FMCSA Data Repository

10th International Conference on Managing Fatigue
March 23, 2017

Rebecca Hammond
Miguel Perez
Richard Hanowski

Virginia Tech
Transportation Institute

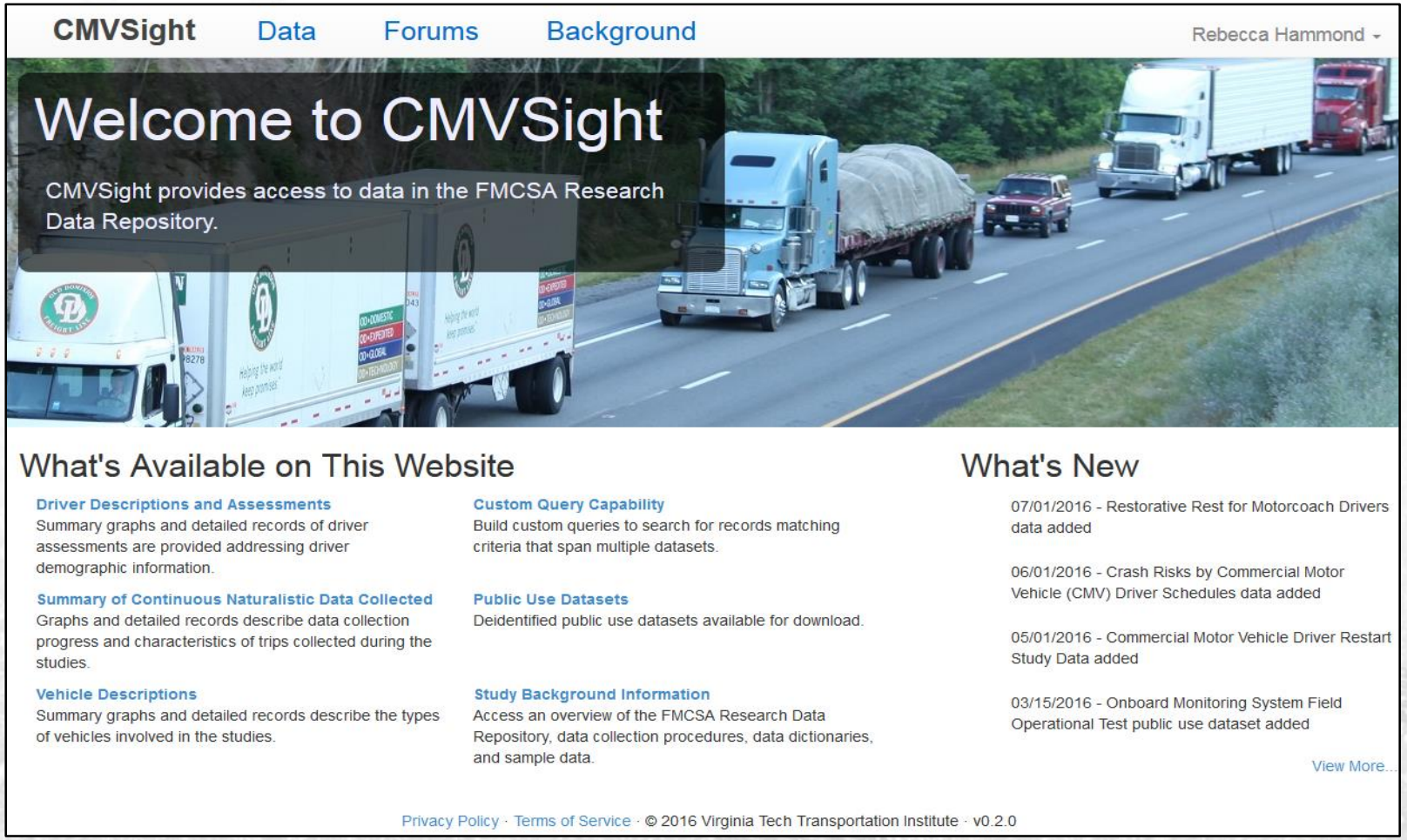
Project Overview

- Develop, maintain, and operate a data repository.
  - Include public-use datasets and complete raw study data.
- Develop web infrastructure for data access.
- Develop data access and management that includes oversight, support, and tracking.
- Provide ongoing support.

First Year Datasets

- Commercial Motor Vehicle Driver Restart Study
- Crash Risks by Commercial Motor Vehicle (CMV) Driver Schedules
- Restorative Rest for Motorcoach Drivers

User Website



- Data – Includes Query Builder which allows users to access data within a dataset or across multiple datasets.
- Forums – Used to post FAQs and facilitate interactions among users and host.
- Background – Provide background information about the studies and data collection processes.

Query Builder

Vehicles

- Vehicle types (car, truck, van, etc.)
- Vehicle ages and condition
- Amount of data collected per vehicle
- Quantities of vehicles installed
- Vehicle technologies and equipment

View…

Query Builder

- Select variables and conditions
- Submit query, assess results
- Build cross tabulations
- View graphs of output
- View table of individual records

View…

Trips

- Summary measures describing trips
- Trip length, duration, start time, stop time
- Min, max, mean for speed, acceleration
- Trip summary record table
- Trip density maps

View…

Drivers

- Numbers of participating drivers
- Amount of data collected per driver
- Driver demographics and driving history
- Driver physical and psychological state
- Driver participation experience

View…

Events

- Crashes, near crash, and baseline event records
- Events by type and severity
- Event viewer

View…

Access Control

- Security Levels – vary for Levels 1, 2, 3
- Login Account
  - Basic login – email address required only, no proof of IRB training; grants access to Level 1 (de-identified data).
  - Qualified Researcher login – proof of IRB training required; grants access to Level 2 and Level 3.
- Data Use License – Specifies which datasets a researcher has access to, the period of time over which the data can be used, and requirements for protecting PII. A data-use license will not be required for access to Level 1 datasets, but will be required for access to Level 2 and Level 3 data.

Data Access – Level 1: Open Access

- Provide de-identified, public-use datasets developed at the end of each study.
- Data may include reduced event and baseline characteristics, subsets of kinematic data from vehicles, Actigraph data, psychomotor vigilance task (PVT) data, and questionnaire data without PII.
- Users required to obtain basic login (no proof of IRB training required).
- VTTI will track all download activity from the website.

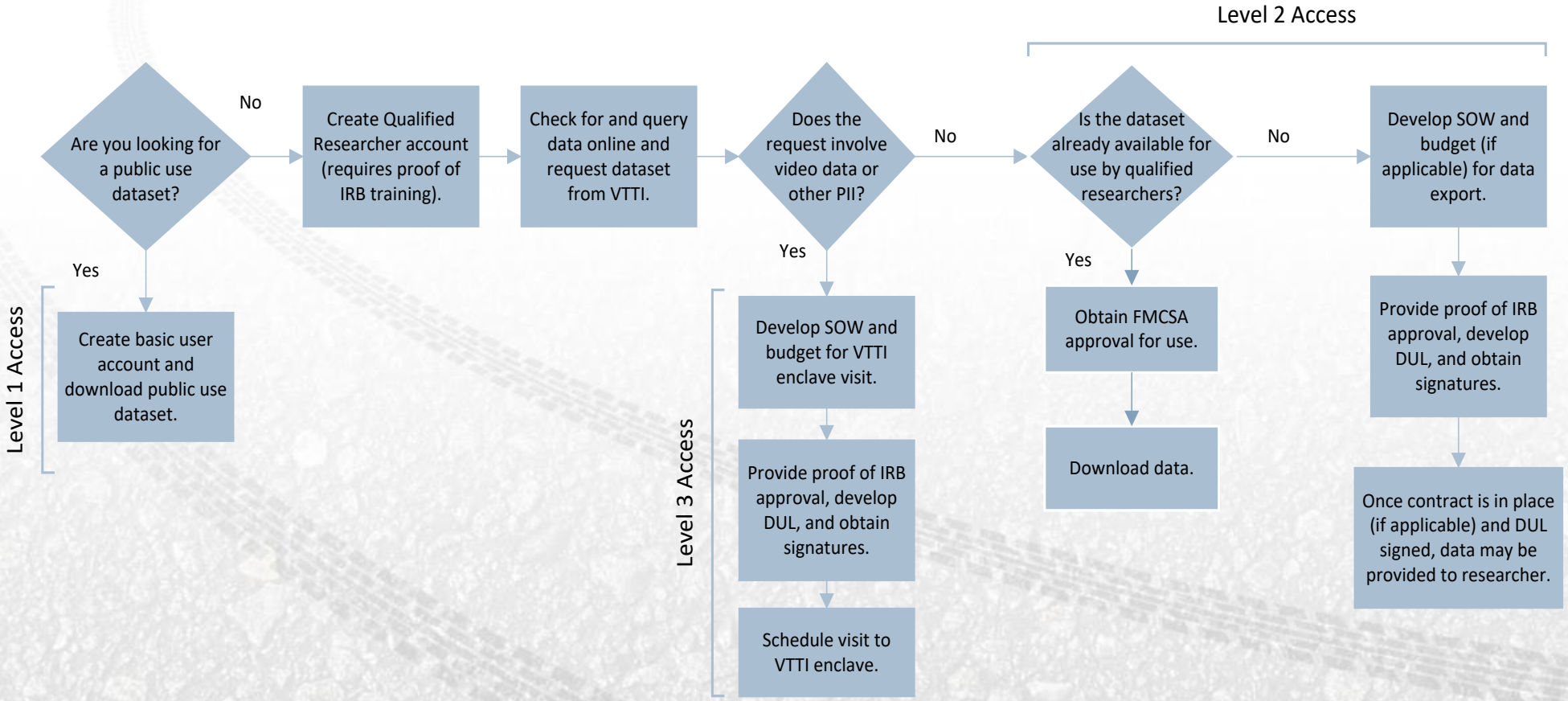
Data Access – Level 2: Identifiable Data with Secured Access

- Provide raw data for each study, except those data that require access from a secure data enclave.
- Users will be able to query data on the website within a single dataset, or across multiple datasets. Results of data will be displayed on the website, but not available for download. Users must request desired dataset from VTTI for research use.
  - PII will not be displayed via the website. Certain PII (e.g., GPS coordinates) may be requested under certain conditions.
- Users required to obtain FMCSA approval and qualified researcher login (requires proof of IRB training).
- Funding may be required depending on data request.
- VTTI will track all data requests and exports.

Data Access – Level 3: Identifiable Data Requiring Enclave Support

- Provide access to identifiable data that can only be accessed in a secure data enclave.
  - Video, GPS coordinates, date/time of SCE.
- Data not permitted to leave secure enclave at any time.
- Users required to obtain FMCSA approval and qualified researcher login (requires proof of IRB training).
- Funding required to cover cost of supporting the enclave visit and providing any de-identified data after the visit.
- VTTI will track all data requests and exports.

Data Access Flow Chart



IRB and Data Tracking Considerations

- VTTI will complete an IRB and tracking assessment for each dataset to be included in the repository.
  - Did the participants agree that data could be used for additional research and by researchers not part of the original study team?
  - When must the study data be deleted? Are there differences in deletion times for PII and non-PII data?
  - What types of PII data were collected?
  - Given the study context, what types of other public-use datasets are available that might allow for re-identification if combined with the study data?
- VTTI will apply its existing data-use license and data tracking processes.

Re-Identification Risk

- Prior to the release of each public-use dataset, the data must be assessed for re-identification risk.
- As new non-research databases become available to the public, existing public-use datasets may suddenly change risk status.
- For datasets determined to pose greater than minimal risk, an in-depth assessment will be performed.

Commercial Vehicle Datasets – Completed or Ongoing

Set	Dataset Title	Collection Dates	Trucks	Busses	Operation Type	Drivers	Miles	Data Collection System	Status/ Reference
1	Drowsy Driver Warning System Field Operational Test (FOT)	05/04 – 09/05	46	0	Line/Long Haul	103	2,300,000	100-Car	See Hanowski et al.
2	Naturalistic Truck Driving Study	11/05 – 05/07	9	0	Line/Long Haul	100	735,000	100-Car	See Blanco et al.
3	Heavy Vehicle Camera/Video Imaging System FOT	07/09 – 09/10	6	0	Long Haul	12	275,000	100-Car	See Fitch et al.
4	Advance System Testing Utilizing a Data Acquisition System on the Highways (FAST DASH) 1	09/11 – 08/12	19	0	Long Haul	21	1,335,000	NextGen	See Schaudt et al.
5	Winter Maintenance	01/13 – 04/13	2	0	Snowplow	4	Not available	NextGen	See Camden et al.
6	FAST DASH 2	09/13 – 08/14	17	0	Long Haul	27	1,450,000	NextGen	See Krum et al.
7	Onboard Monitoring System (OBMS): Motorcoach	05/13 – 07/15	0	44	Motor Coach	73	1,142,000	NextGen	Data reduction in progress
7a	Onboard Monitoring System (OBMS): Truck	02/12 – 03/13	206	0	Line/Long Haul	167	2,516,000	NextGen	Data reduction in progress
8	Crash Avoidance System (CAS) FOT	11/13 – 06/15	150	0	Line/Long Haul	180	3,245,000	MiniDAS	See Grove et al.
9	Canadian Truck Study	12/14 – 01/16	26	0	Long Haul	26	800,000 (est)	NextGen	Data collection complete, no reduction
10	Oil & Gas Operations	07/14 – 10/14	4	0	Maintenance/Service (Medium-Duty Pickup)	4	45,000	MiniDAS	Data collection and reduction complete

Questions?

rhammond@vtti.vt.edu