TREN/06/FP6TR/S07.55976/019851

PROMIT

Promoting Innovative Intermodal Freight Transport

Co-ordination Action
Priority 1.6.2 Sustainable Surface Transport

D 4.3 Recommendations processes and quality improvements in intermodal transport

Due Date of deliverable: February 2009
Actual submission date : 27 February 2009

Start date of the project: March 2006 Duration: 36 months

Main Authors: PDEng, M.Sc., I. Davydenko, Ir. G.Zomer, M.Sc. S. Krupe (all TNO)

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	**X**

Table of Contents

1 Executive summary

This report is the final deliverable for PROMIT WP4 on benchmarking and quality. The report includes condensed relevant information on the PROMIT benchmarking concept, PROMIT European benchmarks, the design and development of the benchmarking self assessment tool and finally it provides information on the current status of the benchmarking database used in the self assessment tool.

The PROMIT benchmarking concept divides intermodal transport into three abstraction levels. This division is necessary to deal with the intrinsic complexity of intermodal transport: it is by its nature more complex than the single mode road transport, because it involves transhipment operations and require more extensive preparation as trains depart according to a schedule. The division starts with level 1, the highest level of abstraction, which looks at intermodal transport as a door-to-door activity, hiding underlying complexities. Measurements at this level do not recognize explicitly modes used in transport; what counts here is the door-to-door measurements of performance thus making this level more important for end users of transport, namely shippers. Level 2 divides transportation chain into logical blocs, performance of which can be measured separately. The blocks normally constitute such independent operations as line haul, road feeder transport, etc. This level is most important for logistics departments and logistics service providers, who plan the transportation chain. Level 3 is the lowest abstraction level and measures performance of operations inside building blocks of level 2 and is intended to be used by asset owners and service providers.

The most important benchmarks include cost, transit time (operation time), variability of transit time, frequency of service, shipment compatibility (namely if it is possible to use a certain type of intermodal solution for a specific type of cargo and loading unit), damages and theft. The majority of these indicators are similar across all 3 abstraction levels, though there are some level-specific indicators. We also propose to measure quality indicators such as information flow and invoicing accuracy, as well as energy used per operation or per trip.

The benchmarking tool is designed in a way that upon entry of benchmarking data generates feedback in the form of relative performance in respect to the average performance measurements collected in the tool. This is necessary in order to protect sensitive company information as well as to generate meaningful and useful output. The software tool consists of 3 main functional modules: data collection module (provides web-based input for entering benchmarking data, error detection mechanism, and database storage facility); feedback generation module (feedback generation algorithm and web user interface to present benchmarking output); and research functionality module to allow analysis of data in the benchmarking database.

To seed the tool and make it attractive for further use by shippers and other type of non-PROMIT users, the members of the PROMIT consortium have provided the tool with a critical mass of benchmarking observations. Currently the tool's database contains 113 records of benchmarking data, of which 37 records are related to intermodal transport. There are 37 unique user accounts currently registered in the online self assessment tool. The majority of the measurements represent operations of swap bodies as the loading unit (53% of the records report benchmarks with this loading unit).

Currently there are not enough benchmarks in the database to draw statistical conclusions over the available data. However, with a certain degree of confidence we can state that the quality of existing benchmarks is sufficient for the purpose of seeding the tool. The data analysis shows a high degree of variability for the short distance transport assignments, while generally adhering to the industry rules on cost / transit time / distance relationships. Quite sadly, the data do not show a consistent cost advantage of intermodal transport over single mode road one, at the same time reporting generally faster transportation speed and smaller transit times for road transport.

In November 2008 TNO organized a workshop to test interest, usefulness and applicability of the PROMIT benchmarking tool for Dutch industrial companies. The workshop has shown that there is quite a lot of interest from the industrial companies towards intermodal benchmarking data. The workshop has also showed some deficiencies of current implementation of the tool, as it in some instances cannot directly satisfy some requirements of the industry. Though being a research tool, its current implementation is not intended to be a commercial service. Therefore, the main conclusion that we draw from the workshop is that there is quite a lot of demand for an intermodal benchmarking tool. But to become commercially viable and self-sustainable, such a tool needs to be more sophisticated, specifically, the tool should not be blind to lanes, goods flow direction and commodities in question.

To our knowledge, some benchmarking-related questions are being tackled in the BELOGIC project. We suggest that some form of knowledge transfer from the PROMIT project to BELOGIC is of vital interest of the later project. Besides extensive documentation of the PROMIT benchmarking concept and the tool, there is a vast volume of tacit knowledge that can be utilized in BELOGIC. Building upon the existing knowledge would make the advancement of BELOGIC faster and more profound.

2 Introduction

This report presents work that has been done in the framework of PROMIT Work Package 4 on definition of European benchmarks in intermodal transport and development of a benchmarking self assessment tool. This report is the last deliverable, which summarises benchmarking-related activities in PROMIT and describes activities and advancements made in the final year of the PROMIT project.

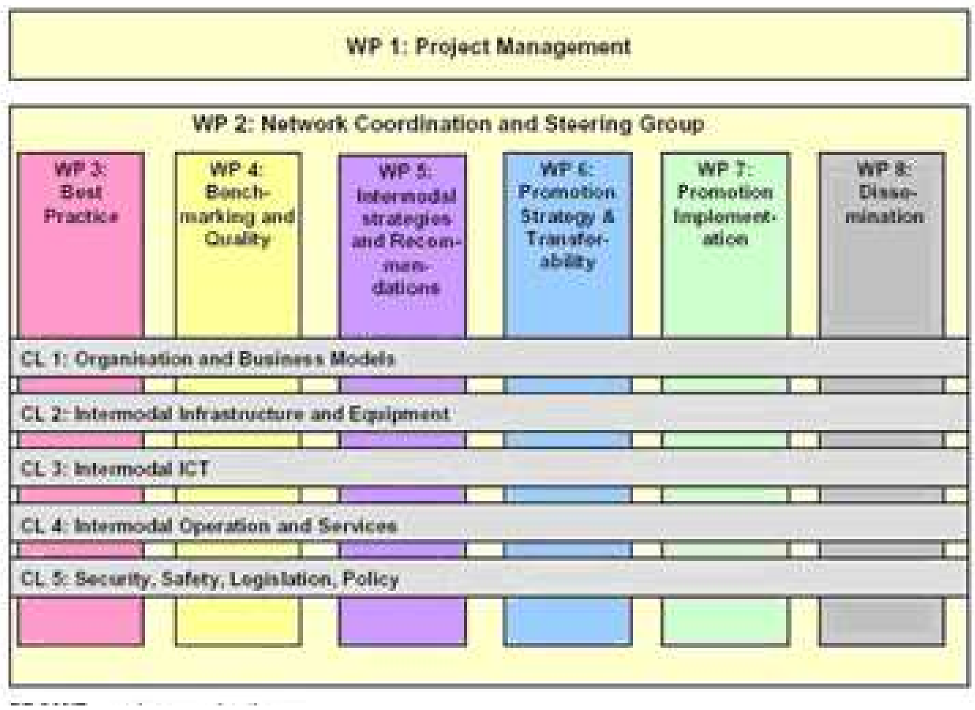


Figure 1. PROMIT project structure

According with the matrix structure of PROMIT, WP4 crosses all clusters in the PROMIT matrix structure. The main emphasis of WP4 is rather different from other work packages of the project: it has been aimed at research on key performance indicators (KPI) in intermodal transport, development of sound framework for measurements and development of a benchmarking self assessment tool.

The tool is a logical extension of the benchmarks that have been defined in D 4.1 European Benchmarks in Intermodal transport. In that report a set of the most important KPI's has been defined as well as logical division of the intermodal transportation into transportation abstraction levels. In deliverable D.4.2 Performance indicators and benchmarks for self assessment we described the design, development process and features of the self assessment tool. This report summarises activities related to benchmarks and self assessment tool and provides new insights on PROMIT benchmarking database (as a part of the tool) and discusses applicability and market potential of the tool.

As we concluded earlier, finding real life benchmarking information is very hard. Private companies see no interest in providing them, unless they can get something useful in return for their effort and information. The intention of the self assessment tool is to gather benchmarking information, at the same time the most important feature of the tool is that it is useful for shippers and transporters. The PROMIT consortium members have provided more than 100 benchmarks to seed the benchmarking database, creating a critical mass to make the tool operational.

Therefore, as the result of PROMIT WP4 activities, the self assessment tool can now help answer such questions as how can a shipper know that its transportation provider offers him a good service? How can it be made aware of other transportation means such as intermodal transport? How can a shipper be assisted with a transportation mode choice and facilitated during change, overcoming the habit barrier?

This report has the following structure. Chapter 3 provides information on benchmarking concept and European intermodal benchmarks; Chapter 4 describes the design of the self assessment tool and documents implementation process as well as implementation compromises. Chapter 5 contains information on current state of the tool and benchmarking database. In this chapter we analyse currently available benchmarks in respect to distribution of transport costs and transport time, as well as draw conclusions on the quality of the data and its usefulness. Chapter 6 contains information on a workshop organized by TNO to test usefulness and applicability of the self assessment tool in the Dutch transport market. We finalize the report with Chapter 7 which contains conclusions and recommendations.

3 PROMIT benchmarks concept and European benchmarks

In the first year of the project we studied and made an inventory of benchmarks in intermodal transport. Intermodal transport is an intrinsically complex undertaking; therefore it became apparent that measuring of some indicators is not sufficient to provide a sound basis for KPIs. In comparison to single mode road transport, the intermodal transport involves a chain of events, companies and operations. Thus the first step in establishing of benchmarks has been logical structuring of operations into different abstraction levels.

3.1 Division of intermodal transportation chain into abstraction levels

Transportation involves movement of goods from the source point to the destination point, and is often called a shipment. If the shipment involves more than one type of transportation, it is called multimodal. We can distinguish three levels of abstraction for intermodal transportation. The levels are the following

- **Level 1:** Transportation from source to destination as a black box. The shipment is considered as a whole, we may call it the "customer level". What is important here is location of the source and destination points, transit time and other shipper-related parameters.
- **Level 2:** Transportation from source to destination as a number of activities (per mode). An activity is treated as a black box. The shipment is considered as a number of constituting "black-box" activities; we may call it the "system integrator" or "4PL level". At this level, emphasis is made on parameters of different transportation activities, such as the road part, train/ship part, etc.
- **Level 3:** Specific operations. The shipment is considered at a very detailed level, taking into account all operations it undergoes on a way from the source point to the destination, we may call it the level of service provider or operator.

Figure 2 summarises the abovementioned abstraction level division.

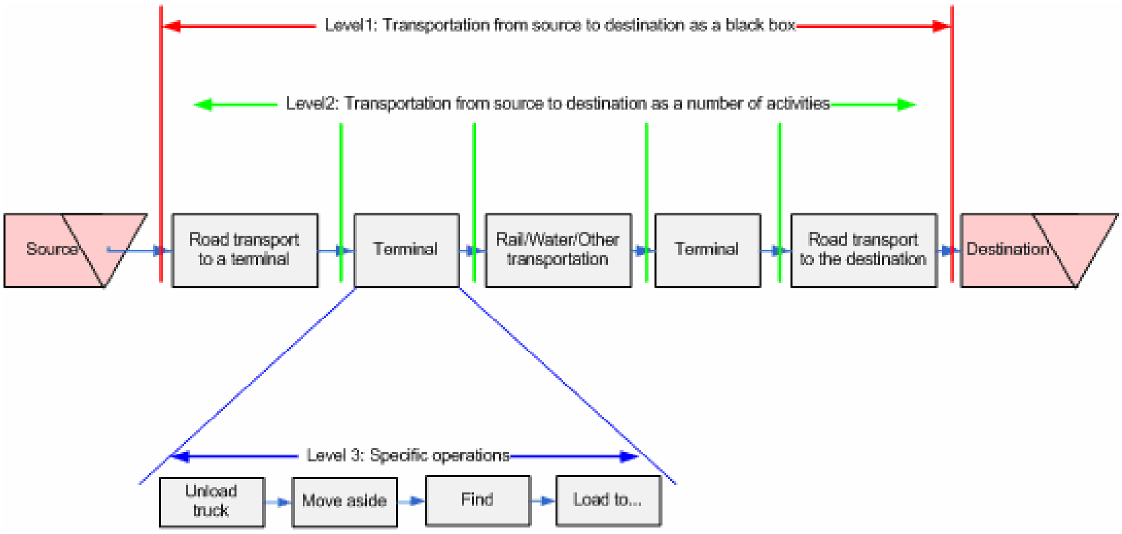


Figure 2 Division of intermodal transportation chain into abstraction levels

3.1.1 *Level 1: Transportation from source to destination as a black box*

The success of intermodal transportation lies exactly at the first level, because the shipper's satisfaction determines whether the intermodal transportation is used. This level is the most important one from the shipper (user) perspective. It can also be seen from the marketing perspective, reflecting such important parameters as shipment cost, delivery time and reliability. At this level we consider shipments from the source point to the destination point. We are not concerned with the processes that take part between the source and destination points. The most important factors at this level are those that are visible to the end-customer. These are shipment feasibility, shipment cost, transit time and transit time reliability.

3.1.2 *Level 2: Transportation from source to destination as a number of black-box activities*

This level considers a shipment as a number of independent black box activities. Each black box represents an activity such as feeder road transportation to a terminal, rail segment or ship segment, activities at a terminal, etc. This level is important for the success of intermodal transportation, however is aimed not at the end-user of transportation, but at the logistics service provider, namely at the responsible entity that chooses the route of the shipment. It could be a company's logistics department, transportation service provider or 4PL logistics provider who matches real transporters with the client's needs. Thus, the service provider makes a route for the shipment combining particular shipment activities as black boxes with corresponding parameters.

The success of a particular route now depends on the properties of constituting activities. A failure at one link of the chain leads to a failure (or substantial worsening) of the whole chain. Therefore, when measuring performance of each link, it is possible to choose the optimal route for a shipment, decide whether to use intermodal transportation or not. Measuring the performance of constituting activities it is also possible to make conclusions on the indicators at level 1.

3.1.3 *Level 3: Specific operations*

The level considers operations within black boxes of which a shipment consists, going deeper to the lowest level of shipping activities. At level three we are interested in specific indicators for activities such as terminals, rail, road and water transport. This is the level of specific service providers: indicators are aimed at assessment of rather technical internal operations of the service providers.

Success of a particular service provider within intermodal transportation chain depends on how well it can perform its operations and to which extent its operations are requested by the market. Competitiveness of a service provider determines whether its services will be required within intermodal chain (the routing questions are done at level two: decision makers at that level are the direct customers) as well as the success of the whole intermodal chain.

The third abstraction level is described and documented in previous PROMIT deliverables (D 4.1 European Benchmarks in Intermodal transport). During the work on the benchmarking tool it has been decided not to include it into the implementation of the tool. The main reason for this decision is that this level is very operational and is related to performance of individual service providers. It is the responsibility of service providers to measure their own performance; to some extent it is also the responsibility of LSPs and logistics departments to monitor their performance as well. However, it has too little added value for shippers and decision makers who might want to compare performance of intermodal transport chains as a whole. Adding third level to current implementation of the tool would greatly increase its complexity and decrease usability of the tool.

3.2 Key Performance Indicators (KPI's).

In this section we present the most important KPI's for the three above discussed levels. The indicators presented are the result of our research: we studied reports of the partners and literature in order to identify important indicators. Here we discuss the most important KPI's that are almost independent from the environment and have been implemented in the benchmarking tool. The complete set of indicators can be found in the PROMIT deliverable D 4.1 European Benchmarks in Intermodal transport.

Table 1 KPI's per transportation level

Key Performance Indicators		
Level 1: Shipper level	**Level 2: 4PL level**	**Level 3: Operator level**
Price	Price	Price
Lead time (Transit Time) + variability	Lead Time (Execution Time) + variability	Transit Time (Execution Time) + variability
Frequency of service	Frequency of Service	Capacity
Shipment compatibility	Shipment Compatibility	Capacity utilization
Damages	Damages	Reliability
Theft	Theft	Damages
		Theft
		Operation-Specific

As it can be concluded from table 1, the indicators are mostly overlapping (marked in light yellow colour), especially for levels one and two. Level three is more operation specific, therefore it has somewhat different indicators and many of them are operation and even equipment specific indicators.

3.2.1 Main KPI's

1. **Price (cost)**. This is the cost of sending a shipment or conduction an activity. The price is often a function of distance and shipment size. Whenever possible, a standardized shipment size should be used, for example twenty feet containers (TEU), otherwise price comparison becomes difficult.
2. **Lead time (transit time)**. This is time that passes from the moment the activity starts (shipment leaves the source point) to the moment the activity finishes (shipment reaches the destination point). It should be specified whether lead time accounts for frequency of service. There can be distinguished several types of this indicator.
    a. **Effective mean lead-time**. This indicator is measured as the average lead time over several executions (shipments)
    b. **Promised lead time**. Promised lead time is the lead time that is promised by the service provider
    c. **95% reliability lead-time**. This indicator measures minimum time that has been required for 95% of all activities (shipments).
    d. **On-time performance**. Measured as percentage of activities (shipments) carried out within promised lead time
3. **Lead time variability**. This performance indicator is often coupled with the effective mean lead time. The pair represents statistical mean and deviation of the lead time and can be used for many technical purposes such as stock level determination, replenishment policy, safety stock levels determination, price negotiation with service providers, etc.
4. **Frequency of service**. It can happen that activities are not possible at an arbitrary moment of time. Limitations could be due to e.g. specific departure times of ships, limited working time of terminals, etc. Frequency of service has to be taken into account when lead time and its variability are measured. The less frequent service is, the higher is the lead time (lead time variability increases even faster).
5. **Shipment compatibility**. This performance measurement provides assessment of the possibility to use intermodal transportation in principle or compatibility of particular activities. Shipment compatibility can be expressed in terms of yes/no or degree of compatibility on a scale from 1 to 10. Compatibility factors include and not limited to
    a. **Loading units' compatibility**. This parameter assesses whether the shippers' and provider's loading units are compatible (i.e. whether they both use the same type of containers)
    b. **Business model compatibility**. This parameter assesses whether intermodal transportation can be used by the sender. Limitations are transit times, availability of service, sufficient stream of goods, etc.
    c. **Regulations compliance**. This parameter assesses whether the shipper's cargo comply with environmental and safety regulations for intermodal transportation

d. **Cargo compatibility**. Does the nature of cargo allow intermodal transportation (e.g. can be transported in containers, can withstand bumps while loaded/unloaded, need special treatment such as temperature controlled environment, etc)

6. **Damages during shipment.** There are several ways to measure damages, for instance, as
   a. Percent of damaged goods measured in terms of value/quantity
   b. Percent of caring units (e.g. 20 ft containers) that contained damaged goods
7. **Theft during shipment.** As in the damages case, there are several ways to measure theft, for instance, as
   a. Percent of goods stolen, measured in terms of value or quantity
   b. Percent of caring units (e.g. 20 ft containers) that has been stolen, broken in or experienced theft

4 PROMIT benchmarking tool design and implementation

4.1 General considerations

The main aim of the PROMIT benchmarking tool is to provide companies with means for assessment of their logistics performance. The secondary aim of the tool is to gather benchmarking information that can be used by policy makers and for standards setting purposes. Given the secondary role of the tool, the participating companies must be assured that their benchmarking information is used properly and is not disclosed, at least not on the company aggregation level.

It is tempting to measure various indicators. However, the large number of potentially useful indicators represents a problem: the tool must be simple enough to be easily understandable by the participating companies; on the other hand, the tool must collect meaningful information. Therefore, the number of indicators must be very limited and at the same time giving a consistent overview. In the current implementation of the tool we stick to the indicators on level 1 and 2, which are described in Chapter 3 of this report and in more detail in the 1st year PROMIT report.

Another issue which got extra attention is comparability of the measurements. For instance, if one company transports goods from A to B and the other one transports goods from C to D (or even from A to D), then a direct comparison of costs and transit times will not be meaningful, because these shipments are carried out over different distances and, possibly, within different countries. Therefore, results of the measurements must be somehow standardized, possibly by measuring cost per kilometer and transit time per kilometer (i.e. shipment speed). All these considerations are included into the design of the tool; however some extra attention should be paid to them while interpreting results.

4.2 Presentation method

When a company enters its data into the system, it gets feedback on its performance in the form of relative position in respect to the aggregate performance of other companies. The relative position of the company is important here; a visualized relative position gives a good feeling in respect to the performance.

Moreover, the logistics performance has to be measured using several indicators. An excellent performance in respect to one indicator is often coupled with underperformance in respect to other indicator(s). This principle reflects a common wisdom in logistics: a higher service level leads to higher costs. The companies are (presumably) more interested in a balanced picture, namely to see their position on the cost/service tradeoff. The following figure uses an example to represent the principle.

Figure 3. Example of presentation method for benchmarking transport operations

As it can be seen at the example on the picture above, the (fictitious) company has a good performance on transit (lead) time, excellent performance in respect to the damages; however the costs are high. This leads to a good overall performance.

4.3 Important issues over the presentation method

There are two major issues concerning the presentation method described above. *The first issue is related to the quality of the input.* If the system operates fully automated, an (intentionally) wrong input may severely distort averages. For example, if the average transportation cost by road is 1 Euro per container / km, and a company enters 200 Euro per container / km, it would increase average cost by the factor of 2 if the database contains 200 measurements. If the number of measurements is lower, the distortion will be even higher. As a solution, we propose to accept data only from authorized users and to have a process of "approval" of the data that is entered into the database. In case a human operator suspects a mistake, the data is not activated in the database; the company in question might be contacted for clarifications.

The second issue is the scaling problem. The scaling problem manifests in two ways. Suppose that we have performance indicators of two companies: company A with an annual goods flow of 1 000 ton km and company B with an annual goods flow of 1 million ton km. Should these two companies be given the same weight in the database (e.g. when the averages are calculated)? The second manifestation of this issue is calculation of overall performance of a company. The question here is whether all indicators should be given the same weight while calculating the integral performance indicator? The scaling issue can be dealt with at later stages of development. See the chapter on implementation of the tool for the chosen solutions.

These two issues above were discussed during the design phase of the tool. Currently implemented version of the tool makes some basic checks of entered information, for instance, the sum of distances specified for level 2 should be equal to the distance specified for level 1. Being a research tool, current software version does not implement full checks of data; neither there is workflow that assigns manual checks to people. Nonetheless, the tool has authorization functionality such that the source of data is always known. Thus it is always possible to sift manually through the benchmarking database.

4.4 Technical design of the PROMIT benchmarking self-assessment tool

This section describes the envisaged design of the PROMIT benchmarking tool, assuming a full-fledged and fully functional system that can handle rather big volumes of benchmarking data and provide a robust user interface for data collection and analysis. Having developed benchmarking framework with division on abstraction levels and specific benchmarks, we describe design of the tool as well as what compromises functionality / usability and implementation complexity can be done.

The main principle of the PROMIT benchmarking tool is that it stores benchmarking data in a database and provides users with feedback on their benchmarking data (i.e. relative benchmarking position of the company under consideration). A full-fledged tool should be based on client-server architecture: the users, using web browsers, should be able to enter their benchmarking data and get back assessments from the system. Further, in a broad sense, the tool should provide functionality in respect to the following 3 major tasks:

1. **Data collection functionality.** Users (companies) should be able to enter their benchmarking data in a convenient and unambiguous way.
2. **Feedback generation functionality.** The users (companies) should receive an immediate feedback on their data, in other words there must be implemented the principle of immediate gratification.
3. **Research functionality.** It should be possible to provide (restricted) access to the database for research purposes.

The following figure depicts the functionality and information flows of a full-fledged PROMIT self assessment tool.

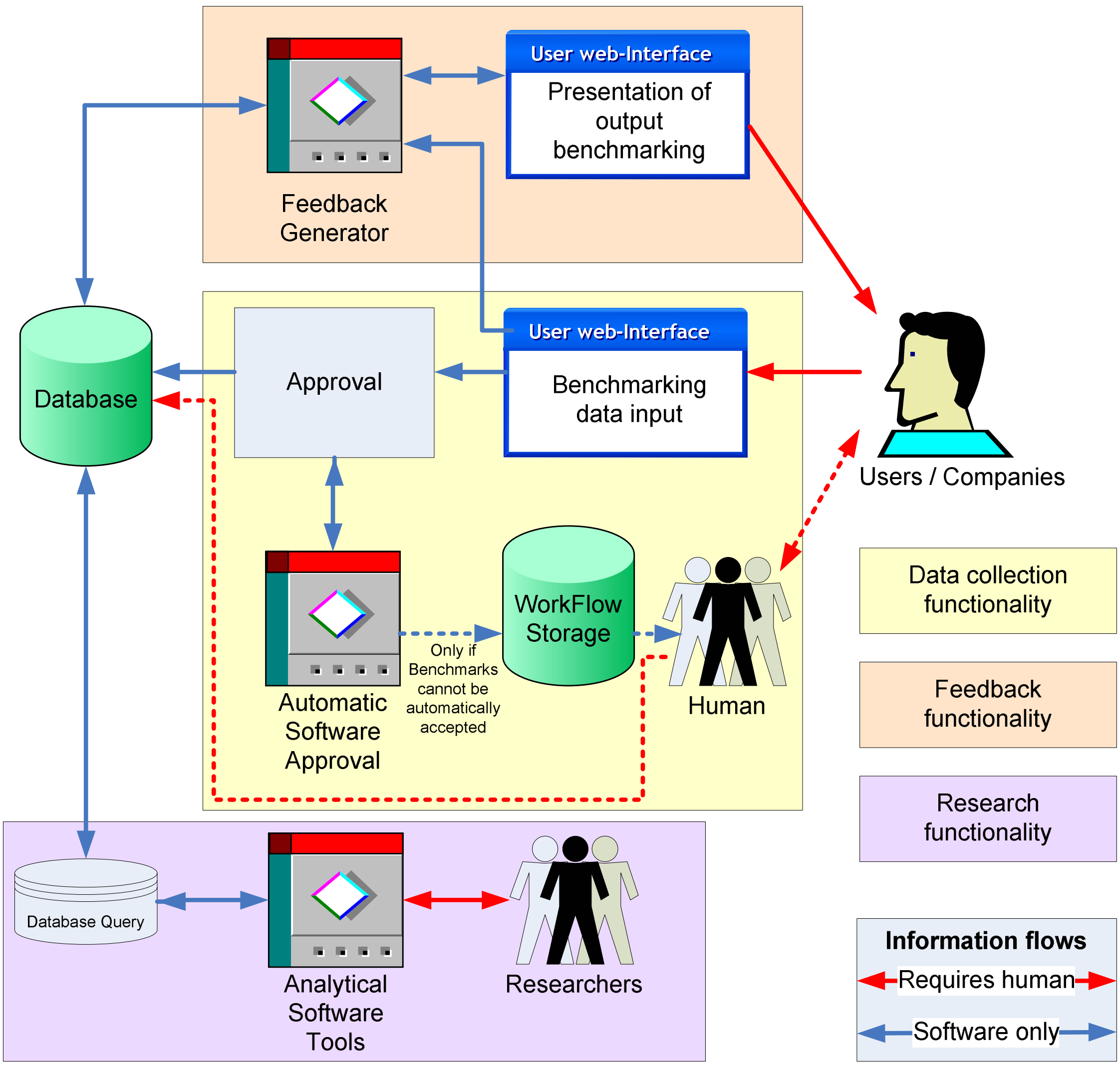


Figure 4, functionality and information flow of PROMIT benchmarking self assessment tool

In essence, the tool is an interface between users (companies), researchers and other groups of users and a database. The following list describes information flows within the tool and software pieces that are required to facilitate the flow.

1. The user enters benchmarking data of his company into a web-based form. There are many examples of such an interface: for instance online shops. They first require users to register (user identification) and then enter their data such as location, payment method, etc. For the benchmarking tool we apply a similar process: first companies should register, and then they should be allowed to enter their benchmarking data. Benchmarking data entry process depends on amount of knowledge that the company has on its operations, thus the application guides the user while he or she enters the data.
2. After entry, the data is split into two processes: approval of the data (before it is actuated in the database) and generation of the feedback (benchmarks) that will be presented to the user. The idea behind such a split is that wrong data should not be allowed into the database; however, the users should be given feedback immediately.

   Because of the possibility of mis-use of the benchmarking system, there was a discussion on whether it is necessary to provide immediate feedback or later, after the entered data has been approved. There are actually two points of view on when feedback is provided to the user
    a. Feedback is provided immediately
    b. Feedback is only provided when approval process is completed

   For more information on approval process, see section 3 of the list, which describes the approval process and motivates it.

   If there is a vigorous user-registration process, namely, not everyone is allowed to register or registering involves application process such that identities of companies (users) are checked, then feedback can be provided immediately. It is due to the fact that the companies would not want to spoil their reputation and "trustworthiness" within the system.

   The second way, namely providing feedback only after an approval, is better if there is an open registration process when every company can register with the benchmarking system without any identity checks. A user (company) in this case may have as many accounts as it wants and may mis-use the system, trying to study benchmarking data just entering garbage data. To prevent such a mis-use, it is indeed necessary to provide feedback only after approval.

   The position of TNO is inclined towards providing feedback immediately. This is because the first selection of users will be known and trusted. When the system is fully operational and if there is an open registration, then for the users that registered without background checks, the feedback data could be shown only after approval

   Note that the current implementation of the tool by PTV does not include the data approval process. In opinion of PTV, it is superficial for a research tool.
3. When the user has entered his benchmarking data, it must be approved. *Note that the text on approval process represents this process form the design point of view. It is not implemented in the current version of the tool.* Unapproved data can cause serious distortions in the database benchmarking data. Thus, there must be some measures taken to prevent outrageously wrong data to

be entered into the database. This can be implemented by a software filter in combination with some human operator assistance.

   a. First, the data entered is assessed by an automatic software assessment tool. The tool can be simple, for instance by allowing only data that is within some range of the averages in the database. The assessment tool can be somewhat smarter, for instance, allowing a balanced assessment of data. However, the main aim of the automatic assessment tool is to allow into the database only valid data: in more difficult cases there will be a need of a human assessment. Thus, the tool just simply reduces workload of human operators who are (in the end) responsible for correctness of the input. If data is OK according to the assessment tool, it is stored in the database. If the data is not OK or questionable, it is stored to be reviewed by humans.

      What important is that the software assessment tool sensitivity should be tilted towards false alarms (as opposed to allowing acceptance of wrong data), guaranteeing only correct data in the database.
   b. If the data is not OK according to the software assessment tool, it is stored as a workflow for human assessors. In principle, the storage physically may take place in the same database where the actual performance data is stored, however marked as needed to be approved (it is technically advisable, because the data is in the same format). The purpose of this storage is to decouple process of entering data from its confirmation, namely to make the processes asynchronous and independent (e.g. working hours, etc)
   c. Human assessors get data for approval from a workflow database. If the human assessor decides that the data is OK, it is immediately actuated in the database. If the human assessor does not trust the data, he or she can delete it or contact the company that entered data for further clarifications.
4. As we mentioned above, after entry the data processing is split into two processes: approval and generation of feedback, thus they can be executed asynchronously. The approval process may require some time, while the users (companies) may expect an immediate gratification: getting results instantly after data entry. For this purpose, there is a feedback generator software module. Main functionality here is to benchmark user's data with the averages from the database. Extended functionality may present some additional work, for instance, giving feedback while taking into account weights, different queries, and so on. The essence of the feedback generator is to perform queries and calculate the benchmarking results
5. The benchmarking results should be presented to the user: user web interface is intended to present data in a user-friendly form. The user web interface is discussed in more detail in Chapter 5 on Current state of the tool and benchmarking database.
6. Benchmarking DATABASE stores the benchmarking data. The database should be a relational transactional database. Ideally, an industrial database should be used: Oracle, MS SQL, or others that are typically used in the IT industry.
7. The benchmarking tool should also provide research functionality. It is difficult to foresee possible research questions that can appear in the future, however, the tool

should have some research and statistics gathering capabilities. It could be also necessary to restrict access to the confidential benchmarking data, namely providing access only to aggregated data, but not to the individual benchmarking records. Currently there is no research functionality implemented: the data is available in a tabular form for further analysis in, for instance, Microsoft Excel.

4.5 Implementation protocol and division of responsibilities

In this section we briefly describe how implementation of the tool was envisaged and what steps were actually taken. The following figure describes the envisaged development process with division of responsibilities.

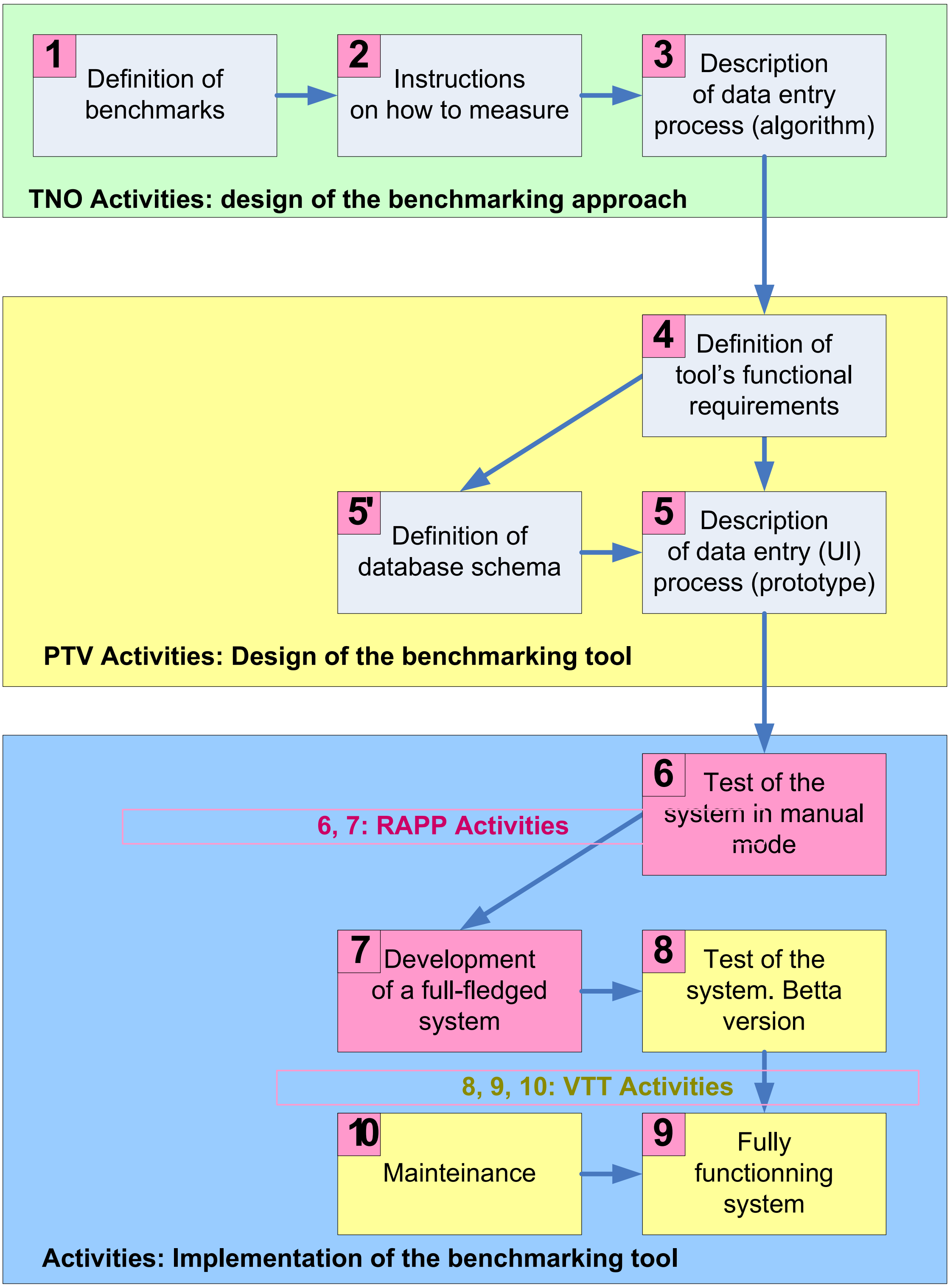


Figure 5. Workflow of the development process

The figure above is separated into three big logical groups. The **first group** concerns the development of benchmarking concept and benchmarks themselves. It has been specified in the first year PROMIT report (D4.1); the description of data entry process has been described in the second year PROMIT report (D4.2). The **second group** concerns the design of the software tool. This activity has been carried out by PTV. The **third group** concerns implementation of the tool, which has also been done by PTV. Actual implementation of the tool has some limitations in comparison to the envisaged functionality, for instance, there is no data approval process implemented. Contribution of other partners has been centered on data provision to get some critical mass of benchmarks inside the tool. RAPP, VTT, TNO and other partners contributed to initial seeding of the tool.

5 Current state of the tool and benchmarking database

In November 2007 the PROMIT benchmarking self assessment tool became operational. The actual implementation of the tool has been carried out by PTV, which developed web-based data entry interface for indicators on the levels 1 and 2. The first version of the tool does not implement all envisaged functionality. However, it allowed a quick start in gathering of benchmarking data. The web interface has some verification logic that tests validity of the data and stores entered data in a database. Also the process of identification of users has been implemented.

The current implementation of the tool represents a compromise over the envisaged design of a full-fledged tool. There are two main compromise points. The *first compromise* is in the validation of data entered: there is no sufficient validation process implemented. Thus data entered by the users of the tool become active in the database immediately upon entrance. The *second compromise* is in the benchmarks: not all benchmarks that have been discussed in the PROMIT first year report [D4.1] are implemented. The main reason for the first compromise is the complexity of implementation and maintenance process. The main reason for the second compromise in the number of benchmarks is the shire complexity of the data entry process and, namely, complexity for the end users of the tool.

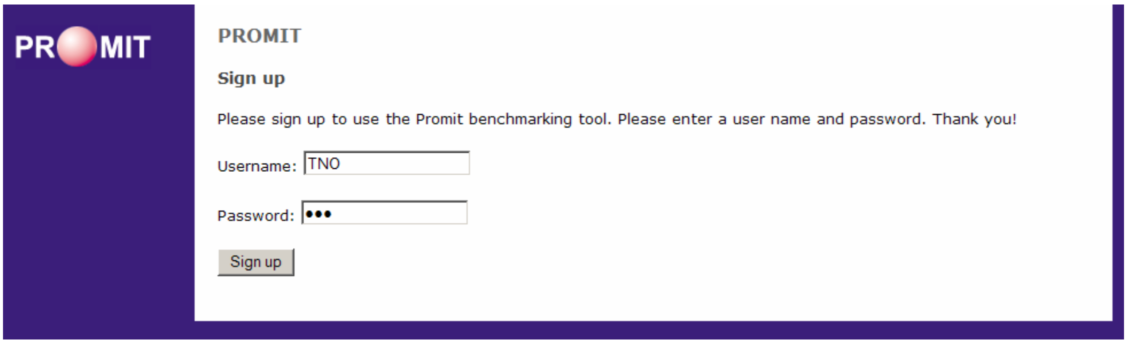
It is important to acknowledge that the tool is in a research phase of life cycle, namely it is operational with all necessary basic functionality implemented, but insufficient for a wide-spread use by third parties. The tool has been tested with entrance of benchmarking data and filled in with data that provides a critical mass of benchmarking information, providing a test environment for the tool and allowing beta testing. The tool has been filled in with some useful benchmarks and is attractive for some private users. However, we cannot say now that the tool has reached sufficient maturity to become self-sustainable.

Conceptually, the tool provides functionality in 3 main areas:

1. Registration of users (account maintenance)
2. Data entry
3. Benchmark generation (feedback)

5.1 Registration of users

The tool allows entering of benchmarking information only by registered users. Therefore, to be able to use the tool, a user needs to create an account. In order to do that, the user has to go to the sign up page, which is located at http://www.ptv.de/promit/register.php (there is a link to it from the main page).



The current implementation of registering is very simple. Only user name and password are required to create an account. No further information is asked about company identity and other information. Such simple registration process is easy for users, because it requires minimum time to execute and there is no worry about keeping company's identity confidential. However, it is not so handy for research and integrity purposes. Thus, we propose that the following versions of the tool implement more extensive registration procedure. When an account is created, the user needs to log in at http://www.ptv.de/promit/index.php. After the user is logged in, he may proceed with the data entry process.

5.2 Data entry

Data entry starts at level 1 indicators. The users must enter the following parameters

1. Origin and destination, which include
   a. Zip Code
   b. Town
2. Type of shipment: the user may choose from
   a. 40-ft container
   b. Trailer
   c. Swap body (13,60 m)
3. Distance between origin and destination points, km
4. Frequency of this type of shipment, in the following brackets
   a. Once per day or more
   b. Once per day – Once per week
   c. Once per week or less
5. Shipment costs in Euros
6. Lead time in hours (from door to door)
7. Reliability of service, defined as perceived deviation from specified arrival, time in the following time brackets
   a. Less than 1 hour
   b. From 1 to 3 hours
   c. From 3 to 6 hours
   d. From 6 to 12 hours
   e. From 12 to 18 hours
   f. From 18 to 24 hours

    g. From 24 to 36 hours
    h. From 36 to 48 hours
    i. From 48 to 72 hours
    j. From 1 to 3 hours
    k. More than 72 hours
8. Assessment of flexibility of the service provider regarding additional services on a scale from 1 to 10, where 1 is low and 10 is high
9. Assessment of quality of the information exchange with the service provider on a scale from 1 to 10, where 1 is low and 10 is high
10. Assessment of the quality of payment procedure on a scale from 1 to 10, where 1 is low and 10 is high
11. Transport modes that the transportation includes (multiple choice)
    a. Road
    b. Rail
    c. Inland waterways
    d. Short sea

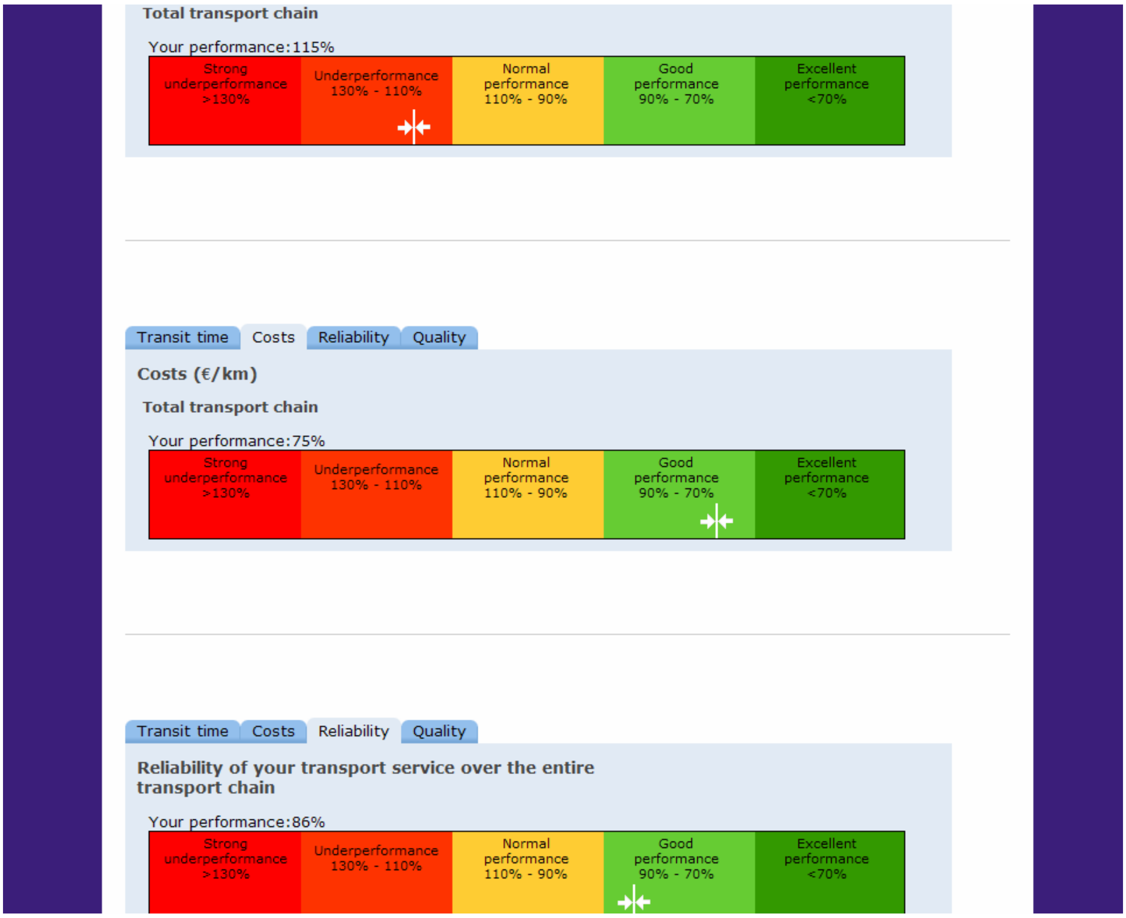
If the shipment involves more than one mode, then the user can enter level 2 benchmarks. He is asked to enter parameters described in the list below. What is important here is that the tool divides the whole transportation chain into segments (operations) specified at the level 1 and asks to enter KPI's consistently. The tool, for instance, checks that the sum of distances (and the sum of costs) specified in the level 2 segments is equal to the total shipment distance / cost specified for the level 1 indicators. The tool also makes sure that all compulsory fields are filled in with appropriate information, for instance, it is not possible to leave them blank or unspecified and it is not possible to enter some text in the fields that require numerical information.

1. Distance
2. Transit time
3. Cost

The process of benchmark acceptance / validation is not implemented in the current version of the tool. Implementation of them could be costly and maintenance of such a process would require a dedicated person(s) who would monitor entered benchmarks. This is a trade-off that is made in order to get the first version of the tool operational. If there is further development of the tool, it would be highly desirable that benchmarks entered are monitored and the acceptance process is implemented. In this way it could be possible to guaranty quality of the benchmarking database. There are also some benchmarks omitted, for example energy use. There is still an ongoing discussion on what should be measured. On the one hand, it is highly desirable to measure all indicators specified in the 1st year PROMIT report [D4.1]; on the other hand, complexity of such benchmarks could scare off potential users of the tool.

5.3 Benchmarking feedback generation

When the user correctly completes entering data, he immediately gets assessment of the entered benchmarks. The assessment is based on comparison per indicator to the averages of respective indicators in the database. For instance, if the average road cost per kilometer is 1,00 Euro per km and the user entered his cost of 1,20 Euro per km, then his performance is 20% worse than average and falls in the category of underperformance.



In addition to feedback generation, the tool adjusts the database with the newly entered benchmarking information. Thus, the database contains a new record(s) and following users will get their indicators compared to new averages.

5.4 Analysis of the database current state

In this section we describe the current state of the database. First, we present a general state of the database, analysis of the number of created accounts and describe information entered into the database. In the second sub-chapter we discuss the quality and properties of information entered into the database. As a general remark, we state that the database has reached the necessary ‘critical mass’ in order to make benchmarking meaningful, however, the number and quality of the benchmarks is still not sufficient to be a real tool for companies to get a detailed benchmarks of their operations.

5.4.1 Benchmarking database highlights.

Currently the database contains 113 benchmarking records, of which the majority of records are provided by one of the PROMIT partners, RAPP-TRANS. There are 37 user accounts in the database, which represent unique users of the benchmarking tool. Of 113 currently entered benchmarking records, only 5 contain information on level 2 indicators. It means that only these 5 records have more detailed information on operations of which door-to-door transport consists. These records benchmark all operations of intermodal transport.

The tool supports 3 types of loading units. These loading units can be though of as ‘full truck loads’, making comparison with single mode road transport quite straightforward. In the benchmarking database there are 29 records (26% of total) over transport assignments that use 40-ft containers, 24 records (21% of total) for transport assignments carried out in swap bodies and 60 records (53% of total) for transport assignments carried out in trailers.

There are 37 benchmarking records that describe intermodal transport assignments (5 of which have level 2 indicators), while remaining 76 cases represent single mode road transport. The distribution of modes used in intermodal transport is the following

- Road + Rail: 18 cases
- Road + Rail + Short Sea: 1 case
- Road + Short Sea: 6 cases
- Rail + Inland Waterways: 2 cases
- Rail only (intermodal, though not purely intermodal): 10 cases

5.4.2 Detailed analysis and quality of benchmarks

In this sub-chapter we take a close look at benchmarking data, which is currently available in the database. First, we look at the cost data, normalized per loading unit kilometre and then we look at transportation speed. These measurements have been made for each transport mode and as a function of transport distance. We do not distinguish loading units in this analysis, as all three loading units fall into ‘full truck load’ category and do not differ much in their cargo load capacity.

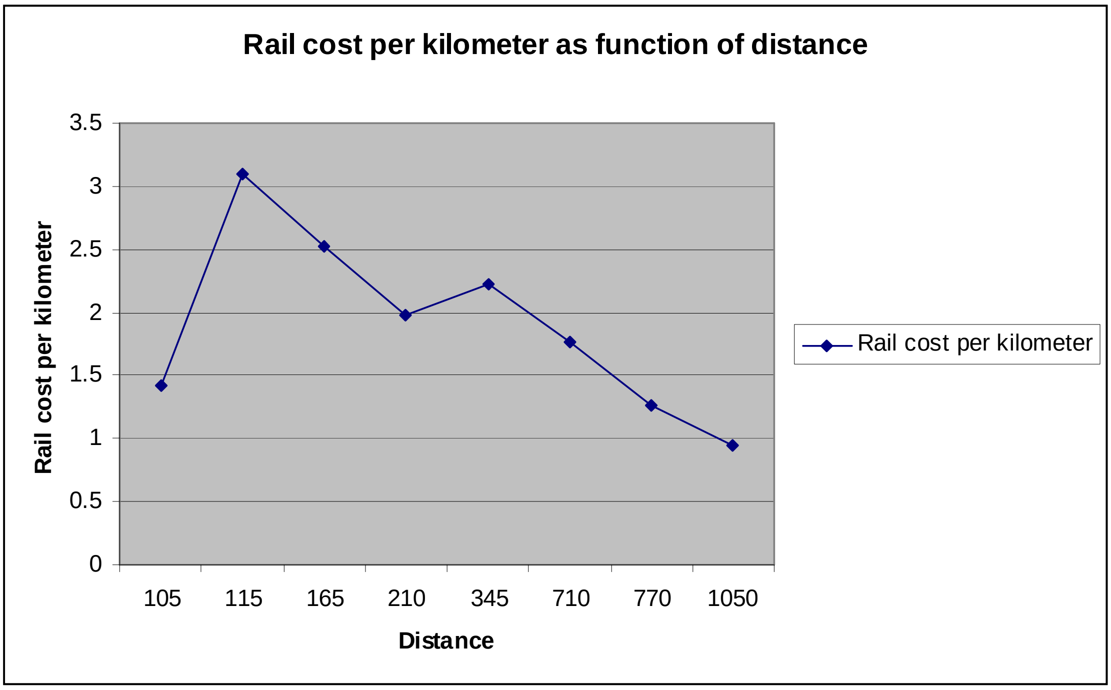


Figure 6. Loading unit rail transport cost as a function of distance

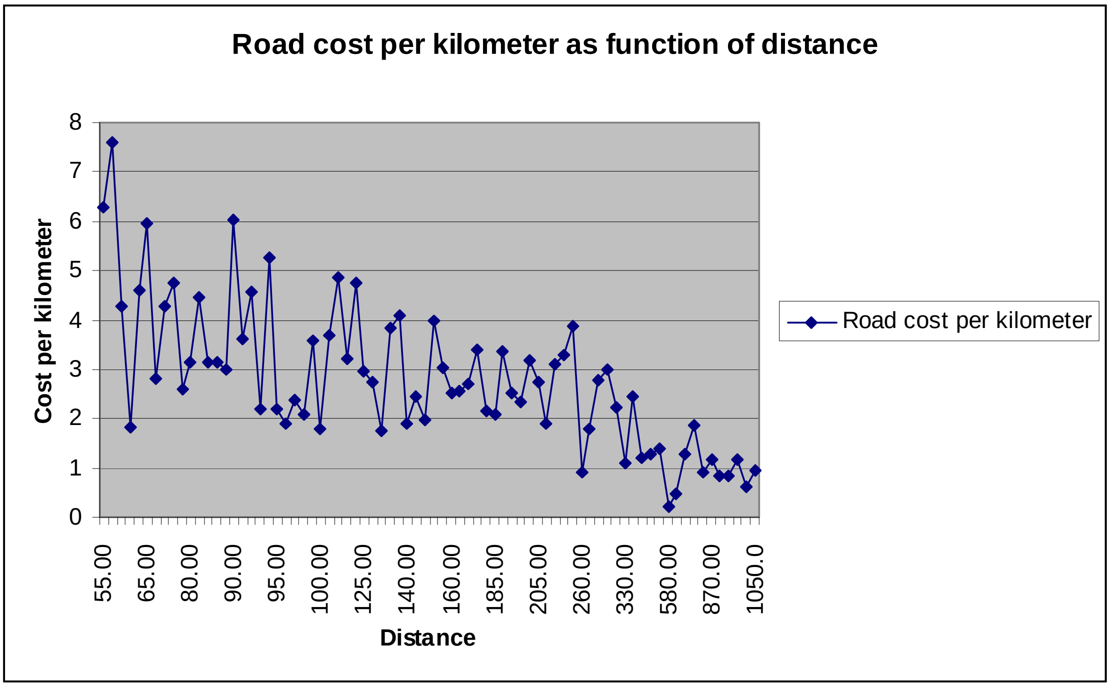


Figure 7. Loading unit road transport cost as a function of distance

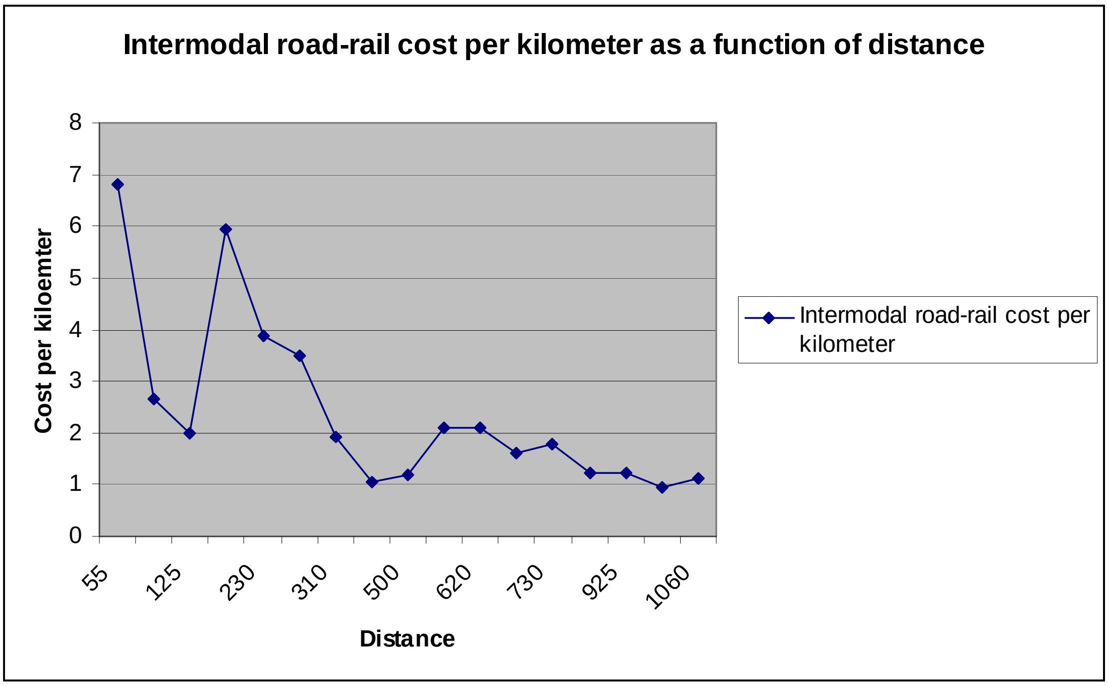


Figure 8. Loading unit intermodal road-rail transport cost as a function of distance

All three graphs show a general trend in decrease of transport cost per kilometre as the distance of transport assignments increases. The graphs show a relatively high degree of "nervousness" for short-distance transport, which is quite explainable in terms of different environments: some transport assignments might involve substantial waiting time, city environments, etc.

The reader can draw an observation that intermodal transport is not generally cheaper than the single mode road transport. Even single-mode rail transport on distances 500-700 km is somewhat more expensive than the road one. This fact might explain general difficulties with intermodal operations: they appear to be more expensive than road transport, while having other disadvantages in comparison to road. Though we do not possess enough data to draw statistically sound conclusions, the available observations confirm general cost disadvantage for rail and intermodal solutions.

A close look at rail intermodal transport costs shows surprising cost behaviour in the distance range of 300-600 km. The cost curve dips for distances around 500km, while for longer trips it starts rising again. There are a number of possible explanations for such behaviour: it could be due to "out shooters" or due to international border crossing process in train transportation, which brings substantial extra costs.

Generally we conclude that the data is within the range of expected values and thus can be trusted. Having observed a number of phenomenon and out shooters in the graph, we stress that the tool needs a data acceptance module, which would eliminate erroneous data and provide explanations for unexpected benchmark values.

The following three graphs show average transport speed as a function of transport mode and trip distance. The graphs are drawn in the same way as the graphs for transport costs.

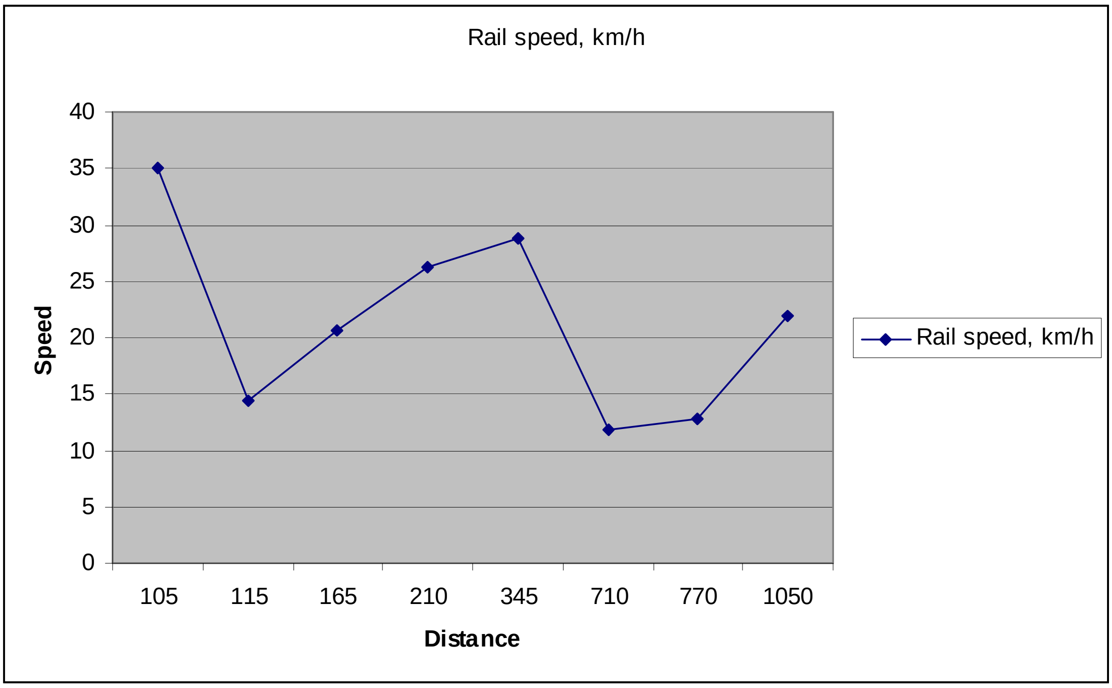


Figure 9. Average rail transport speed as a function of trip distance

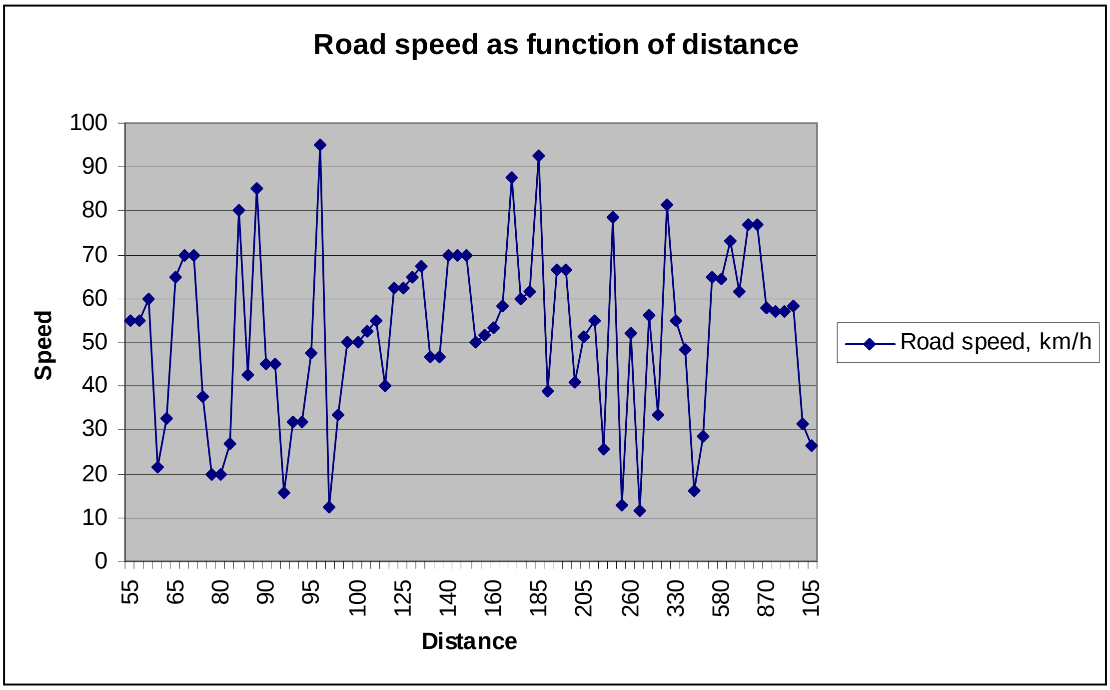


Figure 10. Average road transport speed as a function of trip distance

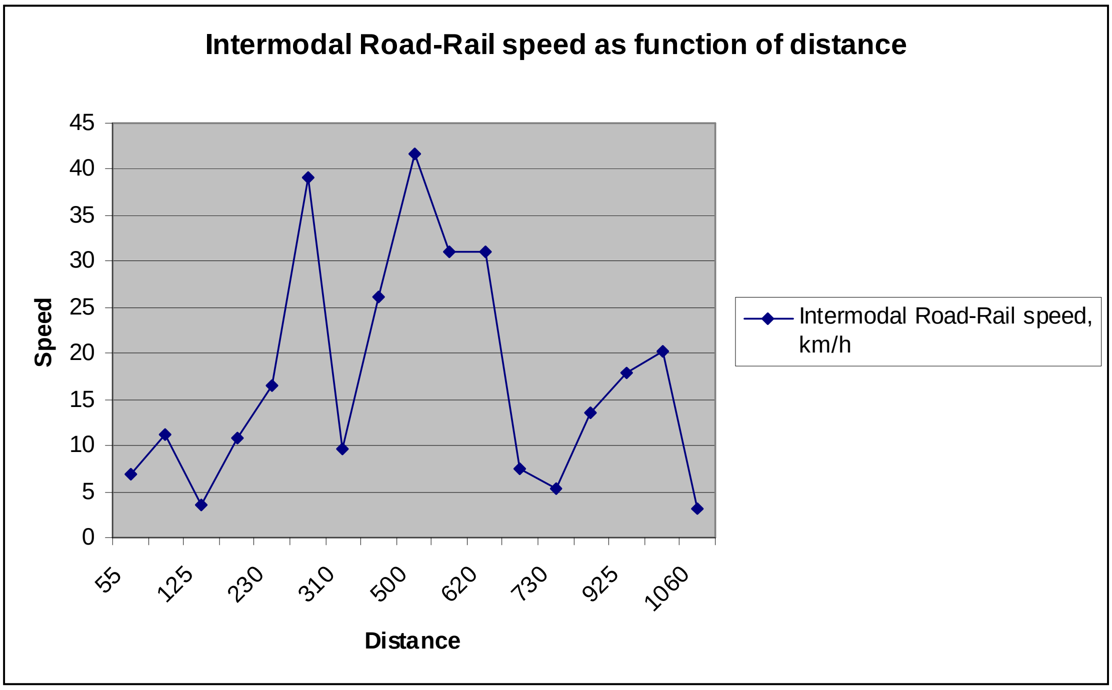


Figure 11. Average intermodal road-rail transport speed as a function of trip distance

Analyzing average speed of rail and rail intermodal transport, there is no clear message coming out from the graph: the data show a high degree of volatility. Moreover, the database does not contain enough measurements of transit time, given such a degree of volatility. Only one conclusion is clear: rail transport is generally slower than road transport.

The database contains more measurements on road transport time: there are some conclusions possible. First is that average speed increases with the distance of the trip, however, up to a certain point. On sufficiently long distances, for instance of 500 km and more, the average speed may decrease as a result of the driving time regulation: truck drivers must spend sufficient resting time. Theoretically, resting time regulation should make long distance transportation time quite comparable between single mode road and intermodal transport.

As a conclusion over data analysis, we state that the current 113 records are short of the goal of 150 records, but the records give a reasonable seeding of the tool, allow testing of the concept and give a 'taste' of the data. There is an apparent lack of level 2 indicators data in the database. The quality of data is uncertain as we observe a big degree of deviations and nervousness in the benchmarks.

6 PROMIT benchmarking tool test at the Dutch market

TNO has invested a considerable effort in development of the PROMIT benchmarking tool. As the project is approaching its end, it has been decided to test applicability and usefulness of the tool in the Dutch market. To do that, TNO has invited representatives of industrial companies that are active in the Netherlands. In total there were 12 representatives of industrial companies present and the Ministry of Transport in the workshop.

During this PROMIT benchmarking tool workshop, TNO has given a presentation on the PROMIT project, design of PROMIT European benchmarks and showed a test case of entering benchmarks and obtaining feedback in the online benchmarking tool. The goal of these presentations has been to hear opinions of the workshop participants on the usefulness of the approach and tool, as well as to test the interest of the participants in further development of the tool.

6.1 Feedback from workshop participants

The presentations were followed by a lively discussion. Here we report on discussion, critics and suggestion points. The following list summarises the discussed issues, after which the reader will find main conclusions that we draw from the workshop.

1. **Concerning feedback generation on entered benchmarks**
    a. How the entered benchmarking values are compared to produce output comparison? Could it be that the tool compares incompatible values (oranges and apples)?
    b. What is the usefulness and value of generated feedback information? Can it be that comparison of services of intermodal service providers gives more information, especially on lane / corridor basis? For instance, if the tool had information on services and schedules of Railcargo.
    c. Probably a simulation module in such a tool would be useful: database of the tool should accept only real-life benchmarking data, while there is also data on, for instance, other corridors, which is generated by a kind of simulation module. In this way the tool could achieve a bigger geographical coverage and per lane comparison.
    d. There are some very specific transport needs, for instance transporting potato starch from Northern Netherlands to Italy by train. Parameters of such shipments could be very specific and differ substantially from other transport assignments on the same lane.
    e. The quality of feedback generated by the tool depends very much on the quality of data in the benchmarking database.
    f. The focus of the tool could be directed at specific industrial regions
    g. There are much more important factors for meaningful comparison of performance indicators: specific origin-destination pairs, volume of the

shipment, imbalance observed in the corridor or trade lane (i.e. empty return trips), nationality of drivers, etc

   h. Big organizations such as expeditors and logistics service providers can have much better agreements with providers of intermodal services, such that the price and quality of intermodal transport could depend on the size and buying power.
   i. How will the data be kept up-to-date? For instance, fluctuations in the fuel price are decisive for the final transport cost within a lane

2. **Concerning KPIs implemented in the tool**
   a. More hard indicators on reliability of the service are necessary, for instance, damages and theft
   b. The way in which indicators are presented is not always clear. For instance, if reliability is more than 100%, is it better or worse than the average?
3. **General remarks**
   a. Some loading units simply not fit for intermodal rail transport. This would make comparison nonsensical.
   b. Services offered on the market are often line-services. If connections from hubs such as Duisburg are taken into account, the tool would get the functionality of a scheduler.
4. **Conclusions of workshop participants**
   a. The participants found a benchmarking system for intermodal transport to be an important initiative
   b. The current state of the tool does not satisfy direct needs of companies, which sent representatives to the workshop
   c. There are good chances that a niche benchmarking tool would be directly of use for industries at which it is aimed
   d. Development of a universal benchmarking tool is a very complex and challenging undertaking, which can require a very substantial effort

6.2 Conclusions from the workshop

The workshop has shown that there is quite a lot of interest from the industrial companies towards intermodal benchmarking data. A clear indicator of this interest has been the willingness of the companies to be present at the workshop and to get acquainted with the recent developments in this field.

The workshop has also showed deficiencies of current implementation of the tool, as it cannot satisfy now many important requirements of the industry. Though being a research tool, its current implementation is not intended to be a commercial service. On the other hand, the workshop has shown rather narrow-minded interests of the companies, which are willing to benchmark their performance on very specific lanes, cargo types, loading unit types, etc. This means that only a fully mature benchmarking tool can capture interests of the companies, at least this is true for representatives who took part in the workshop.

Therefore, the main conclusion that we draw from the workshop is that there is quite a lot of demand for an intermodal benchmarking tool. But to become commercially viable and self-sustainable, such a tool needs to be much more sophisticated and much more developed. Specifically, the tool should not be blind in respect to lanes, goods flow direction and commodities in question. Further, the envisaged functionality of entered data checks must be implemented.

To our knowledge, these benchmarking-related questions are being tackled in the BELOGIC project. Therefore, we suggest that some form of knowledge transfer from the PROMIT project to BELOGIC is of vital interest of the later project. Besides extensive documentation of the PROMIT benchmarking concept and the tool, there is a vast volume of tacit knowledge that can be utilized in BELOGIC. Building up on the existing knowledge would make the advancement of BELOGIC faster and more profound.

7 Conclusions and recommendations

During the course of PROMIT WP4 a benchmarking concept, a benchmarking methodology and a self assessment tool have been developed. This chapter draws conclusion and provides recommendations based on three year experience of PROMIT.

First, intermodal transport is an intrinsically complex form of transport, which requires more synchronized operations than the single mode road transport. It is also less flexible, as trains and ships travel according to schedules. Yet another prominent feature of intermodal transport is that carrying capacity of a train or a ship is much bigger than capacity required for one loading unit, for instance, 40-ft container. This feature requires bundling of cargo to utilize transport capacity.

Given these features of intermodal transport, we used division of it into three levels of abstraction. At level 1, indicators related to door-to-door transport are measured: they look holistically at the whole transport chain and are of great interest to the shipper. At level 2, indicators are related to individual operations in the transport chain, such as line haul, road feeder transport, etc. These are of interest to logistics operators, who do plan the trip. At level 3, indicators are related to low-level operations that occur within logical steps of level 2, for instance measures of such terminal-related operations as truck unloading, movement of container, etc.

We propose to measure performance at each of the three levels of abstraction: these levels imply different nature of the measurements and different applicability of the indicators. However, the indicators themselves are very similar across the levels of abstraction. We have identified that at all 3 levels the following indicators are of vital interest: price, lead time and its variability, shipment compatibility, theft and damages. At level 3 we have identified some extra indicators such as capacity, capacity utilization and reliability.

In addition to the main KPI, we have proposed to measure quality indicators and indicators of eco-performance. These include flexibility, information flow, invoicing accuracy, relationship with shippers, market assessment, energy consumption (and hence CO_2 footprint). Despite the fact that there is quite a lot of interest in measurement of these indicators, some of them are very subjective, while assessment of energy consumption is often too difficult for companies to grasp.

The self assessment benchmarking tool has been designed to let shippers and logistics providers assess performance of intermodal transport operations. On the one hand, the tool is a research undertaking, on the other it is intended to provide interested parties with a practical way to compare performance of their operations to those of others, often their competitors. Thus, the tool has been designed in a way that allows companies to safely benchmark their operations in respect to nondisclosure of their benchmarks. The tool is not information hungry, while capturing main performance indicators of intermodal transport.

During the last year of the project, the PROMIT consortium managed to seed the tool with a critical number of benchmarks which make it possible to assess performance of commercial companies, should they be interested in doing that. Currently the tool contains 113 benchmarking observations, which give reasonable average values to benchmark operations.

During a workshop on the PROMIT benchmarking tool it appeared that there is a substantial interest coming on the part of shippers and logistics service providers in benchmarking intermodal operations. However, industrial parties require a tool with a higher level of sophistications: it is desired to have functionality that distinguish benchmarks on the basis of lanes, goods flow direction and commodities.

It is desired to extend the tool and benchmarking concept with at least lane and commodity functionality. Such advanced tool would not be able to rely only upon benchmarks as empirical observations, because the amount of data necessary would be beyond the realm of collection. Thus, it would need advance modelling and statistics methods which are currently being developed in macro transport models, for instance such as TRANS-TOOLS. The BELOGIC project works on advanced benchmarking methods, thus there is a clear link between RTD activities which have been done in PROMIT and current activities in the BELOGIC project. Knowledge transfer from PROMIT to BELOGIC would further advance development of European benchmarks of intermodal transport.