Livewire Data Platform – A Solution for Energy Efficient Mobility Systems (EEMS) Data Sharing

Lauren Spath Luhring
National Renewable Energy Laboratory
June 2, 2020

DOE Vehicle Technologies Program
2020 Annual Merit Review and Peer Evaluation Meeting

Project ID: eems066

This presentation does not contain any proprietary, confidential, or otherwise restricted information.

Overview

Timeline

- Project start date: 10/01/18
- Project end date: 09/30/21
  - Go/No-Go: 6/30/20
- Percent complete: 60%

Budget

- Total project funding: $3M
  - DOE share: $3M
- Funding for FY 2019: $1.5M
- Funding for FY 2020: $1.5M

Barriers

- Expansive community of relevant stakeholders
- Difficulty in sourcing empirical real-world data applicable to new mobility technologies, such as connectivity and automation

Partners

- Interactions/collaborations
  - Oak Ridge National Laboratory, Carnegie Mellon University, FOA awardees, Lab Call awardees
- Project leads
  - National Renewable Energy Laboratory (NREL), Pacific Northwest National Laboratory (PNNL), Idaho National Laboratory (INL)

Relevance – Historical Data Challenges

Data is a critical need for EEMS, but technical and cultural challenges stand in the way of sharing

Platform

- Already stored in home institutions
- Vulnerable to loss
- Difficult to find
- Access is not standardized

People

- Data is hoarded for competitive advantage
- Disinterest in sharing data
- Licenses/ non-disclosure agreements (NDAs)/ restrictions limit data sharing
- Data is generated by researchers, not computer scientists

Livewire Data Platform (LDP) focuses on both the platform and people to build a usable system for sharing EEMS data

Milestones

Go/No-Go 6/30/20

Name	Criteria	Date
(NREL, INL, PNNL) Livewire Data Platform Viability; Determination that the Livewire Data Platform is facilitating easy and secure data sharing that is enabling EEMS research.	Demonstrate growth on all usage metrics compared to the baseline	6/30/20

Completed Milestones

Milestone Name/Description	End Date	Status
Develop Beta Livewire platform and implement API platform, exposing at least 3 APIs	6/30/19	Complete
Implement first iteration of data catalog through the DataHUB including 10 datasets	9/30/19	Complete
(NREL) Slide Deck and WebEx briefing summarizing meetings, accomplishments, lessons learned and progress status for the quarter including establishing a baseline for project usage and providing one testimonial on usefulness of platform.	12/31/19, 3/31/20	Complete
(INL) Complete and share expanded version of detailed metadata for DOE/UMich CAV dataset shared through LDP, including characterization of data quality	12/31/19	Complete
(PNNL) Develop the server side capability for users to easily upload files to the Livewire Data Platform.	12/31/19	Complete
(NREL, INL, PNNL) Obtain agreement between Livewire Team and DOE HQ on metrics growth targets and update 6/30/20 Go/No-Go Decision accordingly	1/31/20	Complete
(INL) Develop and share initial versions of detailed metadata for 3 additional datasets shared through the LDP, based in part on prototype tools to automate or streamline metadata creation from existing information sources.	3/31/20	Complete
(PNNL) Develop the user interface of the Livewire Data Platform for users to drag and drop files.	3/31/20	Complete

Milestones

Upcoming Milestones

Milestone Name/Description	End Date	Status
(NREL) Annual Merit Review Presentation or Poster summarizing accomplishments, lessons learned, and progress status for the project.	6/30/20	On Track
(INL) Develop prototype tools to automate/streamline data quality characterization for inclusion in detailed metadata for data sets shared through LDP.	9/30/20	On Track
(PNNL) Develop and test the capability to integrate api.data.gov and Livewire Data Platform for principal investigators (PIs) to manage (add, edit, or delete) groups and users.	6/30/20	On Track
(NREL) Slide Deck and WebEx briefing summarizing meetings, accomplishments, lessons learned, progress status, and usage metrics for the quarter, including one testimonial on usefulness of the platform.	9/30/20	On Track
(INL) Complete and share expanded versions of detailed metadata for three data sets shared through LDP, including characterization of data.	9/30/20	On Track
(PNNL) Release the capability within Livewire Data Platform for principal investigators to manage (add, edit, or delete) groups and users. This capability will allow PIs to manage access controls for active users and deletion of inactive users that are not associated with an institution.	9/30/20	On Track

Approach – Livewire Data Platform Tasks

Task	Description	Partner
1	Create Data Management Platform and Expose Data Sets	
1.1	Develop DataHUB	PNNL
1.2	Develop API platform	NREL
1.3	Develop metadata structure and data quality assurance processes	INL
1.4	Provide data to Livewire	NREL, INL, and EEMS partners
2	Build Complex Data-Management Capabilities	
2.1	Evaluate and gather EEMS data needs	NREL and INL
2.2	Maintain and share Fleet DNA data	NREL
2.3	Maintain and share Transportation Secure Data Center (TSDC) data	NREL
2.4	Evaluate complex EEMS data sets in order to create data-sharing heuristics and prototypes	NREL
3	Facilitate Partnerships, Data Collection, and Resolution of Data-Sharing Challenges	NREL and INL

Timeline overview:

- Throughout – Maintain and extend the existing successful FleetDNA and TSDC data efforts
- Year 1 – Emphasis on developing the new LDP and inventorying desired datasets to include; initial outreach for data and users
- **Year 2 – Launch the LDP; shift emphasis to growth of datasets, users and tracking platform impacts/benefits; support complex data**
- Year 3 – Pending successful go/no-go in Year 2, transition to operation of LDP: continuing to grow data/users and track impacts/benefits

Approach – Enable Data Sharing

- The Livewire Data Platform (LDP) addresses historical challenges by providing:
  - A **platform** – allows data sharing and discovery by linking disparate data sources to a single website through
    - Hosting data on and allowing direct download from Livewire
    - Application programming interfaces (APIs)
    - Links to data sources
  - A **community** – builds partnerships and collaboration rather than competition
  - A **system** – allows shared data to increase in size and complexity as EEMS evolves

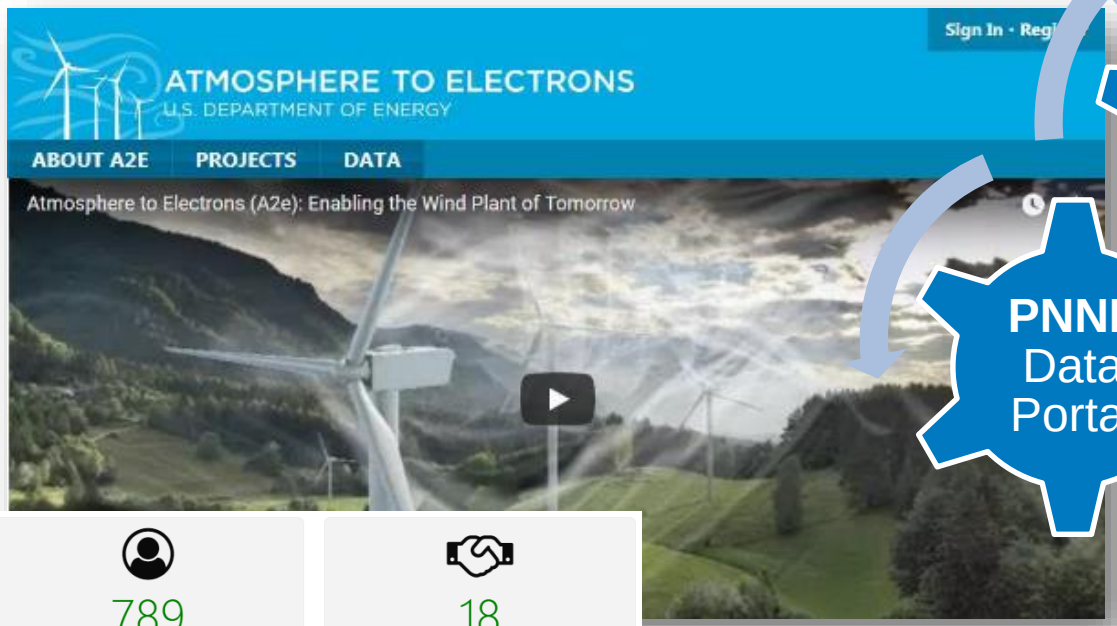
- The Livewire Data Platform development is built around data owners' needs

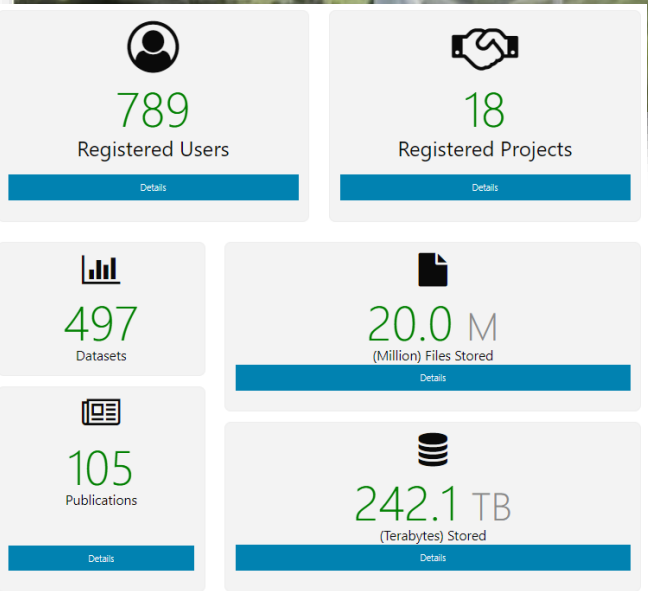
- New features enable data owners to:
  - Upload their own data and validate file names
  - Request information on how data will be used before granting access
  - Manage visibility and access to data

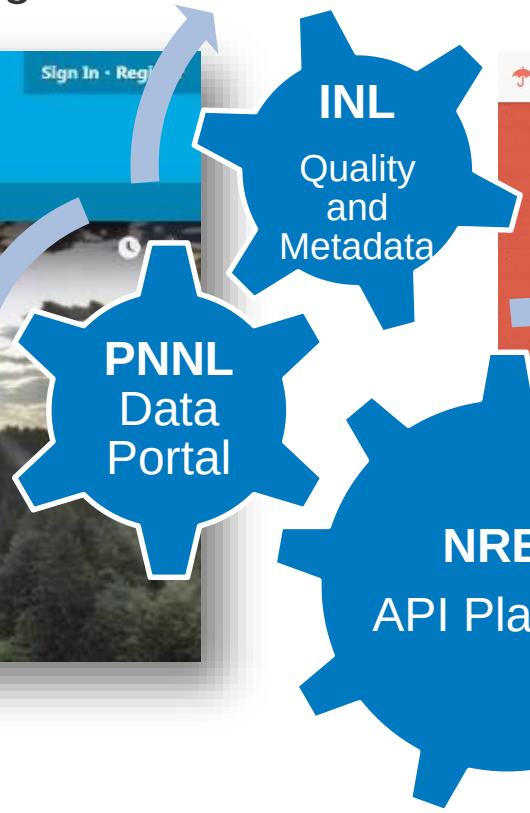
Goal: Remove barriers and give EEMS researchers the data they need to answer big transportation questions

Approach – Leverage Existing Successful Data Platforms

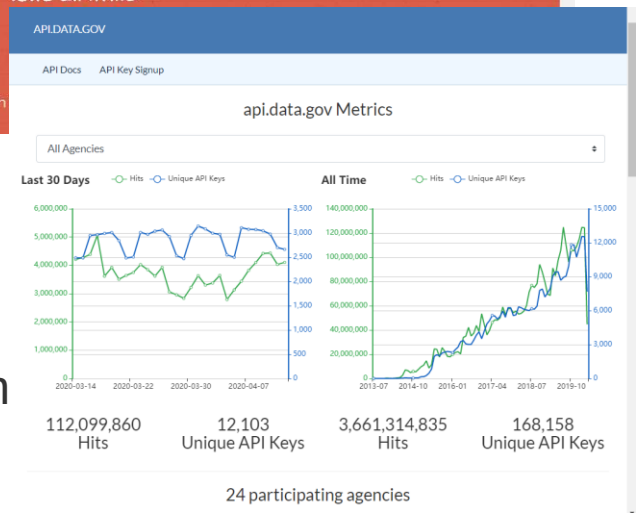
Atmosphere to Electrons – a2e.energy.gov





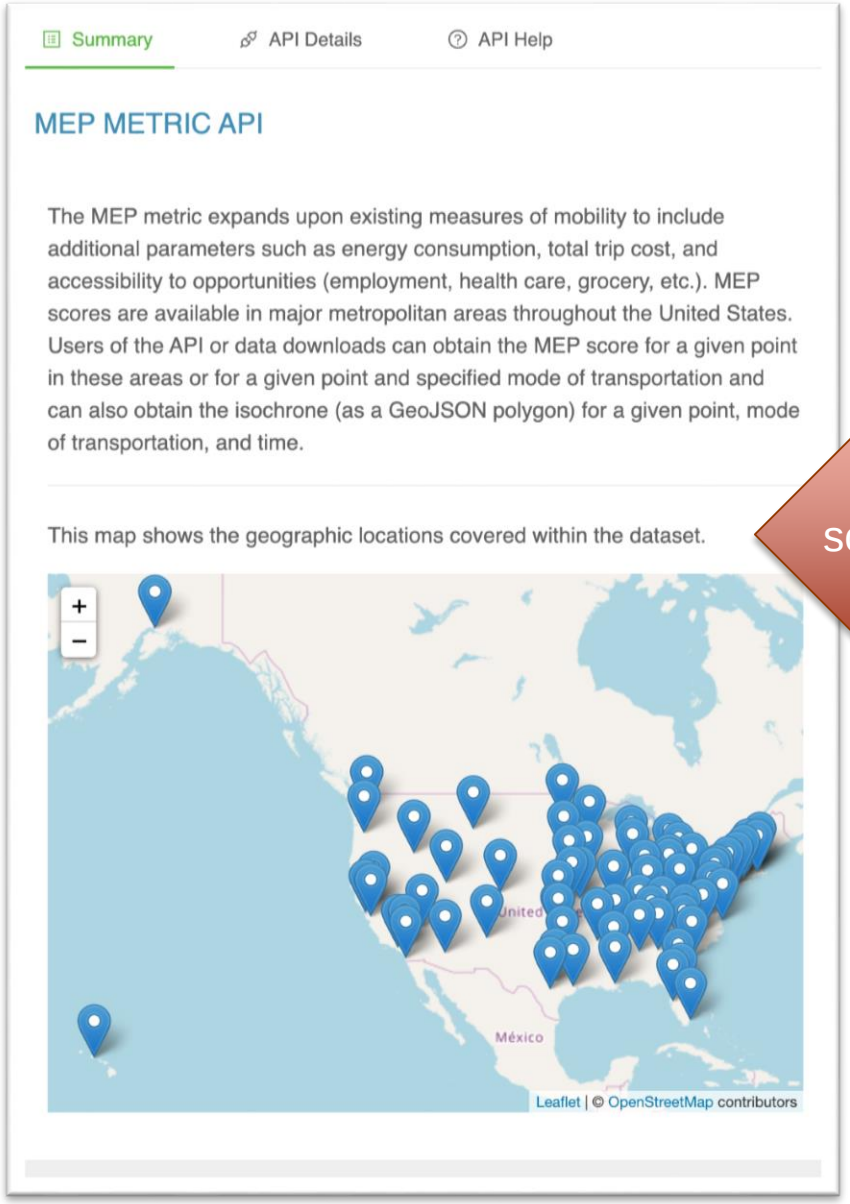


api.data.gov and API Umbrella



- Livewire can host (store) data to enable direct download
- Livewire can make data accessible via Application Programming Interface (API)
- Livewire can link to data on other websites

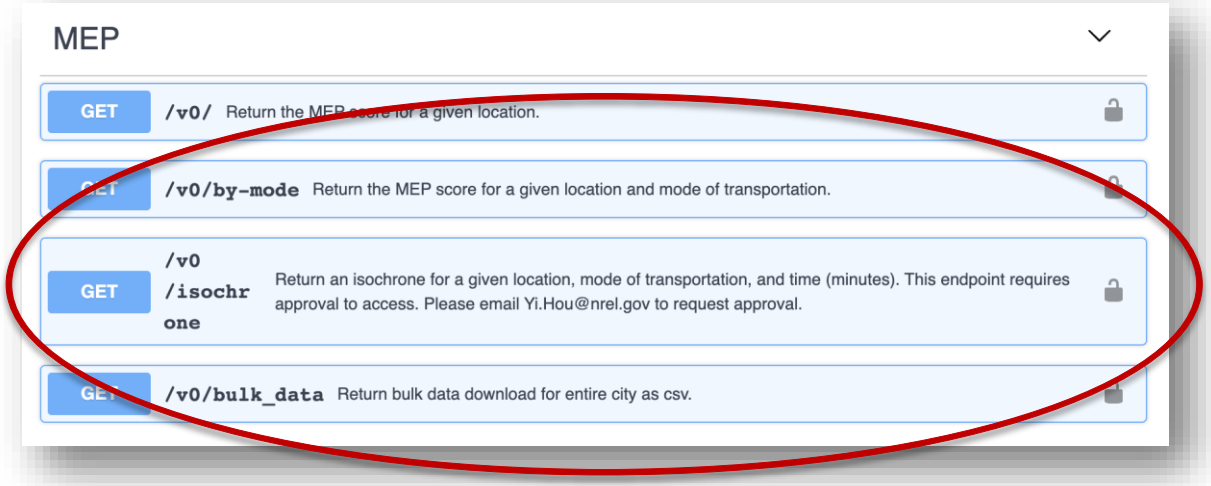
Accomplishments – Improved Usability of Existing Data Set



Added MEP scores for over 60 cities

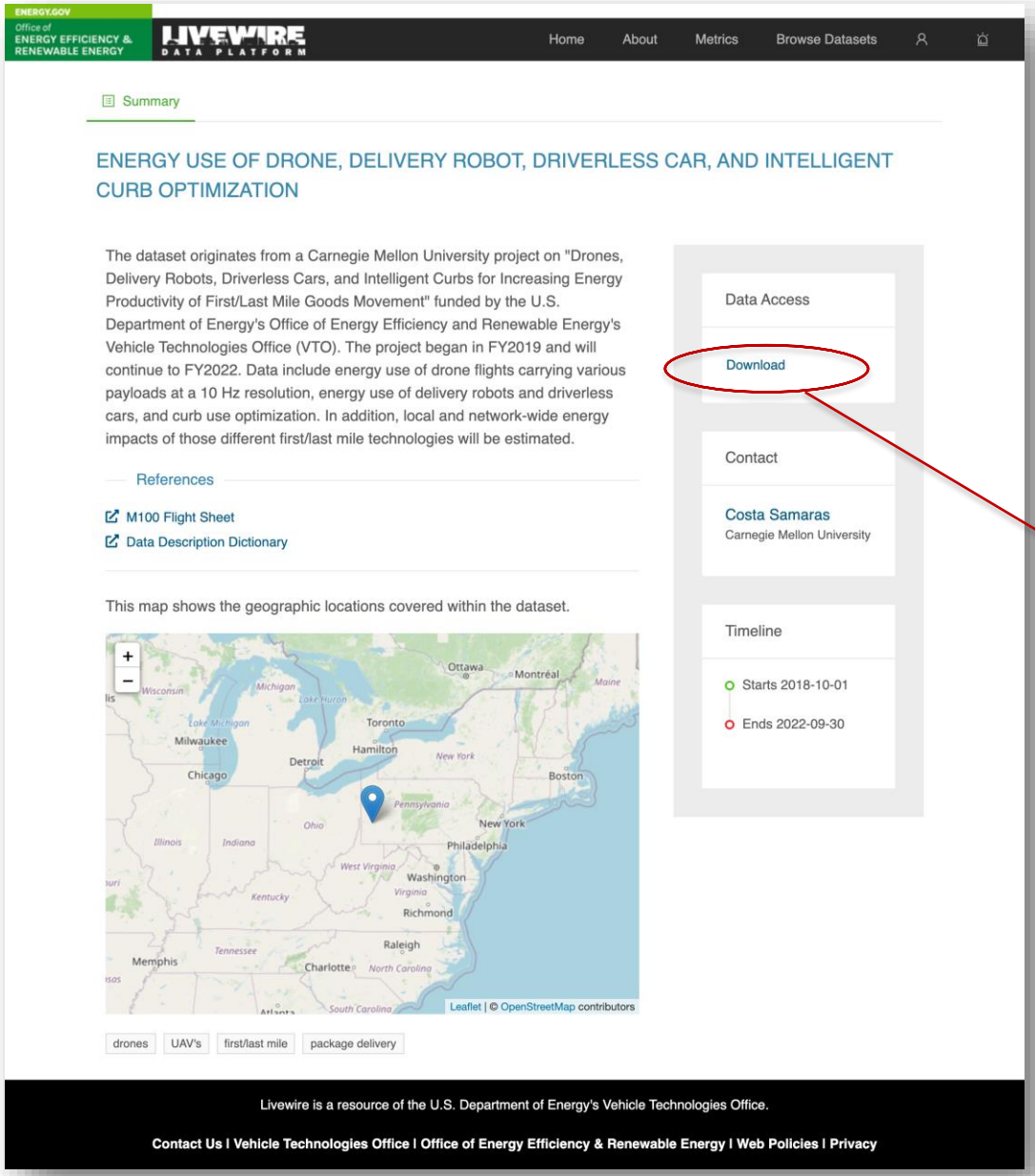
An API, or application programming interface, is a way for one application to provide data from its database to another in a standardized, machine-readable format.

- Accessing data via API automatically updates a user's content as database content changes

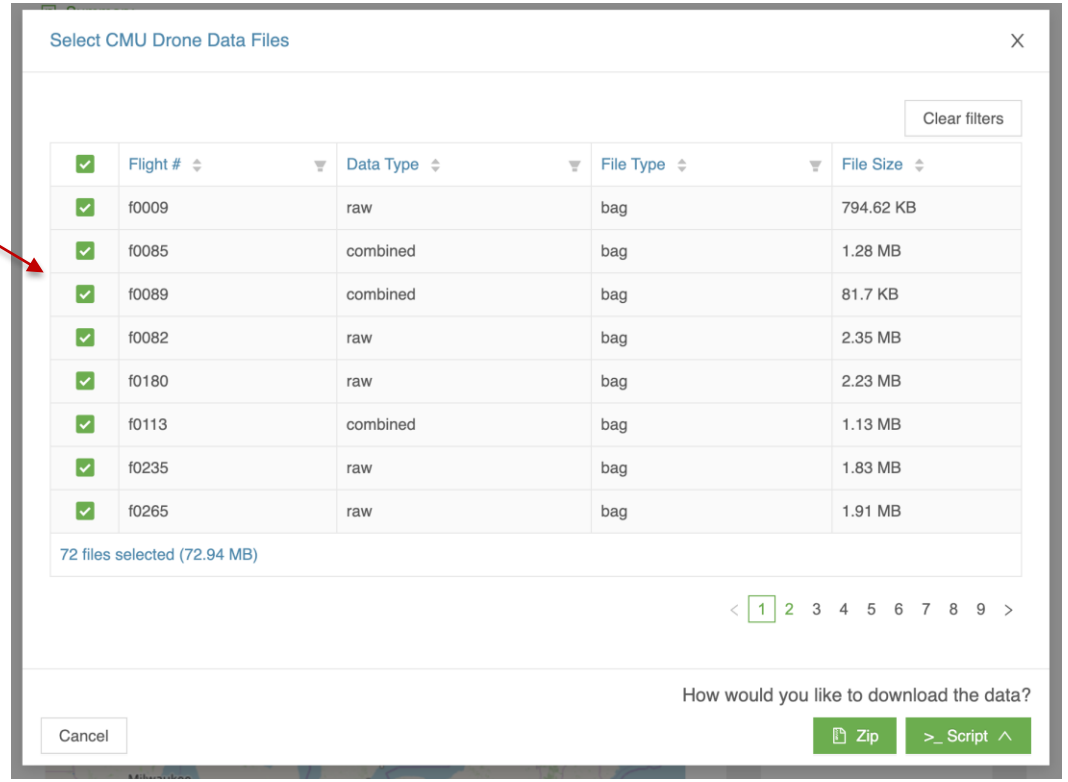


Added 3 endpoints to the Mobility Energy Productivity (MEP) API. Each endpoint represents a specific collection of data that a user can access.

Accomplishments – Data Download Added



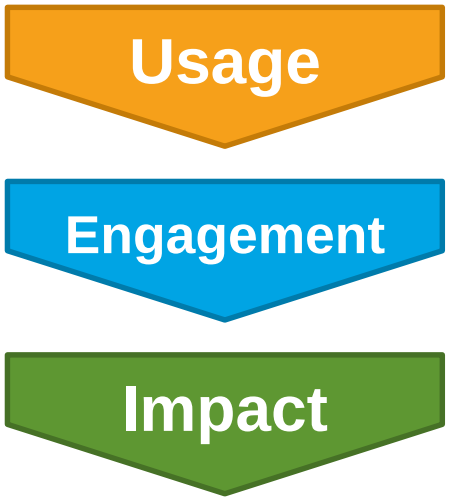
Data hosted on the Livewire Data Platform can be sorted, filtered, and downloaded directly by logged-in users.



Approach – Measure and Report Platform Success

Livewire Data Platform success is measured through growth of:

- **Usage** metrics – tell how many people use the platform
- **Engagement** metrics – show how people interact with the platform
- **Impact** metrics – demonstrate the impact Livewire has on a user's work



June 30, 2020 Go/No-Go decision is based on growth since 12/31/19

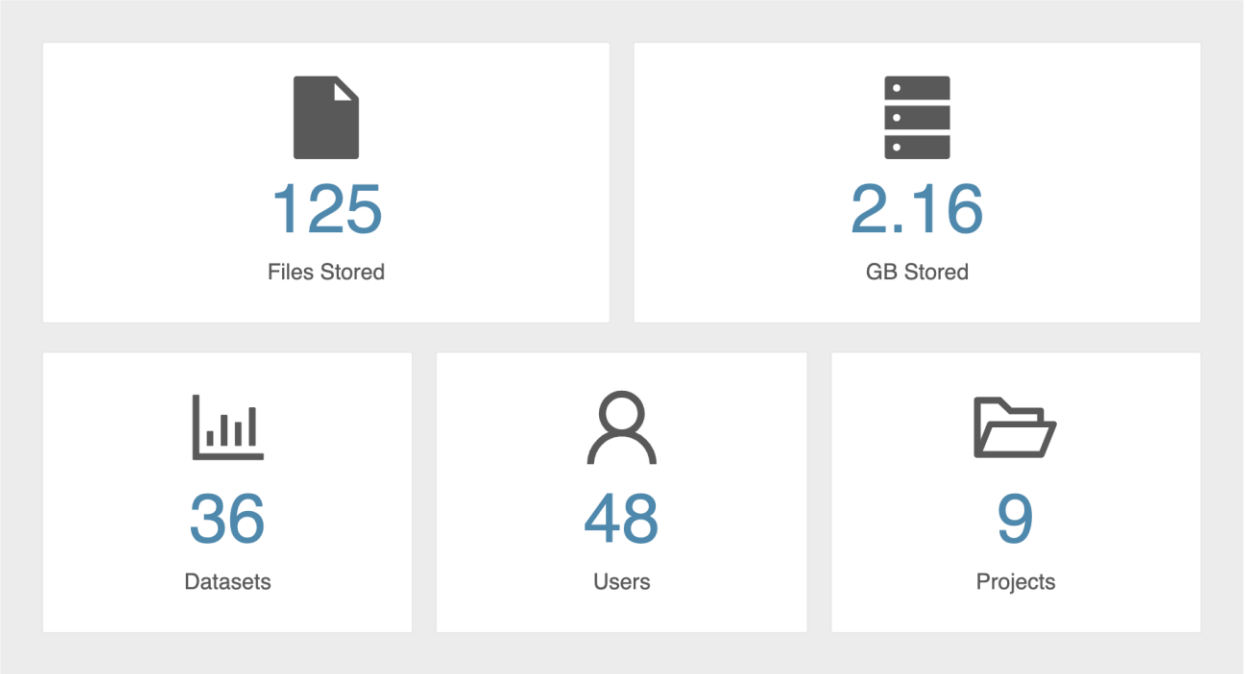
- 50% increase in projects contributing data = 12 projects
- 100% increase in datasets = 62 data sets
- 50% increase in users = 48 users

Usage, Engagement, and Impact metrics are reported quarterly

Engagement Metric	12/31/19 Baseline	FY 2020 Q2
Data sets downloaded	1	4
Size of files downloaded	650 MB	3,797 MB
Number of users downloading data	1	3
Number of files downloaded	23	108

Accomplishments – Growth in Users

- Added 35 data sets from 9 projects since platform launch
- Grew registered users by 50%
- Conducted over 15 interviews with future data contributors from state, SMART Mobility, and EEMS partners
- Grew catalog of data sets for inclusion in Livewire Data Platform based on researcher interviews



Current metrics from livewire.energy.gov/metrics

LIVEWIRE DATA PLATFORM

livewire.energy.gov

Registered users come from 18 organizations

Accomplishments – Growth in Data Available

Expanded data available from existing projects:

- Mobility Energy Productivity (MEP)
  - 4 API endpoints
  - Scores for over 60 cities
- Transportation Secure Data Center (TSDC)
  - 24 data sets; 50+ distributions
- Fleet DNA
  - 2 API endpoints
  - Downloads for 3 commercial fleets

New projects and data sets:

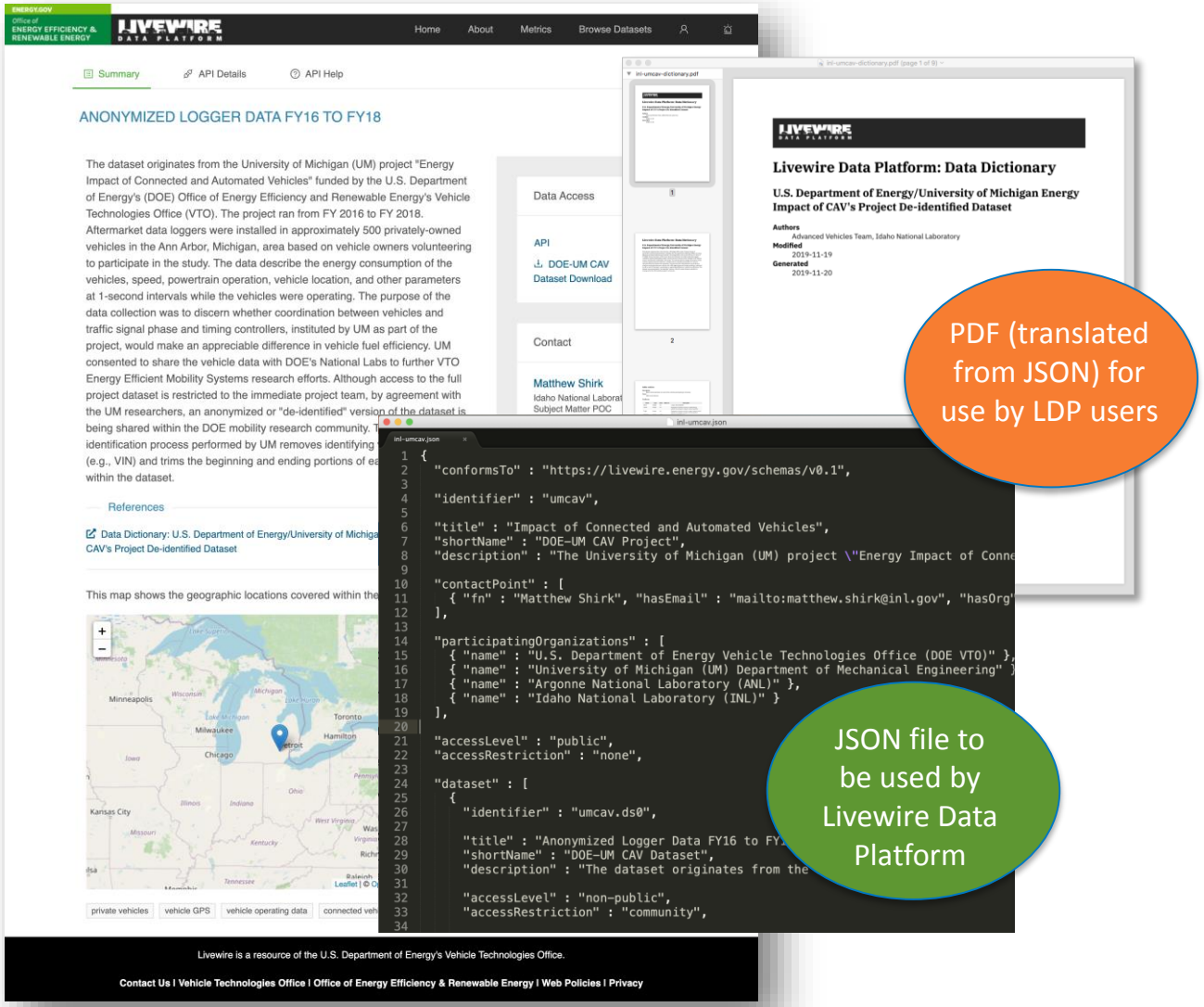
- U.S. Department of Transportation Freight Analysis Framework (FAF4)
- Ride Austin Transportation Network Company
- Carnegie Mellon University Drone Data

Over 12 active data pursuits in pipeline, including:

- Athena
- Additional Fleet DNA data
- American Center for Mobility (ACM) Connected and Automated Vehicle (CAV) Model and Simulation Validation
- Argonne National Laboratory (ANL) Downloadable Dynamometer Database (D^3)
- INL Advanced Vehicle Testing Activity

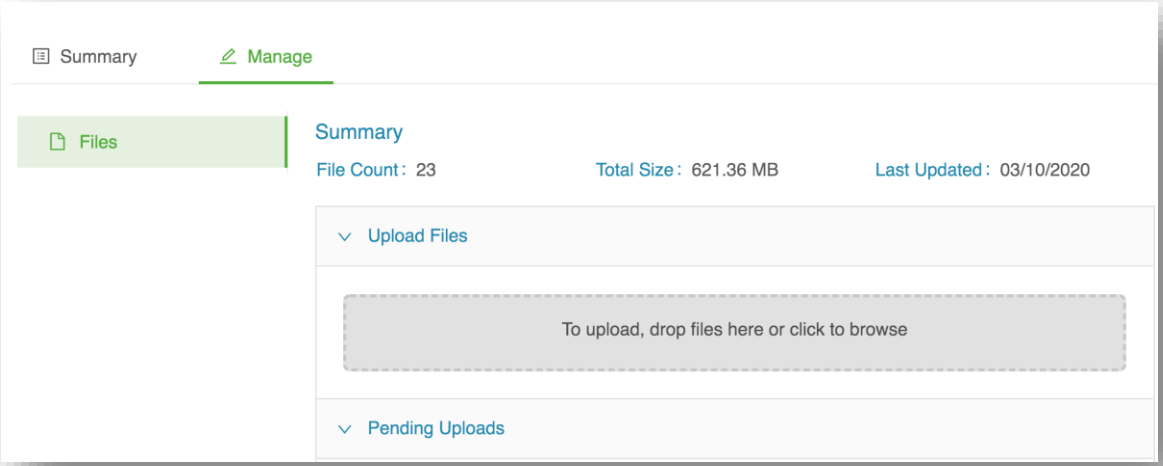
Accomplishments – Metadata Creation Tools Developed

Metadata provides information about characteristics of data sets

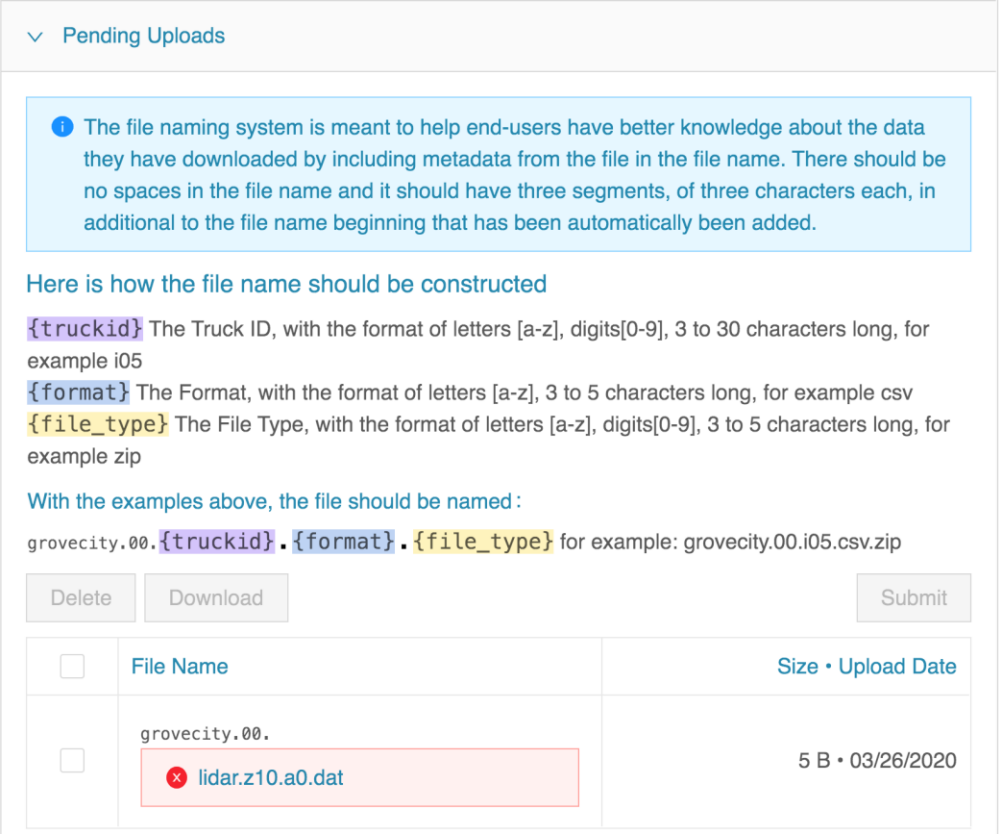


- Created and published low-level metadata (JavaScript Object Notation and Portable Document Format) for all data
- Developed prototype automation tools used to create and share detailed metadata for 4 data sets
  - Automation tools will be used to create detailed metadata for all LDP data
  - Use of automation tools will standardize metadata creation and capture aspects of data quality

Accomplishments – File Upload Feature Launched



User interface allows data owners to upload their own files and validate file names



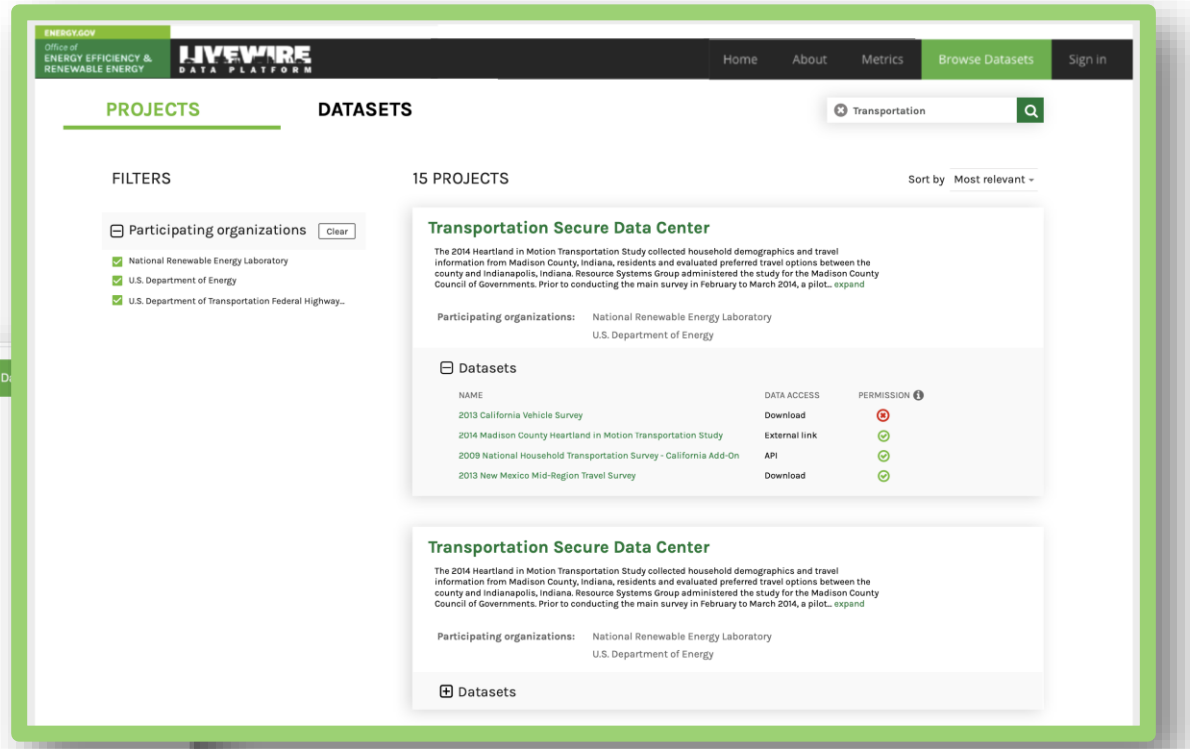
File name validation defined by metadata

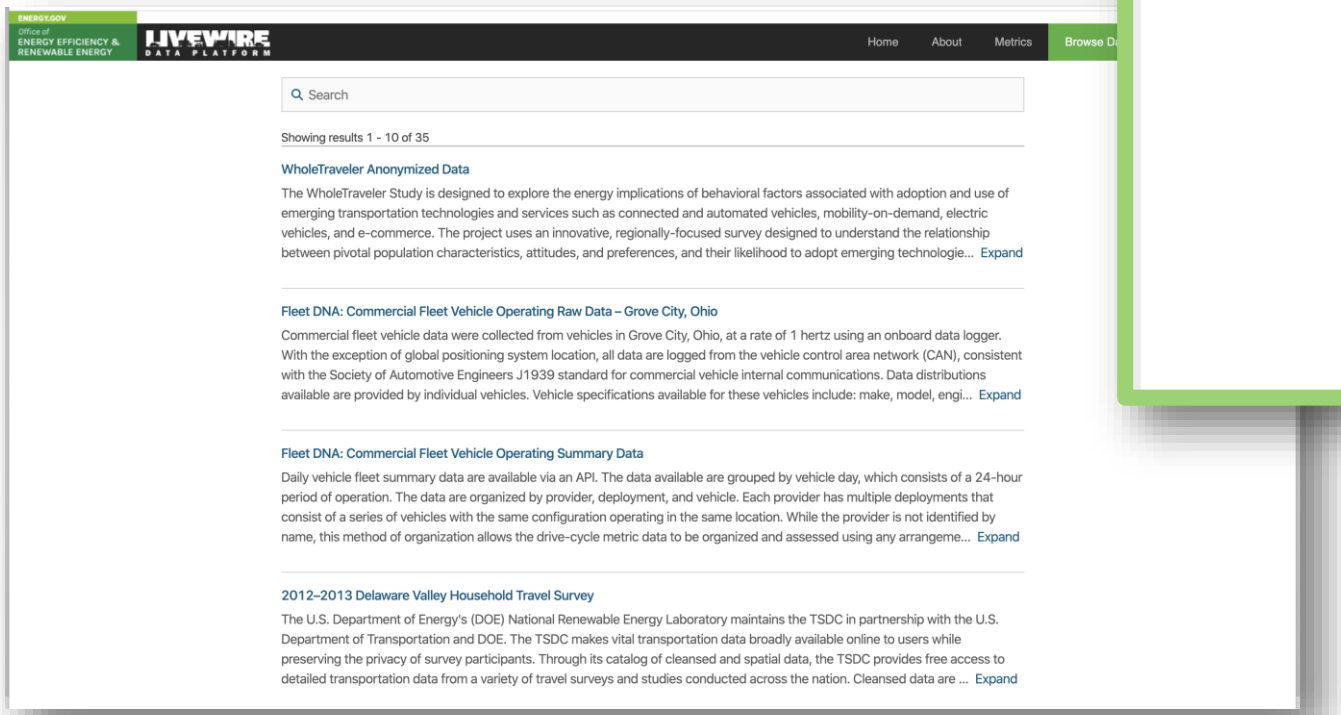
- Ensures standardization
- Allows users to sort and filter data

Accomplishments – Site Redesigned for Better User Experience

The Livewire Data Platform is constantly improving to meet the needs of our users.

- Updated browse page (right) will group data sets by project for an easier search experience
- Updated look and feel coming Q3





livewire.energy.gov/search

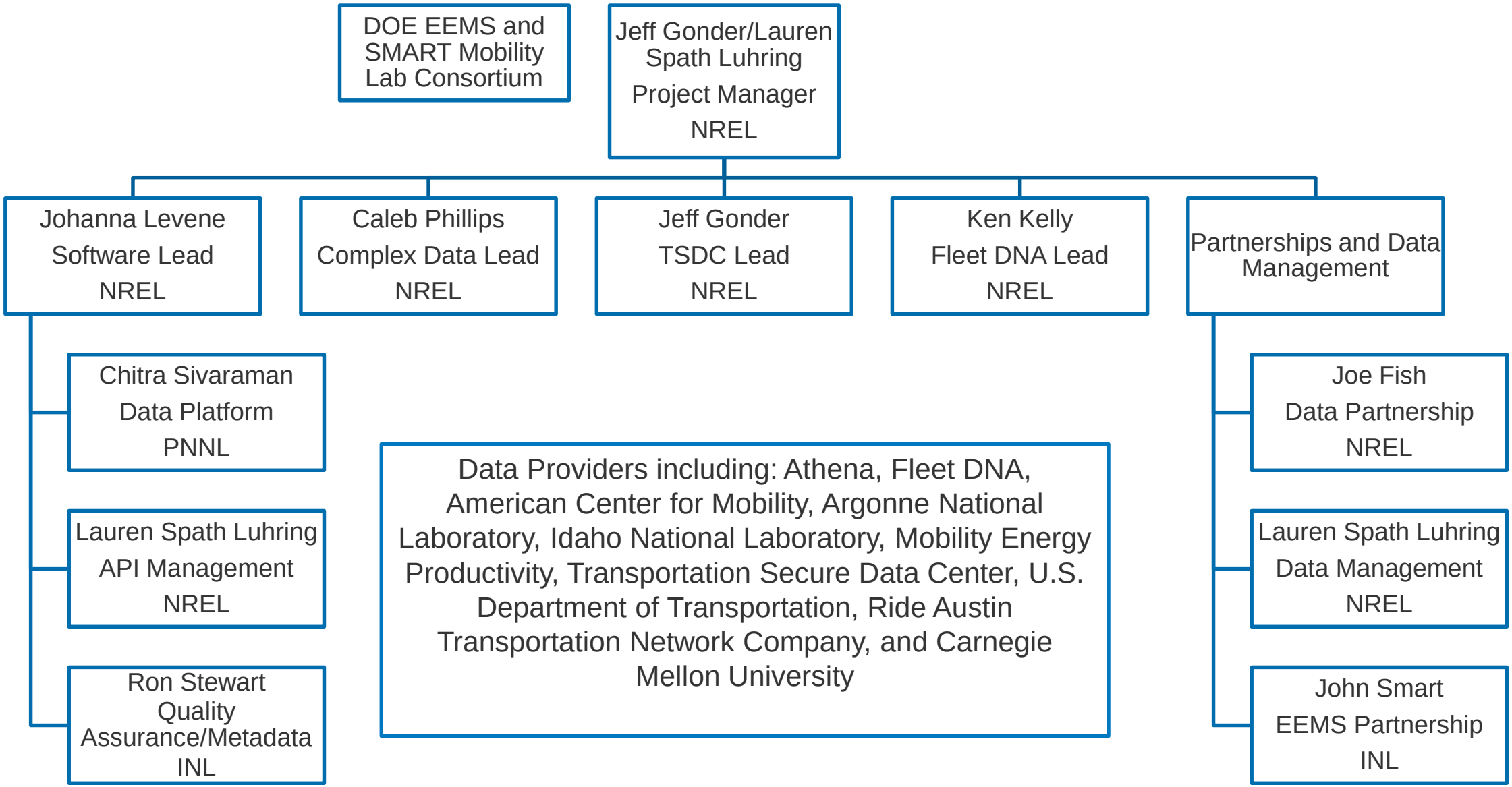
Responses to Previous Year Reviewers' Comments

Comment	Response
"…it would help to provide reassurance if firm commitments from numerous researchers upfront were obtained to contribute to and utilize the Livewire Data Platform prior to development."	The Livewire team worked with NREL, PNNL, and INL Laboratory Program Managers to encourage language about sharing data via Livewire in EEMS FOA and Lab Call concept papers. FY 2020 VTO FOA requirements stated that data generated by awarded FOA projects must be shared with National Labs via LDP.

Responses to Previous Year Reviewers' Comments

Comment	Response
"The reviewer noted that this is a social-technical problem. Success will depend on both building a user-friendly technical system and transforming institutional cultures to make use of the tool. Cultural transformation is hard to measure."	The Livewire Data Platform is constantly evaluating and evolving to meet user and data needs. Feature design and implementation are executed with user experience in mind. Through reporting on Usage, Engagement, and Impact metrics we aim to capture this "cultural transformation." In parallel, VTO is driving cultural change by requiring data sharing via LDP.

Collaboration and Coordination



Remaining Challenges and Barriers

- Legal challenges around non-disclosure agreements and licenses make sharing data difficult
- Manual processes impede sharing by rapidly evolving, complex modeling projects and data sets
- Human factors remain a challenge; social acceptance lags technical implementation

Proposed Future Research

- Implement capabilities for PIs to manage (add, edit, or delete) groups and users*
- Address human factors impeding adoption of and sharing data with Livewire Data Platform
- Continue evolution of metadata structure and content in order to meet platform needs and capture aspects of data quality*
- Develop data access controls that enable protection of controlled data

*indicates progress toward milestone

Any proposed future work is subject to change based on funding levels.

PI = principal investigator

Summary

- The Livewire Data Platform (**livewire.energy.gov**) focuses on both the platform and people to build a usable system for sharing EEMS data
- LDP leverages two existing successful platforms (a2e.energy.gov and api.data.gov) to provide data to users via API, download, and outside links
  - The project also supports the legacy FleetDNA and TSDC efforts + LDP integration
- LDP handles data storage and management, allowing PIs to focus on research
- LDP engages data owners and prioritizes features through interviews
- Over 35 data sets from 9 projects are available on LDP
- The LDP catalog of projects and data continues to grow: Chicago and San Francisco SMART Mobility MEP Scenarios, ACM CAV Model and Simulation Validation, ANL Downloadable Dynamometer Database (D^3), ATHENA, and INL Advanced Vehicle Testing Activity data will be available this fiscal year.
- Contact Livewirecontact@lyris.pnnl.gov if you have data to share or questions.

Thank You

www.nrel.gov

NREL/PR-5400-76694

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the DOE Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office (VTO). The authors would particularly like to thank David Anderson and Erin Boyd from the VTO Energy Efficient Mobility Systems Program for their on-going guidance and support. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Technical Back-Up Slides

Accomplishments – Metadata Creation Tools Developed

- Metadata automation tools capture basic aspects of data quality
  - Structure, content, and relationships
  - Size and completeness
  - Statistical measures of distribution and skew
  - Some aspects of reasonability
- Created detailed matrix for assigning project- and data set-level access restrictions

Metadata Access Level and Restriction				Metadata Visibility Within the Liveware Data Platform					
Project		Dataset		Unauthenticated LDP User			Authenticated & Approved LDP User		
accessLevel	accessRestriction	accessLevel	accessRestriction	Project Metadata	Dataset Metadata	Distribution Metadata	Project Metadata	Dataset Metadata	Distribution Metadata
public	none	public	none	Visible	Visible?	Visible?	Visible	Visible?	Visible?
		restricted public		Not visible: not supported in LDP					
		non-public	none	Visible	Not visible: not supported in LDP		Visible	Not visible: not supported in LDP	
			community	Visible	Not Visible	Not Visible	Visible	Visible	Visible
			project	Visible	Not Visible	Not Visible	Visible	Visible	P-Group
			project-mfa	Visible	Not Visible	Not Visible	Visible	Visible	P-Group + MFA
			dataset	Visible	Not Visible	Not Visible	Visible	Visible	D-Group
			dataset-mfa	Visible	Not Visible	Not Visible	Visible	Visible	D-Group + MFA
restricted public				Not visible: not supported in LDP			Not visible: not supported in LDP		
non-public	none			Not visible: not supported in LDP			Not visible: not supported in LDP		
	community	public	none	Not visible: not supported in LDP			Not visible: not supported in LDP		
		restricted public		Not visible: not supported in LDP			Not visible: not supported in LDP		
		non-public	none	Not visible: not supported in LDP			Not visible: not supported in LDP		
			community	Not Visible	Not Visible	Not Visible	Visible	Visible	Visible
			project	Not Visible	Not Visible	Not Visible	Visible	Visible	P-Group
			project-mfa	Not Visible	Not Visible	Not Visible	Visible	Visible	P-Group + MFA
			dataset	Not Visible	Not Visible	Not Visible	Visible	Visible	D-Group
			dataset-mfa	Not Visible	Not Visible	Not Visible	Visible	Visible	D-Group + MFA
	project	public	none	Not visible: not supported in LDP			Not visible: not supported in LDP		
		restricted public		Not visible: not supported in LDP			Not visible: not supported in LDP		
		non-public	none	Not visible: not supported in LDP			Not visible: not supported in LDP		
			community	Not visible: not supported in LDP			Not visible: not supported in LDP		
			project	Not Visible	Not Visible	Not Visible	Visible	P-Group	P-Group
			project-mfa	Not Visible	Not Visible	Not Visible	Visible	P-Group	P-Group + MFA
			dataset	Not Visible	Not Visible	Not Visible	Visible	P-Group	D-Group
			dataset-mfa	Not Visible	Not Visible	Not Visible	Visible	P-Group	D-Group + MFA
	project-mfa	public	none	Not visible: not supported in LDP			Not visible: not supported in LDP		
		restricted public		Not visible: not supported in LDP			Not visible: not supported in LDP		
		non-public	none	Not visible: not supported in LDP			Not visible: not supported in LDP		
			community	Not visible: not supported in LDP			Not visible: not supported in LDP		
			project	Not visible: not supported in LDP			Not visible: not supported in LDP		
			project-mfa	Not Visible	Not Visible	Not Visible	Visible	P-Group + MFA	P-Group + MFA
			dataset	Not visible: not supported in LDP			Not visible: not supported in LDP		
			dataset-mfa	Not Visible	Not Visible	Not Visible	Visible	P-Group + MFA	D-Group + MFA

TSDC Background

High-resolution travel data (e.g., GPS trajectories, geo-coded trip ends)

- Very valuable for research
- Misuse could violate individual privacy

Secure data center makes data available for legitimate research while preserving privacy of participants

- Maximizes value from limited public funds
- Benefits data providers and users
  - Shifts responsibility of archiving data and responding to data requests
  - Data accessible from a central location

As recommended by this National Research Council report: http://books.nap.edu/openbook.php?record_id=11865

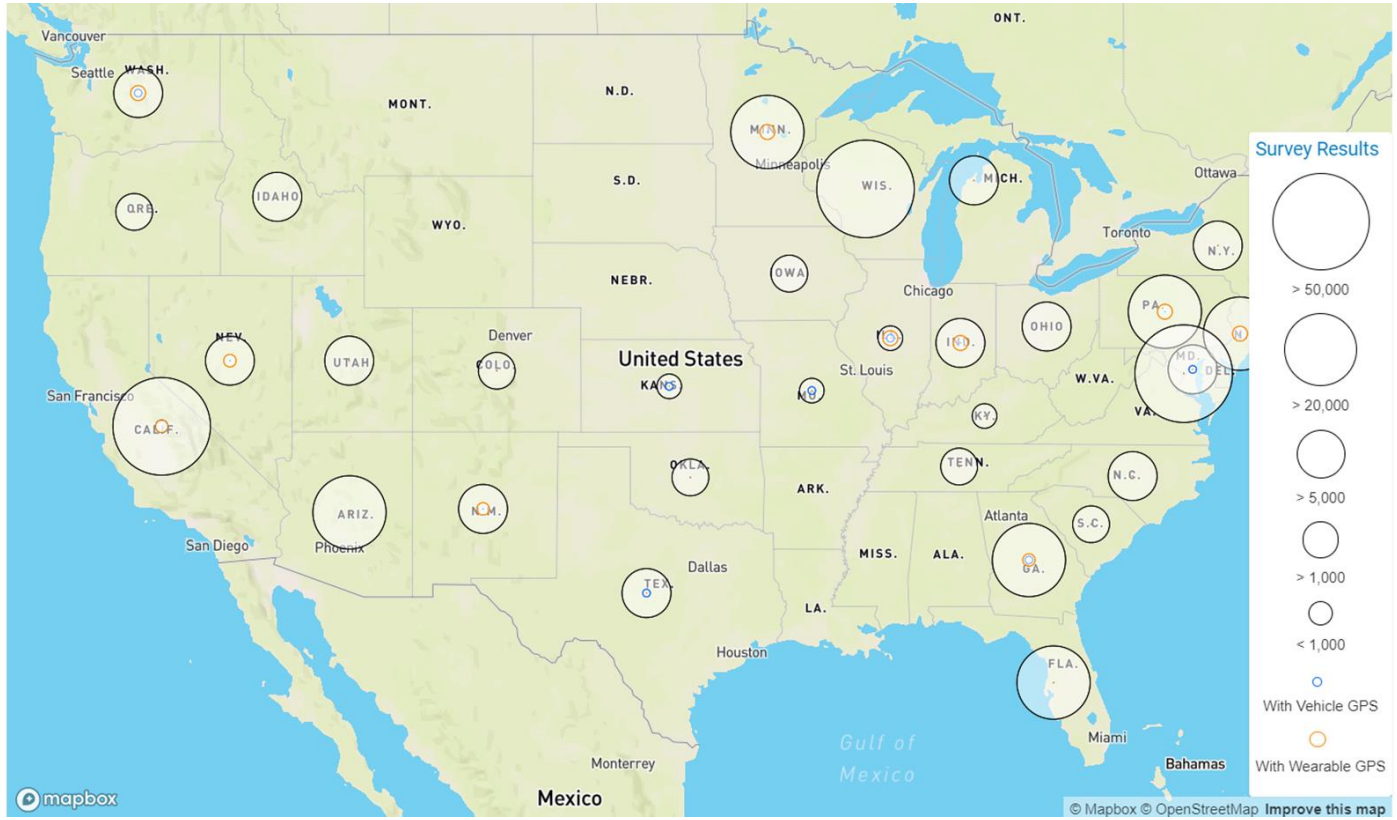
The TSDC has been supported by NREL, U.S. DOT, and U.S. DOE since 2009

- Department of Transportation, Federal Highway Administration
- Department of Energy, Vehicle Technologies Office

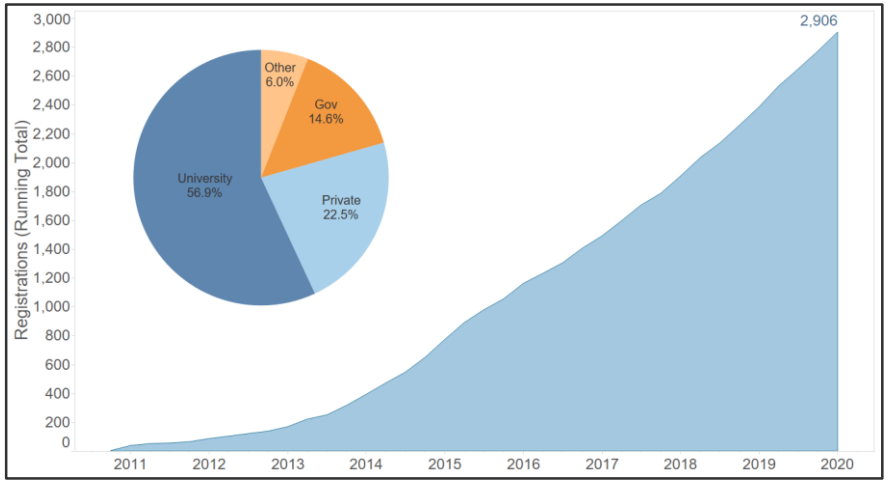
Scope & Growth of the TSDC

TSDC

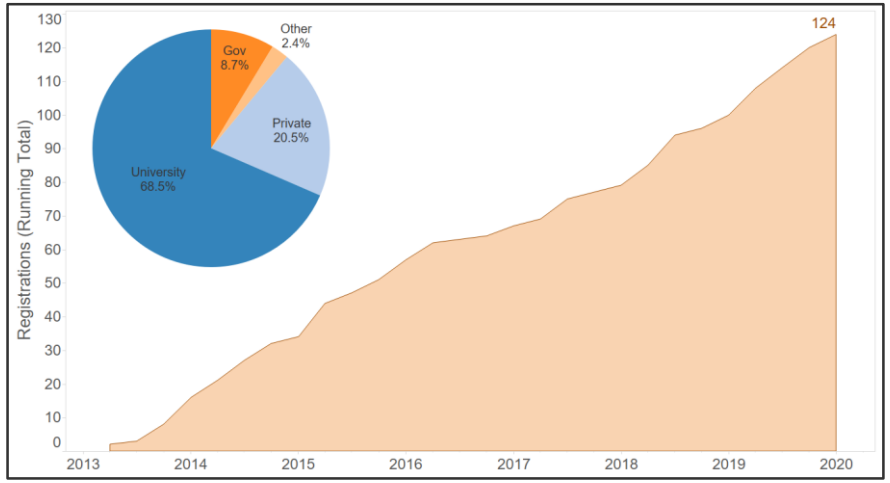
Number of Surveys by State



Web Users



Secure Portal Users



Other TSDC Updates

TSDC

- Completed and in-progress data additions include
  - Taking over hosting of data previously archived in the University of Minnesota Metropolitan Travel Survey Archive (which is no longer being maintained)
  - Addition of National Household Travel Survey add-on data from many participating states
  - Engaged with many other regional and state agencies
- TSDC website updates in progress to include more sophisticated filtering and new data additions
- Improving data processing to facilitate cross-study analyses and reduce staff time via automation
- Ongoing discussions with the Federal Transit Administration about hosting transit passenger surveys and with the Federal Highway Administration about a pooled fund study

Fleet DNA Background

Fleet DNA clearinghouse of commercial fleet vehicle operating data helps vehicle manufacturers and developers optimize vehicle designs and fleet managers choose advanced technologies for their fleets. https://www.nrel.gov/transportation/fleettest-fleet-dna.html

Fleet DNA is one of the foundational building blocks for the Livewire platform

In FY19, data from approximately 200 vehicles were added to the Fleet DNA database including:

- Shuttle buses from Dallas Fort Worth and Los Angeles International airports;
- Yard tractors at the Port of Long Beach and the Port Authority of New York and New Jersey;
- National Park Service shuttle buses at Zion, Bryce Canyon, and Yosemite National Parks;
- UPS line-haul truck supporting the PACCAR Super Truck II project;
- Regional-haul trucks from North American Council for Freight Efficiency's Run on Less Regional program.

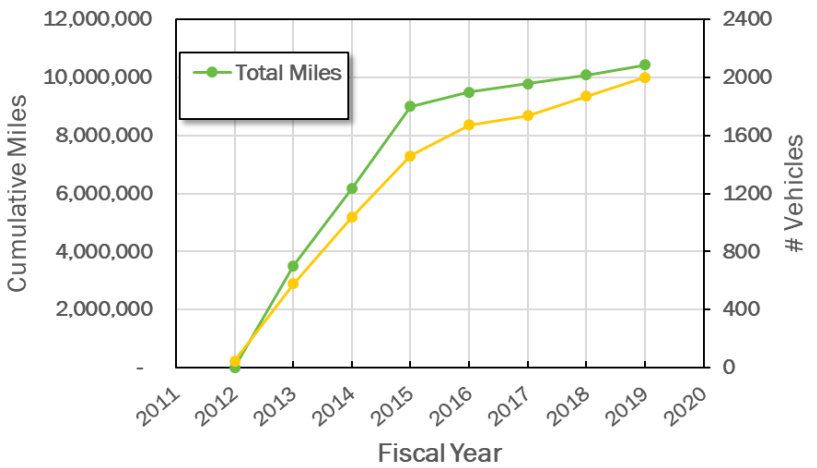
NREL researchers have developed scalable high-performance-computing architectures and employed data fusion, multivariate analysis, and innovative visualization techniques using Fleet DNA to help multiple government and industry partners understand the operational characteristics of commercial vehicles and to develop and optimize efficient vehicle systems and components.

In FY19, Fleet DNA data and NREL analytics made significant contributions to DOE-funded activities including: 21st Century Truck Partnership, EEMS Livewire Data Platform, Athena, Super Truck II, DOE-funded industry truck development projects, and numerous NREL Strategic Partner projects

Fleet DNA

	Miles	Vehicles
Local Delivery	2,752,542	352
Class 8 - Long Haul	2,395,228	149
Food Delivery	1,688,650	158
Parcel Delivery	899,636	212
Utility Truck	438,113	249
Transit Bus	428,911	140
Class 8 - Day Cab Tractor	402,460	27
Tanker	377,703	26
Beverage Delivery	216,164	89
Airport Bus	183,797	39
School Bus	125,334	276
Other Vocational	116,901	56
Drayage	79,690	31
Transfer Truck	70,390	29
Refuse Hauler	70,239	86
Port Yard Tractor	65,384	52
Towing	51,030	6
Dump Truck	34,981	7
National Park Shuttle	13,015	6
Snow Plow	12,887	11
Concrete Mixer	11,670	3
TOTAL	10,434,729	2004

Fleet DNA Data Inventory & Timeline



Fleet DNA Progress

Fleet DNA

Adding Fleet DNA datasets to Livewire

- Regional-haul Grove City, OH
- Regional-haul, Sanger, TX
- Regional-haul, Loveland, CO
- Line-haul, cross-country
- Airport Shuttle buses, TX
- Port yard tractors, CA and NY
- Airport shuttle buses, TX
- CNG Transit buses, CA
- Package delivery, TX
- Food Delivery, WA
- Package Delivery, OH
- Refuse Trucks, FL

- Data cleansing of EV data sets
  - Navistar EVs - nationwide
  - Smith EVs – nationwide
- EV Data represents a <u>novel, advanced powertrain</u> addition to Livewire
- Developing Relational database modeling software (RDBMS) platform
- Revamped data cleansing to data sharing pipeline
  - Fleet DNA deployment descriptions summarize:
    - Deployment Location
    - Vehicle Type
    - NREL Goal/Context, etc.
  - Better practices provide data descriptions that are searchable, provide dates, specify vehicle vocation, city of collection, etc. - provided partner information is protected

CHARGING LOCATIONS

This map shows the charging locations—which range in population density and climate—for the vehicles under study. There are 35 charging locations for the Navistar vehicles and 81 for the Smith vehicles.

Navistar eStar	
Weight Class	Class 3
Gross Vehicle Weight Rating	5,488 kg
Payload	2,313 kg
Battery Capacity	80 kWh
Motor Power	70 kW
Top Speed	80.5 km/h
Advertised Range	Up to 160 km
Total Distance	558,987 km
Number of Vehicles	101

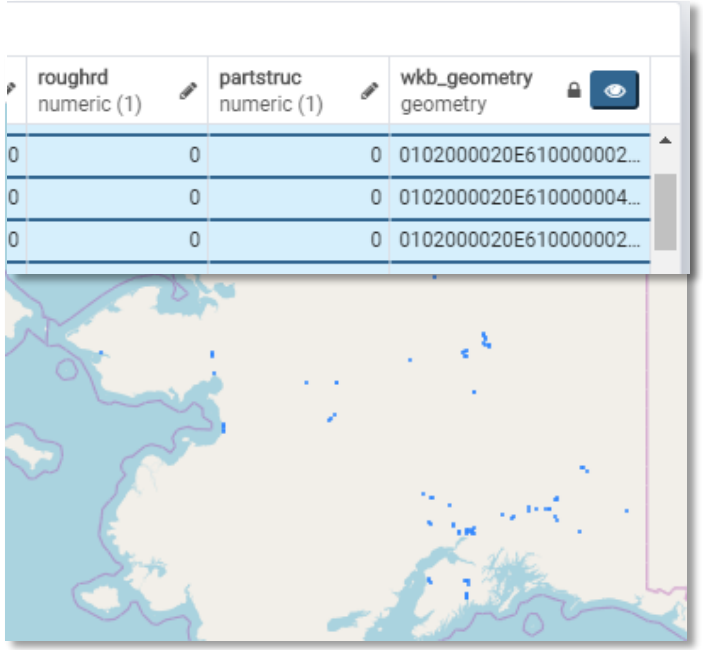
Photo from Navistar, NREL 18624

Fleet DNA TimescaleDB

Fleet DNA

TimescaleDB: PostgreSQL based
Specialty is time series data

Bonus: GIS extensions available



roughrd numeric (1)	partstruc numeric (1)	wkb_geometry geometry
0	0	0102000020E610000002...
0	0	0102000020E610000004...
0	0	0102000020E610000002...

Data access and computation:
Goes from hours to minutes

Data compression:
Data compresses by factor of ~10

Partition on Vehicle ID, Vocation, etc. to speed up queries

Schema so far: Simple table of “common” or “core” signals (DateTime, Speed, Fuel Rate, etc.)

Feeder deployments

- Dallas Ft. Worth Intl. Airport
- Foothill Transit CNG
- Frito Lay, Federal Way WA
- PACCAR UPS Super Truck 2
- Los Angeles Intl. Airport
- UPS Solazyme
- Port Authority NY NJ
- Port Operations Long Beach