

SURFACE TRANSPORTATION BOARD

Docket No. EP 724 (Sub-No. 4)

UNITED STATES RAIL SERVICE ISSUES—PERFORMANCE DATA REPORTING

Summary of Ex Parte Meeting between Union Pacific Railroad Company (UP) and
Surface Transportation Board (STB) StaffHeld December 3, 2015, 9:10 AM – 11:11 AM

UP Attendees: Jeremy Berman (General Attorney), Thomas Haley (Vice President, Network Planning & Operations), Jamie Houton (Assistant Vice President, External Relations), Louise Anne Rinn (Associate General Counsel), Robert Schmidt (General Director, Resource Planning, Network Planning & Operations)

STB Attendees: Michael Higgins, Stephanie Lyons, Ronald Molteni, Lisa Novins, Nderim Rudi, Jason Wolfe

UP expressed its appreciation for the opportunity to meet with STB Staff, noting that it tried to be clear and thorough in written comments, but interaction facilitates additional productivity. UP then led the discussion by going through its prepared presentation. (Ex. 1.)¹

UP stressed that a few key measures provide more insight into service levels than a large number of secondary metrics. (Ex. 1, at 2.) UP focuses on a handful of key measures to indicate how the railroad is performing. If an issue arises, UP does more research into that particular matter. A lot of customer-specific information about shipments flows to UP's individual customers. Its customer service website has 17.5 million queries (year-to-date).

UP also stressed that railroad networks are not comparable. The networks are different, but metrics lead the public to make comparisons that frequently are not valid. Railroads dedicate a lot of time trying to explain why those comparisons are not valuable. UP stated that the Board should focus on metrics that provide meaningful information, are reasonably available from all carriers, and do not reveal confidential, proprietary information.

UP then explained the differences between its three major service networks: manifest, intermodal, and bulk. (Ex. 1, at 3.) Each network operates very differently, but they share resources. UP must overlay the networks in order to serve them all. When looking deeper into the service networks, one sees very different numbers and very "different UPs." Below system-level metrics, the results differ depending on the network. The networks are a product of how UP defines them. For example, Slide 3 (describing the three service networks) does not include

¹ Pursuant to the December 2, 2015 protective order entered in this proceeding, UP submitted public and highly confidential versions of its presentation, with confidential information removed from Slide 3 of the public version.

the automotive network, (Ex. 1, at 3), which includes both unit trains and manifest trains. How a railroad defines manifest or bulk affects how an automotive train is classified, and those definitions can vary by railroad. Metrics below system-level that roll up these differences get very complicated, particularly when considering specific commodities.

STB Staff asked whether there is a way to ascertain the degree to which the separate networks use common resources. UP responded that operating a railroad requires five resources: locomotives, freight cars, crews, line capacity, and terminal capacity. The degree to which the internal networks share those resources varies. For example, there is a high degree of overlap in line capacity, but certain car types are very unique and not shared.

With respect to the carloading mix of each railroad, UP emphasized that metrics such as cycle times and miles per day are not comparable across railroads. (Ex. 1, at 4.) In addition to carloading mix, the eastern and western railroads have extremely different geographies, which impacts operations. For these reasons, UP suggested that one cannot directly compare numbers between railroads.

UP explained that the map on Slide 5 of its presentation illustrates its large and intensive manifest network that it considers core to the railroad. (Ex. 1, at 5.) UP observed that its network is impacted by customers on either end of the national network. The interconnectedness of the network means that even if UP appears to be having service issues, the cause may be with a connecting partner or a customer. UP's metrics are sensitive to connecting carriers and how rapidly its customers turn and release its cars.

Next, UP stated that velocity, specifically system train speed, is a very useful metric. (Ex. 1, at 6.) It also stated that track conditions can result in reduced speed, and freight car terminal dwell is an indicator of terminal congestion and health. Overall, service metrics vary highly. Speed and congestion provide a good look over time at the fluidity of the railroad. UP uses seven day carload rates so it has an equal number of each day of the week in its calculation. UP does not deliberately hold or stop trains or not use equipment; the railroads are in the service business and so they monitor held trains very closely.

STB Staff inquired as to whether there are standardized metrics that would benefit the Class I railroads when the railroads are holding to wait for other carriers. UP responded that the Class I railroads communicate in real time about whether they are holding trains for other railroads. Requiring a specific reporting metric would therefore not help the railroads. UP added that nearly 40% of its traffic interchanges with other carriers.

UP then explained that there is a very strong relationship between customer satisfaction and speed. (Ex. 1, at 7.) Speed is an indicator of service and provides a good idea of how the railroad is turning equipment and resources. UP also stated that freight car terminal dwell can be a function of transportation plan; it is about optimizing train lengths. Sometimes a railroad can increase dwell at a yard and, as a consequence, can decrease cycle time, which leads to an overall better customer experience. External issues also may require an increase in terminal dwell, such as the recent west coast labor issues. Rather than moving additional volume west, traffic was

held in eastern terminals. Thus, dwell increased but not because of an issue at those terminals, nearby, or even on the same network.

UP explained that throughput is about how much traffic the railroad moves for its customers. (Ex. 1, at 8.) Over time, UP has tried to increase velocity and throughput, thus achieving greater service levels. A better service product will lead to a more efficient railroad. Capacity investment is one way to increase throughput.

UP next stated that freight car inventory is a measure of network consumption and possibly congestion. (Ex. 1, at 9.) It is impossible to take UP's freight car inventory and compare it to other railroads; it is a function of many different factors. UP believes that the trains held methodologies in the October 2014 interim data order in Docket No. EP 724 (Sub-No. 3) and the Notice of Proposed Rulemaking (NPR) in this docket are misleading. UP stated that it has had considerable dialogue about how to calculate the metric.

UP stated that surge capacity could also be a controversial metric because different railroads have different philosophies on this resource, and they do not all agree. (Ex. 1, at 10.) UP explained that surge capacity is the reserve capability that railroads have to overcome demand spikes, network interruptions, or other disturbances during normal operating conditions. The chart on Slide 10 shows that velocity and service improve when reserve capacity is within a normal range. (Ex. 1, at 10.)

UP emphasized that more detailed metrics do not necessarily advance the Board's objectives in the NPR. (Ex. 1, at 11.) Many factors may impact car orders and loading volumes. Commodity- and geography-specific metrics might be normal for reasons not driven by rail operations and not within railroad control. UP explained that, while networks are not comparable on a high level, comparisons at more detailed levels are even less valid. If metrics become too detailed, there is also a danger that confidential information could be revealed.

UP then discussed its customers' access to their specific data. (Ex. 1, at 12.) It noted that between customer-specific data and system-level metrics, the commodity- and geography-specific metrics are cloudy and problematic.

UP discussed many of the structural differences between railroads and emphasized that this means each railroad may measure the same aspects of performance differently. (Ex. 1, at 13.) Normalization may help, but the data is really only relevant to one railroad over time. UP explained that the graph on Slide 14 illustrates how different railroads are at below system level numbers. (Ex. 1, at 14.) UP emphasized that fewer measures are better; more data creates more room for invalid comparisons (Ex. 1, at 15).

UP concluded its presentation by stating that the four metrics it recommends, where there is a small area of overlap between the railroads, are the key. (Ex. 1, at 16.) It stressed that the high level metrics in the chart on Slide 16 provide a good idea of how the railroad is doing; anything more specific is not useful and may create too many issues.

STB Staff followed up on the presentation by inquiring how the Board would use high level metrics to detect service issues affecting agriculture shippers in the western United States in circumstances comparable to 2013-14. UP responded that the Board and customers would see all four high-level metrics [on Slide 16] moving adversely. A follow-up dialogue would reveal a lot of additional information, particularly when combined with internal Board information from shippers. It suggested that the Board could then impose a more specific temporary metric or requirement. UP questioned how the Board would anticipate where to collect more granular data and on what commodities and stated that it is hard to anticipate where the next problem will arise and where to stop collecting data. UP also noted that the Board can ask the railroads for historical data, which could be a powerful tool in the future.

STB Staff asked whether UP has internal metrics specific to the agriculture market which allow UP to monitor its performance serving its agriculture customers. UP replied that it has a regular internal meeting to discuss various issues that touch on network health and how the railroad is meeting demand. The operating side uses this information to ensure that resources are being properly allocated. UP has frequent affirmative outreach to customers about service and that information is shared with business group leaders. This is accomplished through conversation and a forward looking dialogue, not by simply looking at data.

In the context of monitoring service to the agriculture market, STB Staff inquired whether UP looks at turns per month. UP replied that its network geography is different, and because turns per month is radically different geographically, it quickly becomes a more detailed analysis. UP also noted that using grain as an example is problematic because most of its grain moves in shuttle train service, which uses a disproportionately lower proportion of the fleet and is impacted by the secondary market. STB Staff asked approximately what percentage of UP's grain moves in shuttle service. UP answered about 68%, and stated that it uses a much smaller percentage of equipment.

UP also noted that cycle time relates to service lanes. For example, cycle times to the Pacific Northwest compared to the Gulf of Mexico can be extremely different. UP suggested that it would be difficult for the Board to collect enough data to understand all of these nuances in terms of the national rail network.

STB Staff referred to UP's comment that it does not maintain a coal loadings plan like the other railroads and asked if there is a metric, such as cycle time, that is more informative than trainloadings and carloadings and that tracks performance in meeting utility customers' needs. UP responded that velocity and throughput are the best measures. Cycle times and carloadings may provide considerable information, but do not reveal anything about unmet demand. Carloadings might appear low, but UP might be meeting the entire demand at that time. UP can provide more service when there is more demand. It suggested that a metric from the Energy Information Administration, such as utility stockpiles, might allow the Board to monitor stockpile levels.

STB Staff then asked whether there is a metric that UP values that drills down further than the basic trains held metric. UP explained that it calculates train hours held, which show the total

time held as it affects trains. It values some measure of duration, so it measures hours held in a 24 hour day, and it can also look at the number as a percent of total train hours. UP said that it also tries to capture a reason for trains held. The proposed Board metric would require railroads to report a train held at one location and then a second location as being held twice. This is very different than the 0600 snapshot UP currently uses. The impact of the change is that it creates an incentive to hold a train at one location for a longer duration, which might be detrimental to the system. UP stated that a related issue is how to define a “train,” because railroads may define it differently to reflect their business and operations.

STB Staff next inquired whether UP sees value in the type of metrics that some other carriers make available on their websites or in public presentations, such as on-time departure, on-time arrival, connection performance, and plan adherence. STB Staff also noted that such metrics are somewhat different than the ones UP has discussed. UP responded that it does not see value in those metrics and is skeptical of service indices across railroads. UP explained that these indices are a function of what is counted and how it is measured; these differ by the type of train or commodity. It emphasized that there is a strong connection to each railroad’s carloading mix, which differs dramatically by railroad. There is a significant danger in making comparisons across railroads. UP also stated that it would be possible for railroads to manipulate their individual indices and questioned why the Board might take this route when there is such a strong connection between velocity and customer satisfaction.

STB Staff then asked whether the Board should require all carriers to measure metrics using the same methods. UP advised against such a requirement and noted the differences between railroads’ markets, customers, and systems. UP is concerned that requiring specific reporting methods would discourage innovation in the market to meet customer need.

Next, STB Staff asked whether there would be worthwhile metrics regarding surge capacity. UP responded affirmatively. Locomotive and freight cars in storage could be meaningful, and UP noted that freight cars in storage is information that it often shares with the Board. UP also noted the limitations of a freight cars in storage metric due to the numerous varieties of freight cars that it has available.

STB Staff next inquired about the Chicago metrics and whether UP could advise on additional ways for the Board to look at fluidity in Chicago. UP responded it had not prepared to comment on Chicago, but its impression is that the Board’s current analytics provide a fairly good look at Chicago. It also noted that the status level changes are beneficial and that its own terminal dwell numbers for Chicago yards can also be informative.

STB Staff asked why UP did not recommend reporting on trains holding for locomotives and crew, since it examines those metrics internally. UP responded that the metrics it recommended are the most relevant, and are fairly consistent and developed across railroads. While UP may look at other data, it does not know if the other railroads look at the same type of data, or if they would be similarly valuable. The four metrics UP recommended explain 90% of what is happening on the railroad.

STB Staff also asked whether UP sees value in reporting a trains held metric, including the reason held. UP responded that it looks closely at trains held because there is tremendous incentive for railroads not to hold trains. However, it does not see the value in publishing that data; a lot of the dialogue between shippers and railroads is about crews, locomotives, and freight cars.

STB Staff then asked UP to confirm that its preferred reporting period remained 12:01 a.m. Saturday through midnight Friday, to which UP replied affirmatively. STB also asked when UP would prefer to submit data to the Board, and UP replied that the Wednesday reporting week has worked out well for it; with a Monday holiday they are able to get the numbers compiled and checked by Wednesday. In closing, UP asked that the Board consider its written comments that discuss some of the issues with the proposed metrics.



December 3, 2015

Railroad Performance Data Reporting

EP 724 (Sub-No. 4)

Tom Haley – VP Network Planning
Bob Schmidt – Gen Dir Network Dev
Lou Anne Rinn – Assoc Gen Counsel
Jeremy Berman – Gen Attorney
Jamie Houton – AVP External Relns

Overview



1. **A few relevant measures can provide more insight to service levels than a large number of secondary metrics.**
2. **The customer-specific information that customers obtain from their rail carriers is more valuable for logistics decisions than metrics reported across the rail industry.**
3. **Railroads are not comparable. Larger numbers of public metrics lead to exponentially more comparisons that frequently are not valid.**
4. **An appropriate report for the Board's purposes should include metrics that:**
 - a) Provide meaningful information
 - b) Are reasonably available from all carriers, and
 - c) Do not reveal confidential, proprietary information
5. **UP will review today several core metrics it has found useful in managing its network to best serve its customers. Because rail networks are different, some of UP's core metrics may not be relevant to other railroads.**



Three very different service networks at UP

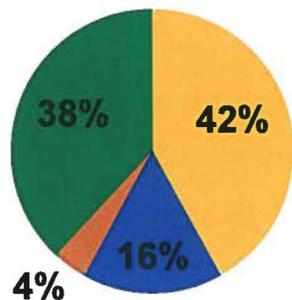
<u>Product</u>	<u>Principal Model</u>	<u>UP Dimensions</u>	<u>Avg Train Speed (mph)</u>	<u>Avg Car Inventory</u>
A. Manifest (Carload)	Hub & Spoke Local / Yard / Thru Freight	38% of Carload Units	22.6	170,000
B. Intermodal	Ramp to Ramp Interchange to Ramp	42% of Carload Units	30.6	14,000
C. Bulk	Mine – Utility Elevator – Dock Stockpile – Stockpile	20% of Carload Units	24.3	109,000



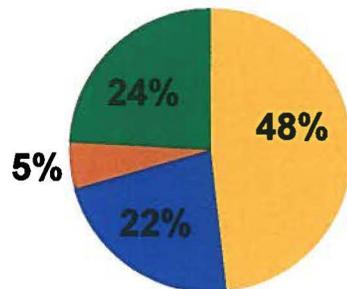
The rail networks are not comparable

2015 year-to-date carloading mix

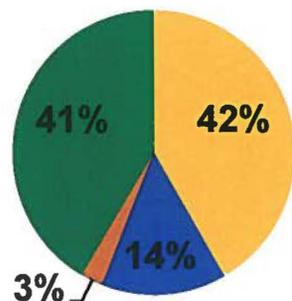
Union Pacific



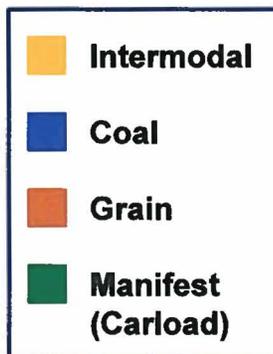
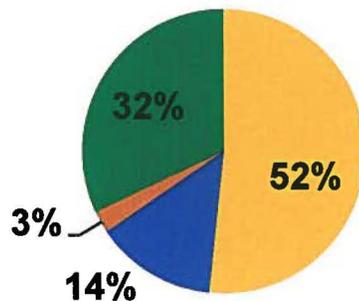
BNSF



CSXT



NS



Implications

- **Most comparisons are not valid across railroads**
- **Best to compare within a railroad over time**
 - Even that becomes shaky as mix and geographical demand change
- **Differences in the way railroads construct measures make comparisons even less valid**

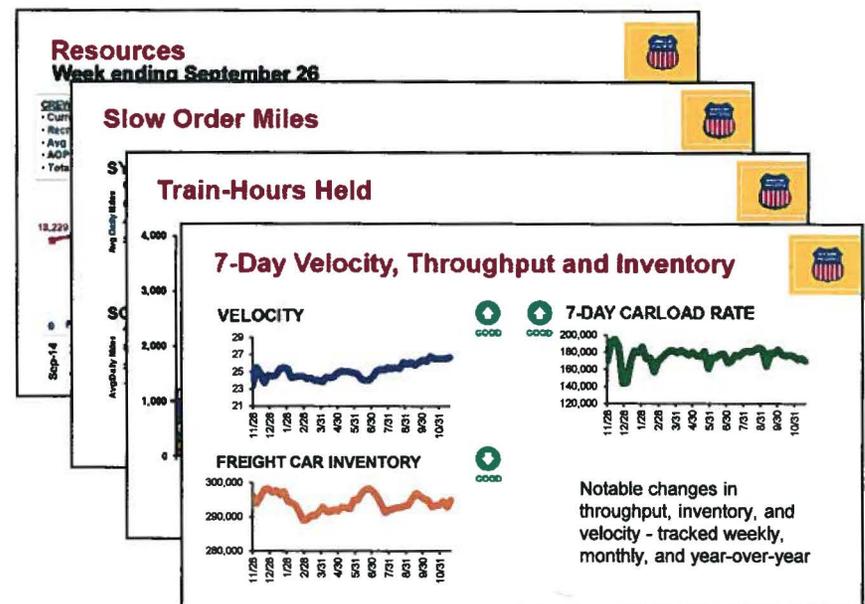
UP focuses on metrics proven to reflect its network health and capability



Union Pacific Core Operational Metrics

- **Velocity**
 - System train speed
 - Miles of track condition slow orders as % of total
 - Freight car terminal dwell
- **Throughput**
 - 7-day carloading rate
- **Congestion**
 - Freight car inventory
 - Train-hours or trains held as % of total
- **Surge Capacity**
 - Stored locomotives
 - Stored freight cars
 - Train-hours or trains held for crews

UP has found these very relevant over time



Velocity

1. Average Train Speed (Velocity)

- The most significant measure of railroad operating health
- Drives both service and efficiency

2. Track Condition Slow Orders

- Impede network velocity and capacity
- Indicates the degree to which a railroad is maintaining its track infrastructure over time

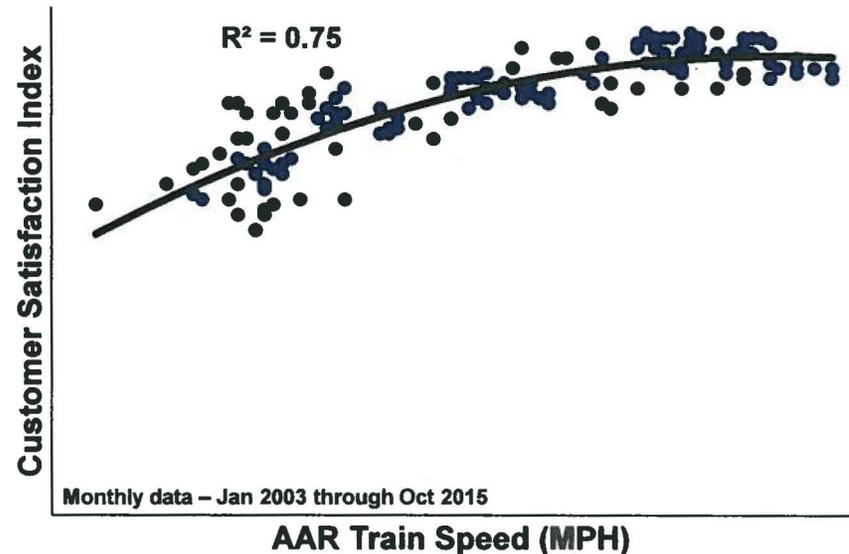
3. Freight Car Terminal Dwell

- An indicator of terminal fluidity

- **Average Train Speed and Freight Car Terminal Dwell are currently reported through the AAR and the Interim Report**



Customer Survey Results vs Train Speed

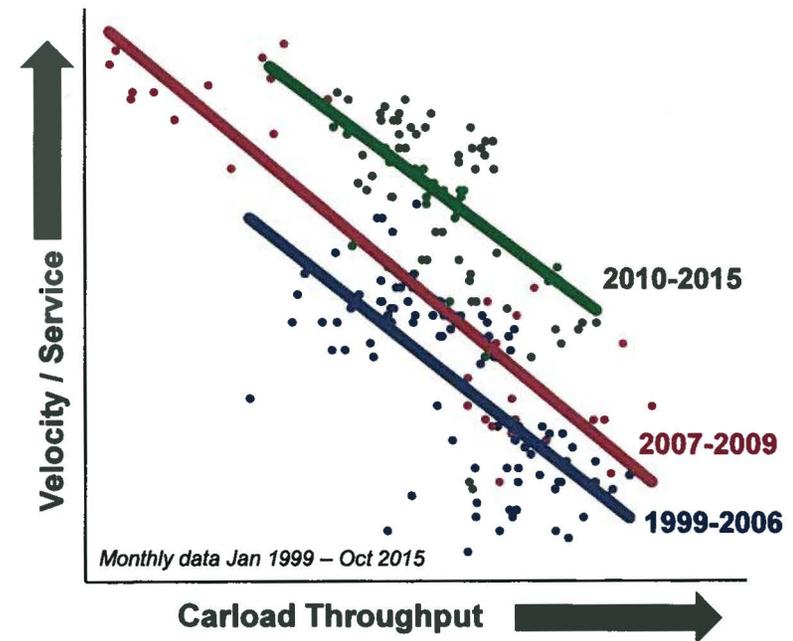


Throughput

- **Critical measure of network performance and fulfilling customer demand, especially viewed with velocity**
- **As carload volumes increase, service levels may decline**
 - Downward slope of regression lines
 - Consumption of road and terminal capacity
- **7-Day Carloading Rates are publicly available on the internet from all Class-I railroads**



UP Carloading Rate vs Service Performance



Congestion

1. Freight Car Inventory

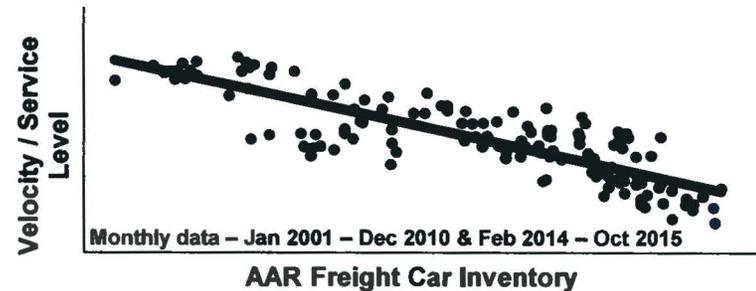
- A measure of network consumption and possibly congestion beyond thresholds that vary by railroad
- Freight Car Inventory is a function of multiple factors, including volume, velocity, terminal dwell, customer issues, gateway fluidity, weather and interruptions

2. Average daily % trains or train-hours held

- Includes all trains holding both at origin and en route
 - Calculate as a percentage of total trains or train-hours
 - Also capture reason for holding
- **Freight car inventory is currently reported through the AAR and the Interim Report**
 - **UP believes the trains held methodologies in both the Interim Report and NPRM are misleading**



Freight Car Inventory vs. Service Performance



Train Hours Held vs. Service Performance

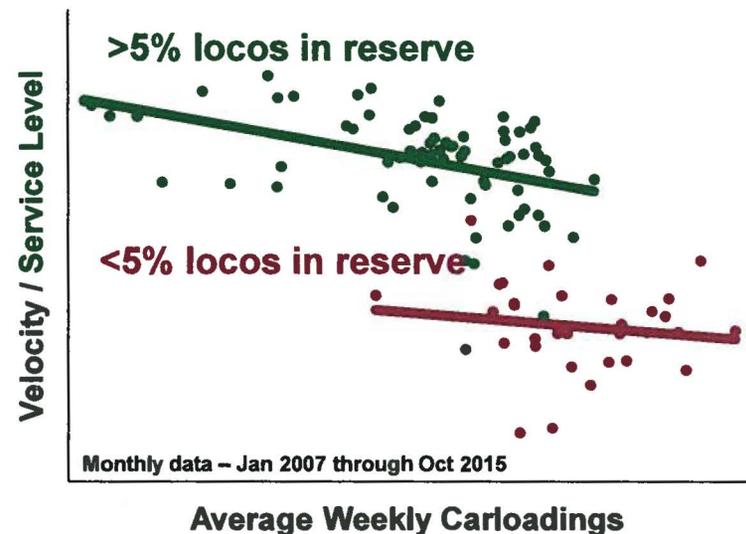




Surge capacity

- **During normal operating conditions, some reserve capability helps overcome:**
 - Demand spikes
 - Significant network interruptions, such as major storms
 - Disturbances in the national rail network
- **Acquiring additional resources often takes 9–12 months or more**
- **UP monitors and manages surge capacity**
 - Locomotives, freight cars, crews
 - Locomotives are the most fungible critical resource
 - Must manage productivity

UP Locomotives In Reserve vs Velocity/Service Level



More detailed metrics are not meaningful for the NPRM objectives



- **Railroad performance is just one of many factors that impact car orders and loading volumes. Other important factors include:**
 - Seasonality
 - International and domestic market demand
 - Strength of US dollar
 - Price of substitute commodities
 - Bottlenecks in other links of the logistics chain (shippers, receivers, interchange partners)
- **Fluctuations at commodity-specific and local levels may be normal for reasons not driven by rail operations, so do not indicate rail issues**
- **Comparability becomes even less valid at more detailed levels**
- **Too many metrics distract attention from the critical few, and potentially reveal customer-specific information**

Customers have access to their specific data for insight and planning



UP provides significant amounts of private information to its customers for their logistics and operations planning

- **Local variations communicated directly**
 - Weather
 - Maintenance projects
 - Network interruptions
 - Specific line and terminal issues
 - Interchange partner performance
- **Union Pacific's customers can monitor service performance and cycle times for each origin, destination, and commodity**
 - Real time and historical data available through the Internet
 - 17.5 million queries of customer web applications YTD 2015
 - Shipment-specific and summary level metrics
 - Quarterly performance-quality reviews with key decision makers / local levels
 - Actionable data is too specific to be public

Comparing railroads is of limited value

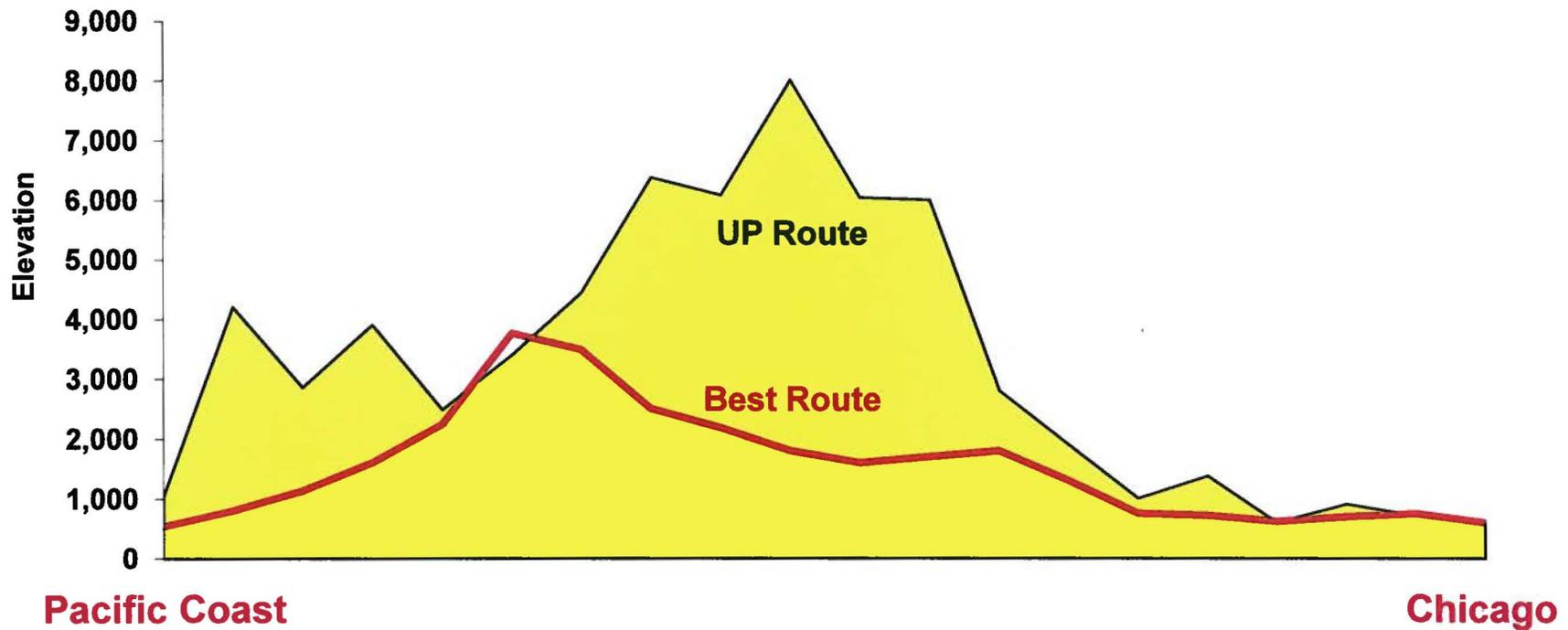
Each network is fundamentally different



- **Structural differences cause one railroad's metrics to be higher or lower than another's**
 - Traffic volumes
 - Commodity mix
 - Geography
 - Population density
 - Infrastructure design & engineering
 - Train schedules
 - Operational design
 - Information systems (data collection, processing & reporting)
- **Data normalization can reduce some of these differences**
 - Percent of mainline track miles with track condition slow orders
 - Train-hours or trains held as a percent of total
- **Best to track each railroad's operations over time**
 - Fewer metrics tracked over time would be far more valuable than adding more metrics

Chicago – Pacific Coast route comparison

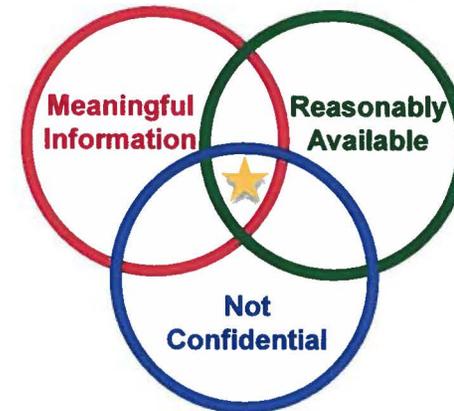
UP to Pacific Northwest vs best route



Summary



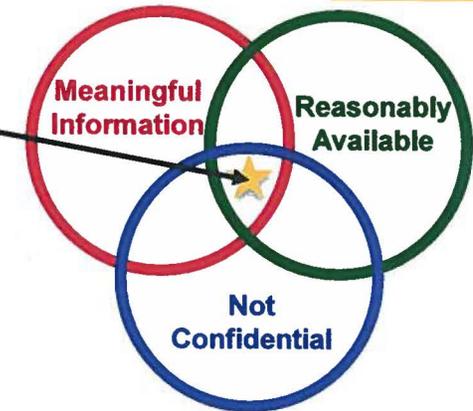
1. **A few relevant measures can provide more insight than a large number of secondary ones**
2. **The railroads are fundamentally different from one another**
3. **What's relevant to one railroad may not be relevant to another**
4. **The need for consistency in reporting rules requires fewer high-level metrics that satisfy:**
 - a) **Provide meaningful information**
 - b) **Reasonably available**
 - c) **Does not reveal confidential, proprietary data**



Summary



Metric Level	UP Input	Comment
High-Level Network	Few broad, network-level metrics provide good insight	7-Day Carloadings, Velocity, Terminal Dwell, Freight Car Inventory
Commodity & Geography-Specific	Not useful Not comparable Misleading Distraction	Fluctuation is normal depending on a multitude of factors
Detailed Customer-Specific	High degree of interaction already exists	Between individual customer and railroad (confidential and proprietary)



Most relevant metrics, from UP's perspective

- 7-day carloading throughput rates (total plus 22 major commodities)*
- Average train speed (system plus 8 major train types)
- Freight car terminal dwell average hours
- Freight car inventory (total plus 8 major car types)

Eliminate data elements 4 – 9 from the NPRM

4. Average dwell time at origin
5. Number of trains held short of destination
6. Number of loaded and empty cars in revenue service that have not moved
7. Grain cars loaded and billed, by state and train service
8. Outstanding grain car orders, average days late, new orders, filled orders, cancelled orders
9. Coal unit train loadings

* New or modified metric