Annex 2
## Benchmarking

1 Introduction

Benchmarking is the process of comparing performance of one entity (the subject railway) to the performance of other entities (the benchmark companies) to identify best practices and opportunities for improvement. Often, benchmarking begins with a high-level comparison to identify areas of greatest potential, followed by detailed analysis of these high-potential areas. In the railway industry, benchmarking may compare financial measures such as operating ratio and revenue per ton-km, or productivity measures such as traffic units per employee, and traffic units per track-km.

During benchmarking, after comparison of statistics identifies the high opportunity areas, the business processes of the subject and benchmark railways are compared to identify changes necessary to close the performance gap. Next, an implementation plan is developed for the subject railway to adopt improved business processes. For example, the high-level analysis may identify a gap in locomotive operating costs. The detailed analysis might reveal that the subject railway's locomotives work fewer hours per day than the benchmark railways and that the sub-

Figure 1   Benchmarking Locomotive Operating Costs
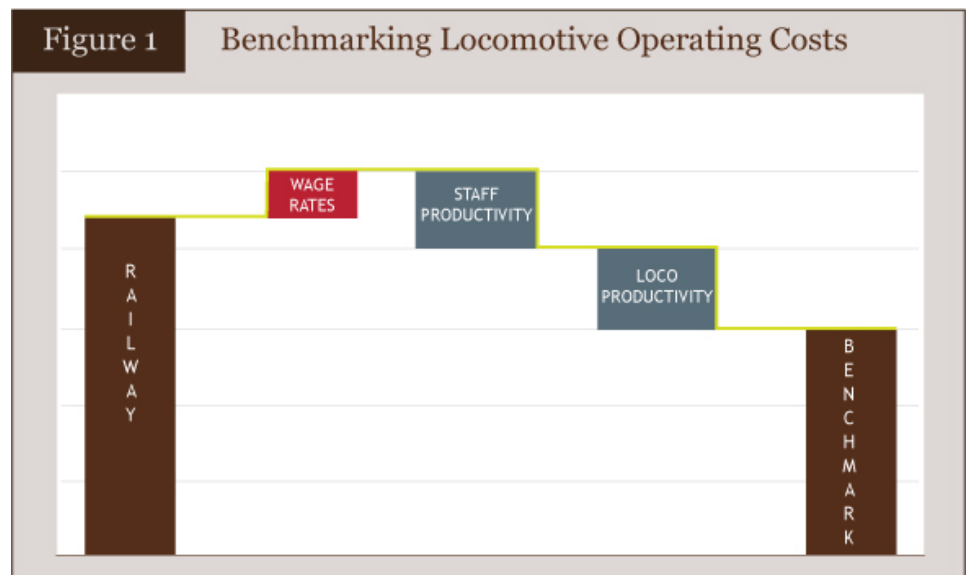
ject railway's staff have lower productivity, which is partially offset by lower wages (Figure 1). The business process analysis would reveal a difference in locomotive assignment practices—the subject railway restructure changes the locomotives every 150 km, while the benchmark railways change every 2,500 km—which affects staff and locomotive productivity. A performance improvement plan would be implemented to change locomotive assignment practices and adjust staffing. This

would subsequently be evaluated by whether costs and locomotive productivity had improved.

Benchmarking against both railways and companies outside the railway industry can be useful. Railway comparisons are more useful for operational issues. External comparisons are useful to examine how competing transport providers and logistics companies handle their markets, corporate culture, and strategic issues. This annex focuses on benchmarking using statistical comparisons between railways. It explains the steps for statistical analysis, provides definitions, highlights data issues, and explains the typical ratios used.

The statistical analysis for benchmarking starts with selecting benchmark railways and indicators.[151] Next, data are collected and adjusted to improve comparability. Then indicators are calculated. Finally, results are interpreted.

2 Selecting Benchmark Railways

Benchmarking is most usefully carried out using high performing railways with fairly similar characteristics and operating conditions. This controls for factors that management or government cannot influence and focuses analyses on factors that can be changed. Thus, to the degree possible, benchmark railways should be similar in the following characteristics: (i) size; (ii) traffic volume and type; (iii) traffic mix and journey types, such as passenger vs. freight, and originated vs. transit journeys; and (iv) traffic density. Other factors, such as having a similar technology level, may be considered.

3 Selecting Indicators

Indicator choices depend on the benchmarking objective. If benchmarking aims for a general review of railway operations, the process would likely begin with key financial indicators and an overall productivity measure for each primary railway resource—labor, track, locomotives, wagons, and coaches—followed by detailed statistical analysis of any areas with large gaps. Typical indicators appear in the table below.

Box 1 Railway Benchmarks		
Name	**Definition**	**Interpretation**
Financial Measures		
Average tariff	Freight revenue/ton-km	A measure of the railway's ability to generate revenue from freight traffic. Most tariff level variations are due to competition, commodity, and haul length. But low tariff levels may indicate a tariff policy issue.
Average fare	Passenger revenue/passenger-km May be calculated by type of service (e.g., commuter vs. intercity)	A measure of the railway's ability to generate revenue from passenger traffic. Most fare level variations are due to competition, type of service, and average distance traveled. But low fares may indicate a fare policy issue.

---

[151] The annex draws heavily on the World Bank Railways Database Update 2007, *Users Guide*.

Box 1 (cont.) Railway Benchmarks

Name	Definition	Interpretation
Average passenger subsidy	Passenger subsidy/passenger-km	A measure of railway ability to obtain compensatory revenue from government in exchange for providing loss-making passenger services.
Ratio of passenger fares to freight rates	(passenger revenue/ passenger-km)/ (freight revenue/ton-km)	Rough measure of the degree to which railway revenue structure depends on freight services to cover fixed costs and/or to cross-subsidize passenger services. This indicator must be used with great caution when comparing railways because either freight or passenger yield can be heavily influenced by traffic mix in each market.
Operating ratio	Operating costs/operating revenue May be calculated with and without operating subsidies	A measure of railway ability to cover its costs and generate investment funds. Operating ratios for reasonably profitable US Class I railroads range from 80-85 percent. Limited data available on operating ratio including operating subsidies for EU railways typically show ratios around 95 to 100 percent, indicating inability to cover all costs, even after receiving government PSO payments.
Labor share of revenue	Total wages/total revenue	A measure of the share of revenue from customers that is paid to workers. It excludes subsidies to focus on the direct relationship between wages and revenues. Typically, profitable US freight railways have a ratio of about 0.30 (China is even lower). Many EU railways have ratios approaching 1.00 or higher.
Productivity Measures		
Track density	(passenger-km + ton-km)/ track-km	A measure of the volume of traffic produced with railway infrastructure. Railways are capital intensive and infrastructure is a substantial proportion of total assets. Railways with high utilization of this expensive asset such as those in China, U.S., and Russia, have an advantage in reaching economic viability.
Locomotive productivity	(passenger-km + ton-km)/ locomotives Passenger-km/passenger locomotives Passenger-km in multiple unit(MU) service/MU powered coaches Ton-km/freight locomotives	A measure of the volume of traffic produced with railway locomotives. High utilization of this expensive asset gives the railway an advantage in reaching economic viability. If data permit, the locomotive fleet should be separated into passenger, freight, and shunting services to calculate separate productivity measures. If a substantial proportion of passenger service is provided with MU equipment, the figures should be adjusted to reflect this.
Wagon productivity	Ton-km/wagon	A measure of how much freight traffic is produced with the railway wagon fleet. High utilization of this expensive asset gives the railway an advantage in reaching economic viability. This measure should be used with caution because wagon ownership practices can differ. In some countries, customers own a substantial proportion of the wagon fleet. In others, the railway may handle substantial traffic moved in wagons owned by other railways.

Box 1 (cont.) Railway Benchmarks

Name	Definition	Interpretation
Coach productivity	Passenger-km/coach	A measure of how much passenger traffic is produced with the railway coach fleet. High utilization of this expensive asset gives the railway an advantage in reaching economic viability. If the railway operates both coaches and MU equipment, MU coaches should be added to the coach total. Higher figures are usually associated with high-speed rail or commuter services; lower numbers indicate longer-haul services with lower-density seating, and significant coach space allocated to dining or sleeping. Thus it is important to benchmark against railways with similar types of passenger services.
Employee productivity	(passenger-km + ton-km)/ employee	A measure of how much traffic is produced with the railway labor force. Labor is the largest single-cost item for nearly all railways, so output per employee is fundamental to financial and economic viability.
Other Measures		
Locomotive availability	Locomotives available/total locomotives in the fleet Units may be number of locomotives or locomotive hours	A measure of the technical capacity of the railway to maintain its locomotives and to provide funding for spare parts. U.S. Class I railroads expect a diesel locomotive availability ratio of 90-95 percent. In the developing world, good performance would be 70-90 percent. A ratio below 70 percent or a ratio that is deteriorating over time indicates a management problem.
Locomotive-km/ day	Locomotive-km/locomotives/365	A measure of the work performed by railway locomotives.
Wagon availability	Wagons available/total wagons in fleet Units may be number of wagons or wagon hours	A measure of the technical capacity of the railway to maintain its wagons and to provide funding for spare parts.
Wagon-km/day	Wagon-km/wagons/365	A measure of the work performed by railway wagons. A low figure implies low use—that wagons linger in shunting yards, or that the wagon fleet is too large.
Wagon cycle time	(wagons*365)/loads May be independently measured in railway's operation system.	A measure of wagon use intensity. A high figure may indicate too much time in shunting yards, inefficient redistribution of empty wagons, or unused wagons spending a lot of time in storage because the fleet is too large.
Load-to-empty ratio	Loaded wagon-km/empty wagon-km	A measure of how much of wagon movement is revenue generating. A low ratio of loaded to empty may indicate inefficient redistribution of empty wagons. Unit train movements have a load-to-empty ratio of approximately 1, so the measure is strongly affected by the nature of traffic.

Box 1 (cont.) Railway Benchmarks		
Name	**Definition**	**Interpretation**
Coach availability	Coaches available/total coaches in the fleet Units may be number of coaches or coach hours	A measure of railway technical capacity to maintain its coaches and provide funding for spare parts.
Coach-km per day	Coach-km/coaches/365	A measure of work performed by railway coaches, which is strongly affected by the nature of railway traffic. A railway with long-distance service will have higher coach-km per day than a railway with commuter service.
MU availability	MU trains available /total MU trains in fleet 'Units' may be number of MU trains or MU train hours	A measure of railway technical capacity to maintain its MUs and to provide funding for spare parts.

4 Collecting the Data

Sources for benchmarking data include the following:

- **World Bank Railways Database** (updated in 2007), provides a set of indicators for the railway transport that includes size, scale and productivity measures over a sufficient time frame for adequate cross-sectional and time series performance evaluations, including financial and physical measures.
  http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTTRANSPORT/EXT RAILWAYS/0,,contentMDK:22614614~menuPK:7260743~pagePK:210058~piPK: 210062~theSitePK:515245~isCURL:Y,00.html

- **UIC** maintains the **Railisa** database with railway physical indicators, including all UIC railways over many decades, allowing cross-sectional and time-series evaluations.
  http://www.uic.org/spip.php?article1352

- **Other Rail Industry Associations** may compile member statistics. For example, the Association of American Railroads produces several statistical publications on the North American railway industry.
  http://www.aar.org/StatisticsAndPublications.aspx

- Railway **websites** offer selected physical indicators and/or railways financial reports. For example, Turkish State Railways publishes statistics on its website (http://www.tcdd.gov.tr/home/detail/?id=305) Deutsche Bahn publishes its financial statements on its website.
  http://www.deutschebahn.com/ecm2-db-en/ir/financial_reports/re- ports_2010_2009.html)

- **Government Statistical Agencies** usually collect statistics on railway traffic and infrastructure size. For example, the Ukrainian Statistics Agency provides traffic volumes by commodity, track length, and electrification (http://www.ukrstat.gov.ua/)

- **Regulatory Agencies** often require railways to provide detailed statistical reports. For example, the U.S. Surface Transportation Board requires large railways to file R-1 reports with detailed operational and financial statistics. (http://www.stb.dot.gov/stb/industry/econ_reports.html

Comparing statistics from multiple data sources requires great care to ensure consistency. Definitions, pitfalls, and interpretation issues are discussed below.

Infrastructure

Typically, infrastructure statistics such as line-km, track-km, km of electrified line, km of double track line, are accurate. Definitions to note are the following:

- Track-km. Track length in kilometers, counting every track. 100 km of double track = 200 track-km. Track-km may be disaggregated by gauge, electrification, main lines, secondary lines, or station track.

- Route-km. Railway line length, regardless of single or multiple tracks. 100 km of double track = 100 route-km. Route-km may be disaggregated by gauge or electrification. Route-km is often referred to as line-km.

- Track gauge. Distance between the rails, disaggregated into four categories: (i) narrow gauge (NG) is less than one meter; (ii) meter gauge (MG) includes both true meter—1000 mm, and "Cape Gauge" —1067 mm; (iii) standard gauge (SG) is 1435 mm; and (iv) broad gauge (BG) includes all gauges greater than standard. Broad gauge is found in former Soviet countries and the Baltic Republics (1520 mm), and India (1676 mm).

Rolling stock fleets

Some railways report only locomotives, coaches, or wagons that are serviceable; others report their entire fleet and distinguish between 'total' fleet and 'serviceable' fleet. The difference, referred to as the availability ratio, can be significant. Some railways record three categories of rolling stock: in service, inoperable but repairable, and beyond repair. In some railways, equipment beyond repair is still included in the fleet, distorting the apparent size of the fleet, availability ratio, and measures of rolling stock productivity.

Locomotives

Often, locomotive fleets are reported by type of power—steam, diesel or electric—and purpose—main line or shunting. Each railway has its own definition of 'main line', so this term covers a range of locomotive sizes. The lightweight locomotives used by smaller railways as main line would be used only for marshalling (switching) on larger railways with heavier flows.

In former Soviet countries, railways often permanently coupled two locomotives, and counted the two units as one locomotive. A single locomotive used for lighter

work (e.g., passenger) would *also* be counted as one locomotive. Therefore, when comparing locomotive productivity statistics, analysts must adjust figures so that multiple-unit locomotives are counted in a consistent way.

Multiple units

The rolling stock that is most difficult to measure is multiple unit (MU) passenger equipment. This equipment forms passenger trains that are not hauled by a locomotive. Instead, some or all of the coaches include power units. The ratio of powered coaches to un-powered trailer coaches can range from 1:1 to as high as 1:3. Published statistics do not always clarify whether MU equipment consists of individual coaches or sets of MU trains. If a railway has substantial MU operations, the number of coaches may be unclear. The use of MUs must be considered when compiling statistics, because MUs may substitute for locomotives and coaches and may be responsible for some or all of the production of passenger service.

Passenger coaches

Passenger coaches vary in seating density—more seats per car for shorter distances, fewer for longer distances—and in the numbers of sleeping or dining coaches. Counting multiple unit (MU) coaches is often complex in countries with significant commuter services.

Freight wagons

Wagons vary by size and type, and one freight wagon might carry up to four times the gross weight of another wagon. For example, some railways, operate economically obsolete, two-axle wagons with maximum axle loadings as low as 15 metric tons (30 ton maximum gross weight), whereas a few state-of-the-art heavy haul freight railways uniformly use four-axle wagons with axle loadings as high as 35 or even 40 ton (140 to 160 tons maximum gross weight). Also, in many countries, customers own a substantial part of the fleet. Thus, in calculating productivity statistics, care must be taken to match output (ton-km) and production (wagon-km) with the wagons that produced it.

Passenger traffic

Accuracy of passenger counts is improving due to advances in ticketing systems. However, inaccurate figures are common if high numbers of seasonal or multi-ride tickets are sold. This is particularly an issue for railways with substantial suburban passenger traffic, and in countries such as Russia, where many passengers have social privileges and therefore ride without being ticketed. Also, railways with multiple passenger interchanges, for example between long-haul and short-haul trains, often record two trips rather than one trip, thus inflating passenger numbers. For example, every day Indian Railways in Mumbai transports more than five million commuters using seasonal or multi-ride tickets. Passenger sampling yields estimates of actual ridership and trip length, but sampling must be designed and executed to yield accurate estimates.

Freight traffic

Statistics for railway freight tonnage are usually reliable because tariff revenues are based on tonnage. Major sources of inaccuracy are (i) weighing, because shippers have incentives to report lower weight and (ii) traffic interchanged between rail-

ways, where tonnages can inadvertently be double counted. Generally, ton-km reports are accurate, but errors can occur if multiple routes are available because shippers insist on being billed for the tariff route, but for operational reasons the railway may use a longer route. As with passenger traffic, on larger and more modern railways, computers and automated shipping documents have improved the accuracy of freight reporting.

During a benchmarking exercise, when comparing across regions, the units of measurement should be checked to ensure they are the same. American railways measure outputs in 'short' tons and miles, not metric tons and kilometers. Conversion rates: one short ton = 0.907 metric tons and one mile = 1.609 kilometers.

Traffic units

Productivity is measured by the ratio of outputs to the resources used to produce the output. For example, freight wagon productivity can be calculated by dividing ton-km by the number of wagons. However, when track and locomotives are shared by passenger and freight services, resource productivity must be calculated using an output measure that combines figures for passenger and freight traffic.

Calculating a combined measure is problematic, but the most common measure used is Traffic Units (TU). TU is the sum of ton-km and passenger-km, using a 1:1 weighting of passenger and freight-km. Most specialists would agree that the resources used to produce a passenger-km and a freight ton-km are not equal, but no agreement exists on what a more accurate weighting should be. Earlier World Bank research indicated that labor inputs associated with a passenger-km are at least twice that of a ton-km. Outputs per freight locomotive tend to be higher than for passenger locomotives. Lighter passenger trains may generate less wear on infrastructure than heavy freight trains, but passenger trains' higher speed may consume more capacity than freight trains. Since an ideal weighting is undefined, benchmarking should involve railways with relatively similar passenger-freight mix.

Financial measures

Benchmarking using monetary figures or financial reporting requires great caution due to wide differences in accounting standards. Financial statements prepared according to International Financial Reporting Standards (IFRS) and audited by qualified external auditors will have consistent definitions across entities, clear descriptions of accounting policies followed, and footnotes that provide details about the figures provided. Ratios that include depreciation or amortization, such as the operating ratio, can be problematic because asset valuation varies widely, especially in countries with high inflation. When IFRS financial reports are unavailable, financial data should be used cautiously.

Financial comparisons between countries require a common currency. Typically, local currencies are converted at the official exchange rate to a common international currency such as the U.S. dollar or Euro; or, currencies are converted using purchasing power parity (PPP). Using international currency at the official exchange rate has the merit of familiarity but can seriously misrepresent local resource use. Instead, PPP dollar conversion offers more accurate estimates of the activity being valued in terms of local resource consumption, within the limits of the calculation of PPP values. However, both methods are imperfect since railway

services use a mix of local resources, such as labor, and international resources, such as locomotives.

Some monetary measures, such as the ratios of wages/revenues and average passenger fare/average freight tariff (this is revenue/passenger-km divided by revenue/ton-km), are relatively robust regardless of currency value because numerator and denominator are affected equally by conversion. Use of time series can reveal useful trends, even if absolute value is questionable due to differing definitions.

Average fare and freight yields

Traffic mix and average distance affect comparisons of passenger fares and freight tariffs. Typically, railways charge lower tariffs for some low-value bulk commodities, such as coal, and higher tariffs for higher-value goods that require higher service levels, such as assembled automobiles. For this reason, two fully comparable railways could report vastly different average freight tariffs if one hauls mostly coal, and the other mostly assembled automobiles. A similar market-mix phenomenon occurs in passenger services—commuter travel has high passenger volume, low prices, and simple coaches. Longer-haul journeys have lower passenger volume, higher ticket prices, and can require more complex coaches including sleeping and dining cars.

Average distance per journey can raise or lower unit price because railways incur costs not only during hauling freight or passengers, but also at the start and the end of each journey. Thus, average freight tariffs and passenger fares are lower in large countries such as China, Russia, and the U.S.A. where starting and ending costs are a smaller proportion of much longer average journeys than, for example, in smaller countries such as Belgium. Without complete data on tariffs and fare schedules for both subject and benchmark railways, adjusting for this type of unit price differential is impossible. Nonetheless, it should be kept in mind.

5   Analyzing Results

Benchmark calculations will identify areas of performance differences—better and worse—between the subject railway and the benchmark railways. Further statistical analysis is then made of the areas in which the benchmark railways are superior to the subject railway. For example, the preliminary analysis may show that the subject railway has lower wagon productivity. Next, additional benchmarking analysis might be done on wagon-km per day, wagon fleet availability, and wagon cycle time.

Using these benchmarks as guides, subject railway operational practices would be compared to those of the benchmark railways to identify differences that account for different results. For example, the benchmark railways may use computer models to distribute empty wagons and the subject railway does not. Or, the benchmark railways may allow customers a single day to load and unload but the subject railway allows customers three days. Or, the benchmark railways may have efficient shunting operations but the subject railway's marshalling yards average 36 hours dwell time.

The analysis aims to identify which of the benchmark railways' good practices are responsible for better performance, and then implement these practices in the subject railway.