the document level, employing coarse-grained mining techniques, whereas content mining works at a deeper content level, analyzing syntactic and semantic aspects of the text. The web portal for the RETRACK KB includes both an information side and a content management side (see figures 1 to 3). This web portal obtains information from a wide variety of sources, to meet the needs of variety of users.

Antunes et al. (2007, p. 187) warned that 'As software systems become bigger and more complex, software developers need to cope with a growing amount of information and knowledge'. With this in mind, the web portal used for the RETRACK Knowledge Base adopted Microsoft Sharepoint, version 2010, which has:

- Public areas and restricted areas. (The public areas are accessible to anyone, for the restricted area a password is needed.)
- A main topic navigation system that consists of tab pages on top of the client area. Each tab page provides access to a different topic. In the RETRACK KB, the topics are corridors. Selecting a corridor, the web portal provides access to all the corridor documents, table data and geo-data.
- The left hand side of each topic has a corresponding quick link navigation to provide corridor specific quick access to all documents, table-data and geo-data of the topic (corridor) in question. The quick link navigation has, amongst others, links to five topics defined in the previous chapter: Infrastructure supply and demand; Market

condition; Institutional conditions; Connectivity of the corridor; and Environmental conditions.

Document repository and document search engine

A document management system (DMS) is a computer system (or set of computer programs) used to track and store electronic documents and/or images of paper documents. The Microsoft Sharepoint version 2010 portal includes an industry standard document management system, which provides necessary storage, versioning, metadata, security, as well as indexing and retrieval capabilities. For each corridor and for each main topic a document repository is provided via the left hand navigation pane. The Sharepoint document repositories have a clear folder structure for easier navigation once the list of documents becomes very long. Registered users can upload documents to the repository. The document can be any type of file, but recommended is MS Word, Excel or PDF. During the upload the user is asked to provide (Dublin Core) metadata for the document.

Table data software

The data-viewer and data-edit capabilities of the system are provided through the existing ETIS-VIEW and ETIS_EDIT client-server software with a central SQL compliant database. The ETISplus project software is used to implement a separate RETRACK KB data viewer and data editor. This database can handle two types of data:

- Multidimensional data, e.g. zonal data (table data with one geographic dimension)
- Origin-Destination type data (table data with at least two geographic dimensions).

The RETRACK KB web application consists of two parts: RETRACK-VIEW and RETRACK-EDIT. The RETRACK-VIEW application is a web based data viewer designed to easily select a table for viewing the original data in terms of tables, charts and thematic maps. The RETRACK-EDIT application supports a Check-Out/Check-In procedure that allows for editing of data from the SQL Server in any of the supported exchange formats (MS Excel or MS Access). It allows an easy selection of tables and, if necessary, definition of a subset for one or more countries and for one or more time steps. If the user is logged in with proper credentials, he/she can check out this data. This means that each selected cell of data is flagged and the cell is locked for editing. The next step for the user is to select the exchange file type and then the user can download the data to his/her computer for editing. Once done with editing, the user can log-in again and upload the edited file for updating the central database. As part of the update process, a check of the data validity is made, to ensure consistency with the data dictionary of the collection. The system keeps a history of all the values and who has uploaded the values. This provides accountability for the database contents. For more information on the functionality see ETIS (2011d) in www.etisplus.eu.

Geo data software

The Geo data viewer and editor is based on the available on-line GIS network editing tool (InterNetter) used in the projects REORIENT and ETISplus. It allows for viewing and editing of the network topology and related attribute data, via the internet. RETRACK KB software has the ability to define a corridor, nodes and links. Considering this, the underlying Geo data is similar to the ETISplus project. The functional specifications for the software are discussed in ETIS (2011f) in www.etisplus.eu.

Web portal and content management system

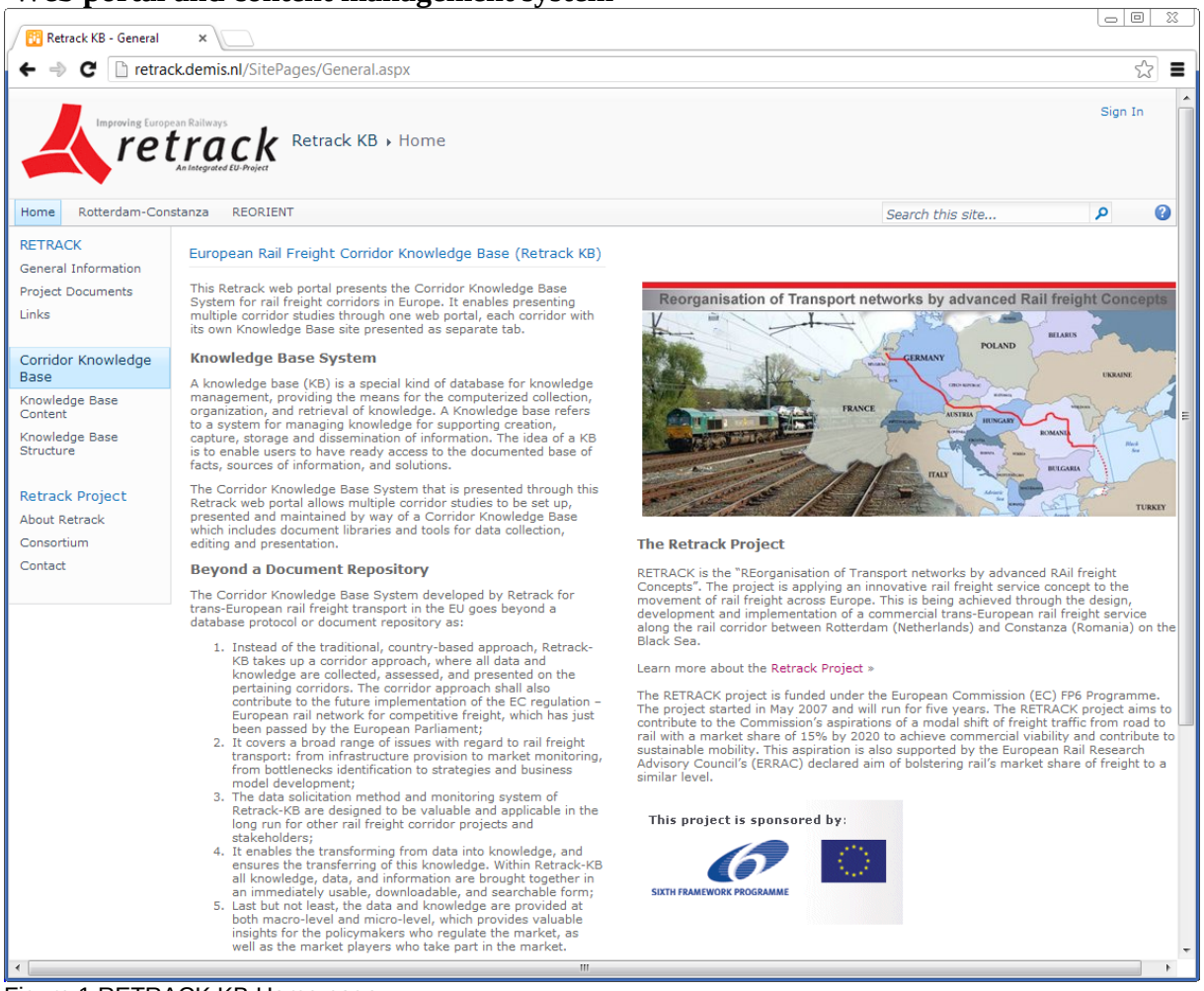

Figure 1 RETRACK KB Home page

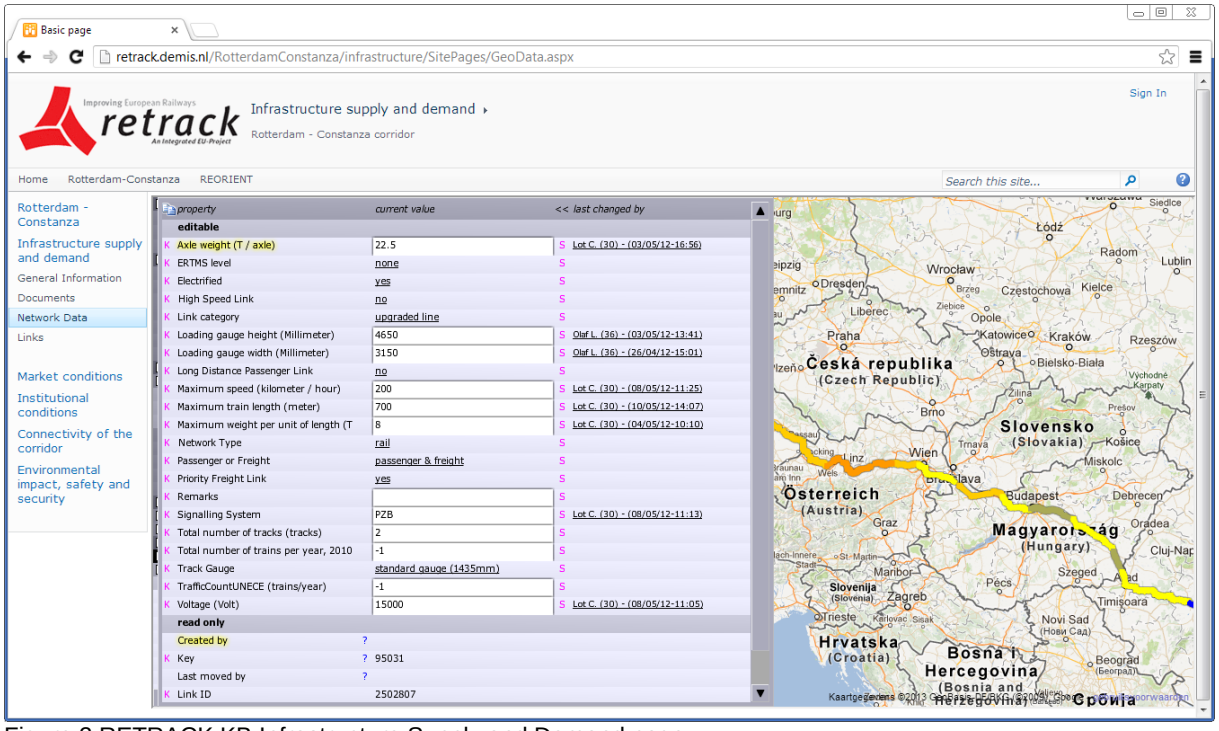

Figure 2 RETRACK KB Infrastructure Supply and Demand page

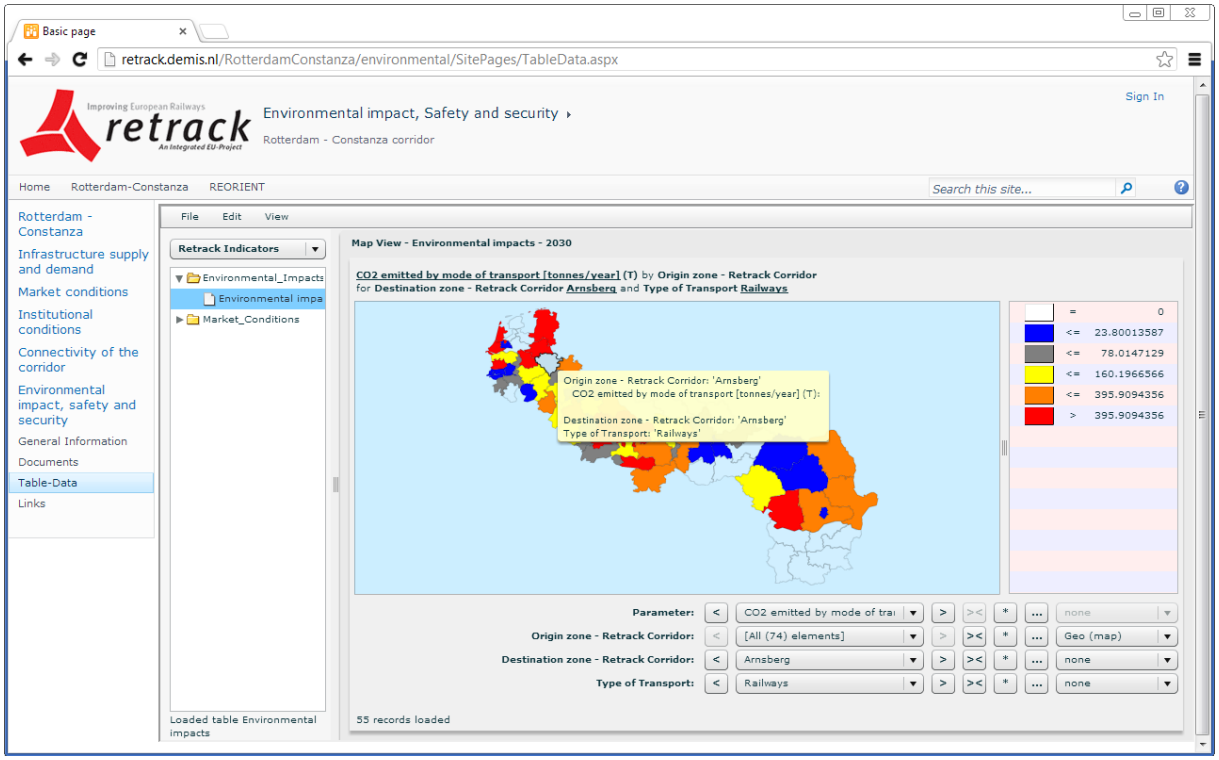

Figure 3 RETRACK KB Market Conditions page with Map view of table data on CO2 emissions

SUMMARY

The RETRACK KB is an information centre for pan-European rail freight transport with information from RETRACK and other relevant European projects such as REORIENT and ETISplus. The RETRACK KB transforms data into knowledge and transfers knowledge to relevant users. Within RETRACK KB, all knowledge, data, and information are brought together in an immediately usable, downloadable, and searchable form. Instead of the traditional country and company/organisation based approach, RETRACK KB takes a corridor approach, where all data and knowledge are collected, assessed, and presented on the pertaining corridors. These data (input) and knowledge (input and output) are systematically structured and linked with each other and are updated by gearing them to the actual development of the EU rail freight sector.

The data solicitation method and monitoring system of RETRACK KB are designed to be valuable and applicable, in the long run, for other corridors and stakeholders in the rail freight sector. The RETRACK KB covers a broad range of issues with regard to rail freight transport, from infrastructure provision to market monitoring; from bottleneck identification to strategies and business model development. Last but not least, the data and knowledge are provided at both macro-level and micro-level, which provides valuable insights for the policymakers who regulate the market, as well as for the market players. The users of RETRACK KB gain valuable contribution to the following:

- Quality improvement of various infrastructure modelling work with regard to (among others) capacity (e.g. maximum allowed speed, maximum allowed weight) and infrastructure management (e.g. scheduling, path/slot management);
- Development of business (and management) models, with integration between different modes of transport and strategies through collaboration between players;
- A platform for a web-based monitoring system of market conditions on corridors.

ACKNOWLEDGEMENT

The authors thank the European Commission for part-funding the research, under the 6th Framework Programme, within the project “RETRACK: REorganisation of Transport networks by advanced Rail freight Concepts”. The authors would also like to thank all RETRACK consortium partners and other contributors (in particular sub-contractor DEMIS.NL). The content and the opinions expressed in this article are the full responsibility of the authors.

REFERENCES

Antunes, B. Seco, N. Gomes, P. (2007). Knowledge Management using Semantic Web Technologies: An Application in Software Development, p. 187.

Bell, T. Shegda, K. M. Gilbert, M. R. Chin, K. (2010). Magic Quadrant for Enterprise Content Management, Gartner RAS Core Research Note G00206900, 16 November 2010 in http://www.gartner.com/technology/mediaproducts/

reprints/microsoft/vol14/article8/article8.html, last accessed 29.09.2011

BusinessDictionary.com (2011): Knowledge Base in
http://www.businessdictionary.com/definition/knowledge-base.html, last accessed 29.09.2011.

Castorini et al (2010). Castorini, E.; Palazzari, P.; Tofani, A.; Servillo, P., "Ontological Framework to Model Critical Infrastructures and their Interdependencies," Complexity in Engineering, 2010. COMPENG '10, pp. 91-93, 22-24 Feb. (doi: 10.1109/COMPENG.2010.25)

Cheng et al (2009). Yung-Hsiang Cheng, Li-An Yang, A Fuzzy Petri Nets approach for railway traffic control in case of abnormality: Evidence from Taiwan railway system, Expert Systems with Applications, Vol. 36, Issue 4, May, pp. 8040-8048, ISSN 0957-4174, 10.1016/j.eswa.2008.10.070. (http://www.sciencedirect.com/science/article/pii/S0957417408007732)

Chiang et al (1998). Tewei Chianga, HaiYen Hau, Hwan Ming Chiangb, Su Yun Kobb, Chao Ho Hsieh, Knowledge-based system for railway scheduling, Data & Knowledge Engineering, Volume 27, Issue 3, 1 Oct. pp. 289–312.

Chklovski, T and Gil, Y. (2005). Improving the Design of Intelligent Acquisition Interfaces for Collecting World Knowledge from Web Contributors, in The ACM Digital Library, in http://dl.acm.org/citation.cfm?id=1088622.1088630&coll=DL&dl=ACM&CFID=43421804&CFTOKEN=97407042, last accessed on 19.09.2011.

CREAM (2012) Final Report - The CREAM Project – Technical and operational innovations implemented on a European rail freight corridor, HaCon Ingenieurgesellschaft mbH, 30163 Hannover, Germany, July. Retrieved 13 August 2012 from http://www.cream-project.eu/home/index.php

Dictionnary.com (2011). Knowledge Base, in Computing Dictionary, in http://dictionary.reference.com/browse/knowledge+base, last accessed on 19.09.2011

DG Mobility and Transport (2012). TENtec Information System, in http://ec.europa.eu/transport/infrastructure/tentec/tentec_en.htm, last accessed on 05.09.2012

ETIS (2011a). ETISplus D2 Draft Specification Report - Main report , in http://www.etisplus.eu

ETIS (2011b). ETISplus D2 Annex Report – ETISplus Architecture, in http://www.etisplus.eu

ETIS (2011c). General_concept_for_ETIS_BASE_data, in http://www.etisplus.eu

ETIS (2011d). ETIS-VIEW / ETIS-EDIT User Manual, in http://www.etisplus.eu

ETIS (2011e). ETISNetter Quick Reference Guide, in http://www.etisplus.eu

ETIS (2011f). ETISNetter User Manual Annex: Overview of ETISNetter link and node attributes, in http://www.etisplus.eu

Fay (2000). A fuzzy knowledge-based system for railway traffic control, Engineering Applications of Artificial Intelligence, Volume 13, Issue 6, December, pp. 719–729

Zou Hui, Z., Xia, Z., Qiong-yan, Z., and Xun, L., (2012). "Research and design of expert system for urban transit train signal system," Intelligent Control and Automation (WCICA), 2012 10th World Congress, pp. 3983 -3986, 6-8 July, (doi: 10.1109/WCICA.2012.6359139)

Kolesa, P. (2007). Muf: Tool for Knowledge Extraction and Knowledge Base Building, The ACM Digital Library.

Lepreux, S., Kolski, C., Abed, M., (2004) "Decision support systems design as knowledge-based tool integration," Information Reuse and Integration, Proceedings of the 2004 IEEE International Conference, pp.516-521, 8-10 Nov.

Maalel, A., Mejri, L., Ben Ghezela, H., Mabrouk, H. H., (2010). System based knowledge for the help in exploitation of field data feedback in railroad transport, Ecologic Vehicles & Renewable Energies (EVER) Conference, MONACO, 25-28 March.

Merali, Y. and Davies, J. (2001). Knowledge Capture and Utilisation in Virtual Communities, The ACM Digital Library, in http://dl.acm.org/citation.cfm?id=1088622.1088630&coll=DL&dl=ACM&CFID=43421804&CFTOKEN=97407042, last accessed on 19.09.2011.

Peng He (2011), "Research on Urban Railway Knowledge-Based CBTC System," Circuits, Communications and System (PACCS), 2011 Third Pacific-Asia Conference, pp.1-3, 17-18.

Rajaraman, K and Tan, A-H. (2002). Knowledge Discovery from Texts: A Concept Frame Graph Approach, in The ACM Digital Library, p.669.

REORIENT(2012). Knowledge base at http://www.reorient.org.uk

RETRACK (2012). REorganisation of Transport networks by advanced RAil freight Concepts, http://www.retrack.eu/, last accessed on 05.09.2012.

Scholz (2002). Roland W. Scholz and Olaf Tietje, Embedded Case Study Methods: Integrating Quantitative and Qualitative Knowledge, ISBN 0-7619-1945-7

SearcCRM.com (2011). Knowledge Base, in http://searchcrm.techtarget.com/definition/knowledge-base, last accessed on 28.09.2011.

Liu et al (2008); Ziyu Liu; Lei Huang; Dongyun Xu, "Research on Semantic Retrieval System for High-Speed Railway Knowledge Based on Ontology," Computing, Communication, Control, and Management, CCCM '08, pp.303, 307, 3-4 Aug.

Wigan, M. Kukla, R. Benjamins, M. Grashoff, P. (2007). RKB: a knowledge base to support research documentation, data, GIS communications and data for a major rail freight project, ETC Conference, 17-19 October, Leeuwenhorst Conference Centre, The Netherlands.

Wikipedia (2012). Knowledge Base, in http://en.wikipedia.org/wiki/Knowledge_base, last accessed on 05.09.2012.