

# Secure Data Commons

## Data Provider User Guide

[www.its.dot.gov/index.htm](http://www.its.dot.gov/index.htm)  
Draft Report — March 26, 2020  
FHWA-JPO-18-xxx

## **Notice**

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. The U.S. Government is not endorsing any manufacturers, products, or services cited herein and any trade name that may appear in the work has been included only because it is essential to the contents of the work.

---

## Revision History

#	Name	Version	Revision Date	Revision Description
1	REAN Cloud	1.0	08/02/2018	Initial Draft
2	Volpe	1.1	08/08/2018	Formatting
3	REAN Cloud	1.2	09/13/2018	Add Export Functionality Documentation
4	SDC Team	1.3	01/08/2020	Updated links to Data Analyst user guide, updated Chapter 4 title, fixed grammar throughout
5	SDC Team	1.4	03/26/2020	Added Auto-Export Permissions narrative to Chapter 5

# Table of Contents

- Table of Contents ..... iii
- Chapter 1. Introduction and Document Overview ..... 5
- Chapter 2. Initial Setup and Validation ..... 6
- Chapter 3. Accessing Workstations ..... 7
- Chapter 4. Data Ingestion ..... 8
  - Ingesting Data to S3 ..... 8
  - Ingesting Data to Kinesis Firehose ..... 8
- Chapter 5. Exporting Data Out of SDC ..... 10
  - Approving Export Requests ..... 10
  - Automatic Export Permissions ..... 11
  - Review Actions ..... 12

# List of Figures

Figure 1: SDC Portal – Export Requests Tab ..... 10  
Figure 2: SDC Portal – Data Analyst Auto-Export Request Form ..... 11  
Figure 3: SDC Portal – Approval List for Auto-Export Requests ..... 12

# Chapter 1. Introduction and Document Overview

The Secure Data Commons (SDC) is a United States Department of Transportation (U.S. DOT) sponsored cloud-based analytical sandbox designed to create wider access to sensitive transportation data sets, with the goal of advancing the state of the art of transportation research and state/local traffic management.

The SDC stores sensitive transportation data made available by participating data providers, and grants access to approved researchers to these datasets. The SDC also provides access to open source tools and allow researchers to collaborate and share code with other system users.

The SDC platform is a research environment that allows users to conduct analyses and do development and testing of new tools and software products. It is not intended to be an alternative to any local jurisdiction's traffic management center or local data repository. The existing SDC provides users with the following data, tools, and features:

- **Data:** The SDC is ingesting several datasets currently. Additional data sets will be added to the environment over time. Users can bring their own data into the environment to use along with the Waze data.
- **Tools:** The environment provides access to open source tools including Python, RStudio, Microsoft R, SQL Workbench, Power BI, Libre Office, and Jupyter Notebook. These tools are available on a virtual machine in the system enabling data analytics in the cloud.
- **Functionality:** Users can access and analyze data within the environment, save their work to a virtual machine, and publish processes and results to share with others. Virtual machines can be configured with additional CPU/RAM and scheduled uptime during global workstation shutdown periods.

The SDC platform supports two major roles:

- **Data Providers** – These are entities that provide data hosted on the SDC platform. The data provider establishes the data protection needs and acceptable use terms for the data analysts.
- **Data Analysts** – These are entities that conduct analysis of the datasets hosted in the SDC system. Note that analysts can bring their own data and tools into the SDC system.

This document provides guidance for the **Data Provider** role. A similar guide has been prepared for the Data Analysts which can be accessed by clicking [here](#).

## Chapter 2. Initial Setup and Validation

This chapter provides guidance on the initial setup and validation of the user into the SDC system. Please refer to Chapter 2 of the Data Analyst's user guide, which can be accessed by clicking [here](#).

# Chapter 3. Accessing Workstations

Users are assigned cloud-based workstations to perform analysis on the datasets. This section provides a description of how to launch and use these workstations. Please refer to Chapter 3 of the Data Analyst's user guide, which can be accessed by clicking [here](#).



# Chapter 4. Data Ingestion

This chapter provides guidance on how to perform data ingestion to the SDC platform by different providers. Two approaches are currently supported by the SDC platform to ingest Connected Vehicles (CV) data:

- Ingesting Data to Simple Storage Service (S3)
- Ingesting Data to Kinesis Firehose

## Ingesting Data to S3

Data providers can ingest data to S3 by following the below steps:

- Data providers will receive an instructions e-mail from [support@securedatacommons.com](mailto:support@securedatacommons.com) with set of instructions, shell script as attachment and below set of values:

```
API_KEY          = REAN CLOUDREAN CLOUD
AUTH_CODE        = REAN CLOUDREAN CLOUD
ROLE_NAME        = REAN CLOUDREAN CLOUD
API_END_POINT    = REAN CLOUDREAN CLOUD
```

- Download the shell script attached to instruction e-mail.
- Update the above provided values in the shell script.
- Please ensure you have the following folder and files under your home directory:

```
\--aws
  |--config
  \--credentials
```

- Please make sure **jq** python library is installed on your machine.
- Run the downloaded shell script which will generate a temporary access/secret keys and update `~/.aws/credentials` files by creating a new profile named **sdc-token**. If you already have the profile named **sdc-token**, it will overwrite by updating the credentials.
- Use the **sdc-token** profile in your aws commands or python code through which you are uploading data to S3.
- Make sure to run the above shell script before you start uploading data to S3.

## Ingesting Data to Kinesis Firehose

Data providers can ingest data to S3 by following the below steps:

- Data providers will receive an instructions e-mail from [support@securedatacommons.com](mailto:support@securedatacommons.com) with a set of instructions, a shell script as attachment, and the set of input values below:

```
API_KEY          = REAN CLOUDREAN CLOUD
AUTH_CODE       = REAN CLOUDREAN CLOUD
ROLE_NAME       = REAN CLOUDREAN CLOUD
API_END_POINT   = REAN CLOUDREAN CLOUD
```

- Download the shell script attached to instructions e-mail.
- Update the above provided values in the shell script.
- Please ensure you have the following folder and files under home directory:

```
\--aws
  |--config
  \--credentials
```

- Please make sure **jq** python library is installed on your machine.
- Run the downloaded shell script which will generate a temporary access/secret keys and update `~/.aws/credentials` files by creating a new profile named **sdc-token** and export access key/secret keys as **AWS\_ACCESS\_KEY\_ID / AWS\_SECRET\_ACCESS\_KEY** environment variables. If you already have the profile named **sdc-token**, it will overwrite by updating the credentials.

Make sure to run the above shell script before you ingest data to Kinesis Firehose.

# Chapter 5. Exporting Data Out of SDC

Data Analysts should be able to export the data out of the system, based on the compliance and data usage policies set forth by a Data Provider. Upon successful export request submission, the request will be sent to appropriate Data Providers. Data Providers are responsible for accepting or rejecting the export requests for the Data Analysts so that they are authorized to get the data products out of the SDC system.

## Approving Export Requests

Data Providers can see the requests in the Export Requests tab of the SDC web portal, as shown in Figure 1 below.

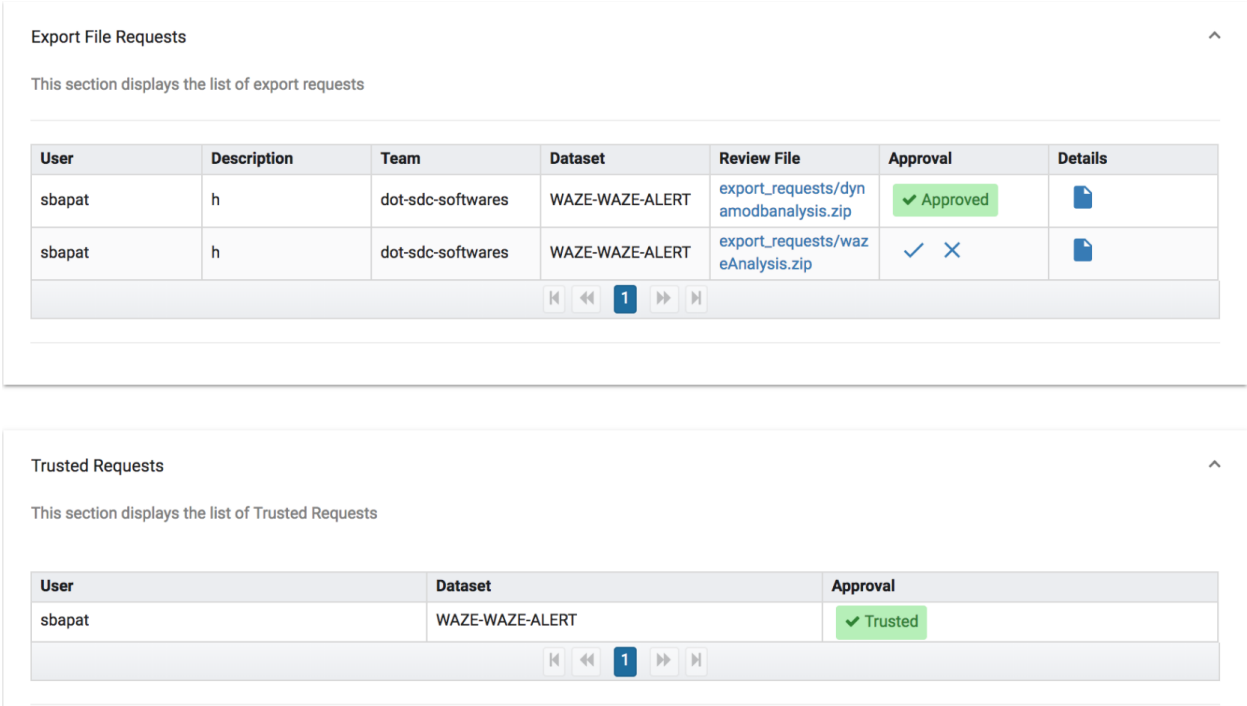


Figure 1: SDC Portal – Export Requests Tab

There are two sections for the requests made by the Data Analysts:

1. **Export File Requests:** These requests correspond to the approval form submitted by the Data Analyst for exporting any data file out of the SDC system.

The Data Provider can accept the request by clicking on the checkmark and reject the request by clicking on the “X” under the Approval column as shown above.

To view all the details regarding the request, the Data Provider can click on the file symbol under the Details column for each request. The Data Provider can click on the file

link to download the file related to the request and review it before giving the access to the Data Analyst.

2. **Trusted Requests:** These requests correspond to providing the trusted status to Data Analysts. The Data Provider can accept the request by clicking on the checkmark and reject the request by clicking on the “X.”

## Automatic Export Permissions

Trusted Data Analysts can request permission to automatically export exclusively derived data from the SDC for a specific Data Provider’s data. This allows the Data Analyst to export data without having to use the SDC Portal. When this permission is requested on the form as shown below, the Data Analyst is required to provide a name for the derived data they wish to be able to automatically export and a justification for why they are wanting this permission.

< Select Project Approval Form Trusted Status **Auto-Export Status** >

Auto-Export Status is a mechanism for analysts to obtain a passport from a data provider for automatic export of data. Obtaining this passport allows analyst to export their data automatically (for subsequent similar requests), as opposed to manually approving the request for export yourself (as a trusted user). If approved, this passport may be revoked by the data provider at any time. This status is acquired per Project + Data Provider + Sub-Dataset/Data Type.

Note - Based on the dataset and datatype selection, you currently do not have Auto-Export Status from this Data Provider. We will notify the Data Provider about your request and send it for approval. Your request will be processed based on the decision from the Data Provider.

Do you wish to request Auto-Export Status from the Data Provider?

Yes  No

Specify the derived dataset name \*  
Monthly\_Reports

Reason for requesting permission \*  
Daily automated machine

Submit

Figure 2: SDC Portal – Data Analyst Auto-Export Request Form

With this information, it is up to the Data Provider to choose whether to approve the Data Analyst for automatic export permissions. If the Data Analyst is approved, the SDC Admin will be informed to provide them with the proper permissions and instructions on how to start automatic exporting.

The Data Provider will be notified via email when a request for automatic export has been made. All requests may be viewed on the SDC Portal under the Export Requests tab. The Auto-Export requests section is located at the bottom of this page. The Data Provider can approve the Data Analyst for auto-export via the checkmark or deny them via the X mark.

Auto-Export Requests

This section displays the list of Auto-Export Requests

User	Dataset	Justification	Approval
tenglish	CVP-WYDOT-TestData	This is a test of the feature	✓ Approved
tenglish	CVP-WYDOT-Monthly reports	For an automated WYDOT machine	✓ ✕

1

Figure 3: SDC Portal – Approval List for Auto-Export Requests

The Data Provider does not have the direct ability to remove the permission for automatic exporting but instead must contact the SDC Support Desk to perform this task. The ability for the Data Provider to do this on their own will be implemented in a future iteration of the automatic export feature.

Whenever a Data Analyst uses the auto-export feature the Data Provider will be notified via email of the specific file that was exported and where they may be able to view it if they would like to check if the content is compliant with what was approved. This data is accessible to the Data Analyst from outside the SDC using their specific auto-export role and is located on a specific auto-export S3 bucket within the SDC.

## Review Actions

1. Notify: The request of the Data Analyst for trusted status or the export request is accepted automatically. The Data Provider is notified with an e-mail for the request that is accepted.
2. NotifyReview: The request of the Data Analyst is sent to the Data Provider for approval by sending him a notification over an e-mail. The Data Provider must accept or reject the request from the SDC web portal under the section named “Export Requests.”

U.S. Department of Transportation  
ITS Joint Program Office-HOIT  
1200 New Jersey Avenue, SE Washington, DC 20590

Toll-Free "Help Line" 866-367-7487

[www.its.dot.gov](http://www.its.dot.gov)

FHWA-JPO-18-XXX



U.S. Department of Transportation