EUROPEAN COMMISSION

DG TREN

SIXTH FRAMEWORK PROGRAMME

THEMATIC PRIORITY 1.6

SUSTAINABLE DEVELOPMENT, GLOBAL CHANGE & ECOSYSTEMS

INTEGRATED PROJECT – CONTRACT N. TREN-06-FP6TR-SO7-69821

RETRACK

REorganization of Transport networks by advanced RAil freight Concepts

Deliverable no.	D 12.1
Title	Methodology, design and development of the RETRACK Knowledge-Base
Dissemination level	Public
Work Package	WP 12
Author(s)	Poul Grashof, Michiel Benjamins
Co-author(s)	Dewan Islam
Status (F: final, D: draft)	F-15112011
File Name	D12.1-Public-Methodology design and development of the RETRACK Knowledge-Base-Final-v1-Grashof-15112011.doc
Project Start Date and Duration	01 May 2007 - August 2012

Contents

Tables

Figures

1. Introduction

This report deals with the methodology, the design and the development of the Retrack-KB. It is deliverable D12.1 of WP 12 of the RETRACK project.

The Retrack knowledge base (Retrack-KB) will be developed in WP12. The knowledge base developed by Retrack-KB for trans-European rail freight transport in the EU goes beyond a database protocol or document repository. First, Retrack-KB enables the transforming from data into knowledge, and ensures the transferring of this knowledge. Within Retrack-KB all knowledge, data, and information are brought together in an immediately usable, downloadable, and searchable form.

Second, instead of the traditional, country-based approach, Retrack-KB takes up 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 itself to the actual development of rail freight sector in the EU. This corridor approach shall also contribute to the future implementation of the EC regulation – European rail network for competitive freight, which has just been passed by the European Parliament.

Third, the data solicitation method and monitoring system of Retrack-KB are designed to be valuable and applicable in the long run for other projects and stakeholders in the rail freight sector.

Fourth, Retrack-KB covers a broad range of issues with regard to rail freight transport: from infrastructure provision to market monitoring; from bottlenecks 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 the market players who take part in the market.

The provision of Retrack-KB further contributes to the following aspects:
a. Quality improvement of various infrastructure modeling work with regard to, among others, capacity (e.g. max. allowed speed, max. allowed weight) and infrastructure management (e.g. scheduling, slot management);
b. Development of business (and management) models, in which integration between different modes of transport and strategies through collaboration between players are assessed;
c. A platform for web-based monitoring system for monitoring the market conditions on corridors.
d. Improvement of market opening enabled by more transparent monitoring of market and framework conditions.

The following chapters will deal with the organization in WP12 of the development of the Retrack-KB; the methodology of the Retrack-KB; the design of the Retrack-KB; and the proposed implementation of the content and structure and the tools to be used.

2. WP12 Development of the RETRACK-KB

The development of the knowledge base in Retrack WP12 consists of five consecutive tasks:

WP 12.1	RETRACK KB - methodology, design, development and maintenance
WP 12.2	RETRACK KB - traffic and infrastructure on the RETRACK corridor[1]
WP 12.3	RETRACK KB - existing services on the RETRACK corridor
WP12.4	RETRACK KB - market condition on the RETRACK corridor
WP 12.5	RETRACK KB - governance and legislation on the RETRACK corridor

Task 1 deals with the methodology -, the design -, the development - and maintenance of the Retrack-KB. The lead for task 1 is with TNO, other partners involved are NEA, TOI and UNEW. In this task the Retrack-KB software is delivered by the subcontractor Demis.

Description of Task 1

Task 1 RETRACK Knowledge Base – methodology, design, development and maintenance (led by TNO, supported by NEA, TOI, UNEW, and Demis)

Subtasks:
1.1   Design the generic technical platform for RETRACK-KB: determine what types of data of a trans-European rail freight will be included and how to present these data; and develop system architecture for building such technical platform (TNO, NEA, TOI, UNEW, Demis)[2];

1.2   Build the web-based platform and dynamic thematic mapping tools for RETRACK-KB. Both geospatial data about traffic and infrastructure (Task 2) and geospatial data concerning the market and framework conditions will be included (Task 3, 4, 5) (NEA, Demis);

1.3   Design uniform procedures of carrying out surveys, and standardise the formats of questionnaires (Demis, TNO);

1.4   Define procedures and requirements for data feeding (data receiving and submitting) (TNO, NEA, TOI, NEA, Demis);

1.5   Conduct data feeding into RETRACK-KB system, update the data, and host this platform (TNO, NEA, TOI, NEA, Demis);

---

[1]   RETRACK corridor refers to several route alternatives linking west hub on the one side and east hub on the other side; routes linking two hubs on only one side are not considered as RETRACK corridor.
[2]   The partners mentioned here are the major partners assigned on this specific sub-task, underlined means lead partner for the subtask.

1.6   Keep regular dialogues with TENTec Team for data interchanging, interface placing, and exploring cooperation towards a common structure; (NEA, Demis, TNO)

1.7   Consider the possibilities of maintaining RETRACK-KB in the long run (NEA, Demis, TNO).

1.8   Draw a road map regarding in which way WP12 and its findings shall contribute to the corridor development in Europe, particularly to the nine priority routes specified in the EC Regulation *European Rail Network for Competitive Freight.* The answer will focus on the future uses of the system architecture, the technical platform, mapping tools, the standardised survey procedures and formats (TNO, NEA, Demis).

1.9   Producing Deliverable 12.1 (all).

As the design of the system is part of the work the terms of reference for the software is seen as flexible and can be changed during the course of the project. For task 1.2 Demis is the lead responsible organisation. This task deals with building the RETRACK KB software as laid down in the terms of reference of WP12. In the remaining chapters of this document the specification of the RETRACK KB software is given.

3. Methodology of the Retrack KB development

The building of RETRACK-KB is made in three consecutive stages: the stage of forming philosophy, the stage of design, and the stage of construct (see below figure).
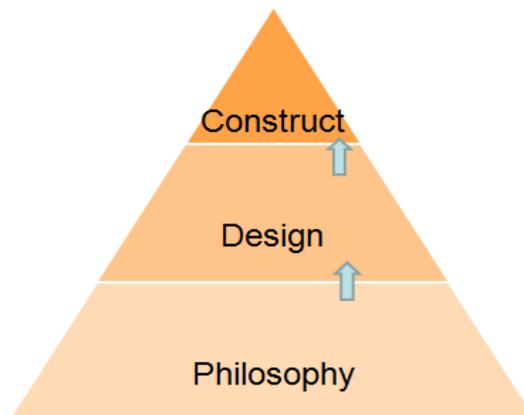

Figure 1 Three stages of building RETRACK-KB

The philosophy is formed along four perspectives:

1) to ensure the system architecture and the technical platform for supporting RETRACK-KB to be built with a generic nature, which will allow them being applicable for the studying of European rail freight corridors in general;
2) to ensure uniform procedures of carrying out surveys and standardised structured questionnaires so that they can be used for the studying of any other European rail freight corridors;
3) to keep regular dialogues with the TENTec Team in order to interchange relevant data into each other's systems for completion, harmonised interfaces between the two systems, and explore possibilities of merging into one structure;
4) to draw a road map in the final phase of Work Package 12 with regard to in which way WP12 and its findings shall contribute to the future corridor development in Europe, particularly to the nine priority routes specified in the EC Regulation *European Rail Network for Competitive Freight.* The answer will focus on several elements including the future use of the system architecture and the technical platform of RETRACK-KB; the future use of standardised procedures of survey conduction and standardised questionnaires; the future use of dynamic thematic mapping which presents and visualises the geospatial data (e.g. transport, economic, technical, operational, institutional), and allows different levels of analyses on the basis of these data; and the maintenance of RETRACK-KB after the execution period of WP12 ends.

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

3.1 What is a knowledge base

Dictionary.com, 2011 enlisted the following definitions of knowledge
1) acquaintance with facts, truths, or principles, as from study or investigation; general erudition: knowledge of many things.
2) familiarity or conversance, as with a particular subject or branch of learning: A knowledge of accounting was necessary for the job.
3) acquaintance or familiarity gained by sight, experience, or report: a knowledge of human nature.
4) the fact or state of knowing; the perception of fact or truth; clear and certain mental apprehension.
5) awareness, as of a fact or circumstance: He had knowledge of her good fortune.
6) something that is or may be known; information: He sought knowledge of her activities.
7) the body of truths or facts accumulated in the course of time.
8) the sum of what is known: Knowledge of the true situation is limited.
9) Archaic . sexual intercourse. Compare carnal knowledge.
10) creating, involving, using, or disseminating special knowledge or information: A computer expert can always find a good job in the knowledge industry.
11) to one's knowledge, according to the information available to one: To my knowledge he hasn't been here before.

The synonyms of knowledge are: information, understanding, discernment, comprehension, erudition, scholarship (Dictionary.com, 2011).

In Dictionary.com (2011) a knowledge base is defined as 'A collection of knowledge expressed using some formal knowledge representation language. A knowledge base forms part of a knowledge-based system (KBS)' (Dictionary.com 1994-10-19).

In relation to information technology (IT), a knowledge base is a machine-readable resource for the dissemination of information, generally online or with the capacity to be put online. An integral component of knowledge management systems, a knowledge base is used to optimize information collection, organization, and retrieval (SearchCRM.Com, 2011).

According to Business dictionary.com (2011) a knowledge base is an organized repository of knowledge (in a computer system or an organization) consisting of concepts, data, objectives, requirements, rules, and specifications. Its form depends on whether it supports an (1) artificial intelligence or expert system based retrieval, or (2) human based retrieval. In the first case, it takes the form of data, design constructs, couplings, and linkages incorporated in a software. In the second case, it takes the form of physical documents and textual information.

According to [Wikipedia](#) a **knowledge base** (KB) is a special kind of database for knowledge management, providing the means for the computerized collection, organization, and retrieval of knowledge. A Knowledge base refers to a system for managing knowledge for supporting creation, capture, storage and dissemination of information. A knowledge base is generally generated via or through an IT based program/department or section. The idea of a KB is to enable users to have ready access to the documented base of facts, sources of information, and solutions.

Knowledge Management (KM) The literature on knowledge management highlights issues of fit between IT-based systems for knowledge management and the socially situated leveraging of knowledge assets by organisations' (Merali and Davies, 2001 p. 92). KM comprises a range of strategies and practices used in an organization to identify, create, represent, distribute, and enable adoption of insights and experiences. Such insights and experiences comprise knowledge, either embodied in individuals or embedded in organizational processes or practice. In the classic meaning the scope is provided by the organization, however with the appearance of the internet this can also be a community of people working in the same field.

From the above discussion we can say that a KB can be any combination of the following:
- Document based i.e. any technology that permits creation/management/sharing of formatted documents such as Lotus Notes, SharePoint, web, distributed databases etc.
- Ontology/Taxonomy based: these are similar to document technologies in the sense that a system of terminologies (i.e. ontology) are used to summarize the document e.g. Author, Subject, Organization etc. as in DAML & other XML based ontologies
- Based on Artificial Intelligence technologies which use a customized representation scheme to represent the problem domain.
- Provide network maps of the organization showing the flow of communication between entities and individuals
- Increasingly social computing tools are being deployed to provide a more organic approach to creation of a KM system.

A KB offers integrated services to deploy KM instruments for networks of participants, i.e. active knowledge workers, in knowledge-intensive business processes along the entire knowledge life cycle. KB can be used for a wide range of cooperative, collaborative, adhocracy and hierarchy communities, virtual organizations, societies and other virtual networks, to manage media contents; activities, interactions and work-flows purposes; projects; works, networks, departments, privileges, roles, participants and other active users in order to extract and generate new knowledge and to enhance, leverage and transfer in new outcomes of knowledge providing new services using new formats and interfaces and different communication channels.

3.2 Specification of a rail-freight knowledge base

The RETRACK KB (RKB) needs to meet operational requirements in the RETRACK FP6 project on railway corridor analysis: this demands a wide range of new and existing data and documents to be brought together in a immediately usable, downloadable and searchable form, plus 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 and restricted timetable of a short work package. It also needs to be achieved with a limited effort to be built, deployed populated, managed documented and delivered.

The RKB has to telescope the time and effort overheads of making complex transport datasets usable. This will be illustrated by the range of analysis, access, time series diagrams, graphs, thematic mapping OD matrix displays, desire-line mappings and many other tools normally only accessible after the time consuming heavy lifting of securing a dataset, setting it up, understanding it, loading it into a specialised analysis system and producing special maps and diagrams to communicate the results.

The RKB needs all these tools inbuilt 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. This is one area where the productivity issues really benefit, and the reductions in user overheads in making use of data sets are cut back to an absolute minimum.

The RKB 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 what already exists – but also to raise the productivity 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. Other key projects with knowledge base components are the FP6 projects WORLDNET (www.worldnetproject.eu) , ROADIDEA (www.roadidea.eu) and the FP7 project ETIS-PLUS (www.etisplus.eu). The lesson learnt and software tools built are all feeding into the design and implementation of the RETRACK KB.

Furthermore, the GIS tools needed to be used to manage and standardise the input of many partners & subcontractors collecting data – so a vector oriented network editor and data collection support system is needed and implemented. A multi layered security system is needed for all aspects, as much of such data is available under limited conditions (mainly due to unresolved final handover to the Commission, or mixed commercial and other interests, plus the similar constraints on the fresh data collected by Reorient team itself.

As the RKB needs to be designed to actively support corridor projects, as well as to provide interactive Knowledge Transfer and needs dissemination to be achieved with markedly lower ongoing costs and support overheads, this ambitious set of capabilities requires a careful system design to allow it to be achieved rapidly and with commercial standards of supportability and solidity.

Any such approach needs three things:
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 as to how it is to make research more productive and technology transfer more effective and economic

The specification of a generic rail freight knowledge base cover Documents, Data, Contributions, Presentation, Overall client requirements, overall system requirements and some of the major points are listed below

DOCUMENT
1) Versioning
2) BLOB storage (video, zipped data, speech whatever)
3) Import/export of Word, Text, Excel, PDF formats
4) Large scale capacity handling
5) Customisable user interface
6) High quality security layering
7) Email services and repository
8) Internal and outbound Weblinking
9) Metadata management including revision and updating

DATA (GEO-DATA or TABLE)
1) Federated multiple data set editor and publisher
2) Crosslinking to document repositories and or web
3) Data cube handling
4) Data exploration (tabulations, charting etc)
5) Selective downloads
6) Thematic mapping
7) Import of data as text, MS access data cubes and other formats
8) Ability to handle and display vector and geospatial data
9) Direct weblinkages within the system and across it

At various levels will all be operational in RKB in the coming months – not just at the very end. Several of these aspects are delivered by the ETIS-PLUS products ETIS-VIEW and ETIS_EDIT and are fully supported by Demis (a commercial support company)

CONTRIBUTION LAYER
1) Document repository for uploads of documents and blobs
2) Uploading of datasets in ETIS-EDIT
3) Metadata editing for documents, blobs and data sets

PRESENTATION LAYER
1) Dynamic thematic mapping of spatial data: handling of standard geospatial data standards such as OPEN-GIS
2) Full data exploration support in ETIS-VIEW

OVERALL CLIENT SIDE REQUIREMENTS
1) W3C compliant browser with no extras
2) Flash client permitted solely for Presentation layer
3) Strong security
4) Full communications (email and wiki support for response to materials located in the repositories
5) Well tested and simple user interface

OVERALL SYSTEM REQUIREMENTS

1) All back end engines to be fully supported by highly reputable companies
2) Based on existing server tools from REORIENT, WORLDNET, ROADIDEA and EDIT-PLUS

To illustrate some of above mentioned aspects, there are some examples below of the REORIENT KB interactive network editor and data manager (Internetter)(for a node in Budapest), interactive menu driven access to the NEA O-D Data (here to Bucharest).

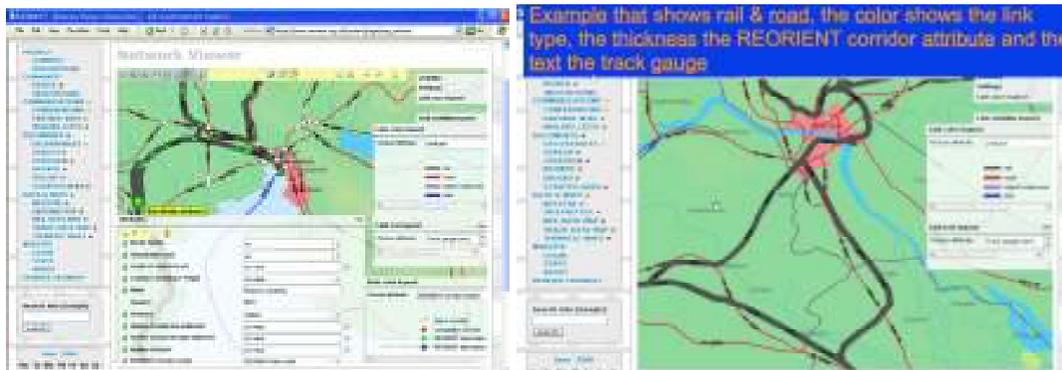Figure 2 Internetter examples as delivered inside the REORIENT KB website

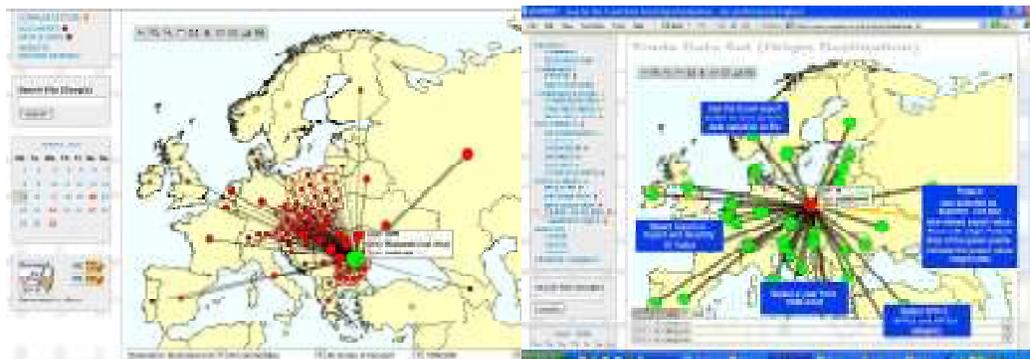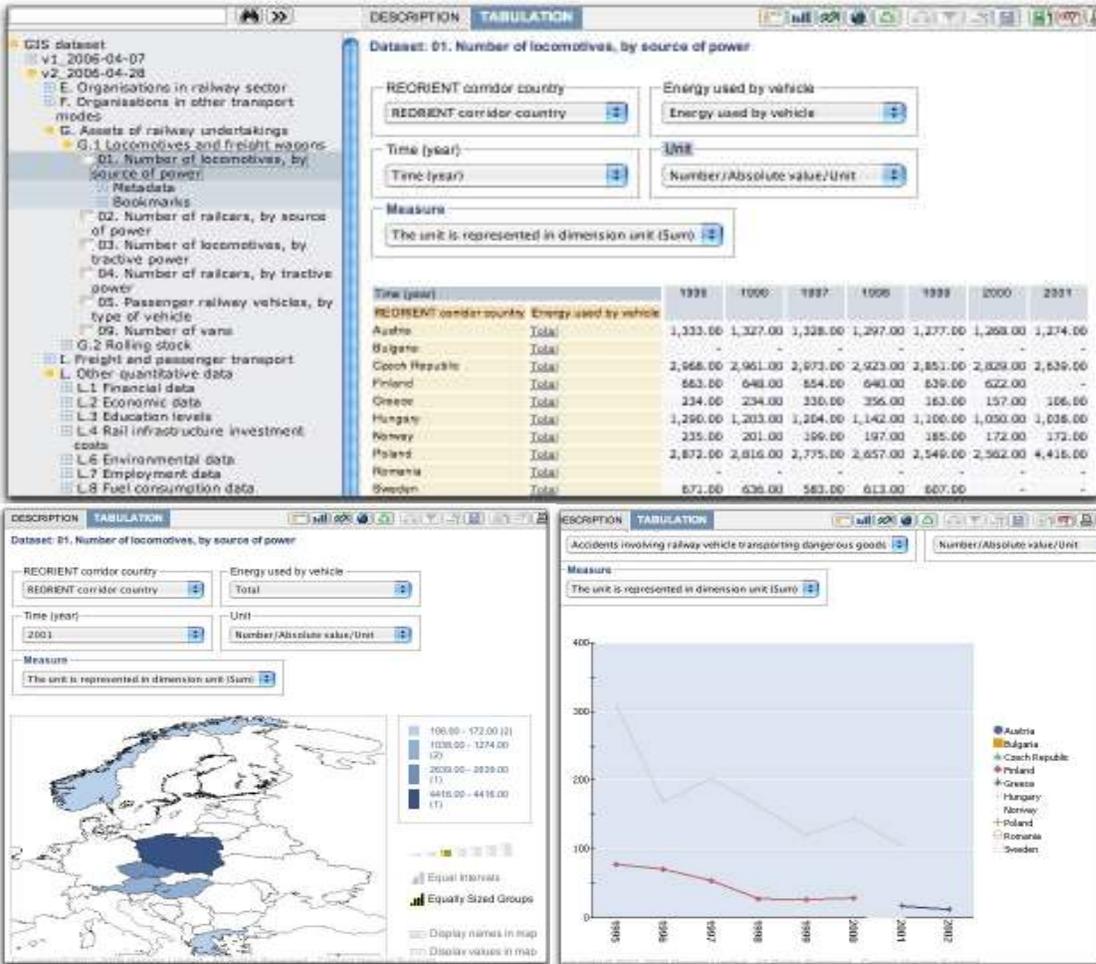*Figure 3 Thematic map examples (Demis mapserver-linked) in the REORIENT KB website*

Figure 4 Some Nesstar examples of the REORIENT KB

4. Design of the RETRACK KB

4.1 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'. Unlike 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.
5) Direct multiple contributors to validate and evaluate previously entered statements.

Considering the necessity of the RETRACK Knowledge Base, the themes and structure of the proposed content are shown in the table below and on the next pages. It shows that the KB is built to capture knowledge on corridors for themes such as:
- Infrastructure supply and demand;
- Market condition;
- Institutional conditions;
- Connectivity of the corridor;
- Environmental conditions.

The definition of the contents of the RETRACK corridor KB will remain under development as the project progresses. An overview of the current proposed content that needs to be accommodated in the KB is shown in the table below and on the next pages. The table is organised in topics, for each a number of questions are defined, a definition of the indicators is given and how the indicators are constructed based on parameters.

Table 4-1 Overview of the RETRACK-KB content and structure

Topic	Questions	Indicators	Parameters
1. Infrastructure supply and demand	What is the present and future status of the rail infrastructure and what is the level of interoperability between the rail lines?	% of length of track with compatible power supply (per corridor segment)	Voltage
			Length of sections (km)
		% of length of track with compatible signalling (per corridor segment)	ERTMS-level
			National signalling system
			Length of sections (km)

Topic	Questions	Indicators	Parameters
		% of length of track with compatible track gauge (per corridor segment)	Gauge: 1435/1520
			Length of sections (km)
		% of length of track with compatible loading gauge (per corridor segment)	Loading gauge
			Length of sections (km),
		% of length of track with compatible axle weight limit (per corridor segment)	Axle weight limits (per link)
			Length of sections (km)
		% of dedicated freight links (per corridor segment)	Freight per link
			Passenger per link
			Mixed per link
		Maximum train length on corridor (per corridor segment)	Maximum train length per link
		% of track per maximum speed class	Maximum speed per link
		% of single track	Number of tracks per link
		% of track electrified	Electrification per link
		% of new track planned or under construction	Forecasted date of opening new situation per link
	What is the technical and operational capacity of the terminals?	Available terminals on the corridor Capacity per terminal	Terminal names
			Location
			Type of node: rail-rail, rail-road, rail-road-iww, sea-rail-road, sea-rail
			Lay-out
			Track length in the terminal
			Connecting modes per terminal
	Which regons are connected by the corridor?	Regions (NUTS-3 level) where individual rail tracks of the corridor are shown	Regions (NUTS-3 level) where individual rail tracks of the corridor are shown
		Regions (NUTS-2 level), where local markets are served by the	Regions (NUTS-2 level), where local markets are served by the corridor.

Topic	Questions	Indicators	Parameters
		corridor.	
		Regions where international markets are servied by the corridor	Regions where international markets are servied by the corridor
	What is the traffic density along the corridor?	Train-km on the corridor (per corridor segment)	Trains per link
		Ton-km on the corridor (per corridor segment)	Tons per link, OD flows.
	To what extend is the capacity of the corridor being used?	Congestion of terminals	
		Utilisation of capacity of the rail links	Number of trains on the links; capacity of the links
		Congestion on nodes (Marshalling, turning, curves, …)	
2. Market conditions	Who are the leading market players?	Overview of the leading market players	RUs, ROs, TOs, LSPs, shipping lines
	What types of (intermodal) services are being provided?	Services on the corridor	Origin terminal per service
			Destination terminal per service
			Duration per service
			Frequency per service
			Supplier information
			Service expansion potential
	What business models and strategies are there?	Types of services by business models	Intermodal /Block train /Single wagon /Multiple wagon groups.
			Type of advanced models to be developed
	What's the competitiveness of rail on the corridor compared to other modes (inter-modal competition)?	Type of models being used	
		Modal split per OD	Traffic flow by mode per OD
		Ton-km per mode per OD	Commodity flow by mode per OD, distance per OD per mode
		Transport costs per mode	Costs per mode per OD
3. Institutional conditions	What are the main institutional actors / policy making bodies?	List of the actors	
	What policies at the EU level are applied?	Overview of most important rail (liberalisation) legislations at the EU level	
		Status of legislation implementation and enforcement	
	What policies at national level are applied?	Initiatives taken at the national levels or between the member states	MoU/LoT signed between the member states
	What are the regulatory conditions and bottlenecks on the corridor?	Implication of lack of coordination between different corridor initiatives (e.g.TEN-T, ERTMS, etc.) with regard to aspects (e.g. investment, management plans) on the flow on the RETRACK corridor.	

Topic	Questions	Indicators	Parameters
		Delays of implementation of the EC legislation at certain states	
		Investment status and completion perspective National institutional structure, policies, or initiatives that negatively affect corridor development.	
	What are their impacts on the functioning of the corridor?	Degree of opening of terminals	
		Cross acceptance of safety certification of rolling stocks	
4. Connectivity of the RETRACK corridor	What is the physical connection between RETRACK and the international corridors?	Number of RETRACK nodes linked with nodes of other corridors	Corridor indicator per node
	To what degree do international corridors interact with RETRACK?	Number of services connecting the relevant nodes to other corridors	Services per node
5. Environmental impact	What is the CO2 emission savings from the rail traffic on the RETRACK corridor, as compared to road traffic?	% reduction of CO2 emission due to modal split from road to rail on the corridor	Present modal split and CO2 emission per mode on the corridor
			Future modal split and CO2 emission on the corridor
		% reduction of CO2 emission due to modal split from road to rail on the EU rail network	Present modal split and CO2 emission per mode on the EU rail network
			Future modal split and CO2 emission per mode on the EU rail network
	What is the PM10 and NOx emission savings from the rail traffic on the RETRACK corridor, as compared to road traffic?	% reduction of PM10 and NOx emissions due to modal split from road to rail on the corridor and EU network	Present modal split and PM10/Nox emissions per mode on the corridor
			Future modal split and PM10/Nox emissions per mode on the corridor
	What is the noise emission savings from the rail traffic on the RETRACK corridor, as compared to road traffic?	% reduction of noise emision due to modal split from road to rail on the corridor and EU network	Future modal split and noise emissions per mode on the corridor

4.2 Retrack KB structure

In the above mentioned content as specified within the RETRACK project three main storage types are defined for the information / data / knowledge:

- Document: information and knowledge captured in the form of readable documents or blobs (e.g. video).
- 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)

In the next chapters the software elements used to implement the Retrack KB content and structure are explained.

5. Main software elements of the Retrack KB

As we can see from the table on the previous pages the presentation form of all the parameters will be as geo-data and/or table and/or document. These three forms will be implemented in a generic way in the Retrack KB. Hence the main elements of the Retrack KB will be:

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
5. **Survey system** to define surveys; conduct the survey on-line and view/export the survey results
6. Support **data exchange** to external systems such as TENTec and Gisco

For each of the RETRACK KB software elements a chapter with specifications is provided in the rest of this document. The required approach followed by Demis is the use of industry standard tools for general ('generic') parts of the KB (portal, document management and searching). For the table data and Geo-data specific web applications need to be developed as the required functionality is not available in industry standard software systems. The selection of the software elements is described in the remaining sections of the chapter.

5.1 Selection of the software elements of the Retrack KB

The elements under 1. 2. and 5. are often bundled together in different content management systems (CMS). Nowadays there is plethora of different CMS, see for example http://en.wikipedia.org/wiki/List_of_content_management_systems.

A comparison of some of these CMS is given by Bell et all, 2010 in http://www.gartner.com/technology/media-products/reprints/microsoft/vol14/article8/article8.html.

In this article the focus is on enterprise content management systems. Strictly speaking the RETRACK KB can be better described as a web content management system. Web content management (WCM) is being transformed as it aligns more with Web channel delivery and extends further into analytics, site optimization, social communities and interactivity, rich media and digital asset management (DAM), and personalization.

Bell et al, 2011 see the market splitting into four main areas of buyer focus and vendor influence: transactional content management, social content management, online channel optimization and content management as infrastructure.

They use a "Magic Quadrant" to assess which ECM vendors have the functional capabilities and vision to support the RETRACK KB requirements, and thus to determine which CMS would therefore suitable.
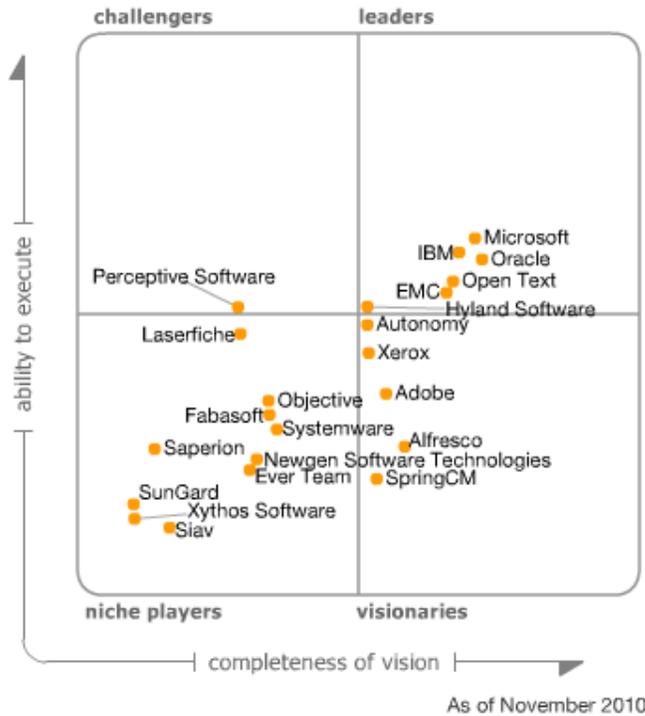

Figure 5 Assessment of content management systems

In the REORIENT EC project different elements were combined to create a CMS based on PHP, including the TERATEXT document management system and search engine; NESSTAR for multidimensional data and Marratech for videoconferencing. On the other hand in the Worldnet, ROADIDEA and ETISPLUS projects the content management was realised using an industry standard solution, namely Microsoft Sharepoint. As can be seen from the figure above the Microsoft solution is amongst the leaders in the field with a high ability to execute and with a complete vision. That doesn't make it necessarily the best option, but in combination with the experience gained in the Worldnet, ROADIDEA and ETISPLUS projects it is advised to use this option. MS Sharepoint has been on the market for a while and has different versions. The EC mentioned projects all used version 3, but the latest version available is version 2010, which Demis recommends for the portal elements under 1. 2. and 5.

Table data software

For the table data software there are some vendors that have solutions covering part of the requirements, such as the NESSTAR system used in REORIENT. However since the ETISPLUS EC project are better solution was developed in house by Demis which will also be used for the RETRACK KB. The major advantage of the Demis system is that it is license free and the source code can be shared, whereas NESSTAR requires a annual license fee and source code is not available. Also the Demis software is specifically developed to allow easy check-out / editing and check-in of the data. This is missing in NESSTAR.

Geo data software

For the GEO data software there are some vendors that have solutions covering part of the requirements, such as the open source system used in www.streetmap.org. However since the REORIENT EC project a better solution was developed in house by Demis. It was also used in ROADIDEA, WORLDNET and ETISPLUS. And not to forget the Dutch bicycle path network is completely covered using hundreds of volunteers using the same Demis software InterNetter. This will also be used for the RETRACK KB with the code name RETRACKNetter. The major advantage of the Demis system is that it can be easily customized and has an optimized user interface for easy editing of network topology and link and node data. The latter is based on the experience and feedback of hundreds of users executing more than 6 million edit actions on the bicycle network data in the Netherlands.

Data exchange software

Here the requirements still need to be decided in cooperation with the EC, TEN-T and GISCO. Therefore no selection of software has been made yet.

6. Web portal and content management system

Knowledge discovery from a knowledge base is an important issue due to its potential of offering 'tools to effectively tackle the information overloading problem' (Rajaraman and Tan, 2002, p.669). They suggested that this is knowledge discovery can be broadly classified: 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 shall include both an information side and a content management side. The Retrack KB will be content-centric. This type of portal definition sees the function as one of obtaining information from a wide variety of sources and displaying that content to users in a way that is based upon the information needs of the target audience.

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'. Keeping this on board, the web portal to be used for the RETRACK Knowledge Base will be Microsoft Sharepoint version 2010. This will have:

- 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 Retrack the topics will be corridor(s). After selecting the the corridor a one-stop portal must provide access to all the corridor documents, table data and geo-data.
- The left hand side of each topic will have 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 will have amongst others links to the topics defined in the previous chapter: Infrastructure supply and demand; Market condition; Institutional conditions; Connectivity of the corridor; and Environmental conditions.
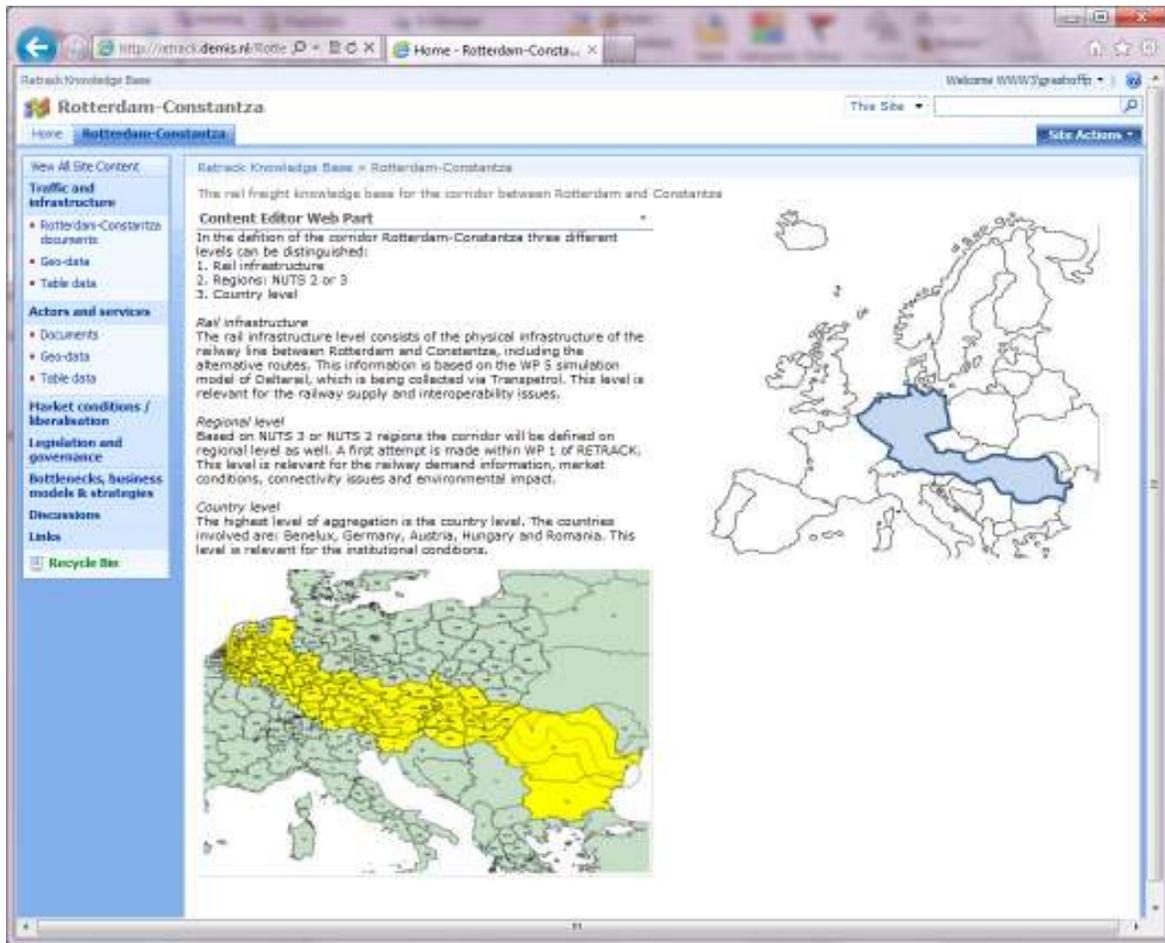
Figure 6 Example of the proposed corridor tab page layout

7. 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 will be used to provide the necessary storage, versioning, metadata, security, as well as indexing and retrieval capabilities. For each corridor and for each main topic a document repository will be provided via the left hand navigation pane. The sharepoint document repositories will 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 any type of file, but recommended is MS Word, Excel or PDF. During the upload the user will be asked to provide metadata for the document. For metadata of the document repository the Dublin Core metadata set as implemented in MS Sharepoint will be used. Metadata can be used in the search for documents. For a description of how to work with the repository see section 7.2. The next section deals with the search engines proposed.

7.1 Document search engine

It is importants that documents are easy to find in the knowledge base. A three-tier approach is followed here:

1. Document repositories are organised in logical folders, as if the user is navigating with the windows explorer;
2. The built-in full text search of the Sharepoint system allows searching for the entire site; for a corridor sub-site or even for a specific document repository inside the corridor sub-site. This way as an user you can choose the appropriate scope of a document search;
3. For a more comprehensive full text search we propose an industry standard tool Google site search.

One of the main requirements is to have a full text search option. We propose that in every page the left hand navigation panel will include a Google site search function. This in addition to the standard Sharepoint document search function

The Sharepoint search service is a tool allowing to index and classify static documents accessible on web sites in various formats, such as for example HTML, Adobe PDF and MS Word. Once indexed, documents can be retrieved through various means.

In addition to the Sharepoint search service we propose to use an industry standard web site search tool, the Google Site Search. With Google Site Search, we can put a Google search box on your website and enable prospective visitors to find information instantly. Google site search supports On-Demand Indexing. This allows a select number of pages of the RETRACK KB web site to be indexed quickly. That way the latest versions of these pages will be included in the Google Site Search results within 24 hours. Google site search has more functionality than the standard Sharepoint search service:

1. Enhanced Index Coverage
2. Synonyms

3. Date biasing
4. Top results biasing

1. Enhanced Index Coverage
Enhanced Index Coverage ensures comprehensive search results on any website by crawling and indexing more content, even for pages deep within a site. Please note that using Google Site Search won't influence the ranking of the RETRACK KB pages on Google.com search results or how many pages are indexed.

2. Synonyms
This lets visitors find documents with related terms without having to issue multiple queries. Demis will upload synonym dictionaries specific to the RETRACK KB contents. For example, (a search for [railway car] can include [railway cars]), because railway cars has been provided as a synonym for railway car.

3. Date Biasing
Administrators can influence the search ranking based on the age of the documents (e.g., a recent document is weighted more heavily than older documents). This feature can also be disabled and administrators can decide the level of influence for their sites (low, medium, high or maximum). We propose to have a high bias for the most recent documents.

4. Top Results Biasing
Google Site Search allows you to target the top search results from specific sections of your website (such as for example policy brochures, programmer descriptions or project deliverables).

7.2 Working with documents in the Retrack KB

Intranet versus Internet site

The RETRACK KB site will be set up to include both an internet site and an intranet site. At first glance there is not much difference between the sites: the content of both sites is the same and making changes to one site will automatically and immediately be available to the other site. Also the same username and password are used.

The main difference between the sites is in the fact that the intranet site offers some additional, very useful functionality for users of MS Internet Explorer. This additional functionality is related to document libraries and include Multiple Document Upload, Documents Explorer View and Edit in Datasheet format. If you think you will not use this additional functionality, you can best keep working with the internet site as it allows you to use any browser you like.

The following technical notes should be taken into consideration by the user when using the intranet site (these do not apply to the internet site):

- The use of the intranet site does requires the availability of a couple of dynamic link libraries (DLL's). The webbrowser will ask for permission from the user for running these DLL's, which should be allowed in order to be able to run the intranet site properly. If you do not have the required local administrative rights to allow the DLL's to run, please contact your system administrator to resolve this issue;

- There are known limitations with using Mozilla Firefox for this intranet site that cannot be resolved. It is therefore recommended to use (the latest version of) the Internet Explorer webbrowser when working with the intranet site;
- The best performance is obtained when using Microsoft Office 2007, because herein the needed DLL's are already installed.

Multiple document upload

When working in the intranet environment, uploading documents can be done for multiple documents at a time. When selecting the usual *Upload* option, there is an additional *Upload Multiple Documents* option.
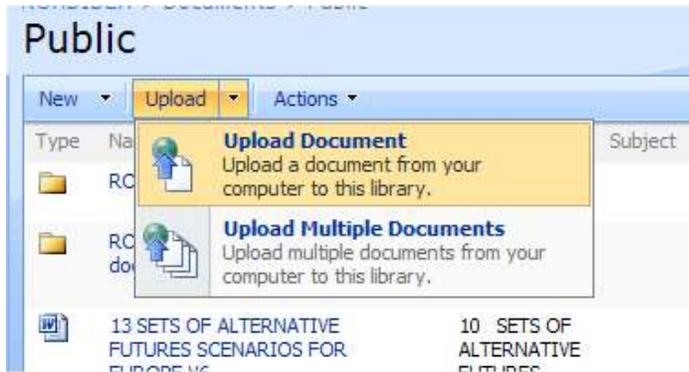

Figure 7 Example screenshot for multiple document upload menu item

Just like when you select *Upload Document* on the internet*,* the intranet window supplies you with a new option **Upload Multiple Files***.* This opens an Explorer-like view where you can browse and select a great number of documents from a single folder to be uploaded at once. Using this option however does not directly allow for adding metadata to the document(s). This can be done by using the [[Editing in Datasheet]] format, which allows for easy editing of document properties (i.e. metadata) of multiple documents.
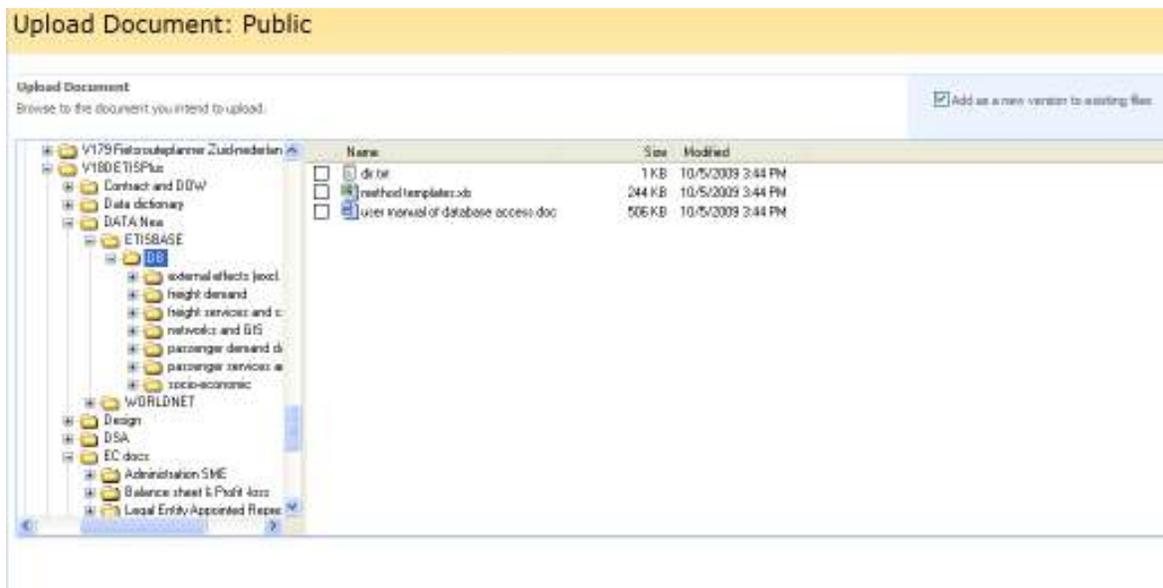
Figure 8 Example screenshot for multiple document upload explorer

Document explorer view

In the intranet environment, it is also possible to use a view on Document Libraries and Lists that is similar to Internet Explorer. This view allows for creating new folders and for drag and dropping / copy and pasting of files. In general however this view tends to be rather slow in Sharepoint, so for uploading documents this probably is **not the fastest way**. You access this menu option via the Actions button, then select Open with Windows Explorer.
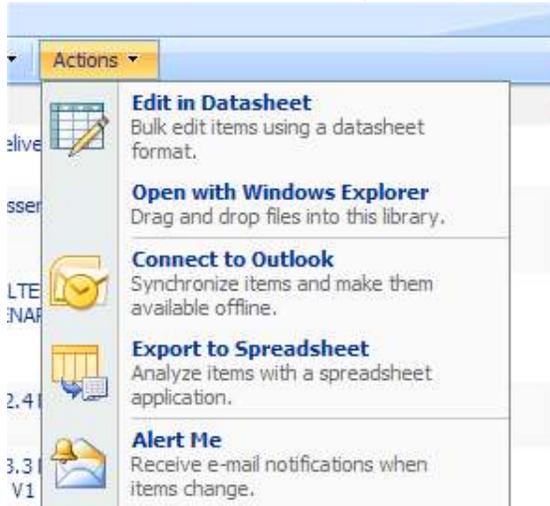

Figure 9 Example screenshot for windows explorer view menu item

Whenever using the Explorer View however, an additional new connection is automatically added to your computer in my Network Places whenever you use the Explorer View. This allows you to browse the RETRACK KB Document Directories in your Windows Explorer. This way you can move or copy files to the RETRACK KB site.

Editing metadata in a datasheet

A very useful option for multiple edit, including copy, cut and paste options is possible on the intranet. Editing document properties (Dublin Core metadata) is usually done by selecting the Edit Properties option for a document. This presents a complete list of the properties for this specific document. When working in the RETRACK KB intranet environment however, it will be possible to use the editing of properties in a datasheet. This allows for easy editing of properties for a larger number of documents instead of having to edit the properties of one document at a time. This option can be selected from the documents list by selecting Actions>>Edit In Datasheet. This will render an Excel-like worksheet that can easily be filled out, also by copying values to multiple fields at a time. In combination with the Multiple Document Upload this allows for easy uploading and properties editing of a large amount of documents.
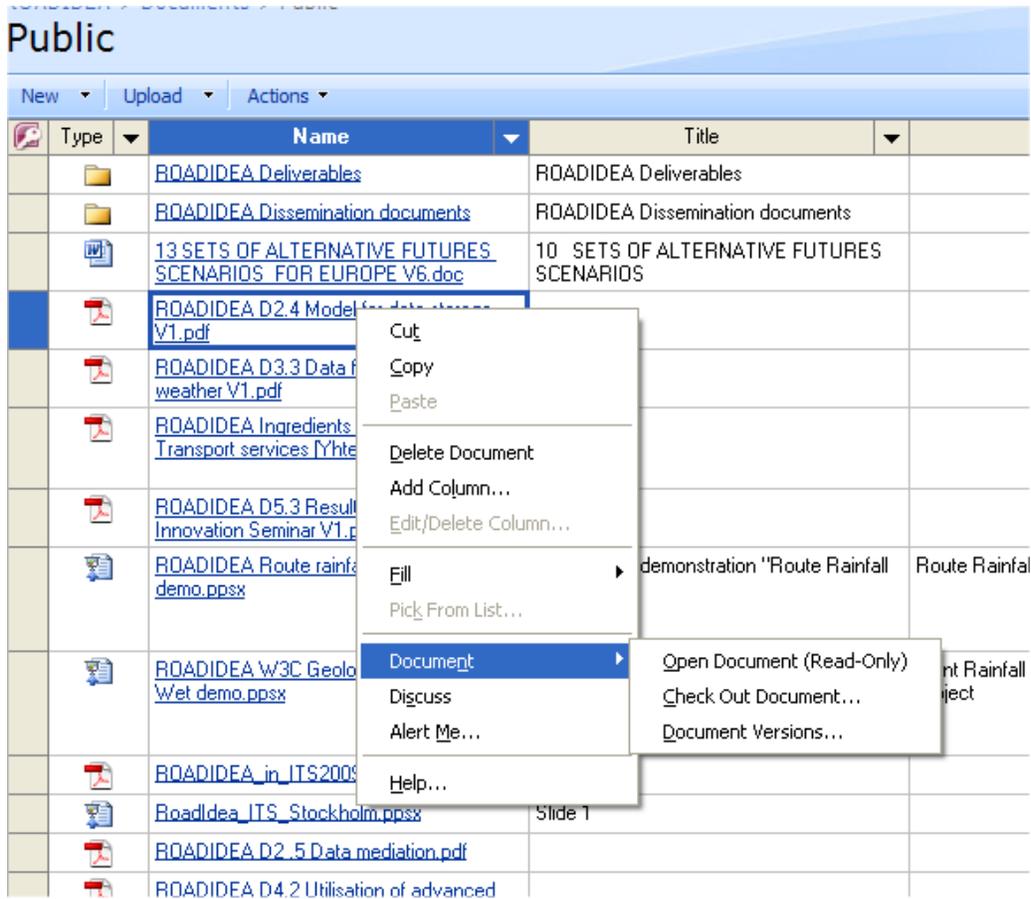
Figure 10 Metadata table for editing

Another useful possibility is to see the document versions by right clicking on the document name and then select document to get a list of options.

Figure 11Example screenshot of the document version history

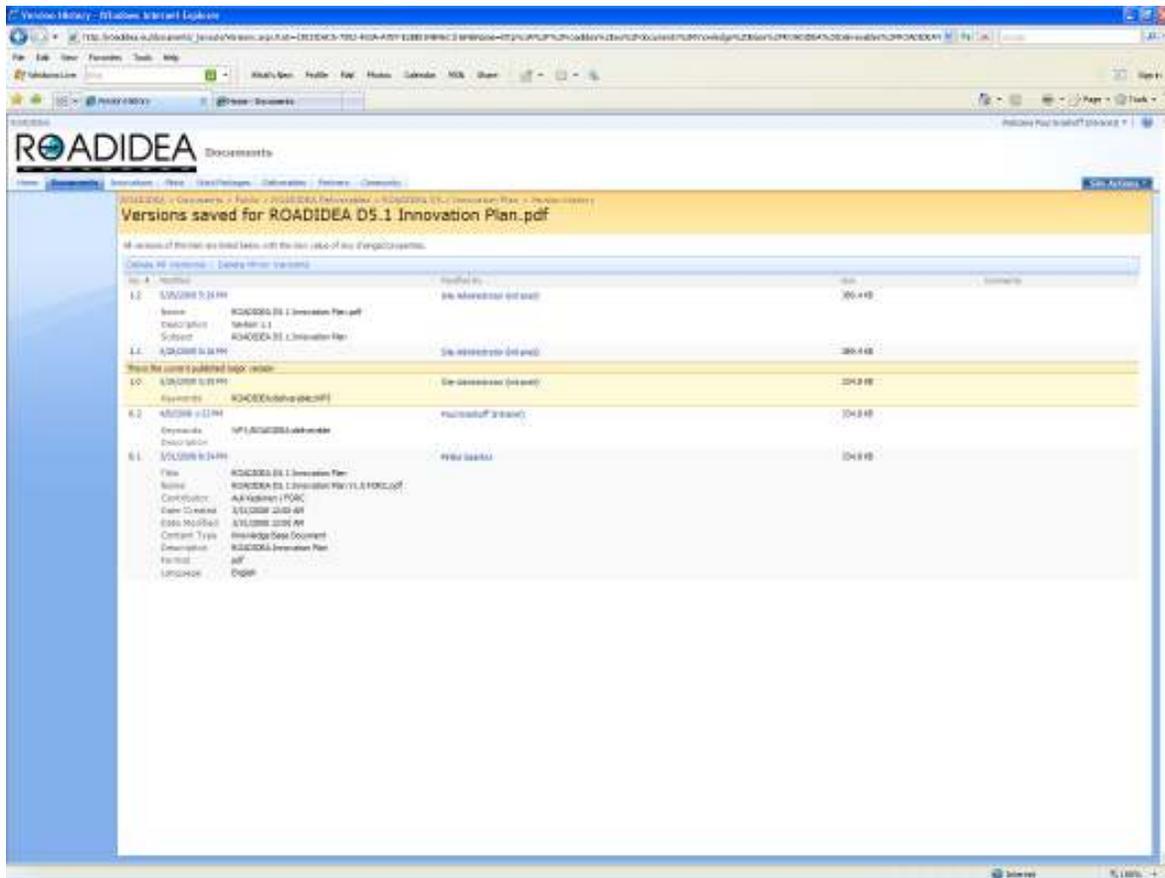A version history of document will look like this:

Figure 12Example screenshot of a version history

8. Table data viewer and editor

The data-viewer and data-edit capabilities of the system will be provided through the existing ETIS-VIEW and ETIS_EDIT client-server software with a central SQL compliant database. The ETIS-PLUS project software will be used to implement a separate RETRACK KB data viewer and data editor. This database can handle two types of data (note that network type data will be stored in the on-line GIS database):

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

The concept of multidimensional data as applied in RETRACK is explained first.

8.1 The multidimensional data model

Within the RETRACK-KB the definition of a Data Dictionary is of great importance. The following definitions will be used throughout this design document and should be fully understood before working with the applications and the data contained therein:

- Data Dictionary: contains the full definition of all available dimensions, dimension elements, tables and parameters for a specific data collection;
- Collection: a self-describing, harmonized and fully consistent data set;
- Dimension: the dimension that makes up the parameter, for instance as TIME (in years) or GEO (country level);
- Dimensions element: all available values for a specific dimension;
- Parameter: the variable for which data is collected;
- Table: the structure containing the actual data.

The conceptual basis of the RETRACK KB Table data is a multi-dimensional data model

e.g. 3 dimensional cube
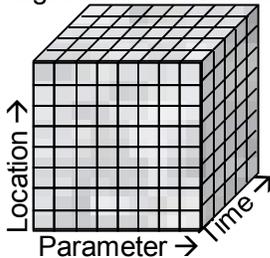
This model is selected for RETRACK (and the preceding projects REORIENT, WORLDNET, ETIS+) because:

- it represents the data very well in simulation models, as they often work with multidimensional arrays,
- it allows easy analysis through so-called Online Analytical Processing (OLAP) in databases,
- the multidimensional data model is very flexible and is self-documenting. Multidimensional data set(s) in combination with a data dictionary can easily be programmed for reading, regardless of the file implementation (database, text file, binary file, Excel file, and so forth).

Key concepts in the multidimensional data model are:

- The basis is that all parameter data values are multidimensional and can be represented as multidimensional data cubes.
- Each dataset (table) has one or more parameters with exactly the same dimensions in it, for example a location and time dimension.
- A collection of one or more tables is defined using a data dictionary, as follows:

- Each dimension needed must be defined as well as the elements to go with that dimension.
- Each parameter must be defined, with metadata (units, default, range, etc.) and which dimensions apply to that parameter
- One can group parameters in tables if they have the exactly the same dimensions. More than one parameter or tables can have the same dimension, but they must all comply with the definition in the data dictionary. In this context the same dimension means using the (parts of the) same dimension elements.
- As this definition is very flexible, it can be used for all sorts of data. In practice it can handle the bulk of data needs for simulation models, as well as all data from Eurostat.
- But it doesn't directly support hierarchical or more complicated relational data. For more complicated data (more strictly enforced relational data models)

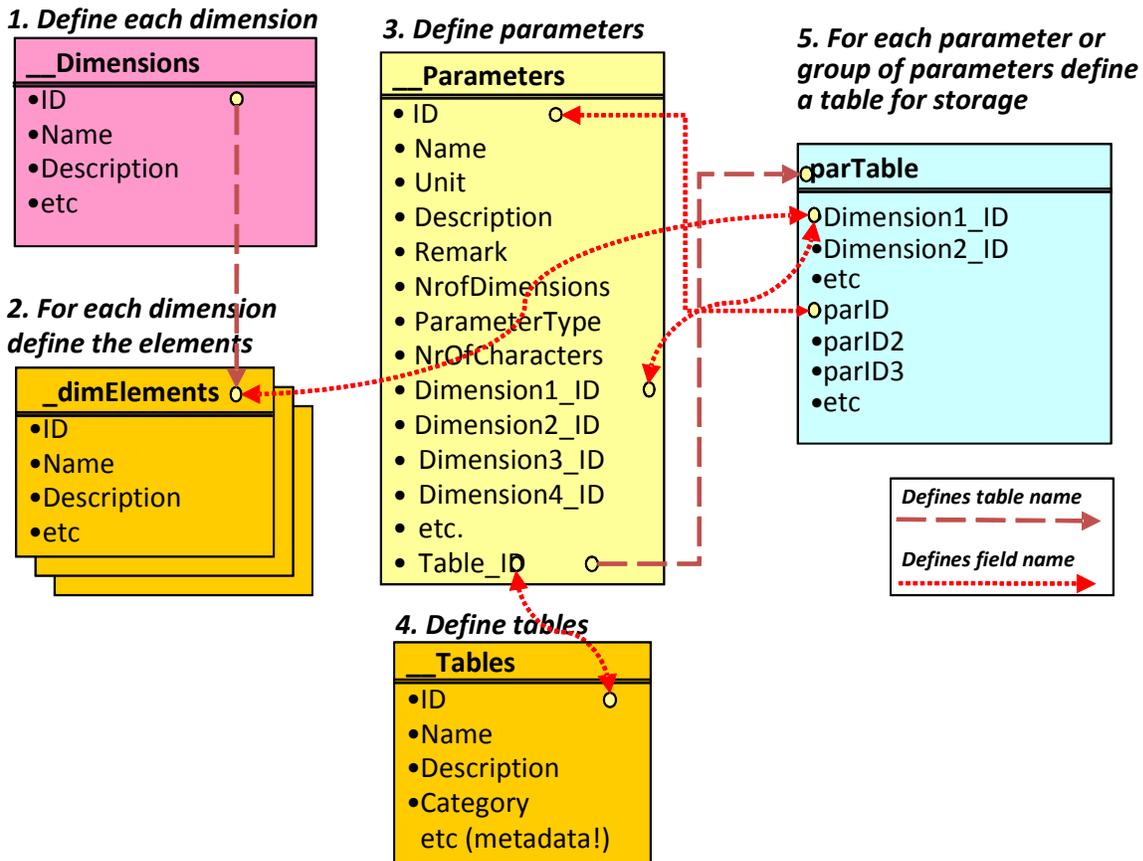The inter relationship of these concepts is illustrated in the figure below.

Figure 13 Construction of a multi-dimensional data model

Collections of multidimensional data

Especially in database applications it is possible to store more than one table with actual data in the system. A collection is defined as a set of data tables that belong together because they are all using the same dimensions and elements for each of the dimensions (in other words the same data dictionary). Some reasons/aspects for working with collections of tables are:

- Some dimensions will change over time (think of NUTS zoning changes; years for which there are data, etc). Within one collection there can be only one definition of a dimension and only one set of dimension elements.
- The central database will handle different collections of data as separate entities, each with their own data dictionary, in the case of RETRACK KB each new corridor can have its own data dictionary depending on the corridor specifics. And of course the data dictionary can also be a copy of an existing one with minor changes (e.g. different NUTS zones and countries that make up the corridor).

Data exchange formats

- Needed to support the planned check-in / check-out process for editing data locally (using subsets of data, e.g. per country etc.
- Three data formats to be specified, each with a header file for the data dictionary:
  - Binary format   BIN   one table
  - Excel format   XLS   collection of tables
  - MS Access   MDB   collection of tables

The web application to be developed for RETRACK to store, view and edit the multidimensional data will consist of two parts:

- RETRACK-VIEW will enable the user to easily select a table from the SQL Server database for viewing the original data by way of tables, charts and thematic topographic maps;
- RETRACK-EDIT will enable the user to edit data locally by way of a check-out / check-in procedure; an (authorized) user can easily select a table, define a subset for countries and time (if relevant and needed) and retrieve the file in the preferred exchange format. Exchange formats supported should include as a minimum an MS Access database and MS Excel file. After editing the file can be checked back in to the central database. The system should store all changes made so that unwanted edits can be reverted by the system administrator.

In the Retrack WP12 Terms of Reference functional specifications were formulated for the data viewer and data editor. These functional specifications need to be fulfilled in the software delivered by Demis. For completeness they are repeated here:

8.2 Functionality of RETRACK-VIEW

The RETRACK-VIEW application will be 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-VIEW application allows users to browse and export data contained within the collections on the website. The full functionality of the viewer is identically available within the RETRACK-EDIT application, where also the editing of the data is enabled through a data check-out-check-in mechanism.

In the following figure an overview is given of the ETIS-VIEW main screen, distinguishing:
1. The menu's File, Edit and View: these pull-down menu's provide functions for export, check-out/in and viewing of the data;

2. The selected collection (from a pull down list);
3. The blue highlighted selection is the table name, for which the corresponding data table is shown on the right side;
4. The data table area, where the table is shown according to selections made:
5. The parameter and dimensions element selections: here the specific (sub) selection can be made to be shown in the data table area.
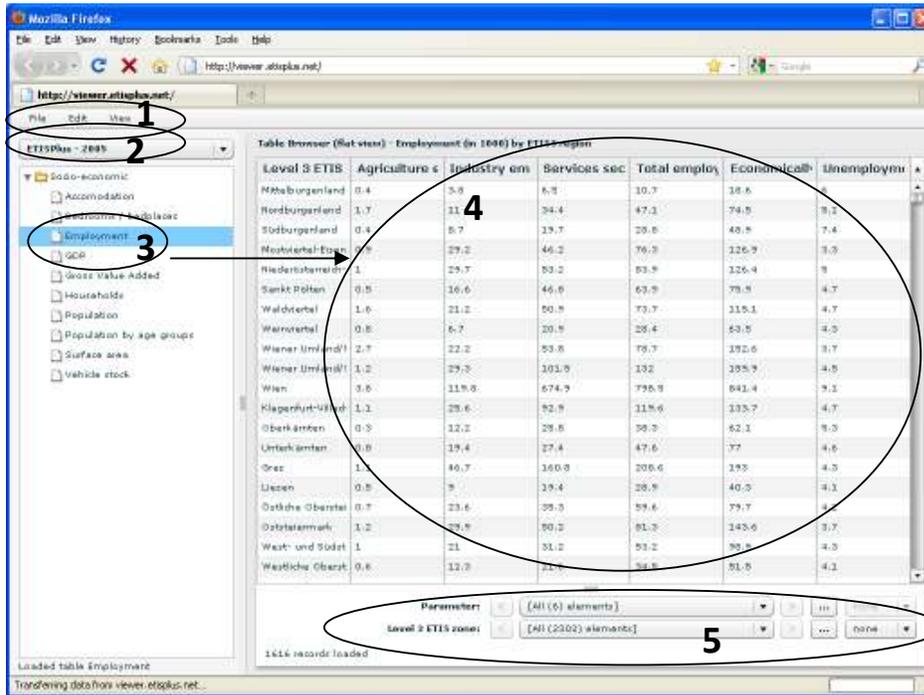

Figure 14 Example screenshot of the RETRACK-VIEW general structure

RETRACK-VIEW will provide the following functionality:
- RETRACK-VIEW allows the user to select a data collection. The name of that collection remains visible in subsequent screens so it is clear for which collection data views are made;
- The viewer functionality does not require a log-in;
- Within a collection a table needs to be selected by the user. Depending on the applicable dimensions selections within the dimensions can be made by the user to create a subset for viewing. A drop down box with check boxes for selection of one or more elements in that dimension. A button allows the user to limit the list of elements to those that actually are found in a table (needed for sparse data sets);
- The resulting data view is provided by default as a table. The user can select chart / thematic map (but here the geographic location must be one of the remaining dimensions) as alternative views. If the table still has more than two dimensions, the user must be able to select the horizontal and vertical axis; and have 'play' buttons to go forward and backward in the remaining dimensions that still have more than one element;
- Metadata for the selected table will be made visible as well, such as versioning information and data characteristics (i.e. unit, type and format) in a memo field for each table;

- The selected subset is also available for download in one of the supported exchange formats (formats to be decided, for example text, binary, XML, MS-Excel or MS-Access format);
- Thematic maps must have the functionality to go through time in an animated mode. For normal zonal (ETIS regions) data the map viewer will have to be similar to the thematic maps used in REORIENT. For Origin-Destination Data the thematic maps must be similar to that used in the WorldNet and REORIENT projects for that type of data.
- All thematic maps in RETRACK must have the ability to show background maps served in the Open GIS WMS 1.1 protocol as well as Google maps

The implementation of the Retrack-VIEW web application will be in the form of a client server application, where the data is stored in MS SQL Server and the client is a Flash application. It will be based on the existing ETIS+ viewer. An example screen from the ETIS-PLUS viewer is given below.

Within the ETIS-VIEW application there are two ways of viewing data: by way of a flat table (record) view and by way of a cube (2-dimensional) view. Please note that the cube view is only available when the table contains 2 or more dimensions. Choosing the Flat Table or Cube View can be done from the View menu.

Flat Table View

The Flat Table View shows a record like structure: for each combination of dimension elements one row is shown with the available parameter(s), such as illustrated in Figure 2.9. Take note that a combination of dimension elements is only shown if at least one parameter has a value for this combination.
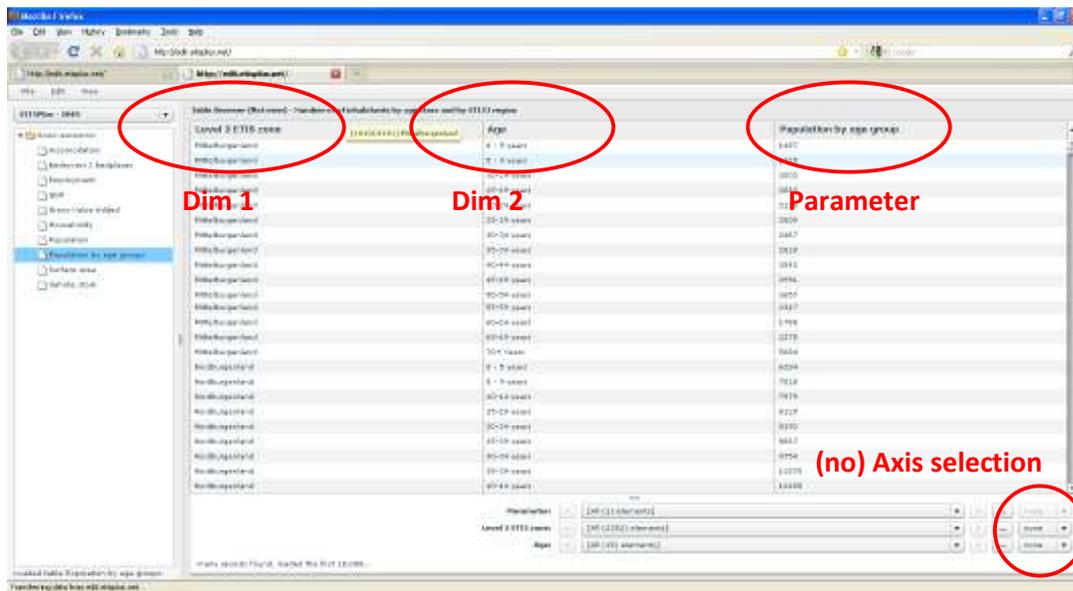

Figure 15 Flat table view - no axis selected

When the table is shown in flat table view, additional table options are available such as sorting the table by dimension value. For this the dimension header of the column needs to be clicked,

after which the icon for sorting (up ▾ or down ▴) is available in the column. Please note that the sorting will be performed based on the ID of the dimension elements, not the name.

Cube View

The Cube View will be available from the top View menu, but also can be activated when selecting a dimension for the x- or y-axis, as illustrated in Figure 2.10. Other than the parameter, the table needs to contain at least 2 dimensions for the cube view to be available. For the cube view to take effect, both dimensions need to be assigned to an axis. This is partly automated: with 2 dimensions, when one is selected, the other is automatically assigned. When there are more than 2 dimensions, both the x- and y-axis will need to be assigned.

Sorting the columns as in the flat table view is not possible here. However, the sorting of the column dimension can be organised by dragging one column to another location and thus arranging the sort manually.
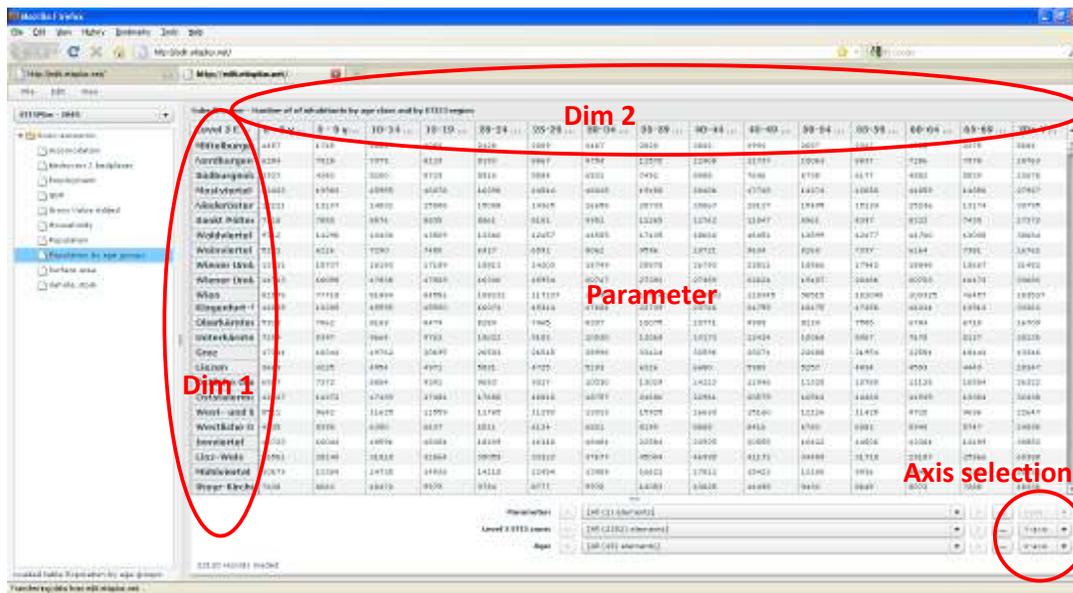

Figure 16 Cube view for multiple dimensions: x and y-axis selected

8.3 Functionality of RETRACK-EDIT

Within the RETRACK WP12 work package, multiple users will be working on completing the data sets for the RETRACK corridor. Checking out data is an absolute requirement for editing data; this ensures data integrity by preventing that multiple users can work on the same data simultaneously. The RETRACK-EDIT tool supports a process of checking out, exporting and checking in selected data online. The following steps need to be taken in order to edit the data:
1.   Make a data selection (based on 'what you see is what you get');
2.   Check out the data that was selected;
3.   Edit the data locally (not online);
4.   Check in the data.

The RETRACK-EDIT web application will support this 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).

RETRACK-EDIT allows easy selection of tables and if necessary defining a subset for one or more countries and for one or more time steps (if relevant). If the user is logged in with proper credentials he can then 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 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.

RETRACK-EDIT will provide the following functionality:

- RETRACK-EDIT allows the user to select a collection. The name of that collection remains visible in all subsequent screens so it is clear for which collection data views are made;
- The check-out – check-in functionality requires a log-in with proper credentials for editing;
- Within a collection a table needs to be selected by the user. Depending on the applicable dimensions selections within the dimensions can be made by the user to create a subset for viewing. A drop down box with check boxes for each element in that selection;
- The view shown to the user now has an extra toolbar that becomes visible once a valid selection has been made (as a minimum the geographic dimension selection must be valid). The toolbar has buttons that allow the user to check if the selected data is free for check-out and if not who has checked out the data. The user can also do "Undo" previous check-out. If you click on the status you can see which person(s) have checked out data in this selection, so you can mail them if needed;
- If you press the download button you get a dialog box for a file name. In this dialog box you can select the file type via the extension (xls, bin, txt, mdb, xml). Once a valid file name is given the download starts. Once the download is successful, the status of the data changes to "Checked out";
- At the table level to store the selected subset with filename of the checkout must be stored for easy retrieval. It must be possible to select a subset via a list of defined subsets and also to remove elements from that list.
- If you press the Upload button you get a dialog box where metadata according to the ETIS standard need to be filled in. In the upload, verification is preformed to see if the data is still consistent with the data dictionary (matching dimension, parameter and table definitions, as well as check on validity of the data itself). If the upload and verification is correct, the data in the central database is overwritten and the previous value per cell is stored, as well as a tag that data was changed, by whom and when.
- Once this upload is complete and the data have been processed these become again available for check-out.

After checking out the data, the checked out data will be highlighted in green rectangles, as shown in the next figure. If data is checked out by another user, the data will be highlighted by red rectangles and the user will know it is not possible to check this data.
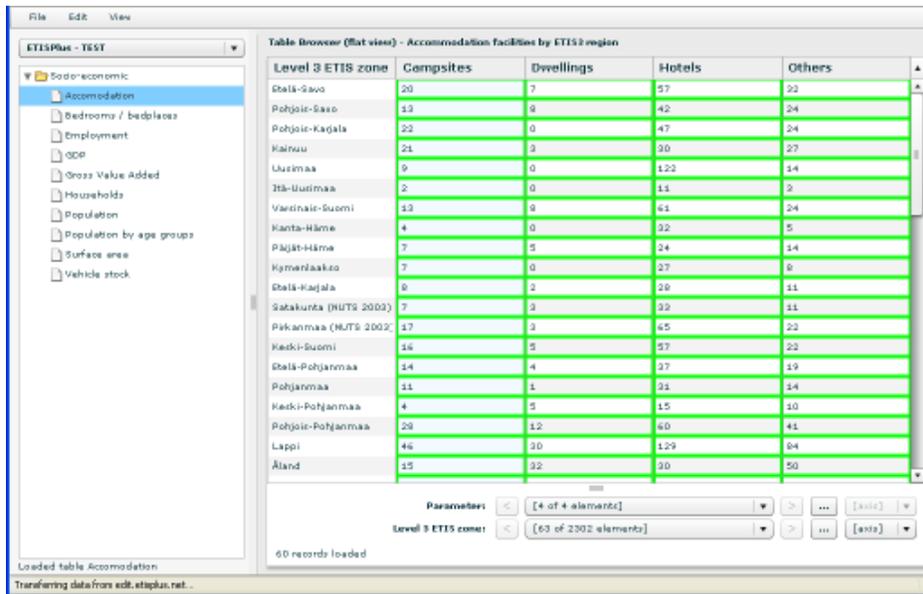
Figure 17 Example screenshot of data marked as checked out by the same user

9. GEO data viewer and editor

The RETRACKNetter client-server application will be a web based software tool that allows for viewing and editing of **network** related data on the internet. It is an on-line GIS that can be used by a community of users for network editing. It could be described as a network Wiki. Within RETRACK KB the RETRACKNetter can be used to gather, update, validate and complete network related data.

For understanding the GEO data viewer and editor concepts, it is imperative that the terminology such as described in the frame below is well understood:

Frequently used terminology

Network objects: the collection of objects of the type network links, nodes and (link) points.

Network link: a segment in the (road or other) network, that connects two nodes. The trajectory of the link is defined by the network points that form the network link. The network link can be of the type sea, inland waterway, road or rail and has a great number of attributes.

Network node: a point where two or more network links meet. The network node is related to the type of network links it connects or can be a terminal or port.

Network (link) point: a point on a network link that determines it's trajectory but has no additional attributes.

Attribute: a characteristic of a network link or node, also referred to as property or parameter. Attributes are dependent of the type of the object; different network links have different attributes (see annexes A and B).

Trajectory: the exact path a network link follows, which is determined by the position of the network link points.

To run the RETRACKNetter tool the user will need to have a PC that has:

- A web browser installed;
- Macromedia Flash player 9 or better installed (Flash 10 is recommended);
- A fast internet connection (ADSL or better).

If you do not have Flash installed, you can go to the RETRACKNetter web address, where a message will appear automatically that asks if you want the required plug-in (Flash) to be installed. Click yes here and Flash will be installed and run automatically. Depending on your specific settings you may have to give a few additional confirmations.

There will be two RETRACKNetter versions: the live version and the test version. The test version can be used for learning how to work with the tool without destroying any data. The live version directly edits the network topology and data of the RETRACK KB.

The network data viewer and editor will be based on the available on-line GIS network editing tool (InterNetter) used in the EC projects REORIENT, Worldnet, ROADIDA and ETIS+. It allows for viewing and editing the network topology and related attribute data via the internet.

Specific for Retrack is that the software should be able to easily define a (new) corridor as well the possibility to only show the network items that are part of the corridor. The underlying network data will be the same as for the ETIS+ project, this will allow both running EC projects to mutually benefit from any work done on the network. For completeness the functional specification for the software to be delivered by Demis as specified in the Retrack WP12 TOR is repeated here:

9.1 System requirements

The system must be client-server based and only require a web browser and Macromedia Flash player for the client. No software licenses or installation should be required at the client's end.

The client should contain the following components:
- Navigation tree (left), including options for log-in, application settings, (navigation and editing) selecting edit actions and network options for appearance of the visualisation;
- Vertical toolbar (middle), providing direct access to system functions (also available under the edit actions in the navigation tree);
- Map client area (right), where the actual navigation, visualisation and editing occurs.
- The system must have the ability to use background maps served in the Open GIS WMS 1.1 protocol as well as Google maps.

9.2 Visualisation options

The first important step in working with the system is to be able to visualise data and navigate through the network data. The visualisation of links (network segment) and nodes (network point) must be set separately in the client in the network menu item of the navigation tree. Here the appearance of the color, size, text, shape (node only) and visibility can be set by the user by:
- Select the node or link box to enable this objects appearance (if not done yet);
- Select the appearance option(s) you want to visualise;
- Select the attribute to show for the appearance: the link color appearance is set to Road Type in the example below; a legend is automatically assigned to the Road Type values.

The system must specifically support showing multiple network attributes, which is a powerful mechanism for combining various appearances. The client must make it possible to create appearance combinations of link color (for example Road type), link size (for example the number of lanes), link text (for example free speed) and node color (for example the Connected roads).
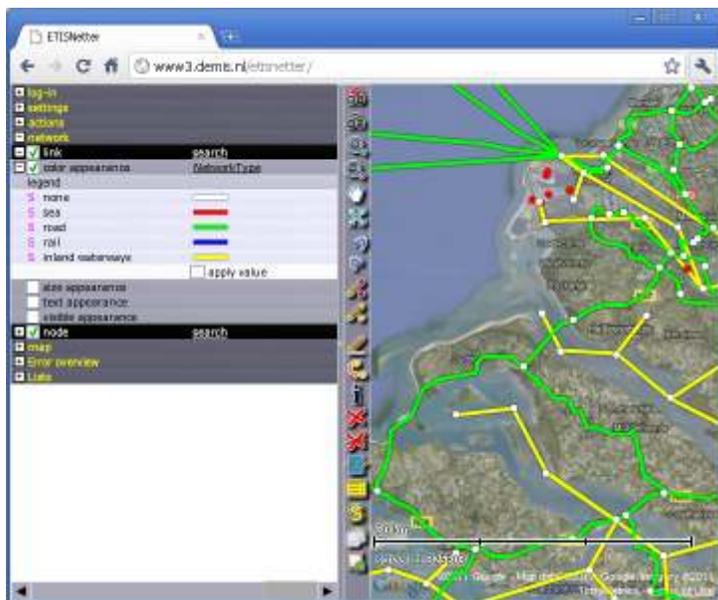
Figure 18 Example screenshot that visualises the network type by colour

9.3 Network editing options

Various editing actions must be available in the system. The actions can be chosen from the toolbar or navigation action menu (mouse / keyboard). However the action should also depend on the object (link, node) that has the focus in the map area. For example, by selecting the *'add point'* action, a point can be added to the network when moving the mouse over a link. When moving over a node or (link) point however, the *'move point'* action becomes active instead, because it is likely that is what the user wants to do.

Editing network attributes must be implemented in three ways:
- Using the link or node properties table from the toolbar to edit one or more attributes: this allows the user to open a table with all attributes of this object by clicking a link or node and edit any attribute presented there;
- Setting the same value to any link or node: this allows assigning the same value for a specific attribute to any object that is selected after clicking the "S" icon next to an attribute value such as in the appearance legend;
- Copy and paste all values of an object: use the copy and paste functionality to copy all values of one object to another.

Editing the network topology must be done by adding, removing or moving links, nodes or (link) points (which determine the trajectory of links but have no attributes as nodes). After selecting an action from the toolbar or by (short cut) keyboard key, the action must be shown as tooltip text when moving over the map. The required add and remove actions are described in more detail below.

Adding a node must be applied by:
- Select the toolbar action *'add node'* (or by typing the keyboard key K);

- Select a position on an existing link (which may be an existing point). The add node tool tip will appear; click the position to add the node as shown below. After any edit action the objects attributes can be edited as described above.

Adding a link must be applied by:
- Select the toolbar action 'add link' (or keyboard key T);
- Click on the map to set the starting point of the link. Every next click on the map will add a point, by which you can determine the trajectory of the link as illustrated below. The link is drawn in gray as long as it is not accepted, showing that editing is in progress;
- When finished with drawing the link, you can either apply the link by pressing [Enter] or cancel by pressing [Esc] on the keyboard. When cancelling the link disappears, when applying the link is drawn in black. The system must automatically turn the begin- and endpoint into a node and all intermediate points into (link) points.

 Removing an object should be rather straightforward, but removing a node must also remove any connected link. Therefore the action 'remove [a node] and join [links]' can be applied by:
- Select the 'remove and join' action from the toolbar (or keyboard key W);
- Select the node you wish to remove; note that this node must connect exactly two links whose attribute values are equal.

10.   Survey system

In the Retrack KB design as it stands now the use of surveys is thought to be **not needed** as a method to contribute to the Retrack KB. Should a survey be needed the required functional specifications are:

1   Allow to design a new survey on-line.
2   Allow respondents to fill out the survey on-line
3   Store the results in the central database
4   Allow viewing of the survey results on-line
5   Allow (registered) users to export the survey data to excel

These specifications can be handled through the built-in capabilities of the Microsoft Sharepoint version 2010 portal.

11. Exchange of data with TENTec and GISCO

For the Retrack KB the European Commission requires that (corridor) data exchange with the TENTec system is possible and that GIS data (e.g. background maps and infrastructure data) are used from GISCO system. The exact requirements for data exchange and definition of the systems to be exchanged to or from will need to be worked out during in consultation with the EC during the (further) design and implementation of the Retrack KB. If standard GIS data file exchange is not sufficient then Demis will develop a custom solution to meet the EC requirements.

12. Implementation plan of the RETRACK KB

The implementation of the RETRACK KB will be organised along the following lines:

1) Preparation of the complete **Table data** model definitions; the **Geo data** definitions and **corridor document repository**   **December 2011**
2) Delivery of the RETRACK KB portal including all software system elements and the Rotterdam-Constantza corridor "empty shelves"   **January 2012**
3) Filling by WP12 partners of their respective "shelves"   **July 2012**

12.1 Data model definitions

Based on the Retrack content as described in Table 1 of section 4.1 the data model definitions of the Table data and of the GEO data need to be finalized. Organising the document repository is easy, just prepare the empty folders and it is ready.

Rules for definition the Table data model

To build multidimensional data tables so that they are self-explanatory rules are needed. The process of defining the multidimensional data model is illustrated in the following figure.
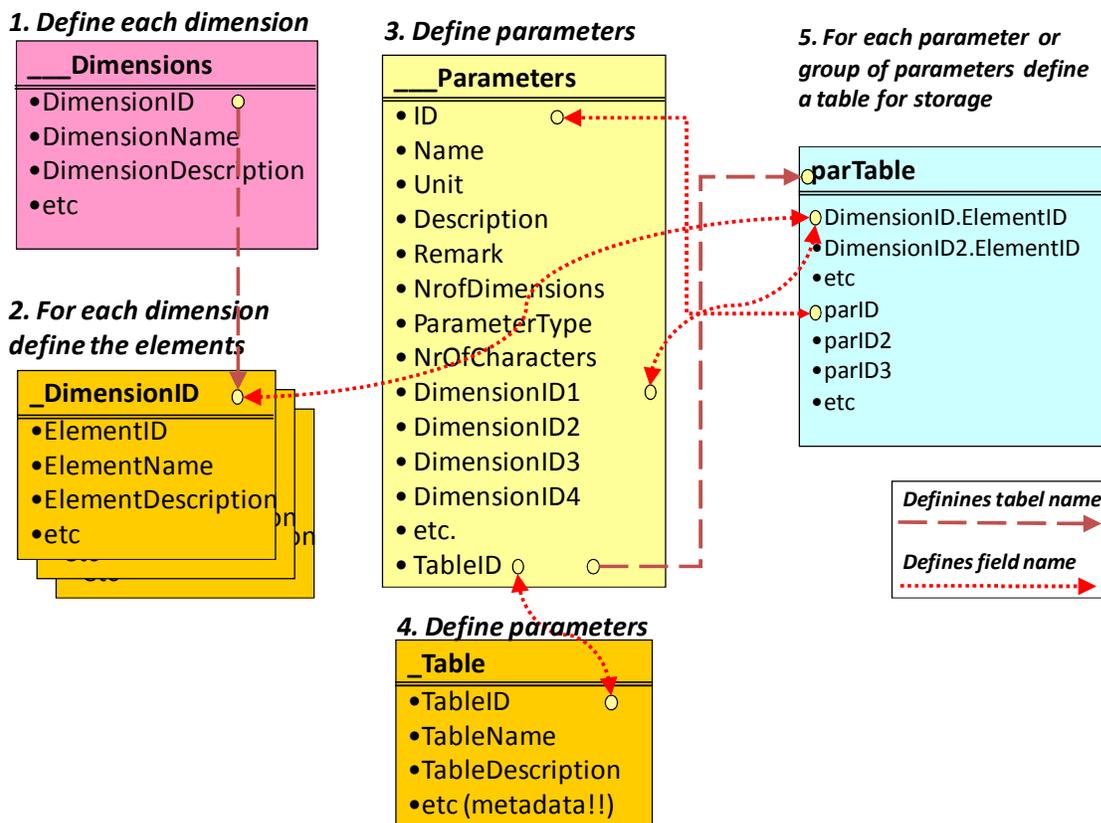

Figure 19 An example for defining a data model of an multidimensional data table

The rules for making a multidimensional model self-reading are as follows:

1. Define all the dimensions needed. If this concerns an implementation in a database give this table the name ___Dimensions. For other file formats this is implemented differently, but the principle stands. For each dimension define a DimensionID, DimensionName, DimensionDescription field. The names of the fields MUST always remain the same (otherwise it is not self-reading), these fields names may not be changed (otherwise all software for reading/writing the files will cease functioning. Extra fields are optional and can be used for additional metadata.

2. For each dimension 'abc' define a table '_abc' to store the elements for that dimension. For other file formats this is implemented differently, but the principle stands. Fill each '_Dimension' table with all elements needed in each dimension by providing a field with the ElementID, ElementName, ElementDescription fields. The names of the fields MUST always remain the same (otherwise it is not self-reading), these fields names may not be changed (otherwise all software for reading/writing the files will cease functioning. Extra fields are optional and can be used for additional metadata.

3. Define all the parameters needed. If this concerns an implementation in a database file give this table the name ___Parameters. For other file formats this is implemented differently, but the principle stands. For each parameters define the major metadata as follows: ID, Name, Unit, Description, Remark, NrofDimensions, ParameterType, NrOfCharacters, DimensionID1, ..., DimensionID7, ChoiceDimension, TableID field. The names of the fields MUST always remain the same (otherwise it is not self-reading), these fields names may not be changed (otherwise all software for reading/writing the files will cease functioning. Extra fields are optional and can be used for additional metadata.

4. In the case of file format that supports more than one table with actual data, such an Access database you can group parameters together in tables (hence the TableID field in the parameter definition, leave it empty for a separate table for that parameter). If this concerns an implementation in a database file give this table the name _Table. For other file formats this is implemented differently, but the principle stands. For each table define a TableID, TableName, TableDescription field. The names of the fields MUST always remain the same (otherwise it is not self-reading), these fields names may not be changed (otherwise all software for reading/writing the files will cease functioning. Extra fields are optional and can be used for additional metadata.

5. For each parameter (when not grouping the parameters in one table) and/or for each table (for groups of parameters with the same dimensions) a data table that defines per record the dimension element IDs of applicable and has data columns for each parameter. A strict naming convention is needed here to make the database self-reading, as in practice in database implementations one must define columns for the elements of each applicable dimension as well as columns for each parameter that is grouped in the table. The naming convention proposed is as follows:

    a. Start with one column per dimension that applies to that table/parameter. If the dimension ID is 'abc' give the column the name 'abcID'. The content of this column must be an **existing ID** in the '_abc' dimension table.

    b. Next add one column per parameter that applies to that table. If the dimension parameter name is 'par_abc' give the column the name 'parabc'. Note that this column must have a datatype that is defined in the column ParameterType of the ___Parameters table.

The naming convention in step 5 is crucial as it links the columns to proper dimension and hence the dimension elements and it links the parameter column data to the metadata stored in the ___Parameters table.

The data model definition will be made Demis in conjunction with the data collection partners for all data that have been flagged as Table data type. The partners are required to help define the parameter list for each the Table data; the applicable dimensions and the dimension elements. Implementation of this in the MS SQL server database is handled by Demis.

Geo data

Based on the Retrack content as described in Table 1 of section 4.1 the data model definitions of the GEO data needs to be finalized. The partner responsible for the data collection is needed to help define the geo-parameter data definitions. Demis will handle the implementation in the RETRACKNetter software.

The deadline to finalise the definitions is December 2011.

12.2 RETRACK KB portal software with empty shelves

The data definitions as mentioned are used to construct the "empty shelves" for the RETRACK corridor KB. Together with the software this will be delivered in January 2012. After this the filling of the empty shelves should start.

12.3 Filling of the RETRACK KB by WP 12 partners

The actual filling of the data needs activity from the partners. The software is delivered by Demis. Demis will also provide manuals on how to fill the data in the system. And end of January Demis will organise a workshop with the partners who are actually doing the data collection and processing so they learn how to work with the system. Demis staff will be available for help on editing / uploading the data up to the end of the RETRACK project.

The actual filling of the system should take place in the period February-June 2012. Data collection can and should start as soon as possible though.

13.   References

Antunes, 2007: Knowledge Management using Semantic Web Technologies: An Application in Software Development

Bell et al, 2010: Magic Quadrant for Enterprise Content Management, Toby Bell, Karen M. Shegda, Mark R. Gilbert, Kenneth Chin, Gartner RAS Core Research Note G00206900, 16 November 2010 in http://www.gartner.com/technology/media-products/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.

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.

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

ETIS 2011a: ETISplus D2 Draft Specification Report - Main report (Chen, Newton et al)

ETIS 2011b: ETISplus D2 Annex Report – ETISplus Architecture (R. Smith, P. Grashoff, F. van der Wouden)

ETIS 2011c: General_concept_for_ETIS_BASE_data (P. Grashoff)

ETIS 2011d: ETIS-VIEW / ETIS-EDIT User Manual (P. Grashoff, F. van der Wouden)

ETIS 2011e: ETISNetter Quick Reference Guide (F. van der Wouden, M. Benjamins)

ETIS 2011f: ETISNetter User Manual Annex: Overview of ETISNetter link and node attributes (F. van der Wouden, M. Benjamins)

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

Merali, Y. and Davies, J., 2001: Knowledge Capture and Utilisation in Virtual Communities, 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

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

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