Trucking Fleet Concept of Operations for Automated Driving System-equipped Commercial Motor Vehicles

Chapter 4: CONOPS Dataset - Dataverse

Authors: Manke, A., Werner, A., Krum, A., Adebisi, A., Hickman, J., Kerns, L., Stojanovski, O.

Partners: Pronto.ai

VIRGINIA TECH TRANSPORTATION INSTITUTE

PRONTO

July 2024

Abstract

Automated Driving Systems (ADS) are set to revolutionize the transportation system. In this project, the research team led by the Virginia Tech Transportation Institute developed and documented a concept of operations (CONOPS) that informs the trucking industry, government agencies, and non-government associations on the benefits of ADS and the best practices for implementing this technology into fleet operations.

The implementation of Automated Driving System (ADS)-equipped trucks as part of the Concept of Operations (CONOPS) involved three main use cases: port queuing, cross-country trips, and fleet integration. Each use case was designed to demonstrate the practical uses and data gathering capabilities of ADS technology. In the port queuing scenario at the Port of Oakland, efforts focused on refining ADS that could operate outside the port to reduce hours drivers lose waiting to enter the port to load or unload intermodal containers, with adjustments made to Pronto's technology to accommodate the unique driving behaviors typical at ports. The cross-country trips saw ADS-equipped trucks traverse various U.S. routes to evaluate infrastructure readiness, gathering data on lane markings, GPS signal strength, cellular connectivity, and road conditions. This information was utilized to develop a road readiness rating system, which is advantageous for government bodies and policymakers. Lastly, the fleet integration use case at the Port of Whittier in Alaska aimed to explore the organizational effects of integrating ADS into trucks that support marine-to-rail port operations, collecting data through observations and interviews to assess tasks, risks, and organizational structures.

These deployments are crucial for stakeholders like government agencies, technology developers, and infrastructure planners who benefit from detailed evaluations and operational insights provided by the extensive data collection.

The following chapter has been extracted from the final report. For access to the full report, see this link: https://www.vtti.vt.edu/PDFs/conops/VTTI_ADS-Trucking_CONOPS_Final-Report.pdf

4. CONOPS DATASET – DATAVERSE

4.1 IMPLEMENTATION

As part of this project, an open-source data repository, the VTTI CONOPS Dataverse, was developed to house the data from the CONOPS roadshows and deployments. The Dataverse, hosted by VTTI, serves as the interface for users accessing the data produced by this CONOPS study. The VTTI Dataverse has four separate collections consisting of 94 different datasets and 185 files. To ensure data privacy and security, VTTI drew on its own experience with naturalistic driving datasets and datasets containing sensitive information. A critical aspect to the success of the VTTI CONOPS Dataverse was its usability, easy access to background project information, and accessible documentation and training related to the website's query tool. The Dataverse features fields that provide an overview of the dataset, including a description of the project, subject, and keywords. The interface also provides information on the datasets, data directories, metadata, terms, and versions.

To complete data migration from the ADS deployments, data from the automated trucks were stored on an encrypted hard drive. The data were then removed from the automated truck by Pronto.ai and stored on its secure servers. Proprietary information was stripped from these data, and the remaining data were sent on an encrypted hard drive to VTTI. Upon arrival at VTTI, these data were first decrypted and stored on VTTI's secure server. Then, personally identifiable information (PII) was removed, and, if necessary, the data were filtered, smoothed, and uploaded to the VTTI CONOPS Dataverse for public access/viewing.

The Dataverse houses all the non-proprietary data collected over the course of the project:

- Data generated from the operation of the advanced driver assistance systems (ADAS)/ADS trucks (including video, kinematic, radar, GPS, and other sensors);
- Driver monitoring datasets from the ADS-equipped vehicles during the three deployments and use cases (port queueing, cross-country trips, and fleet integration); and
- Survey responses obtained from the public during the roadshows and the outreach events that gauge the perceptions and acceptance of ADAS/ADS technologies.

Researchers and decision-makers can access this data for their use.

4.2 DATA FROM ADS TECHNOLOGY ROADSHOWS

The VTTI team developed questionnaires that were used to investigate attitudes toward truck automation, use cases where automation will provide economic and/or safety benefits, and the ways in which truck drivers and the driving public can expect to interact with truck automation. The surveys also gathered demographic data to understand how different segments of truck fleets view ADS. The questionnaires were developed so they could be given at two different time points. In Chapter 2, the data were used to identify current gaps in the industry's understanding of truck ADS and how outreach activities can address this gap to improve attitudes toward truck ADS. Researchers also have access to this data on the Dataverse.

4.3 DATA FROM ADS TECHNOLOGY DEPLOYMENTS

This section provides details on the vehicle variables obtained from the three deployments. The data were collected and uploaded on the VTTI CONOPS Dataverse and can be directly accessed from the following link (https://dataverse.vtti.vt.edu/dataset.xhtml?persistentId=doi:10.15787/VTT1/ZYMSEM). Below is a brief description of the deployments and the data associated with each deployment use case. Figure 33 also summarizes datasets.

- Port Queuing: ADS-equipped trucks offer the potential to allow the vehicle to drive itself while queueing to be loaded or unloaded. The port queueing deployment included various use cases that involved interaction between an ADS and a driver at various stages of port operations. This involved: (1) a human driver manually driving the truck to the back of terminal's queue, (2) a driver engaging the ADS, and the truck automatically proceeding into the queue, through the terminal gate and inside the port, and (3) a driver disengaging the ADS and manually driving to the drop-off spot. Section 4.3.1 details the data collected during this deployment.
- Cross-Country: The purpose of this was to collect detailed inventories of ADS perception of sensory data on roadway features and the quality of supporting communications and location data. The intent was to develop a national dataset on the infrastructure and ADS performance metrics required for ADS operations. This data can be used to provide stakeholders and decision-makers on the infrastructure improvements required to support ADS integration into fleet operations. Section 4.3.2 details the data collected during this deployment.
- Fleet Integration: The Fleet Integration deployment focused on ADS state and safety metrics while being operated for revenue with a participating fleet on public/private roadways. The deployment use case provided video data and real-time vehicle information as an ADS-equipped truck navigated the unusual conditions at the port while also interacting with other vehicles and non-vehicular objects in its vicinity. Section 4.3.3 details the data collected during this deployment.

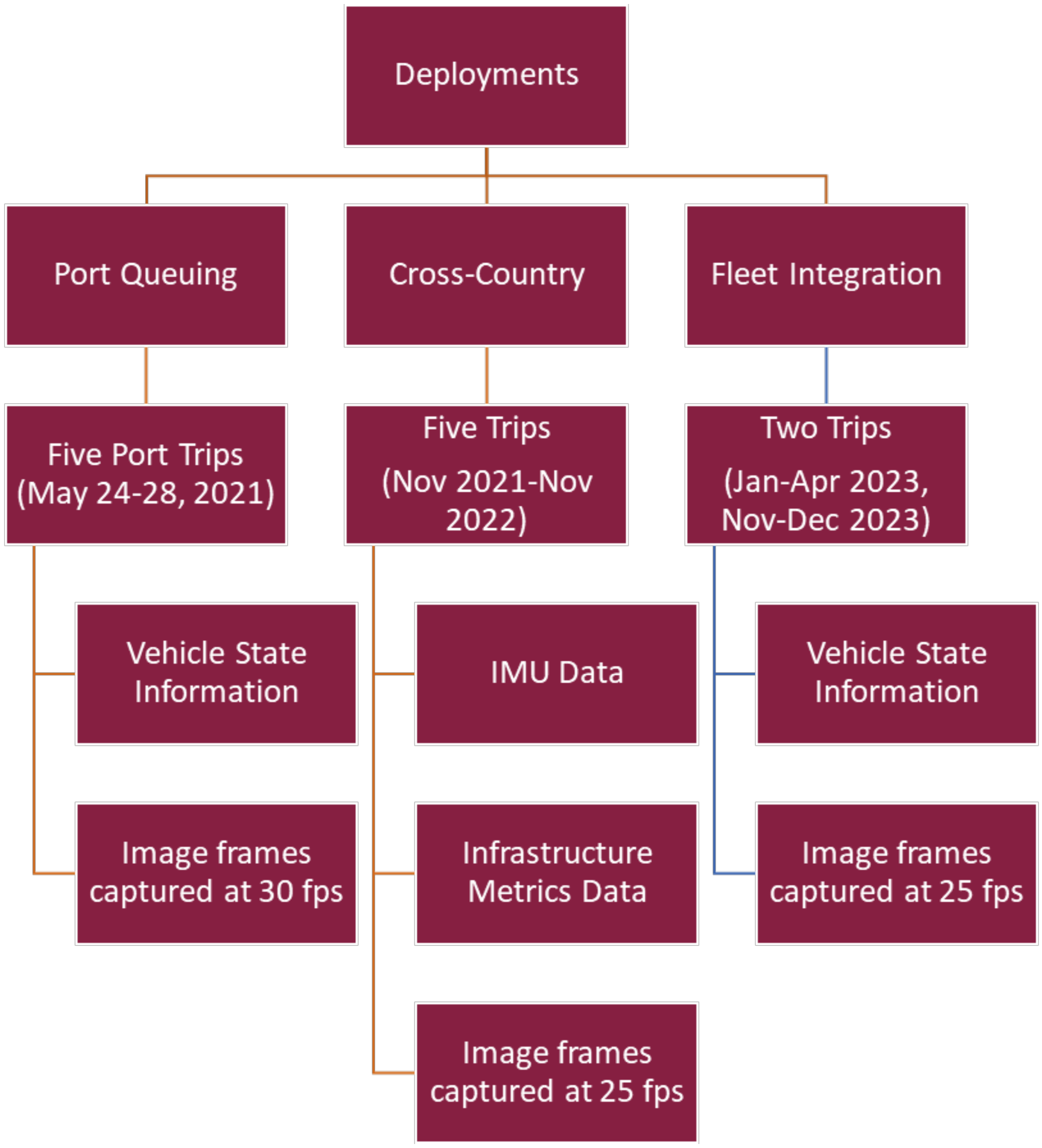

Figure 1. Diagram. High-level summary of data collected by Pronto during each operational use case deployment of ADS-equipped CMV available on the CONOPS Dataverse.

4.3.1 Port Queuing Deployment

The Port Queuing Deployment was conducted from May 24, 2021, to May 28, 2021. A total of 181 minutes of data was collected by the Pronto team from the five trips in Oakland, California. The data consists of vehicle state information, which is stored in CSV format. Images were also captured from the front-facing video stream during the port queueing. These frames are .jpeg format and have a uniquely identifiable name. Table 10–Table 14 provide the data collected at the port queueing deployment, including the information on the variables, the measurement unit, and the data type. The images were collected at a frequency of 30 frames per second (fps).

Table 1. File name: car state.

Variable Name	Measurement Unit	Possible values	Type
Applied Brake Pressure Primary	psi		Numeric
Brake	Percentage		Numeric
Drive State	Categorical	initializing-system booting, homing-Pronto DBW running calibration checks, not Ready-System booted, and calibration checks passed, ready to Engage- system ready for autonomous operation, engage-autonomous operation active, idling-Pronto system disabled	Text
Driver Brake	Categorical	True- safety driver pressed the brake pedal | False-safety driver did not press the brake pedal	Text
Driver Throttle	Categorical	True-safety driver requested throttle | False-safety driver did not request throttle	Text
Gear	Categorical		Numeric
PRNDL	Categorical		Text
Set Speed	m/s		Numeric
Steering Wheel Angle	Degrees		Numeric
System Engaged	Categorical	True- Autonomous system is engaged | False- Autonomous system is not engaged	Text
Throttle	Percentage		Numeric
Vehicle Speed (vEgo)	m/s		Numeric
Time Stamp	Time (in seconds)		Numeric

Table 2. File name: IMU.

Variable Name	Measurement Unit	Type
Accel X	m/s^2	Numeric
Accel Y	m/s^2	Numeric
Accel Z	m/s^2	Numeric
frame	north-east-down	
altitude	meters	Numeric
latitude	decimal degrees	Numeric
longitude	decimal degrees	Numeric
pitch	degrees/second	Numeric
roll	degrees/second	Numeric
yaw	degrees/second	Numeric
vx	m/s	Numeric
vy	m/s	Numeric
vz	m/s	Numeric

Table 3. File name: gpsRTK.

Variable Name	Measurement Unit	Type
altitude	meters	Numeric
latitude	decimal degrees	Numeric
longitude	decimal degrees	Numeric
status	Categorical	Text
frame	earth-centered earth-fixed	Text
x coordinate	decimal degrees	Numeric
y coordinate	decimal degrees	Numeric
z coordinate	decimal degrees	Numeric
Time Stamp	Time (in seconds)	Numeric

Table 4. File name: frames.

Variable Name	Measurement Unit	Type
Time Stamp	Time (in seconds)	Numeric
File name of the jpeg image captured	File name (text)	Text
Latitude	decimal degrees	Numeric
Longitude	decimal degrees	Numeric

4.3.2 Cross-Country Deployment

One of the objectives from the cross-country deployment is to collect, organize, and share data on infrastructure quality by driving an automated truck across various routes under a variety of conditions. Key infrastructure metrics include (1) cellular LTE connectivity, (2) lane marking quality, (3) road bumpiness, and (4) GPS satellite coverage. Based on the dictionary of the dataset collected from the cross-country drive, these infrastructure metrics were quantified by the following variables:

- Cellular LTE connectivity: *Signal Strength* (%)
- Lane marking quality: *Lane Score of Road* (%)
- Road bumpiness: *Road Condition* (Smooth or Bumpy)
- GPS satellite coverage: *GPS Satellites* (Count)

Table 15 summarizes the dataset collected and stored from the cross-country drive. In addition to the four infrastructure measures listed above, the dataset contains information on truck acceleration, motion basics (roll, pitch, and yaw), location, and speed. Table 16 and Table 17 provide more detail on the measurement units, the description, and the data type.

Table 5. Data from cross-country deployments.

Variable	Unit	Mean	Std Dev	Absmean
Acceleration X	m/s^2	✓	✓	✓
Acceleration Y	m/s^2	✓	✓	✓
Acceleration Z	m/s^2	✓	✓	✓
Roll	Deg	✓	✓	✓
Pitch	Deg	✓	✓	✓
Yaw	Deg	✓	✓	✓
Roll rate	Deg/s	✓	✓	
Pitch rate	Deg/s	✓	✓	
Yaw rate	Deg/s	✓	✓	
Latitude	Deg			
Longitude	Deg			
Signal Strength	%			
GPS Satellites	Count			
Speed	mph			
Time	ms			
Lane Score of Road	%			
Road Condition	Categorical			
State				

The Cross-Country deployment occurred in multiple trips. Table 15 shows the list of trips completed during the duration of the CONOPS projects.

Table 6. Five road trips that composed the cross-country deployment data collection.

Trip Name	Duration of the Trip
Nationwide Cross-Country Loop	October 25, 2021, to December 01, 2021
California – Texas (Round trip)	December 14, 2021, to December 21, 2021
Calgary, Canada – California (One way trip)	January 12, 2022, to February 06, 2022
California – Florida (Round trip)	February 28, 2022, to March 13, 2022
California – Oregon – Washington – Idaho – Montana – Wyoming – Utah – Arizona – Nevada – California	November 12, 2022, to November 17, 2022

The completed trips consist of IMU data and infrastructure metrics data. The data collected were stored in CSV format. Images were captured at a frequency of 25 fps from the front-facing camera. The image frames are in .jpeg format and have a uniquely identifiable name. The following variables were collected during the cross-country trips.

Table 7. Data type: IMU.

Variable Name	Measurement Unit	Description	Type
Accel X_Mean	m/s^2	Mean acceleration of the vehicle over 1 second in the x direction with respect to the vehicle body frame	Numeric
Accel Y_Mean	m/s^2	Mean acceleration of the vehicle over 1 second in the y direction with respect to the vehicle body frame	Numeric
Accel Z_Mean	m/s^2	Mean acceleration of the vehicle over 1 second in the z direction with respect to the vehicle body frame	
Roll_Mean	degrees	Mean roll angle of vehicle over 1 second	Numeric
Pitch_Mean	degrees	Mean pitch angle of vehicle over 1 second	Numeric
Yaw_Mean	degrees	Mean yaw angle of vehicle over 1 second	Numeric
Accel X_stddev	m/s^2	Standard deviation of acceleration of the vehicle over 1 second in the x direction with respect to the vehicle body frame	Numeric
Accel Y_stddev	m/s^2	Standard deviation of acceleration of the vehicle over 1 second in the y direction with respect to the vehicle body frame	Numeric
Accel Z_stddev	m/s^2	Standard deviation of acceleration of the vehicle over 1 second in the z direction with respect to the vehicle body frame	Numeric
Roll_stddev	degrees	Standard deviation of roll angle of vehicle over 1 second	Numeric
Pitch_stddev	degrees	Standard deviation of pitch angle of vehicle over 1 second	Numeric
Yaw_stddev	degrees	Standard deviation of yaw angle of vehicle over 1 second	Numeric
Accel X_absmean	m/s^2	Mean of absolute value of acceleration of the vehicle over 1 second in the x direction with respect to the vehicle body frame	Numeric
Accel Y_absmean	m/s^2	Mean of absolute value of acceleration of the vehicle over 1 second in the y direction with respect to the vehicle body frame	Numeric
Accel Z_absmean	m/s^2	Mean of absolute value of acceleration of the vehicle over 1 second in the z direction with respect to the vehicle body frame	Numeric
Roll_absmean	degrees	Mean of absolute value of roll angle of vehicle over 1 second	Numeric
Pitch_absmean	degrees	Mean of absolute value of pitch angle of vehicle over 1 second	Numeric
Yaw_absmean	degrees	Mean of absolute value of yaw angle of vehicle over 1 second	Numeric
Rollrate_mean	degrees/second	Mean roll angle rate of vehicle over 1 second	Numeric
Pitchrate_mean	degrees/second	Mean pitch angle rate of vehicle over 1 second	Numeric
Yawrate_mean	degrees/second	Mean yaw angle rate of vehicle over 1 second	Numeric
Rollrate_stddev	degrees/second	Standard deviation of roll angle rate of vehicle over 1 second	Numeric
Pitchrate_stddev	degrees/second	Standard deviation of pitch angle rate of vehicle over 1 second	Numeric
Yawrate_stddev	degrees/second	Standard deviation of yaw angle rate of vehicle over 1 second	Numeric

Table 8. Data type: infrastructure metrics.

Variable Name	Measurement Unit	Description	Type
Latitude	degrees	Latitude position	Numeric
Longitude	degrees	Longitude position	Numeric
Signal Strength	percentage	Received signal strength percentage for LTE Modem	Numeric
GPS Satellites	Count	Number of GPS satellites that are visible to vehicle	
Speed	mph	Vehicle speed	Numeric
Time	milliseconds	UTC time in milliseconds	
Lane Score of Road	percentage	Score between 0 and 1 indicating the ability to detect lane lines. 1 is the best score whereas 0 is the worst score.	Numeric
Road Condition	Categorical	Road condition of “Smooth” or “Bumpy” calculated over each second	Text
State		US State associated with reported position	Text

Road Lane Score

The dataset has lane scores between 0 and 1. The scores indicate the ability to detect lane lines. Here, a score of 1 or close to 1 is the best score, whereas a score of 0 is the worst score. The road lane score was calculated using Polyscore, which is shown below.

$$PS = 1 - \frac{\sum_{i=1}^{N} \left| |\hat{x}_S - \hat{x}| - \epsilon \right| / \epsilon -}{N} \begin{cases} 0 & if\ PS \leq 0 \\ PS & otherwise \end{cases}$$

$$\begin{cases} N = \text{number of keypoints} \\ \epsilon = \text{constant} \\ \hat{x}, \hat{x}_s = \text{coordinates of the predicted keypoints} \end{cases}$$

The Polyscore is a confidence estimation that is computed from test time augmentation of the image. The Polyscore requires an image and the inference from the PolyNet model. The process starts with a network that has an image and will predict the location of the lane lines in the image. It compares this to the lane line detection camera. This is an internally trained model created by Pronto. After, the confidence in the prediction is computed versus the actual image. A low score means that the prediction is less confident (between 0 and 1). The higher the score, the more confident the prediction is. A negative lane score is corrupted or not usable. The only gaps in Polyscore occur if there was no image collected to analyze, so a score of zero is still providing data.

Road Condition (Bumpy and Smooth)

Road condition was calculated using car state, such as acceleration, yaw, pitch, roll, and speed. The road condition is computed only when vehicle velocity is greater than 40 mph.

Model: Binary Support Vector Machine Classifier

Input features:

$$\begin{cases} \mu_{a_z} = \text{average acceleration} \\ \sigma_{a_z} = \text{standard deviation of the aceleration} \\ \mu_{\psi} = \text{average pitch} \\ \sigma_{\psi} = \text{standard deviation of the pitch} \end{cases}$$

Signal Strength

The raw signal strength is a value in the range [0, 31]. The signal strength as a percentage is computed as: "**percentage_SignalStrength" = "raw_SignalStrength" / 31.** The command AT+CSQ returns signal strength, and the returned values range between 0 and 31. The returned values are mapped to the received signal strength indicator, which is measured in dBm; it is a measure of received power.

4.3.3 Fleet Integration Deployment

The Fleet Integration Deployment was conducted from January–April and Nov–Dec of 2023 in Alaska. The data consists of vehicle state information which is stored in CSV format. Images were captured from the front-facing camera. The image frames are in .jpeg format and have a uniquely identifiable name. The images were collected at a frequency of 20–25 frames per second (fps). Table 18 lists the variables information was collected in car state files, and Table 19 lists the variables information was collected in frame files.

Table 9. File Name: car state

Variable Name	Measurement Unit	Possible values	Type
Brake	Percentage		Numeric
Drive State	Categorical	initializing-system booting, homing-Pronto DBW running calibration checks, not Ready-System booted, and calibration checks passed, ready to Engage- system ready for autonomous operation, engage-autonomous operation active, idling-Pronto system disabled	Text
Driver Brake	Categorical	True-safety driver pressed the brake pedal | False-safety driver did not press the brake pedal	Text
Driver Throttle	Categorical	True-safety driver requested throttle | False-safety driver did not request throttle	Text
Gear	Categorical		Numeric
PRNDL	Categorical		Text
Steering Wheel Angle	Degrees		Numeric
System Engaged	Categorical	True-Autonomous system is engaged | False-Autonomous system is not engaged	Text
Throttle	Percentage		Numeric
Time Stamp	Time (in seconds)		Numeric

Table 10. File name: frames.

Variable Name	Measurement Unit	Type
Time Stamp	Time (in seconds)	Numeric
File name of the jpeg image captured	File name (text)	Text
Latitude	decimal degrees	Numeric
Longitude	decimal degrees	Numeric

4.4 ACCESS AND USAGE

All the data collected by the CONOPS project are accessible via the VTTI CONOPS Dataverse. The datasets are minted with permanent digital object identifier (DOI) citations and published on the VTTI Dataverse. The datasets do not contain proprietary or confidential information, hence, there are no concerns regarding privacy, ethics, or confidentiality. The data management rights have been transferred to the curators of the VTTI CONOPS Dataverse. The data is available for open sharing under the Creative Commons Zero (CC0) universal public domain dedication. Under CC0, data and derivative products are available for reuse and redistribution without restriction. The VTTI CONOPS Dataverse meets the criteria outlined in the Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan. The CONOPS Dataverse promotes an explicit mission of digital data archiving, which is described on the Dataverse website and is listed by the USDOT as a Data Repository Conformant with the DOT Public Access Plan at https://ntl.bts.gov/publicaccess/repositories.html.

4.4.1 Data Organization

Port Queuing: Pronto's logging system records data in 1-minute-long chunks. For each of the trips, Pronto provided CSV files and images. The data for Port Queuing were uploaded on Dataverse in two folders. One folder is dedicated to all the CSV files organized by trip. Each trip folder contains subfolders for each file name (such as carState, IMU, gpsRTK, frames) organized as file name-hr-min. The second folder will have all the images in .jpeg format. The images will also be organized by trip.

Cross-Country: As described for Port Queuing, Pronto's logging system records data in 1-minute-long chunks. A total of five cross-country trips were completed. The trips were as follows: (1) Nationwide Loop, (2) San Francisco–Texas–San Francisco, (3) San Francisco–Calgary–San Francisco, (4) San Francisco–Orlando–San Francisco, and (5) California–Montana–California. The data were uploaded on the Dataverse by State, followed by the trip name. Within the State folder, there are two folders. One folder has the CSV files containing the data that were collected on a particular day. The second folder has images in .jpeg format. The files in the image folder are organized as day-hr-min.

Fleet Integration: The Fleet Integration use case dataset is similar to the port queuing deployment dataset, with information on the vehicle state information stored in CSV and the front-facing images captured during the deployment. The data is also organized in a similar manner to the Port Queueing data organization.