EPR PILOT APPLICATION SUMMARY OF THE DATA QUALITY RESULTS

1. Introduction

This document aims to illustrate the outcomes of the Pilot Application carried out in the framework of the European Performance Regime (EPR) project. The Pilot Application (PA) consisted of a test of the whole EPR system designed by the EPR project team since 2004 and described in the EPR Handbook version 2009. The purpose of the test was to verify the feasibility of the system and the fairness of the commercial model, in order to provide the information basis (both quantitative and qualitative) to update the Handbook (whose new version has been delivered in December 2012).

The test was carried out through the practical application of the system to a chosen sample of international trains by the so-called "Early Implementers", i.e. IMs and RUs who, in May 2010, voluntarily agreed to participate in the Pilot Application:

- Austria :ÖBB and RCA
- Belgium :Infrabel and SNCB Logistics
- France :RFF and SNCF
- Germany: DB Netz, DB Fernverkehr and DB Schenker
- Italy :RFI and Trenitalia
- Netherlands :Prorail and Keyrail
- Switzerland :SBB and BLS (IMs and RUs)
- Luxembourg: CFL

Some partners actively contributed to the Pilot Application by nominating a representative in the EPR Operation Working Group (OWG), others made some of their trains available for monitoring.

The PA implied two major phases: preparation (2010/2011) and actual test with production of reporting (2011/2012).

1.1 Preparation

The preparation of the PA consisted in the nomination of the Corridor Coordinators, in the development of the supporting IT tools and definition of the reporting structure.

1.1.1 Corridor Coordinators

The monitored trains were distributed according to "Corridor Lines" (normally, Origin-Destination links) and "Corridor Groups" (groups of Corridor Lines).

Each Corridor Group was under the responsibility of a Corridor Coordinator whose tasks can be described as follows:

The EPR Corridor Coordinators (EPR-CC) investigated the data quality on "their" EPR routes, coordinated possible improvements with the other OWG members and reported

the results of their work to the OWG. The EPR-CCs took care, involving the other concerned partners, of keeping the EPR-trains list updated and the EPR trains list renewal (at due time). Moreover, the EPR-CCs steered the process to temporarily suspend the EPR validation activity in case of relevant operational accidents, lasting traffic interruption and/or to other major operational problems (also including IT systems failures) having a significant impact on the corridor traffic management and/or data quality. More detailed can be found in the official document: Corridor Coordinator Task description V2.0-2012

List of corridor coordinators:

Corridor Group or line	EPR – CC
FR-LUX	Dominique LeHong (SNCF)
FR-DE	Christophe Cochelin (RFF)
FR-CH	Christophe Cochelin (RFF)
FR-CH (TGV)	Karl Guntern (SBB Infra)
IT-CH	Karl Guntern (SBB Infra)
C02 – Köln/Gallarate	Siegfried Nierichlo (DBNetz)
C02-Rotterdam/Novara	André Beerthuizen (Prorail)
C02 – RoLa	Alexander Paulus (BLS Infra)
C04 – Bilateral	Christian Svatek (ÖBB Infra)
C04 – Trilateral	Siegfried Nierichlo (DBNetz)
C05	Ann Verstraelen (Infrabel)
C08	Roberto Caruso (RFI)
AT-DE-AT	Christian Svatek (ÖBB Infra)

1.1.2 Development of the supporting IT tools

Data were collected through the RNE "Train Information System" (TIS - for more information visit: http://tis.rne.eu/). Data quality checks and validation of the delay causes (see 1.2) required the development of specific IT tools. The whole OWG was responsible of supporting RNE in the development of such IT tools, in order to verify their working, identify bugs and suggest useful developments. The result of this activity was the delivery of the EPR tools (data quality check, validation and calculation). For more information about the tools please check the EPR Handbook and related annexes.

1.1.3 Reporting structure

The definition of the reporting structure took a relatively long time. Several attempts were made in order to fulfil all the requirements, namely:

- To provide a monthly report
- To respect the time constraints: due to the validation procedure (see Handbook chapter 5), definitive data were available only the second week of the second month after the trains had run
- To choose relevant indicators
- To produce reports which could be understandable at high level but also usable for the working level
- To minimize the (manual) work load

The final decision was to produce (see Handbook- Annex 3 for details):

- High Level Report (HL Report), with the aim of informing management (especially the EPR Advisory Group) of the ongoing activities and results. It was mainly focused on data quality and contained a general part and a Corridor Groups´ related part;
- Detailed reports, with the aim to deeply analyse the reasons for bad data quality and decide the appropriate corrective actions;

The HL report was mainly produced by the OWG Chair at RNE, while the Corridor Coordinators were responsible to provide information on their respective Corridor Groups and to insert their comments in the HL report.

Only the HL Report was periodically distributed while the operational reports were only used for day-to-day business.

Reporting was supported by the Oracle Discoverer IT tool. This report makes use of the new RNE reporting tool: Oracle BI.

1.2 Actual test

The actual test was carried out from October 2011 to November 2012. It consisted on the monitoring of data quality along the chosen corridor routes (including manual exclusion of trains), the train list update when necessary (including suspension for specific trains in case of exceptional events) and validation of delay causes.

2. Results of the Pilot Application – foreword

The following pages describe the results of the Pilot Application in quantitative terms. From a qualitative point of view, the achievement of the Pilot Application can be summarized as follows:

- Support for the development of a really usable IT tool
- Setting up of an effective procedure for the creation and update of the train list
- Identification of major problems for data quality along the line and solving of many of them
- Improvement of the data quality

In this document, the quantitative results of the Pilot Application are summarized. The considered period is from November 2011 to October 2012, in order to have one entire year available (November 2012 could not be considered as the validation period was not concluded when drafting this document).

For all following items figures will be illustrated accompanied by important remarks. Figures regarding punctuality, delay causes and "virtual" penalties produced are not included in this report as they are illustrated in Annex 6 of the Handbook and the document 10_EPR_COMMERCIAL RESULTS FINAL REPORT.

During the Pilot Application it was found out that not all Corridor Groups reached the established data quality minimum target to use the related data in the evaluation of the commercial model (see Annex 6 of the Handbook for more explanations). Therefore, in the monthly HL report, the Corridor Groups that were put in "stand-by" (C02-Rotterdam-Novara and C05) for this reason were illustrated separately from the others. The same structure is maintained here.

3. Results of the Pilot Application – general part

3.1 Targets

In order to evaluate the fairness and feasibility of the Commercial model, the EPR Commercial WG (CWG) needed to have a sample with sufficient quantitative and qualitative features. To this aim, the CWG itself set the following targets, to be reached:

- Quantity targets (total in 5 months):
  - 4.000 passenger trains
  - 3.000 freight trains
- Quality targets (each month for at least 5 months):
  - At least 80% of included trains and passenger trains
  - At least 80% of included trains for freight trains
  - At least 60% of included trains by Corridor Group

For the aims of the HL Report on data quality aspects, after evaluating the first results of the monitoring, it was decided to refer to the following targets:

- At least 80% of included trains for both freight and passenger trains
- At least 60% of included trains by Corridor Group

The difference is due to the fact that, for operational evaluation purposes, in the calculation of the included trains, the cancelled trains are not considered (therefore the percentage is calculated on the trains that actually run, not on the monitored trains).

It must be said that during the pilot application, additional exclusion rules where applied that were not established when the targets were decided. Therefore, even if the freight traffic did not reached the target for 5 months the data quality was deemed sufficient to carry out the commercial model evaluation anyway, because it was normally above 75% (for 11 months on 12).

3.2 Sample monitored

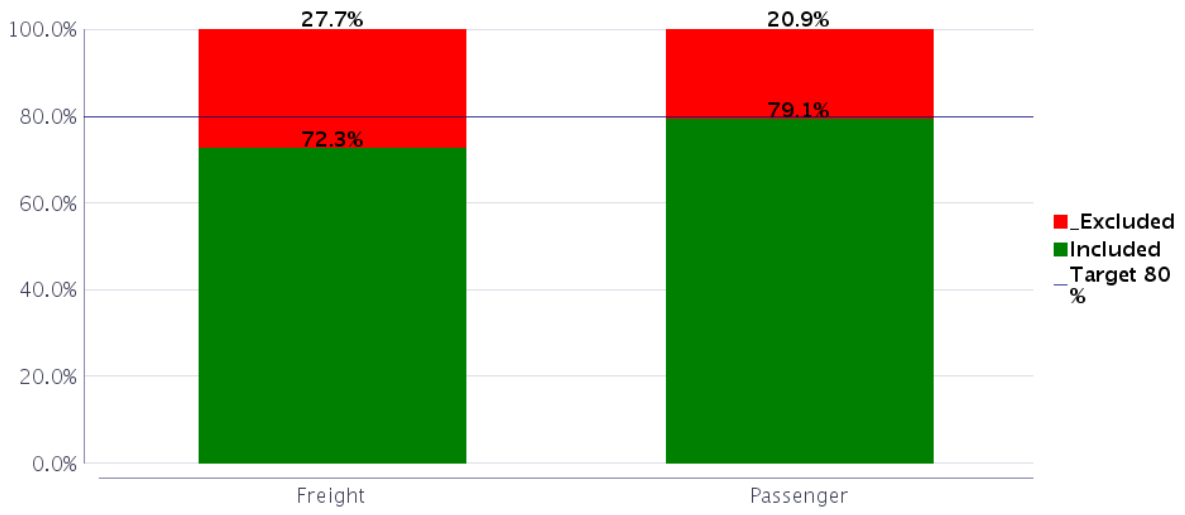
The sample of trains to be monitored had to be agreed among the partners. The criteria to choose the trains were not established but it was recommended to take the possible influence on data quality into account.

During the PA a total of 49.983 trains were monitored, distributed as follows:

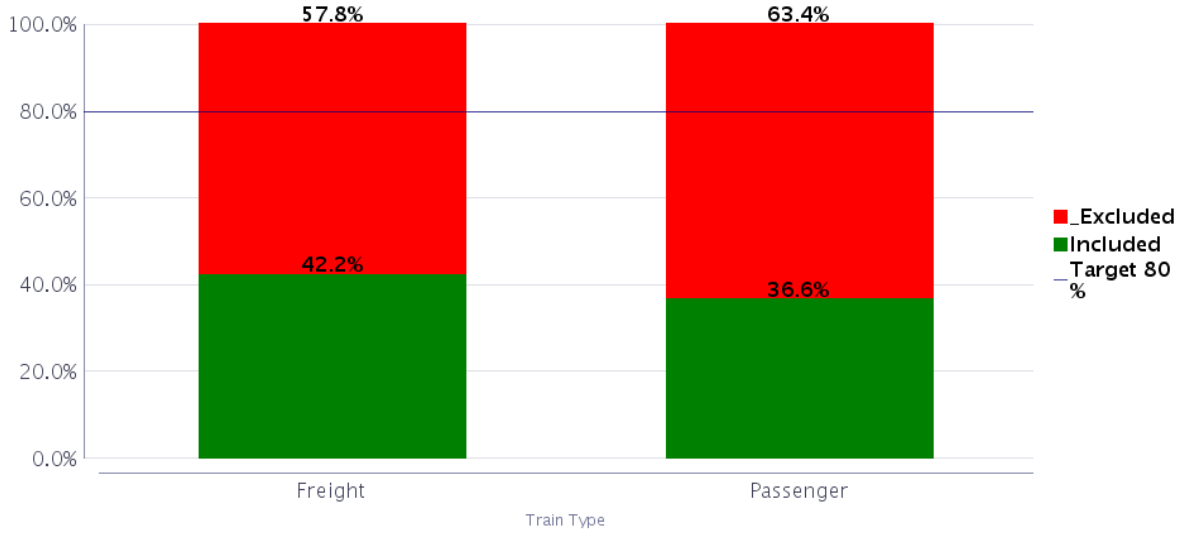
- Passenger: 31.238
- Freight: 12.496
- Standby: 5.249

3.3 Data quality

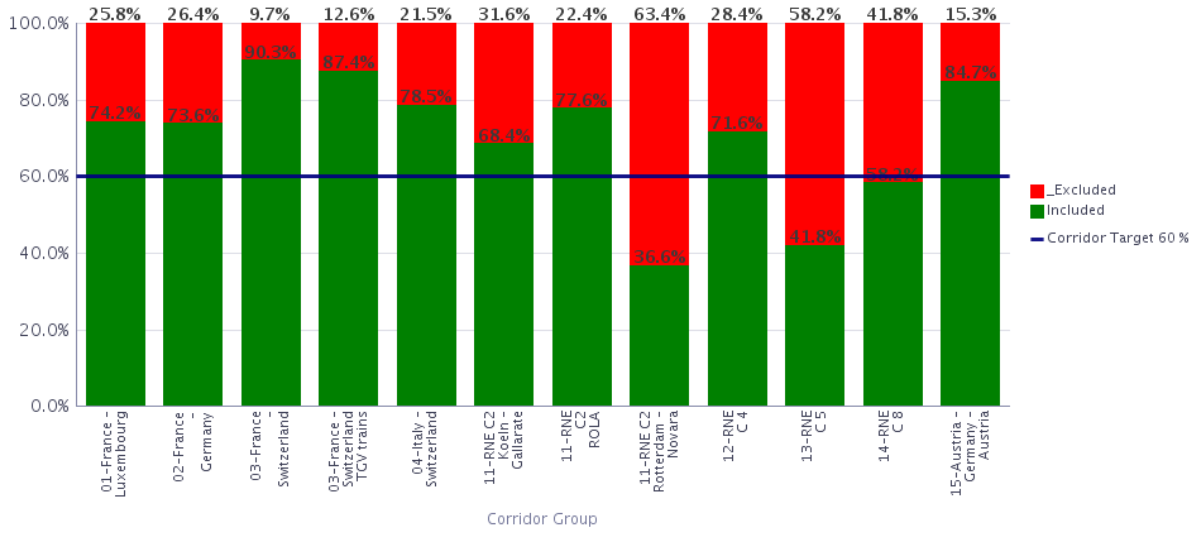
The overall data-quality for the considered period is shown in picture 1 (Corridor Group considered for the evaluation of the commercial model) and picture 2 (standby-corridor groups).



Picture 1- Data quality *per* type- overall November 2011-October 2012 – excluding standby corridor groups

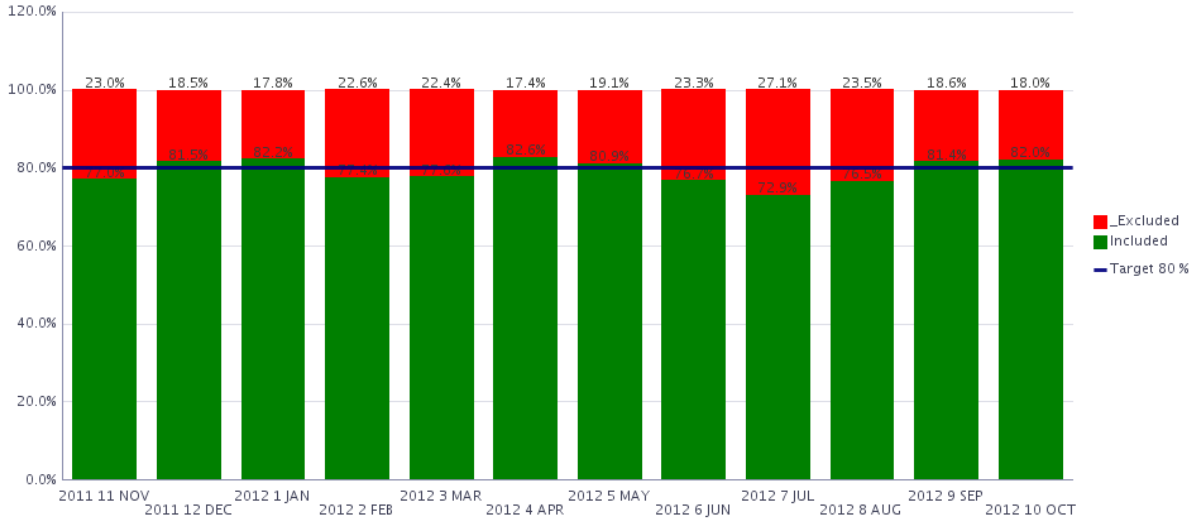


Picture 2- Data quality *per* type- overall November 2011-October 2012 – standby corridor groups

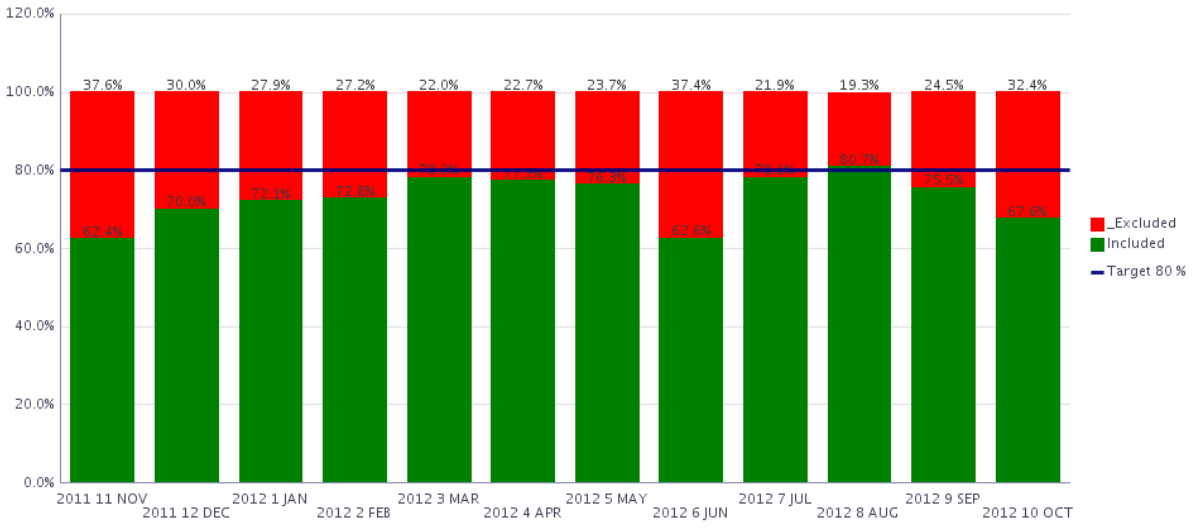


Picture 3 - Data quality *per* corridor group- overall November 2011-October 2012

For C04 and C08 data are partially distorted because in some months the validation was suspended due to force majeure events. In these graphs data could not be "cleaned up" from these distortions. More reliable information on these corridors can be found in 4.9 and 4.11.

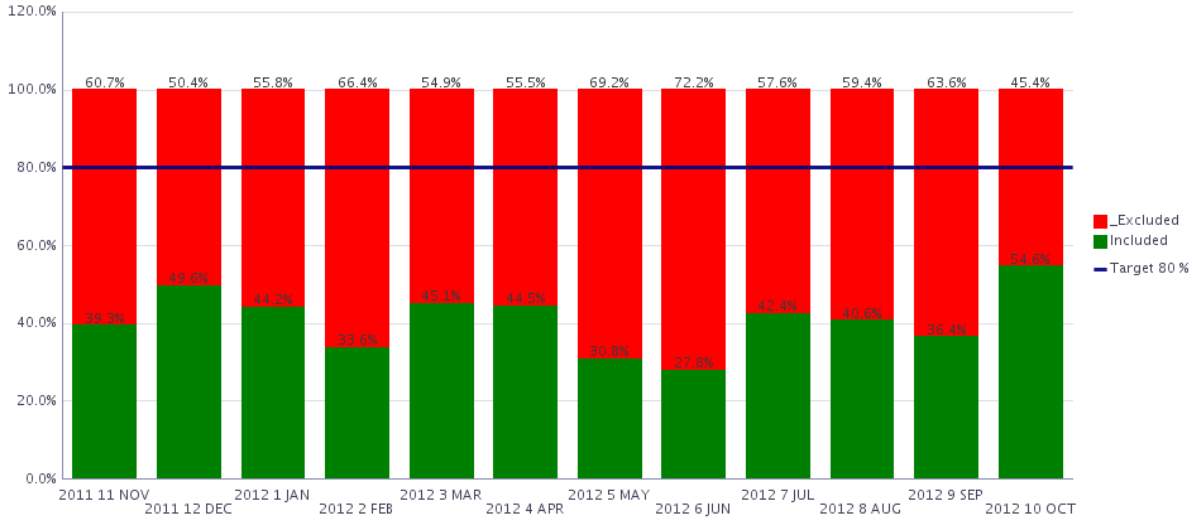


Picture 4- Trend of data quality - passenger traffic - November 2011-October 2012



Picture 5 - Trend of data quality - freight traffic - November 2011-October 2012

An important remark regarding picture 5 must be made, concerning the drop of data quality in June 2012. In fact, during this month big disturbances on the traffic on the Gotthard and Lötschberg lines occurred due to a landslide.



Picture 6 - Trend of data quality - standby corridor groups - November 2011-October 2012

3.4 Reasons for exclusion

Table 1 shows the three main reasons for exclusion, calculated on the overall sample in the whole period. Nevertheless, it must be taken into account that differences exist among corridors and months.

Reason for exclusion	Percentage on total excluded trains
Missing Running Advice	58%
Missing CTT at border	12%
Train run not complete	9%

Table 1 – top 3 reasons for exclusion - November 2011-October 2012

The "Missing Running Advice (RA)" occurrences show a stable trend, with all corridor groups affected. In many corridors groups the problem has rather improved during the PA (for example in C08 the affected trains decreased from 46 in November 2011 to 8 in October 2012), in many case the level remained steady. In October 2012 there was a peak for a couple of Corridors that were improving until the previous month.

Missing CTTs at border mainly influence data quality in Rotterdam-Novara and C8 (33% of excluded trains). Rotterdam-Novara is improving (from 66 trains in November 11 to 6 in October 2012). In C8 seemed this problem seemed solved in September 2012 but again in October many occurrences appear. In the RoLa traffic there was a peak in November 2011 but the problem was fixed. Finally, the bug also appeared in the TGV traffic between Switzerland and France in February/March it has almost disappeared.

Corridor groups mainly affected by the incompleteness of train run are:

- RNE C2 Koeln – Gallarate
- RNE C2 ROLA (peak in June 2012 but sensibly decreasing since then)
- RNE C2 Rotterdam – Novara (where it accounts for 36% of cases, but sensibly decreased since May 2012)
- RNE C 8 (especially September/October 2012)
- RNE C 5 freight (constant trend of around 20% of excluded trains affected)

In some cases, due to wrong/incomplete information contained in the Delay code message (UIC message 2005) the tool cannot identify the partner responsible for the delay (and, in the table containing the information on delay codes, it displays a "?" in the column "company responsible for the delay"), therefore, no calculation of results is possible for trains affected by these problems. The reasons behind such cases can be several. The most diffused are: wrong usage of an international delay code by the responsible IM, wrong time of occurrence in the UIC message, missing responsible RU code in the UIC message, wrong point status in the UIC message.

The excluded trains because of question marks (accounting for 5% in the overall sample) are relevant especially for the following corridor groups:

- C02 Köln–Gallarate (11%)
- C02 – RoLA (15%)
- C04 (15%)
- Italy – Switzerland (9%)

The exclusion of these trains was mandatory because they could not be used for calculations. In the framework of a performance management without penalties the trains can be used anyway in the reporting for punctuality. Nevertheless, it is advisable to monitor the occurrences of the problem because it indicates a wrong use of the delay coding and/or wrong or missing information in the messages and can support an improvement of the national systems or the national procedures for data collection.

The cases of "inconsistent running advice" represent the 8% of all exclusions and occur when the running advice on departure from an EPR point is earlier than the arrival at this point or when the arrival (or run-through) at an EPR point is earlier than departure (or run-through) from the previous EPR point. These cases are more frequent in the following corridor groups:

- France – Switzerland (53%) – when the stations are very close and two RA are sent, it happens that the times are not fitting for few seconds
- France – Switzerland – TGV trains (21%) – same problem as described above
- C02 – RoLa (15%) – the so called "Thun" problem occur (see 4.7)
- C05 –freight trains (16%) – master station concept rules not adaptable to operational reality (Antwerp North)

The EPR tool allows to manual exclude a train when a data problem cannot be detected by the tool or in exceptional circumstances. The manual exclusion of trains accounts for 4% of the total excluded trains but in the following corridor groups it is more relevant:

- C04 (14%)
- C08 (13%)
- Austria–Germany – Austria (15%)

As also mentioned above, during the pilot application, exceptional events or exceptional maintenance works required to suspend the monitoring of EPR trains on these lines for a period. Normally, in these cases, the EPR train list is updated so these trains do not appear in the data quality figures as the trains are in practice cancelled. However, sometimes it is not possible to update the list in short notice so the trains must be manually excluded.

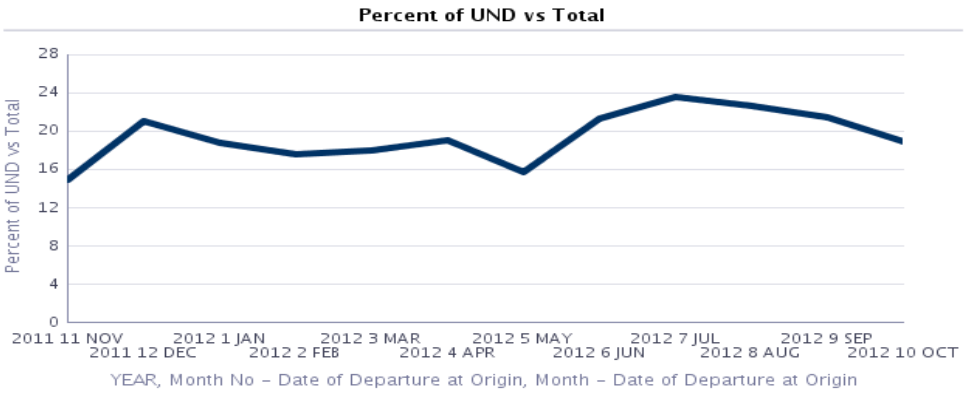
Finally, a major reason for exclusion for the corridor Italy – Switzerland was the disagreement on delay causes (26% of cases). Overall, this reason for exclusion represents 4% of the exclusions.

3.5 Undocumented delays

In the perspective of the use of the data coming from EPR in the framework of the activities for the improvement of railway traffic performance, it is important that the information on the causes of the delays is available as much as possible. Therefore, one of the indicators monitored in the EPR PA was the share of delay minutes to which no delay cause has been attributed. This indicator was measured only for included trains (because according to EPR rules, only the included trains are considered in the delay code validation phase)

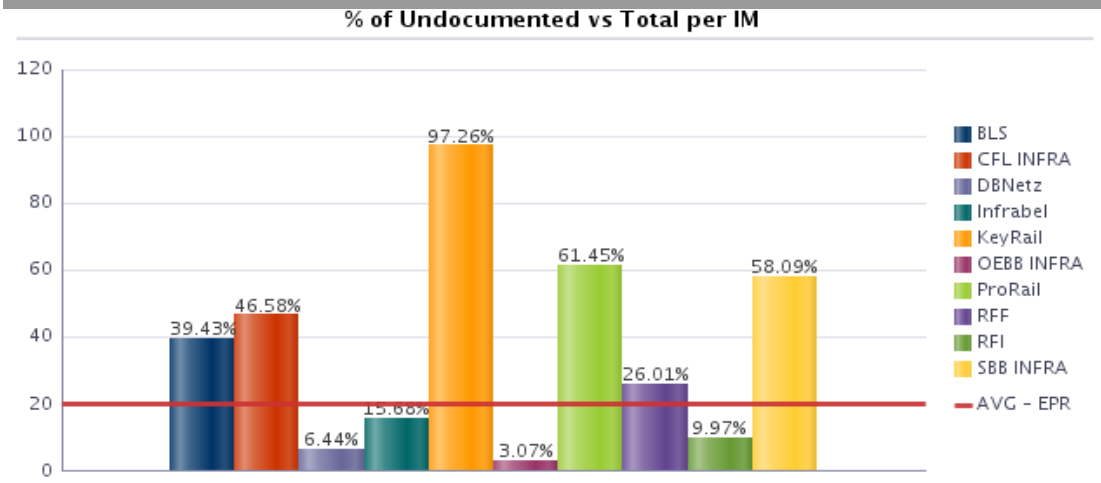
Overall, the undocumented delays for all included trains for the whole PA were 19% (namely 14% in freight traffic and 35% in passenger traffic).

Picture 7 shows the trend along the twelve months



Picture 7 – trend of undocumented delays – included trains- November 2011-October 2012

It must be highlighted that the level of undocumented delays is rather different among the partners IMs, as shown in picture 8.



Picture 8 - share of undocumented delays by IM – included trains

Also in the case of the trend along the twelve considered months, there are differences among the IMs, where some IMs show an improvement (for example, BLS improved by 21%) and others a slight worsening.

DBNetz and RFI maintain the undocumented delays below average since January with trend to increase in the last months. ÖBB keeps undocumented delays well below average (less than 10%) for the whole period. Infrabel trend is similar to the mentioned IMs but there was a peak of undocumented delays in July due to technical problems (delays were documented but UIC message 2005 not sent to TIS). The problem was quickly solved (also thanks to EPR PA).

BLS and RFF keep the undocumented delays slightly above average with improvement in the last months. According to RFF and SBB/BLS´s internal rules, the delays must not be mandatorily coded unless they exceed a threshold of, respectively, 5‘ (RFF, whole train run) and 3‘ (SBB/ BLS, single incident). Taking this into account together with the low average undocumented delay per segment in the French and Swiss network, their percentage of undocumented delays can be explained

Prorail shows data only since March as the points´ configuration was different before and no delays had to be coded by Prorail. Prorail´s high amounts are due to the fact that the timetable (CTT) of heavily delayed trains is changed thus no delay to be coded appears in the national system. CFL´s undocumented minutes are mainly due to the fact that the codification is done manually.

In general, the OWG experts deem these figures of undocumented delays too high, not only for a complete implementation of EPR (including financial consequences) but even more for a train performance management activity. In fact, in the EPR calculations these delays are attributed to (paid by) the responsible IM, therefore the calculation is possible anyway; however, in the analysis of performance, important information remain missing.

3.6 Figures on the international validation procedure

From November 2011 to October 2012 delays coded using the so-called "international codes" represent 6% of all coded delays. Also in this case there are big differences among IMs and among Corridor Groups. Infrabel used international codes in the 25% of the delays (mainly to explain take over delay from the Dutch/Belgian border, because Prorail used to change the timetables unilaterally, This procedure ended in December 1st 2012, so less international codes from Infrabel can be expected), while for all the others use of international codes is under 10%.

On a Corridor Group basis, the international codes are mostly used on the Italy – Switzerland corridor (24% of delay codes) while the level of use is under 10% in all to other corridors.

Normally, this picture can be explained by different operational issues in the different border sections where international codes are correctly used (like in the cases above). However, it is advisable to make further analyses in cases where the use of international codes is rather high or rather low, this is actually due real operational issues or misuse of international delay codes.

Most of the international delays remained untreated (60%-70%). The validation was done more or less intensively by the different partners. Of course, we can also state that many delay codes are correct.

In one specific case, i.e. Italy-Switzerland *via* Chiasso, many trains have been excluded because of delay coded disputed and not agreed. As mentioned above, this Corridor Group was the one where international delay codes were mostly used.

From what said above, it might be stated that the issue of the use of international delay codes and in general of the correct (or commonly agreed) application of the Leaflet 450.2 is still an open point and must be carefully dealt with in the future.

4. Results of the Pilot Application – general part

4.1 France-Luxembourg

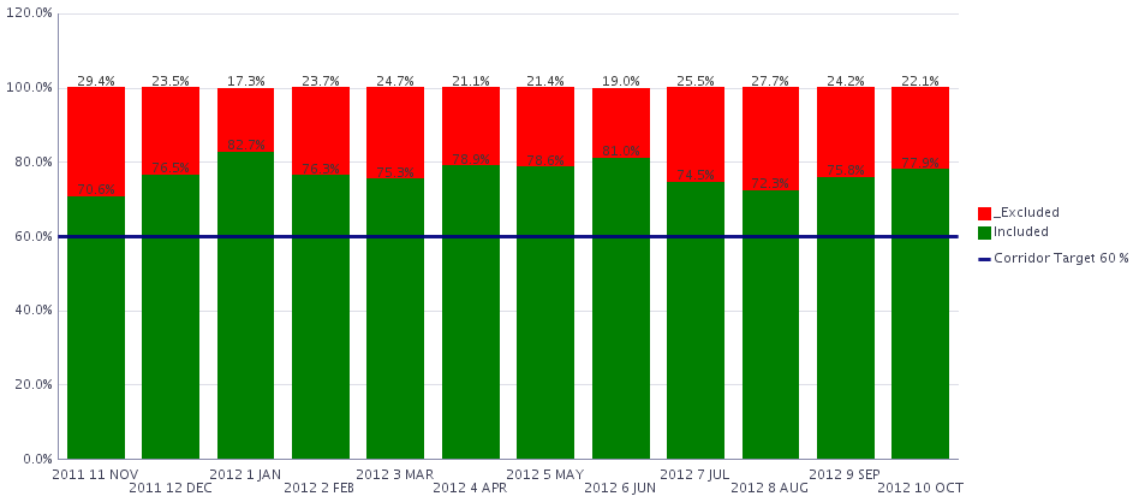
The corridor group is composed by the following corridor lines, both passenger traffic:

- France-Luxembourg – TER (regional traffic), some TER trains also connect Belgium to Switzerland
- France-Luxembourg – voyages (TGV trains)

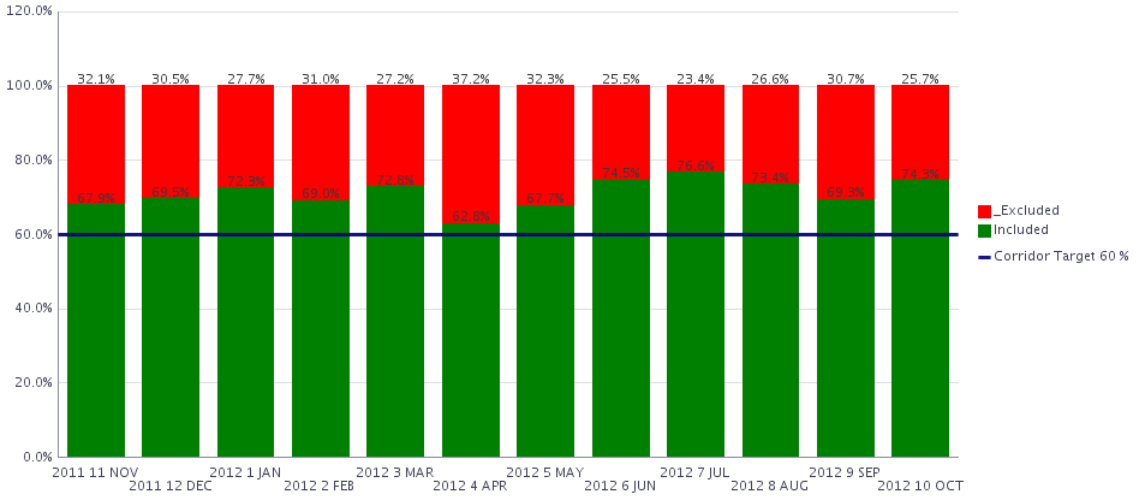
The total number of trains monitored was 9.472, among which 23 were cancelled (0,3%).

The average data quality in the concerned twelve months was 74,2%.

The trend of data quality of the two lines is illustrated by pictures 9 and 10



Picture 9 - data quality trend – France-Luxembourg TER – November 2011/October 2012



Picture 10 - data quality trend – France-Luxembourg VOYAGES – November 2011/October 2012

The main problem in this corridor group is the manual input of traffic data by CFL, leading to a high percentage of missing RA (especially in the points Zoufftgen frontière and Bettembourg-Voyageurs). This explains the lower data quality in case of public holidays and the better data quality for the TER traffic if compared to Voyages because few regional trains are running during the weekend.

In addition to this, a bug in the French IT system led to missing RA in the points Mont-Saint-Martin and Longwy. The bug is expected to be solved by the end of 2012.

4.2 France-Germany

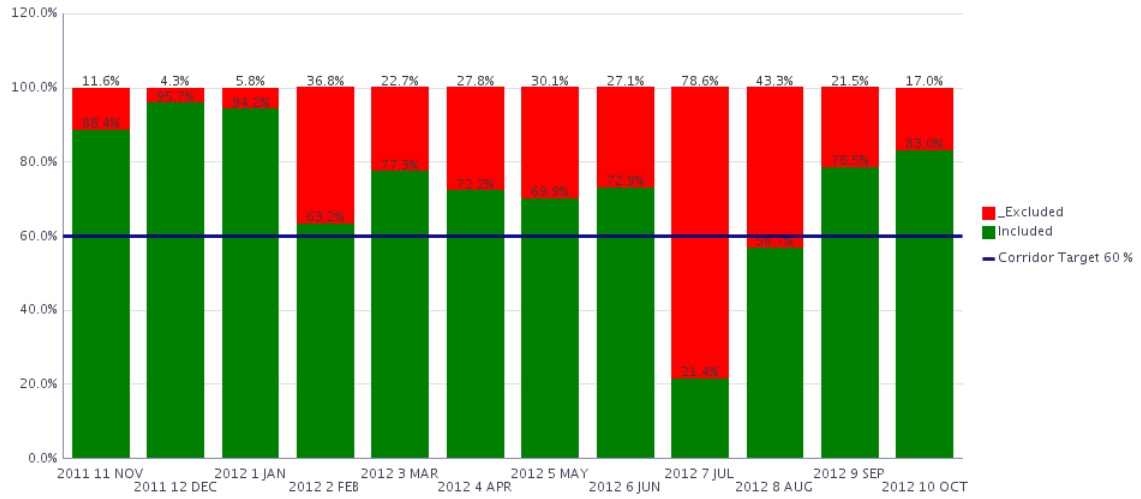
The corridor group is composed by the following corridor lines:

- Passenger traffic *via* Stiring-Wendel
- Passenger traffic *via* Strasburg

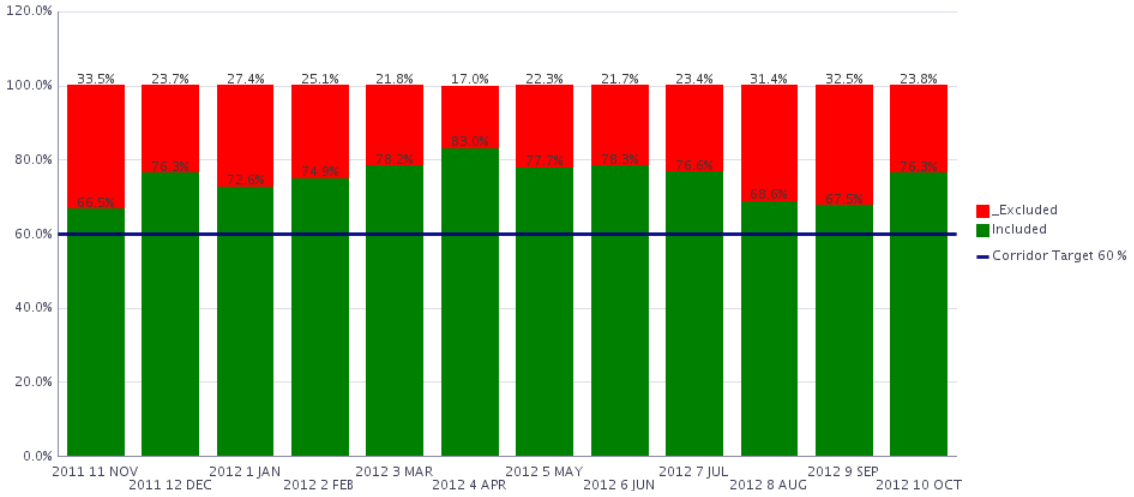
The total number of trains monitored was 6.213, among which 74 were cancelled (1,2%).

The average data quality in the concerned twelve months was 73,6%.

The trend of data quality of the two lines is illustrated by pictures 11 and 12



Picture 11 - data quality trend – France-Germany *via* Stiring-Wendel – November 2011/October 2012



Picture 12 - data quality trend – France-Germany *via* Strasbourg– November 2011/October 2012

Since the first months, a bug (also affecting the Corridor Group France-Luxembourg) in the French system regarding the two points Forbach and Strasbourg-Ville has been identified and not yet solved. The problem in Strasbourg Ville is that RFF´s system is sending new information deleting the previous.

In July 2012, due to a modification on site for National purpose, the data quality has decreased significantly in the line *via* Stiring Wendel, the situation has been corrected the 7th of August.

In the last months many problems have occurred in the point Stiring- Wendel, leading to missing or inconsistent running advices. The solution is under evaluation.

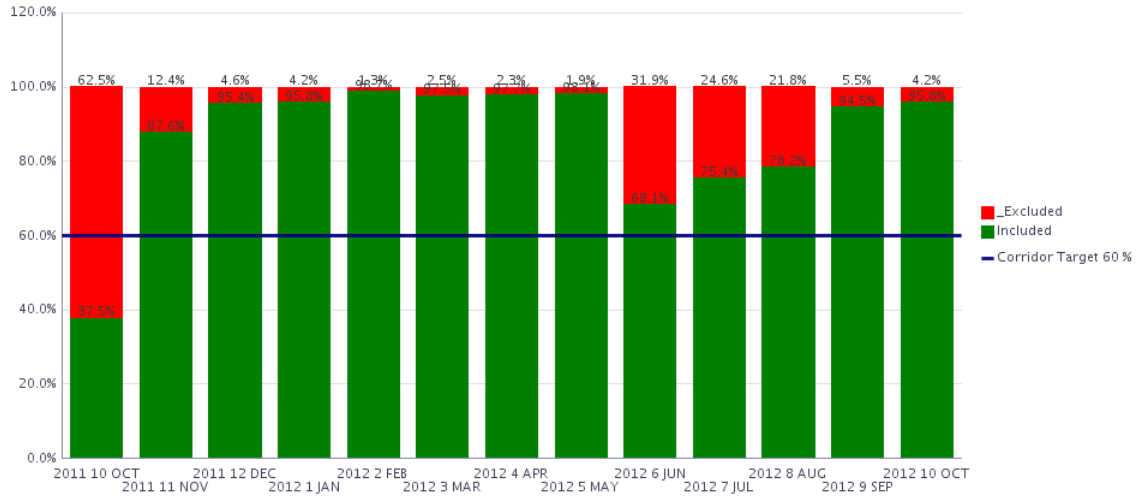
4.3 France-Switzerland

The corridor group includes the passenger traffic between France and Switzerland *via* Saint-Louis.

The total number of trains monitored was 4.087, among which 56 were cancelled (1,4%).

The average data quality in the concerned twelve months was 90,3%.

The trend of data quality of the two lines is illustrated by picture 13.



Picture 13 - data quality trend – France-Switzerland *via* Saint-Louis– November 2011/October 2012

Data quality from November 2011 onwards was rather high, except in June, July and August where inconsistent running advices were detected in St.Louis-Frontière. The bug was fixed and the data quality is again very good.

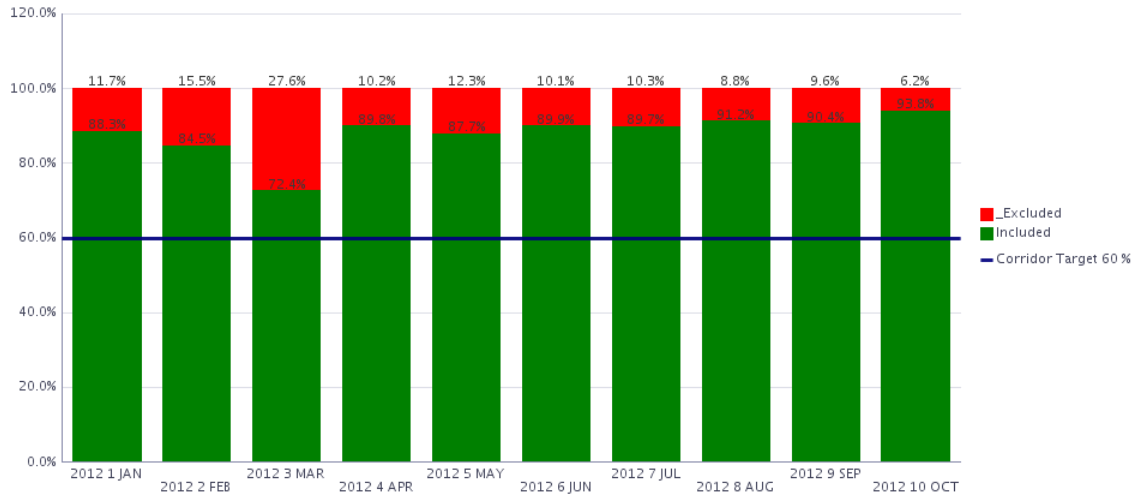
4.4 France-Switzerland- TGV trains

The corridor group includes the high-speed passenger traffic between France and Switzerland *via* La Plaine. The monitoring started in January 2012.

The total number of trains monitored was 4.850, among which 35 were cancelled (0,7%).

The average data quality in the concerned twelve months was 87,4%.

The trend of data quality of the two lines is illustrated by picture 14.



Picture 14 - data quality trend – France-Switzerland, TGV trains *via* La Plaine - November 2011/October 2012

Also in this case the data quality from was always rather high, except in March when missing CTTs were detected in La Plaine Frontière. The problem was promptly solved. Some missing running advices also occurred in the same point but the occurrences are decreasing since June.

4.5 Italy–Switzerland

The corridor group is composed by the following corridor lines:

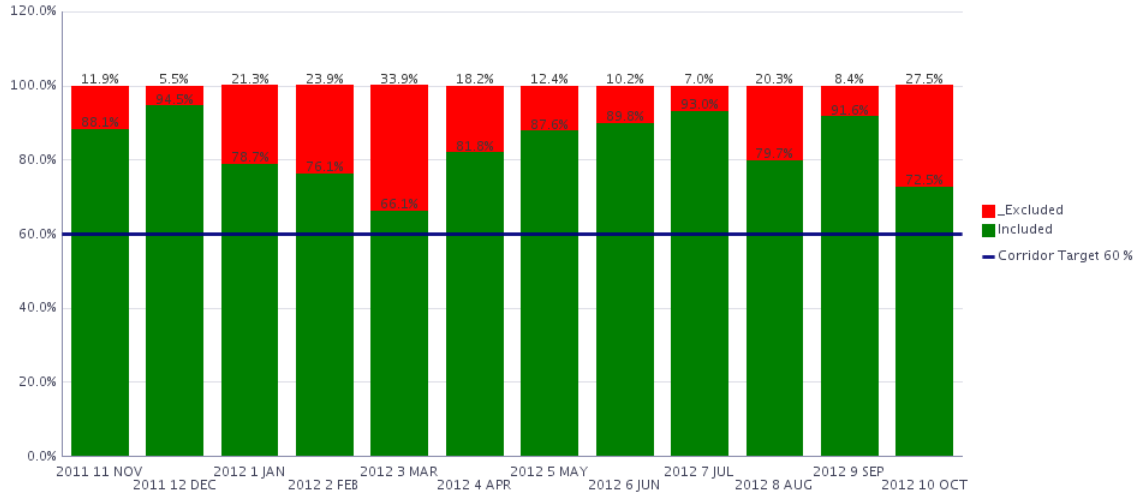
- Passenger traffic *via* Chiasso
- Passenger traffic *via* Domodossola

The total number of trains monitored was 6.616, among which 283 were cancelled (4,3%).

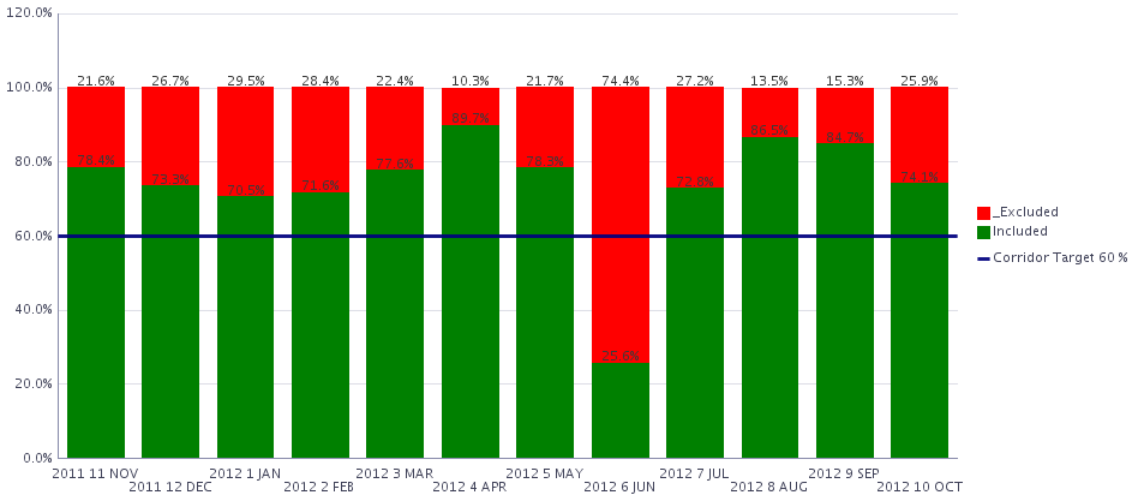
The average data quality in the concerned twelve months was 78,5%.

The trend of data quality of the two lines is illustrated by pictures 15 and 16.

Exceptional events (landslide) occurring in March, June, and August had required the closure of the concerned lines and therefore a temporary suspension of the EPR. Due to the fact that the events were unforeseen, it was not possible to update the train list immediately, so the concerned trains were manually excluded. This means that in pictures 15 and 16, the low share of included trains in the mentioned months, is not due to bad data quality but to operational problems.



Picture 15 - data quality trend – Italy-Switzerland *via* Domodossola– November 2011/October 2012



Picture 16 - data quality trend – Italy-Switzerland *via* Chiasso– November 2011/October 2012

In addition to what said above, a derailment occurred between Brig and Domodossola which caused traffic disturbances between January and March. Notwithstanding this, data quality in the first part of the year in the line *via* Domodossola was rather high.

Real data quality problems concerned:

- Missing running advices in Chiasso and Domodossola, Missing CTTs in Balerna : these problems are linked to the common responsibilities between IMs in part of

the line/border stations. IT and organizational improvements are being implemented and an improvement is expected in the next weeks. In any case, these problems never caused a significant decrease of data quality

- Wrong use of international codes/wrong time indication in the message 2005 (delay cause): these problems cause the exclusion of the trains because the system cannot identify the correct company responsible for the delay cause (see 3.4).
- Disagreement on delay causes

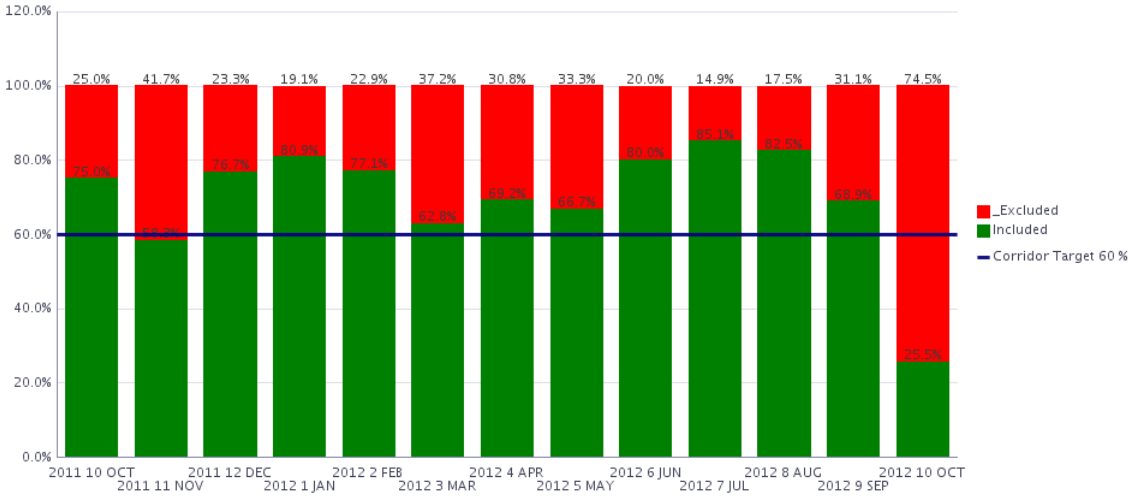
4.6 RNE C02 – Köln/Gallarate

The corridor group includes the freight traffic between Köln and Gallarate *via* Basel/Luino.

The sample was rather small and total number of trains monitored was 563, among which 69 were cancelled (12,3%).

The average data quality in the concerned twelve months was 68,4%.

The trend of data quality of the two lines is illustrated by picture 17.



Picture 17 - data quality trend – C2-Köln/Gallarate– November 2011/October 2012

Data quality was kept on a sufficient level for the whole period. The main problem here is missing RA. The causes are either (partly) cancellations (not correctly sent to TIS) or the train number is not used for the whole train run. The affected point is Luino. In the period until May included other points were affected (Köln Eifeltor, Köln Ehrenfeld Gbf,) but in these cases the problems seems to be solved. In June 2012 also this line was affected by the closure of the Gotthard and few trains circulated. It must be underlined that in this corridor group the sample is rather small so it is sufficient that few trains excluded to sensibly downgrade the data quality. In October, troubles have occurred in Luino. This is probably due to RFI´s modification on the UIC messages sending from border stations. A

new procedure is being tested and until it is not put on production bugs can occur. The deployment on production is expected by January 2013.

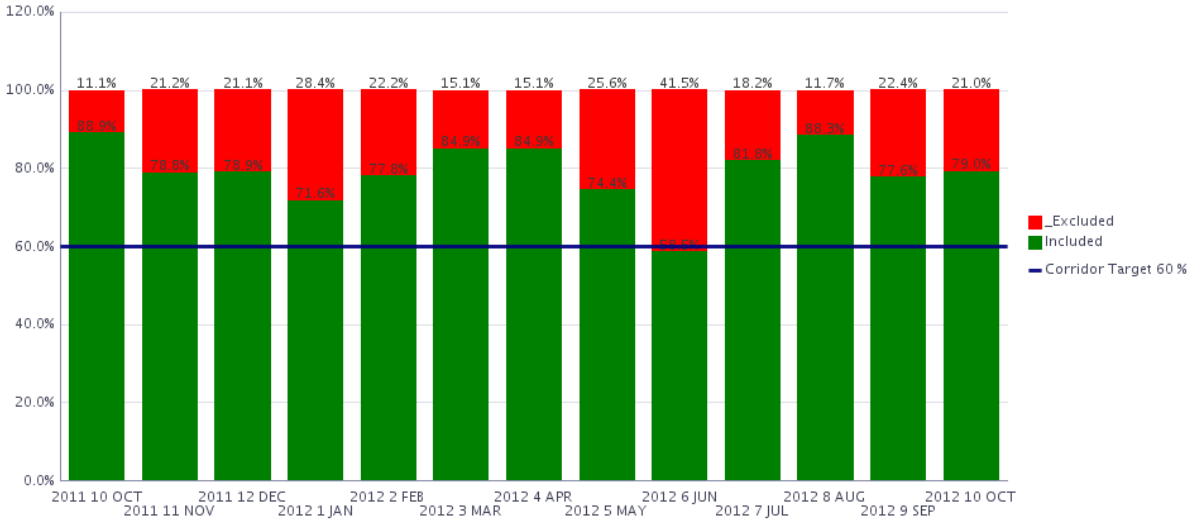
4.7 RNE C02 – RoLa

The corridor group includes the rolling highway between Freiburg and Novara *via* Basel/Domodossola.

The total number of trains monitored was 4.271, among which 207 were cancelled (4,8%).

The average data quality in the concerned twelve months was 77,6%.

The trend of data quality of the two lines is illustrated by picture 18.



Picture 18 - data quality trend – C2-RoLa - November 2011/October 2012

Data quality is normally good reaching the target. In June, operational reasons (not data quality) caused a lower number of included trains: between 05.06.2012 and 02.07.2012 due to a complete closure of the Gotthard Line (rock fall) hundreds of trains have been rerouted via the Lötschberg and strongly affected the RoLa traffic (new timetables / cancellations / partly cancellations / delays with new timetables).

Also the derailment mentioned in 4.5 had consequences on the RoLa traffic in the first months of 2012, even if the data quality level remained high.

As far as data quality is concerned, most of the exclusions are due to known problems for which the solution has been found and it is under implementation (improvement expected in the short term) or already implemented:

- "Thun problem": in this station an inconsistent running advice is detected when a stop is planned but, for operational reasons, it is decided not to stop the train. This problem is solved since mid November 2012.
- Contracted timetables in common stations: the same situation as the EC traffic (see 4.5)
- Question mark problem: solved by a workaround

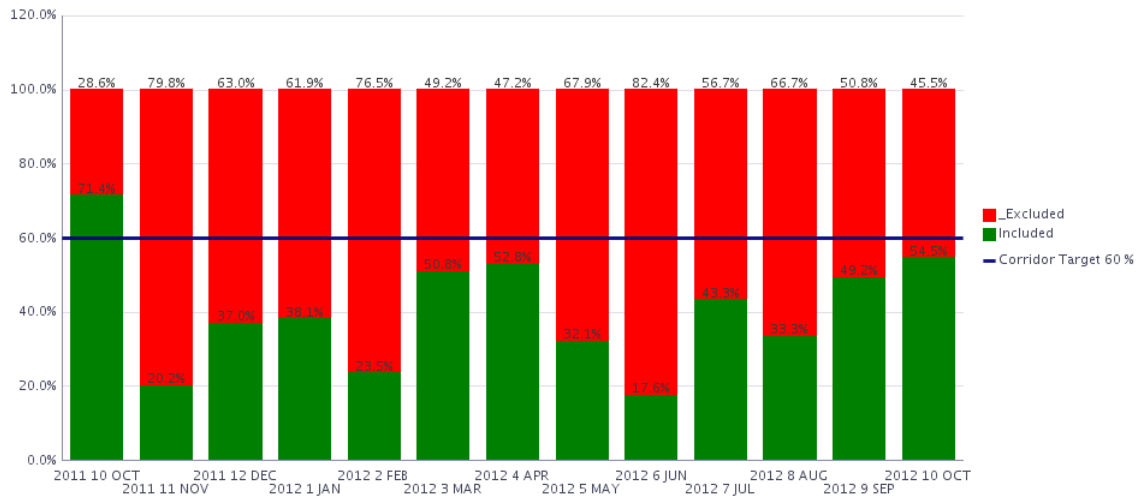
4.8 C2 Rotterdam-Novara

The corridor group includes the freight traffic between Rotterdam and Novara *via* Zevenaar/Emmerich, Basel, Domodossola.

The total number of trains monitored was 1.223, among which 159 were cancelled (13%).

The average data quality in the concerned twelve months was 36,6%.

The trend of data quality of the two lines is illustrated by picture 19.



Picture 19 - data quality trend – C2-Rotterdam-Novara - November 2011/October 2012

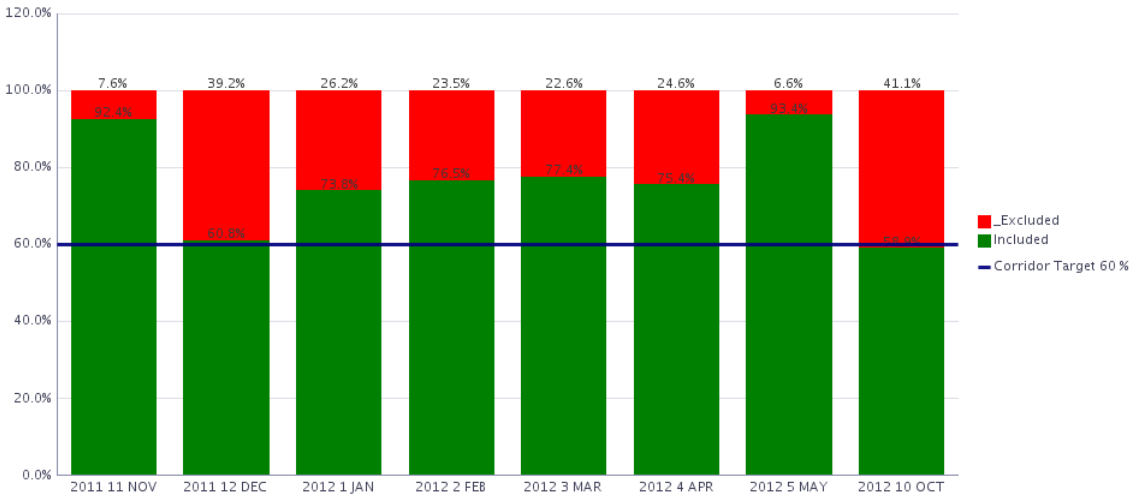
Due to the low data quality and to the fact that no improvements were happening nor planned, it was decided that the data related to this corridor group could not be used for the evaluation of the commercial model. Nevertheless, a deeper analysis is necessary in order to find out the most effective corrective actions improve the data quality on this corridor.

In general, the main problems (missing CTT at border, missing running advice and train run not complete accounting for 85% of the exclusions) are probably connected with operational procedures (load shifting, re-routing) which do not allow to correctly follow

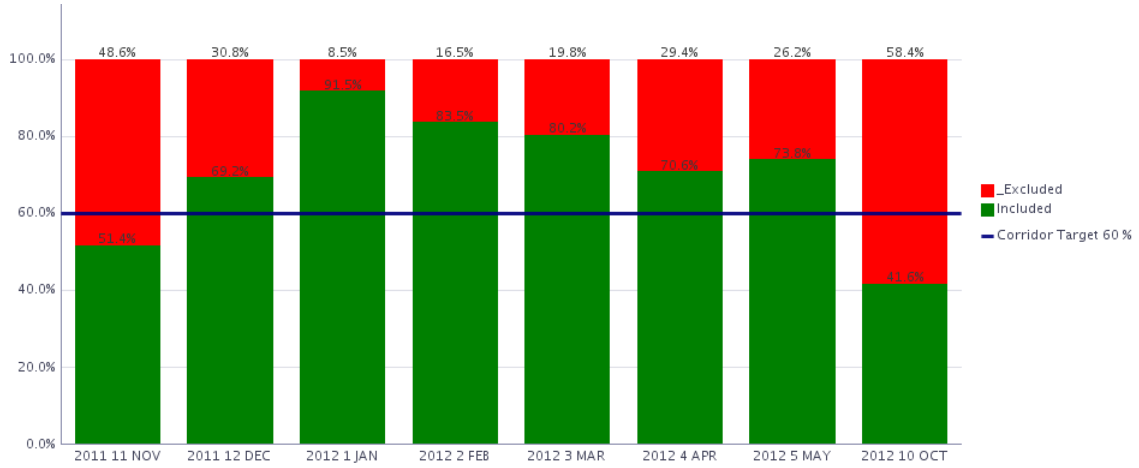
the train from origin to destination using the current IT tools . These issues must be further dealt with.

4.9 C4 Brenner line

The corridor group includes the bilateral (AT-IT) and tri-lateral (DE-AT-IT) freight traffic *via* Brennero/Brenner. The total number of trains monitored was 2.789, among which 263 were cancelled (13%). The average data quality in the concerned months was 71,6%. The trend of data quality of the two lines is illustrated by pictures 20 and 21.



Picture 20 - data quality trend – C4-Brennero/Brenner-Bilateral traffic - November 2011/October 2012



Picture 21 - data quality trend – C4-Brennero/Brenner-Trilateral traffic - November 2011/October 2012

Due to huge maintenance works, with recurring line closure, the monitoring was suspended from June 2012 to September 2012.

The main reason for exclusion is missing RA, mainly due to missing data in the German system. The problem was solved in June. In October data quality was badly influenced by

missing RA for departure from the Brenner station (both directions). The problem caused the exclusion of many trains, but it was solved on October 24th.

The missing responsible company problem was very relevant in the first quarter of 2012, but seems to be solved now.

4.10 C5

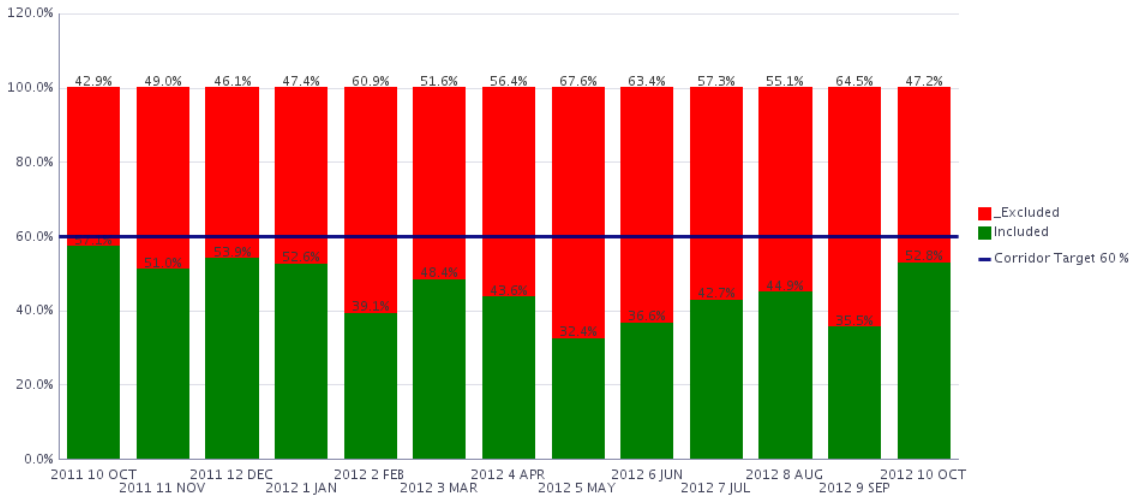
The corridor group includes:

- Freight traffic:
  - Belgium – Luxembourg
  - Belgium – The Netherlands
  - France – Luxembourg
- Passenger traffic: Belgium–Switzerland

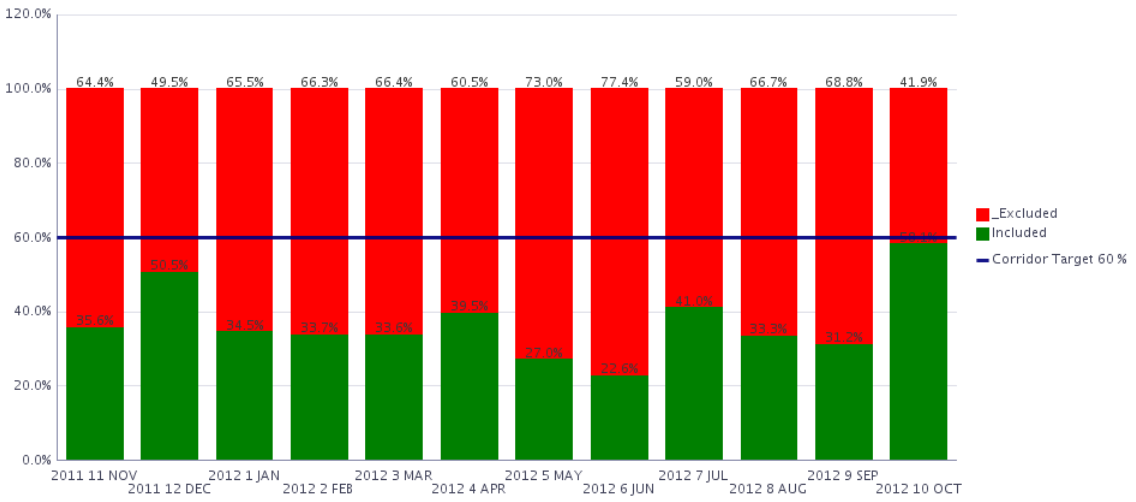
The total number of trains monitored was 4.026, among which 165 were cancelled (4%).

The average data quality in the concerned twelve months was 40,1%.

The trend of data quality of the two lines is illustrated by pictures 22 and 23.



Picture 22 - data quality trend – C5-freight traffic - November 2011/October 2012



Picture 23 - data quality trend – C5-passenger traffic - November 2011/October 2012

The main problem for this corridor Group is the fact that often trains are not run but not officially cancelled by the RUs. This means that the UIC message 2003 – cancellation – cannot be sent to TIS by the concerned IM. The trains are not included in the "cancelled" but have not sufficient data so they are excluded (for missing RA or for Train not complete which represent the 76% of excluded trains) pushing data quality down. Since this problem could not be solved in the short/medium term and still remains open, it was decide not to consider data from this Corridor Group for the evaluation of the commercial model but only for the data quality analysis.

From June onwards, also a problem with inconsistent running advices (points which are mainly affected are ANTWERPEN-NOORD-BUNDEL B1, Bettembourg-Marchandises and Bettembourg-Ouest) has been affecting data quality increasingly.

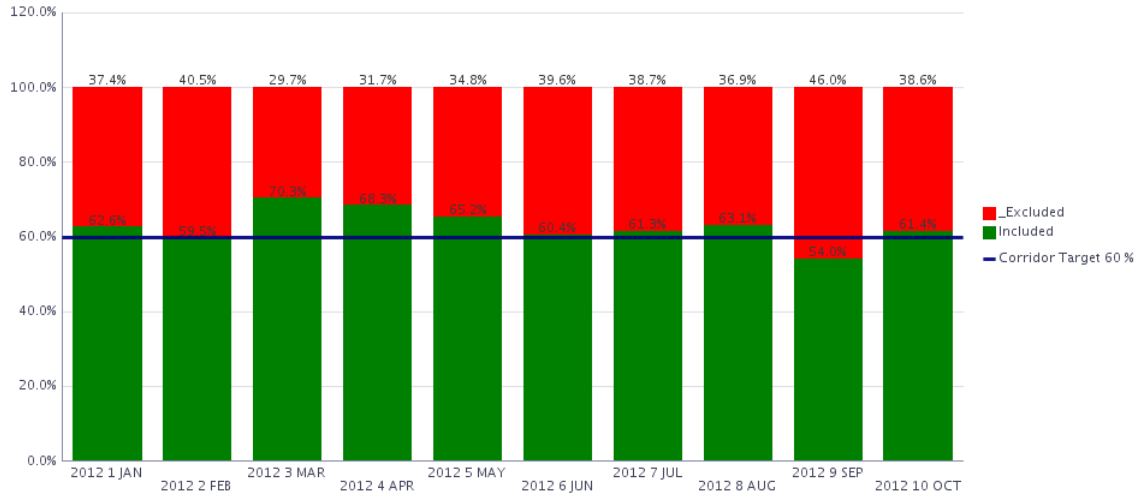
A less relevant case is the problem of responsible company with question mark (null) mainly in several points, usually because of missing RU name/code in the message sent to TIS. The downfall in June is due to 10 cases of mixed up CTTs (partly modified), but this happened only in June.

4.11 C8

The corridor group includes the freight traffic between France and Italy *via* Modane. The total number of trains monitored was 3.089, among which 245 were cancelled (7,9%). The average data quality in the concerned twelve months was 58,2%.

The trend of data quality of the two lines is illustrated by picture 24.

The traffic on this line was suspended until January 2012, when the monitoring could start again.



Picture 24 - data quality trend – C8- November 2011/October 2012

The data quality on this corridor group was always around the requested target, but there is room for improvement.

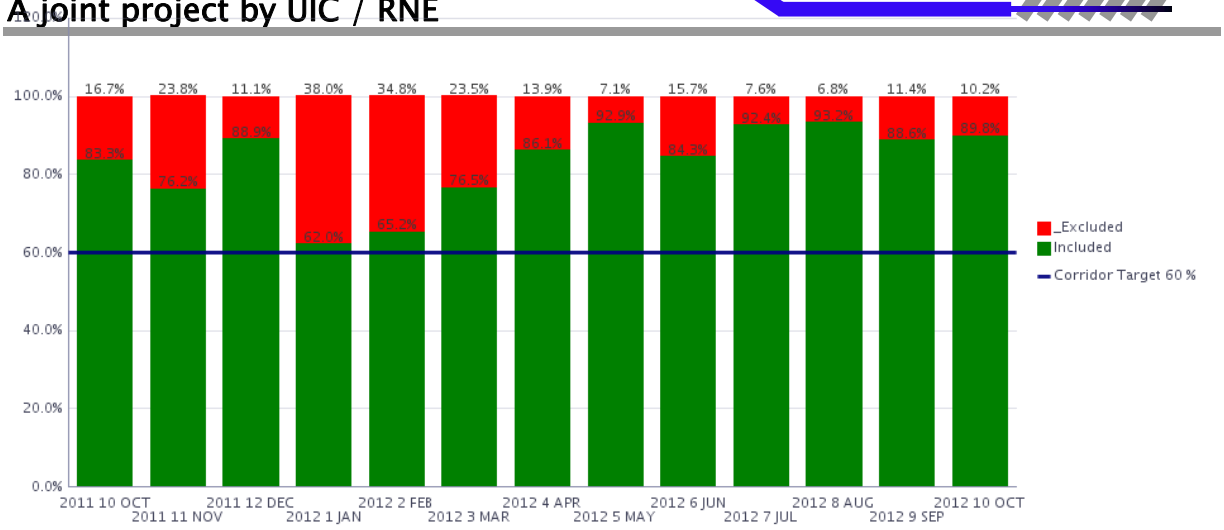
The main problem is the timetable not fitting at the border in Modane. An additional relevant problem is the fact that many train runs are not complete.

In both cases, corrective actions are being taken but no effects are still visible (see also 4.5 for RFI border stations).

4.12 Austria-Germany-Austria

The corridor group includes the freight traffic between Austria and Germany and back *via* Rosenheim. The total number of trains monitored was 1.784, among which none were cancelled. The average data quality in the concerned twelve months was 84,7%.

The trend of data quality of the two lines is illustrated by picture 25.



Picture 25 data quality trend – AT-DE-AT- November 2011/October 2012

There are no particular problems on this corridor. In the first part of 2012 there were more exclusions than usual, partly due to operational problems (several line interruptions on the Arlberg line because of danger of avalanches) and partly to data quality problems (trains in W-E direction were excluded because timetables were sent by DB Netze with a new handover point near Salzburg, which was not implemented in the TIS system). The data quality problems were solved.

5. Conclusions

As already mentioned, the main reasons for exclusion are missing running advice and missing/inconsistent CTT at border. These two cases account around 70% of all exclusion cases. They are connected with national system problems and/or problems affecting their interface with TIS; they are also connected with not harmonized (national) procedures (including cancellations) leading to inconsistencies at the borders.

In most Corridor Groups only few EPR points have problems. Therefore, in order to significantly increase data quality, it is sufficient to solve bugs/ technical problems in few points.

The length of the chosen Corridor does not seem to influence data quality itself: the number of EPR points and especially the number of border points are far more relevant. Of course, the longer a route is the higher is the number of EPR points and borders.

It is important to highlight that the EPR Pilot Application has been an important incentive for data quality improvement.

On the one side, the developed IT tools provided for more detailed data which were automatically available. Manual workload was naturally still necessary but it was mainly focused on reporting and analysis and no more on data collection and processing.

These data helped to bring to light real problems both on the IT and operational side so that these problems could be solved, more easily and more quickly than in the past.

On the other side, the PA provided also for a platform where all stakeholders could confront themselves to find together the better solutions.

From what said above the following recommendations can be made:

- A further care must be taken by IMs when sending data to TIS. Regardless to any future application of EPR, a reliable data set is essential for any activity in data analysing. In addition, no improvement in TIS can be effective on data quality if data sent by national systems are not fully complete and correct. This is even more necessary for Corridors with more border stations.
- A commitment of all partners to align coding behaviour better than today is important: this will imply modifications of internal rules/ behaviour.
- Internal coordination within the involved parties is necessary: sometimes, the poor data quality is not due to technical failure of national systems, but to internal rules (operations, timetable ...) influencing information in the national data bases. Thus the data quality responsible person (in IMs or RUs) needs an overall support from his/her company to solve the problems (IT, TT, Operation departments as an example).

According to all participants in the EPR OWG the analysis of data quality was a real added value as explained above and there are already activities and consultations on how to continue such tasks in different frameworks.

On the other hand, regarding the validation of international delay causes presents a different situation. The systematic validation as described in the EPR Handbook is indispensable in case of application of financial penalties.

In the framework of a train performance analysis, however, the added value of a full scale validation will have to be evaluated.

Of course, a better reliability of delay causes (included those related to border sections) and of delay coding behaviour is still useful. In addition, also it is also important to keep the tool working and to ensure the prosecution of the know-how of operators.

Therefore, the participants suggest to focus on the integration of data quality checks in the current train performance activities as a first step and to evaluate in a second step a procedure to perform the validation of international delay codes not in a systematic way but sufficiently to grant:

- Identification of bugs in the tool if occurring
- Training of new operators or keeping of the know-how of already trained operators
- Use of the information coming from the validation (for example: how many international codes are attributed, how many are accepted, work load necessary for the validation ...) as additional support to performance analysis

This type of validation does not need a structured organization or precise time constraints, but can be organized according to the exigencies of the train performance management activities. Therefore, the use of the validation tool will be evaluated in the framework of the Train Performance activities.