Why we need data infrastructure and tooling to further access to mobility & transportation equity

The goal of the session is to foster a meaningful dialogue on open, multimodal, interoperable accessibility-first transportation data sharing platform, data tooling, and open API's.
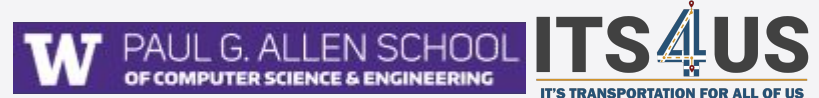Provide hands-on experience using our prototype of the proposed system.

Transportation Data Equity Data Workshop:

Part I– The Transportation Data Equity Platform. Focus on Data CONSUMPTION.

Part II – Hands on exercises (demo and upcoming Alpha)

https://tinyurl.com/TDAdvisory2

TCAT The Taskar Center for Accessible Technology

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

TDEI Introductions

Anat Caspi,
Transportation Data Equity Initiative Development Lead

TCAT The Taskar Center for
Accessible Technology

Anat Caspi, PhD
UW CSE, PI, DDL
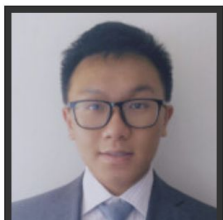
Bill Howe, PhD
iSchool UW, SEL

Alice Marecek
Cambridge Systematics

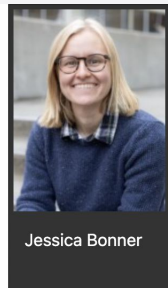Suresh Devalapalli,
Gaussian Systems

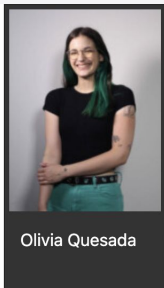Karen Braitmayer,
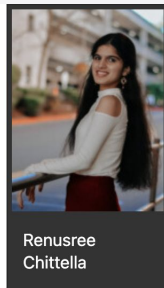Studio Pacifica, Access Lead

Yuxiang Zhang, PhD
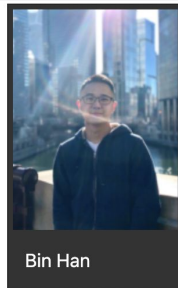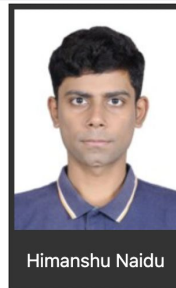UW CSE, TCAT

Kunal Mehta
UW HCDE, TCAT

Jessica Bonner

Olivia Quesada

Renusree Chittella

Bin Han

Himanshu Naidu
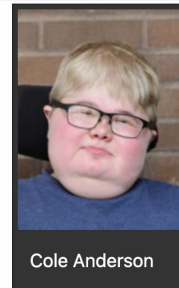
Cole Anderson

Not Pictured:

**Wisam Yasen
Cy Rossignol
Clifford Dalson III
Mario Edgardo Sanchez
Christine Mendoza
Madeline Gugger**

TCAT The Taskar Center for Accessible Technology

W PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

TDEI is an ITS4US Project

ITS4US is a multimodal effort led by the ITS JPO

Visit the ITS4US Deployment Program Website:

https://its.dot.gov/its4us/

https://tinyurl.com/TDAdvisory2

Brief Survey & Agenda

Anat Caspi,
Transportation Data Equity Initiative Development Lead

TCAT — The Taskar Center for Accessible Technology

Agenda

- 10:00 am- introduction and welcome
- 10:10 am- Agenda and Brief Survey
- 10:20 am- Quick review on the purpose and mission of the TDEI project
- 10:30 am- What is the TDEI platform, and how does it support data life cycle for transportation data producers & consumers?
- 10:45 am- Walkthrough & Exercises of issuing a TDEI transportation data consumption API call
- 11:15 am- Data Consumer discussion.

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Review:

Mission and Vision of TDEI

TCAT — The Taskar Center for Accessible Technology

Why is Data Infrastructure linked to Transportation Data Equity?

transitequity.cs.washington.edu

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Today's Session

- Data Equity Gaps and how they connect to open, shared data
  - What are some data producer pain points?
  - What is motivating our Tools for Intelligent Transportation work?
    - (lack of) Data Sharing and Distribution infrastructure Tools are Blockers to Transportation Data Equity
    - Active data use drives better & more data
    - If you're not measuring it, you're not managing it
- Data Equity Initiative– proposed system
 - As a data consumer, how can I apply these ideas?

TDEI Objectives

❖ Create, improve and extend data formats to describe sidewalks and footpaths in the built environment

❖ Create, improve and extend data formats to describe pathways through transportation stations and hubs

❖ Create, improve and extend data formats to describe on-demand travel services

❖ Create and deploy a federated data sharing, management and distribution platform with open API's for sidewalks, transit pathway and on-demand transit data

❖ Publish and maintain for 5 years data infrastructure for six U.S. counties

❖ Deploy mobile applications and experiences consuming standardized data that demonstrate the versatility and scalability of the data infrastructure

❖ Reach goal: Deploy planner dashboards consuming standardized data that demonstrate the versatility and scalability of the data infrastructure

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

The experience of federated data producers for low resource data is important

1. Legal and policy barriers
2. Cultural barriers
3. Institutional barriers
4. Resource constraints
5. Lack of incentives
6. Data Silos and legacy data
7. Void in Data Stewardship
8. Technical barriers

OpenStreetMap

ITS4US
IT'S TRANSPORTATION FOR ALL OF US
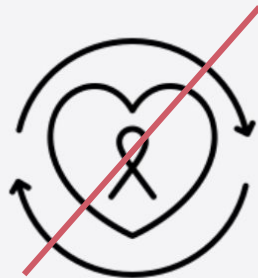
TDEI Principle 1

Behind every useful mobility app is a **complex data pipeline**. To provide reliable & intuitive travel discovery and directions, public agencies and private companies need **shared data and tools to represent all travel environments and services.**

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

TDEI Principle 2

Reliable, **objective, detailed consistent, standardized** shared data and tools can enable neutral data analysis, to better understand travel barriers, and to improve data-driven resource allocation.

•

ITS4US
IT'S TRANSPORTATION FOR ALL OF US
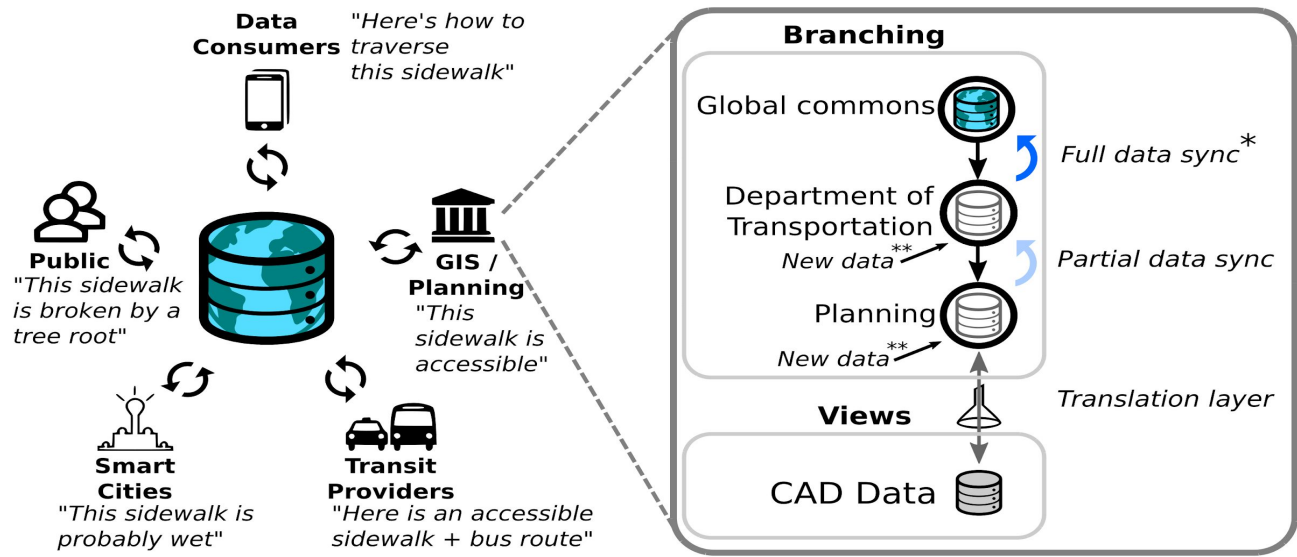
TDEI Principle 3

5. All individuals and the public should be empowered through high-quality, well-distributed mobility data to find, access, and utilize high-quality mobility options that meet their needs as they see fit, while maintaining their privacy.

- Mobility data is critical to helping the public make mobility decisions

- Interoperability helps mobility data get where it is most needed

- **INTEROPERABILITY =** the ability for software or hardware to communicate with other software or hardware using open standards

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Why TDEI?

How do I know the TDEI will help me publish & maintain data within my institutional constraints?

Bolten, Nicholas, et al. "A pedestrian-centered data approach for equitable access to urban infrastructure environments." *IBM Journal of Research and Development* 61.6 (2017): 10-1.

What is the TDEI platform, and how does it support data life cycle for transportation data producers & consumers?

Bill Howe,
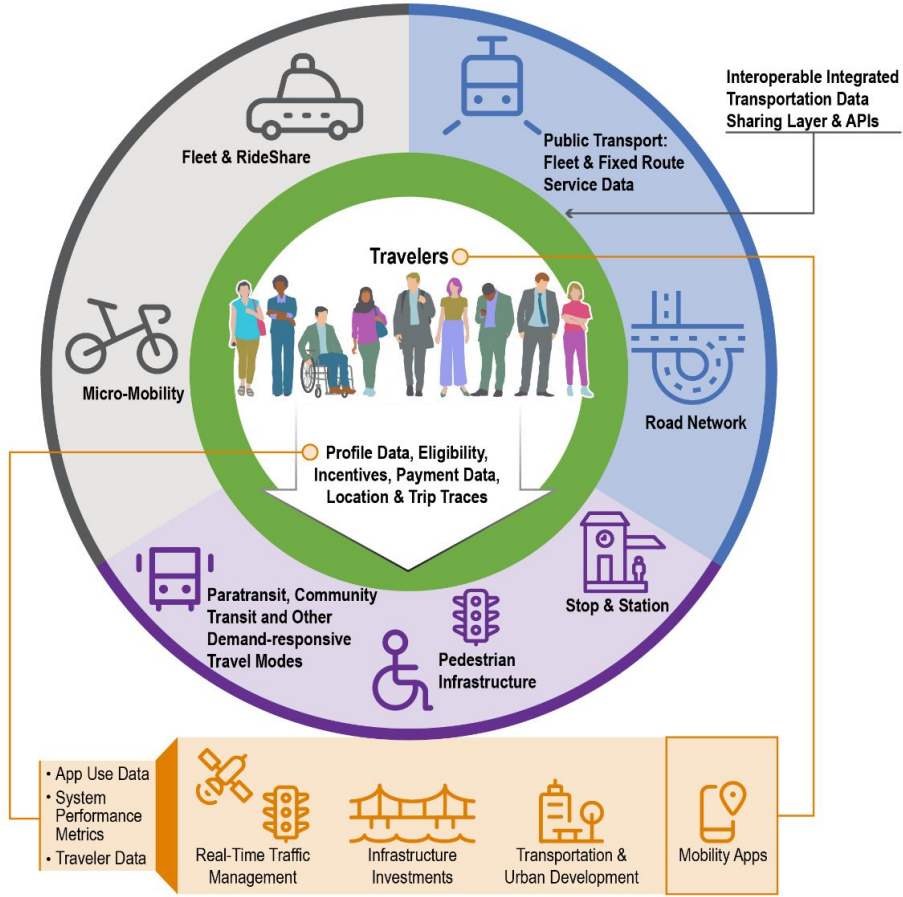Transportation Data Equity Initiative System Development Lead

TCAT The Taskar Center for Accessible Technology

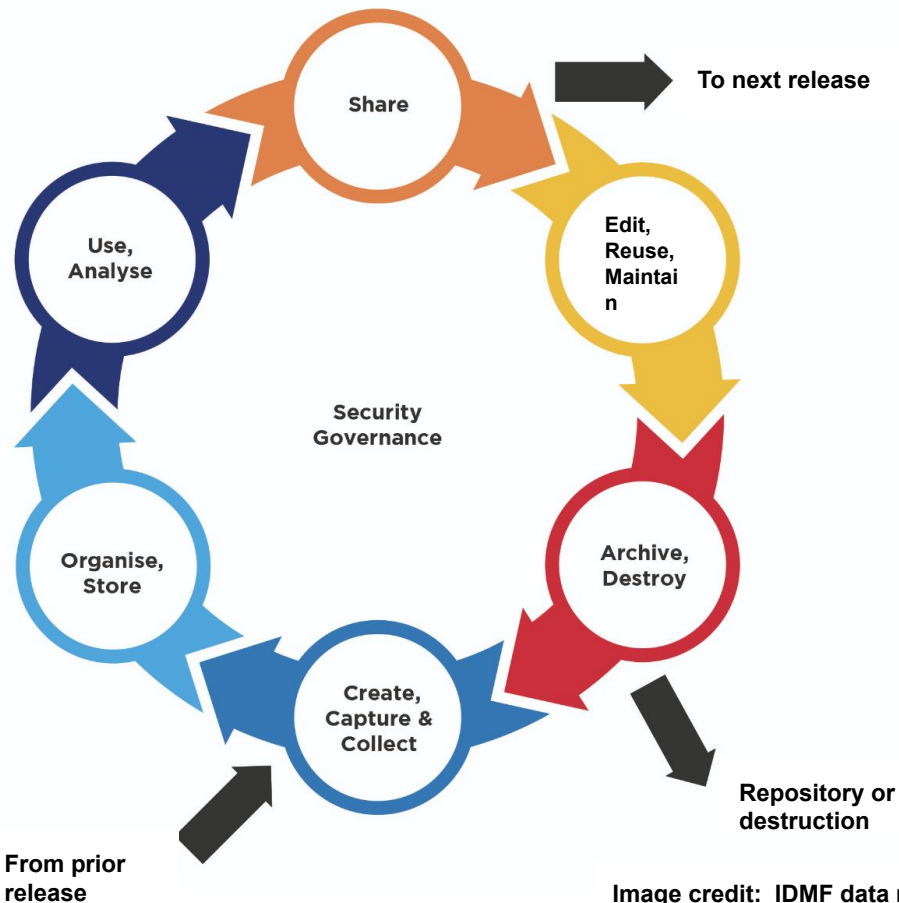# Lack of Transportation Data Infrastructure to Support Standards

All Data Stakeholders need consistent, high quality data they can trust

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Lack of Transportation Data LifeCycle Support

Some Data Stakeholders need reliable & consistent data release management, security, licensing & governance of data they can rely on

To next release

Share

Edit, Reuse, Maintain

Use, Analyse

Security Governance

Archive, Destroy

Organise, Store

Create, Capture & Collect

Repository or destruction

From prior release

Image credit: IDMF data management lifecycle

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Transportation Data Equity Initiative (TDEI) platform

Bill Howe,
Transportation Data Equity Initiative System Development Lead

Extensive and High Quality Data is the key

The success of downstream applications like AccessMap, Find-A-Ride, etc rely on the ability to **produce, access and consume** extensive and **high quality data** about the built environments in a **standardized format**.
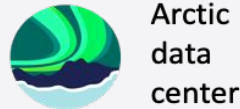
ITS4US
IT'S TRANSPORTATION FOR ALL OF US

In every field of science, with varying degrees of specificity...

"upload any file format"
(2M articles, 800k uploads)

strict metadata standards;
$350M infrastructure project;
~0 datasets ever uploaded

Pangea

NCBI

figshare

SEEK for Science

Chorus

DataONE
Data Observation Network for Earth

Arctic
data
center

USGS
science for a changing world

DRYAD

Galaxy
COMMUNITY HUB

GEO
Gene Expression Omnibus

caBIG™
cancer Biomedical
Informatics Grid™

constraints in format, metadata, quality

(promised) capabilities, automation

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Challenges in obtaining High Quality Data

- Lack of data standards for representation of accessibility related data elements
- Lack of collaboration tools to quantify, validate and enhance the quality of data
- No single place to get aggregated data. Each agency has their own data representation and end point
- Lack of confidence in the available open data sources

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

TDEI: Collaboration Platform for producing and consuming information for complete trips

- Enhance new and existing transportation data standards through a multimodal, accessibility-first approach
- Eliminate barriers to free/secure flow of accessibility-first mobility data
- Empower mobility data producers via tools and **easy to use** data infrastructure to support the **full data lifecycle**
- One stop data repository to seamlessly consume data with no regional bounds

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Functional Depiction of TDEI

Tools to convert proprietary data into standard format

Tools to validate the data

Tools to visualize and edit the data

Data in standard format

Producer 3

Producer 1

Connectors

Producer 2

Data Visualization Tools

Data & Schema Validators

Confidence Score Calculator

Central repository

Reformatters

Consumers
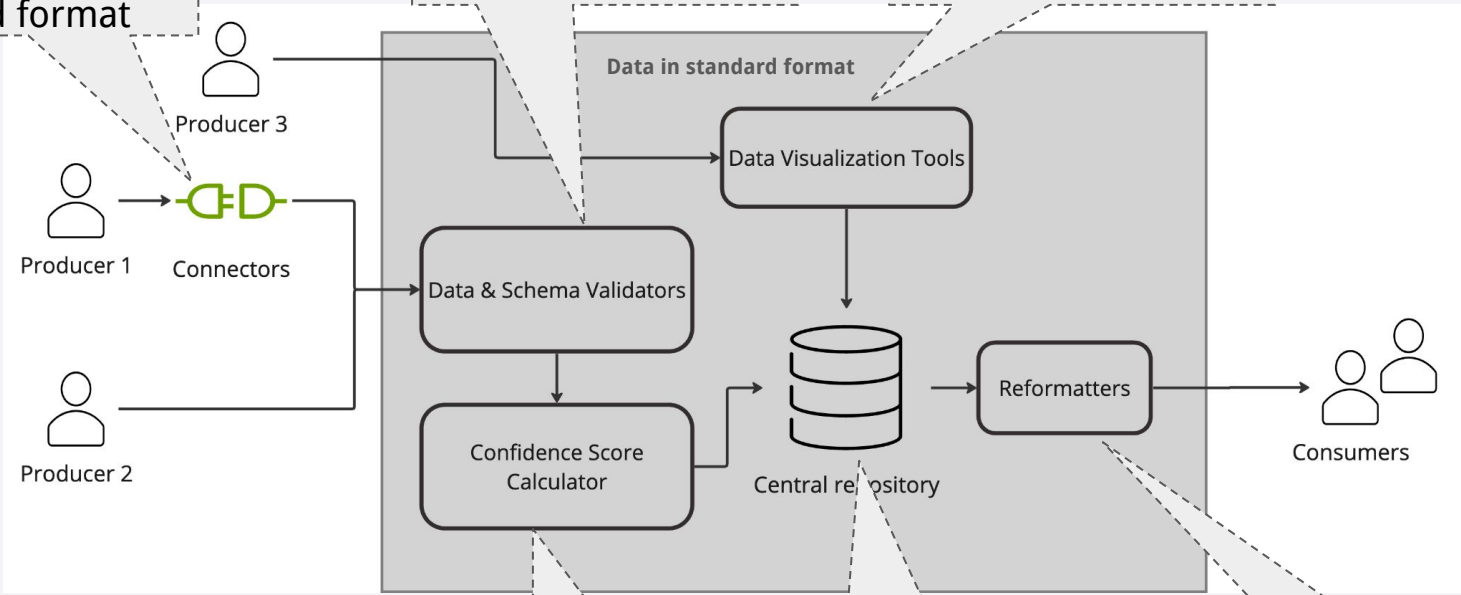
Service to compute the confidence in the data

Central database for all the data versions

Data joins for various applications

ITS4US

IT'S TRANSPORTATION FOR ALL OF US

Deployment Details of TDEI
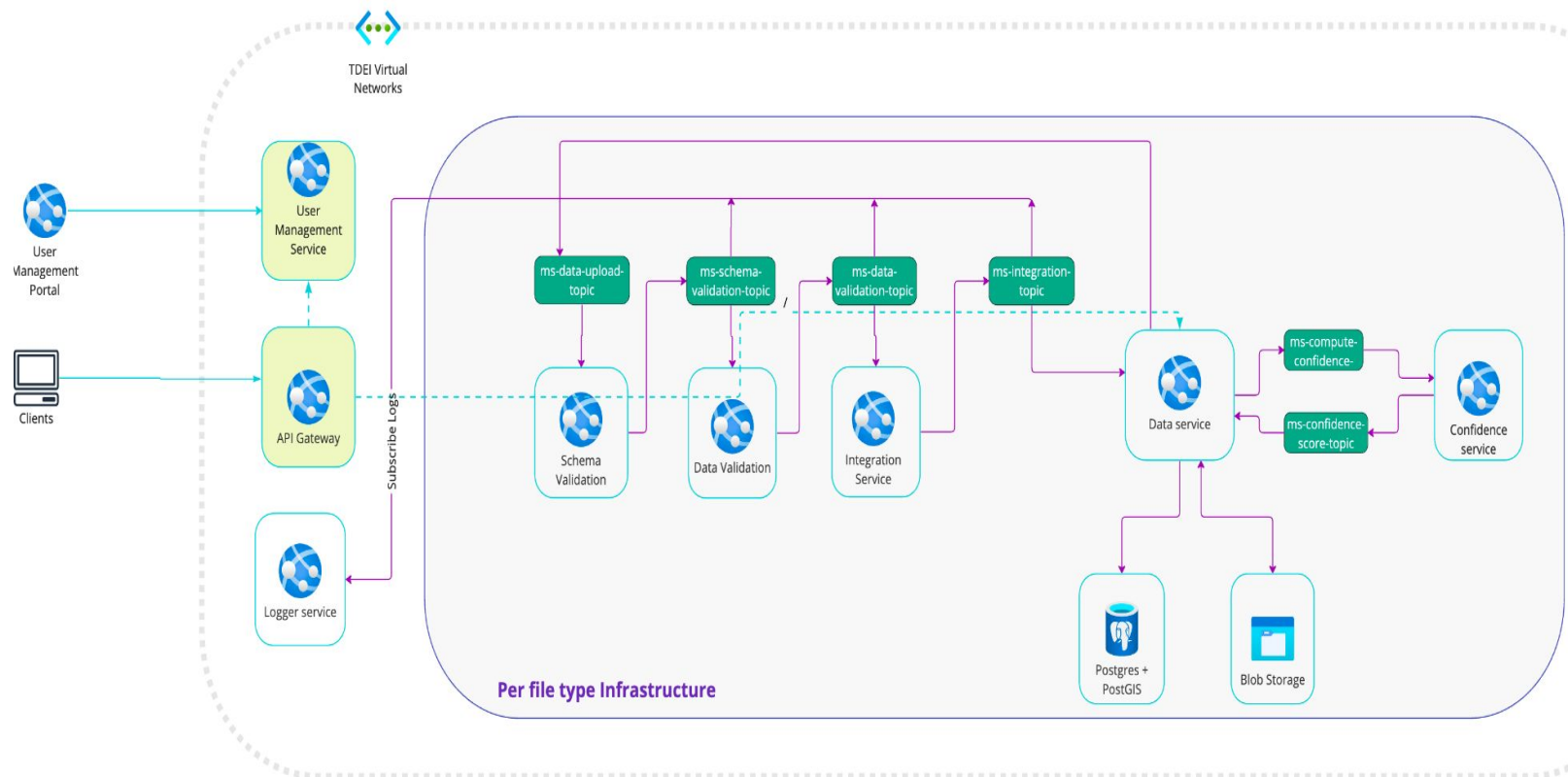
Simplified Deployment Architecture of TDEI

TDEI Components

TDEI Platform

TDEIWorkspaces	TDEICore	TDEIQuery
API (Edits, OSM Integration)	API (User Mgmt, Datasets, Metadata)	API (Bbox, Join, Conflation)

TDEI Tool Ecosystem

RAPID

OSW TaskingManager

GoInfoGame

GTFS Pathway Editor

TDEI Deployment Applications

AccessMapMultiModal

Walkshed Analysis

Audiom

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Deployment Details of TDEI

- Cloud Native development using **microservices** and **event bus** architecture
- Each functional unit implemented as a microservice; microservices communicate through event bus
- Modular, Scalable, Reliable, and high available system architecture
- Functional code is **open source**; deployment in Azure
- One of the few DoT sponsored projects to use the modern microservices software development architecture principles and **agile development**
- Data traceability, System Monitoring and Performance measurements

API & User Management

Authentication ⌃

POST `/api/v1/refresh-token` Re-issue access token 🔒 ⌄

POST `/api/v1/authenticate` Authenticates the user to the TDEI system. 🔒 ⌄

OSW ⌃

GET `/api/v1/osw` List osw files meeting criteria. 🔒 ⌄

POST `/api/v1/osw` Path used to upload/create a new osw data file. 🔒 ⌄

GET `/api/v1/osw/{tdei_record_id}` returns a geojson osw file 🔒 ⌄

GET `/api/v1/osw/versions` List available OSW versions 🔒 ⌄

GTFS-Pathways ⌃

GET `/api/v1/gtfs-pathways` List pathways files meeting criteria. 🔒 ⌄

POST `/api/v1/gtfs-pathways` create pathways file 🔒 ⌄

GET `/api/v1/gtfs-pathways/{tdei_record_id}` returns a gtfs_pathways file 🔒 ⌄

GET `/api/v1/gtfs-pathways/versions` List available GTFS Pathways versions 🔒 ⌄

Welcome!

Please login to your account.

Email Id

Enter Email Id

Password

Enter Password

Sign In

New to TDEI? Register Now

Project Groups, PoCs, Services, Users
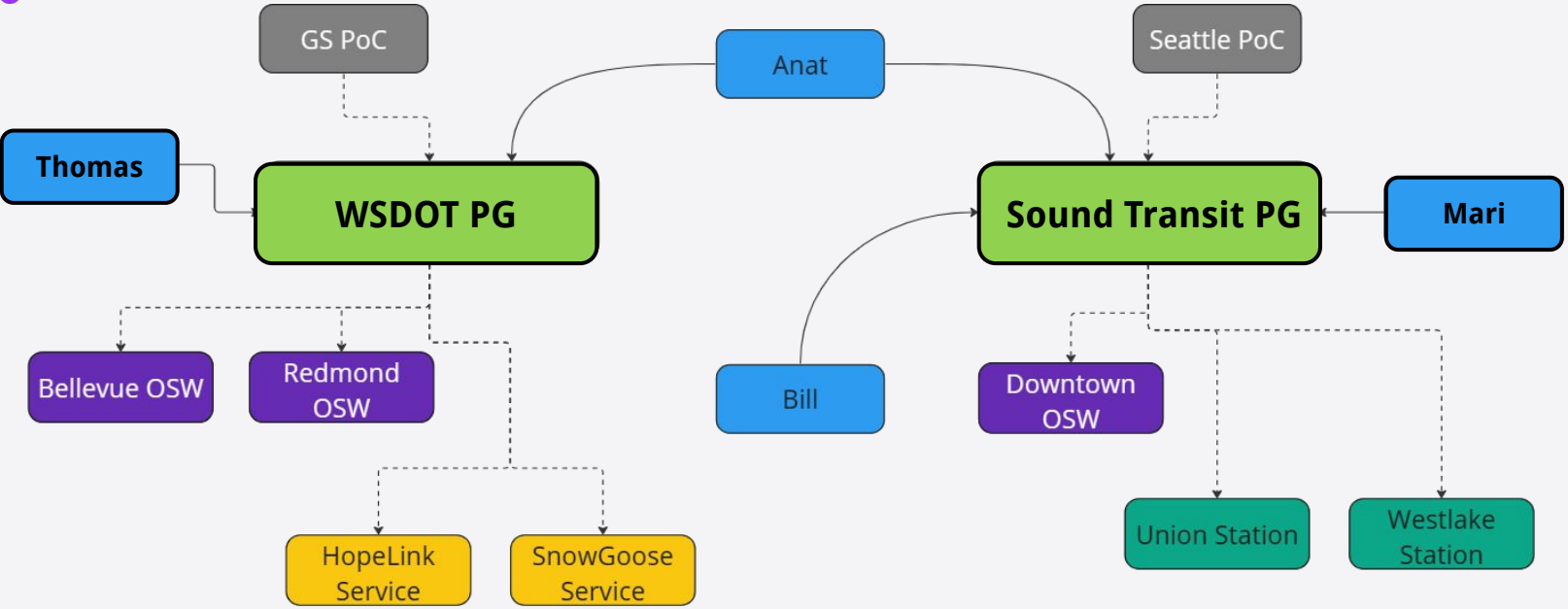
Producer API - OpenSidewalks (OSW) Upload & Download

POST /api/v1/osw

Parameters:

collected_by

collection_date

collection_method

publication_date

data_source

polygon (geojson, optional)

osw_schema_version

GET /api/v1/osw

Parameters (all optional):

bbox

tdei_org_id

date_time

osw_schema_version
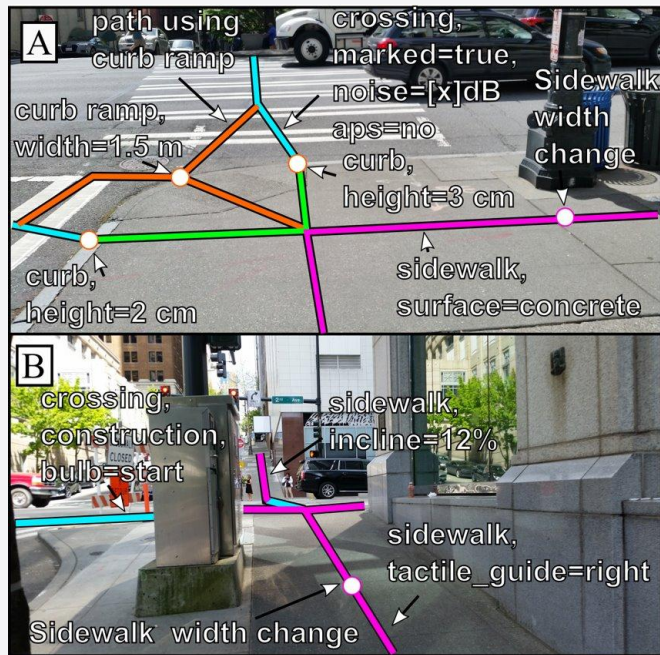
User Management Front End

Allows your organization / agency to:

- Register Users
- Assign a Point of Contact for your organization
- Designate who can upload data for your organization
- Manage Users

Welcome!

Please login to your account.

Email Id

Enter Email Id

Password

Enter Password

Sign In

New to TDEI? Register Now

Questions?

Creating Common Transportation Tools

(or: why not just a data feeds hub?)

TCAT — The Taskar Center for Accessible Technology

Clarifying release governance

- **All producers** can:
  - Govern releases
  - Govern archives and versions
  - Using common tools, perform data audits and visualize history
  - Federate their own data licensing

Consumption at scale

- **All open API consumers** can:
  - Perform daily scrapes
  - Create direct visualizations
  - Build pipelines to consume only federated releases
  - Perform data 'joins' for data integration and flexible applications
  - Perform data 'joins' to allow for specialized, one-off analyses
  - Use QGIS plugins for direct data consumption

ITS4US

IT'S TRANSPORTATION FOR ALL OF US

Vision for producers

- **All mobility data producers** can:
  - Use public data within their agency/municipality and about other agencies/municipalities
  - Use public tools that assist them in collecting, vetting and maintaining their data
  - Use open data formats, and "walk the walk" on data production with ease

Vision for consumers

- **All transportation data consumers** can:
  - Seamlessly consume data with no regional bounds without needing manual intervention or connection strings
  - Easily create innovative travel products using open standards
  - Can implement visualizations, analytics, prioritization schemas, travel apps and service discoverability
  - Aggregate multimodal data for various modes and formats
  - Conduct comparative analyses of travel networks
  - Conduct powerful analyses and comparisons of mobility systems

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Practical questions from consumers

- **Data consumers want to**:
1. Find all datasets available for a region via bounding box
2. Download a given dataset in osw format
3. Edit a dataset using osm tools, without downloading
4. Allow collaborators to edit a dataset in GoInfoGame
5. Allow collaborators to edit a dataset using RAPiD + esri backend
6. (Discussion) Join two datasets to enrich metadata, conflate, etc.

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

As a data engineer at…

Sound Transit…
Seattle DOT…
Washington DOT…
King County Metro…

I want to see all available data in the Seattle region

Practical use cases from stakeholders (1): Data discovery

General

GET	/api/v1/datasets	List Dataset files meeting criteria.
GET	/api/v1/jobs	List job details.
GET	/api/v1/...	
DELETE	/...	

GET `/api/v1/datasets` List Dataset files meeting criteria.

GET	/api/v1/services	List Services
GET	/api/v1/project-groups	List project groups
GET	/api/v1/api	List available API versions

ITS4US

IT'S TRANSPORTATION FOR ALL OF US

Practical use cases from stakeholders (2): Data access

As a data engineer at…

Sound Transit…
Seattle DOT…
Washington DOT…
King County Metro…

I need the current OSW Seattle sidewalks data

…. to join it with a surface disruptions to understand and cost needed access improvements for trainstation walksheds

…. to redistribute disability parking spots along sidewalks

…. to understand who has access to transit

…. to determine alternative paths to scheduled system

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Practical use cases from consumers (2): Data access

Datasets
Here are the list of datasets available

Upload Dataset

My Project Datasets   **All Released Datasets**

Search Dataset	Type	All	Status	All		Sort by

Dataset Name	Type	Collection Date	Status	Action
Pathways Upload testing Stage Test Uploaded at : 4/25/2024, 11:39:45 PM V1.0 Id : 32c48b13233c4a9ab5dc59ed83b53948	Pathways	10 Feb 2019	Released	Open in workspaces
TDEI - Bellevue OpenSidewalks Dataset Uploaded at : 4/25/2024, 2:04:23 PM V1.2 Id : bf3d4f10038444a4b0e7dca5acee3e64	OSW	26 Feb 2024	Pre-Release	Open in workspaces

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Practical use cases from consumers (2): Data access

OSW

POST	/api/v1/osw/upload/{tdei_project_group_id}/{tdei_service_id}	upload a pre-release of OSW dataset.
POST	/api/v1/osw/publish/{tdei_dataset_id}	Publishes the OSW dataset for the tdei_dataset_id
POST	/api/v1/osw/validate	Validates the osw dataset.
POST	/api/v1/osw/convert	OSW reformatting on demand
POST	/api/v1/osw/confidence/{tdei_dataset_id}	Initiate Confidence calculation for a dataset
GET	/api/v1/	
GET	/api/v1/	

GET `/api/v1/osw/{tdei_dataset_id}` downloads the OSW files as zip

| POST | /api/v1/osw/dataset-bbox | Given a dataset tdei_dataset_id returns the subgraph within a given bounding box. |
| POST | /api/v1/osw/dataset-tag-road | Given a target dataset, tags the sidewalks with the road network from source dataset. |

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Practical use cases from consumers (2): Data access

```
curl -X 'GET' \
'https://tdei-gateway-stage.azurewebsites.net/api/v1/osw/612ca2a6acf148e2bc389f0b3c79ba77
?format=osw' \
-H 'accept: application/zip' \
-H 'x-api-key: 7cd7d8c5-18aa-4dbe-91cc-c32ef6eee26c' \
-o bellevue.zip
```

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Practical use cases from consumers: Editing

As a data analyst at…

Sound Transit…
Seattle DOT…
Washington DOT…
King County Metro…

(3) I want to edit TDEI data using the OSM ecosystem and re-upload as a new version

(4) I want to empower collaborators to curate data in the field

(5) I want to empower collaborators to ingest data to OSM

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Practical use cases from consumers (3,4,5):
Synchronize with other dataset (including with general OSM)
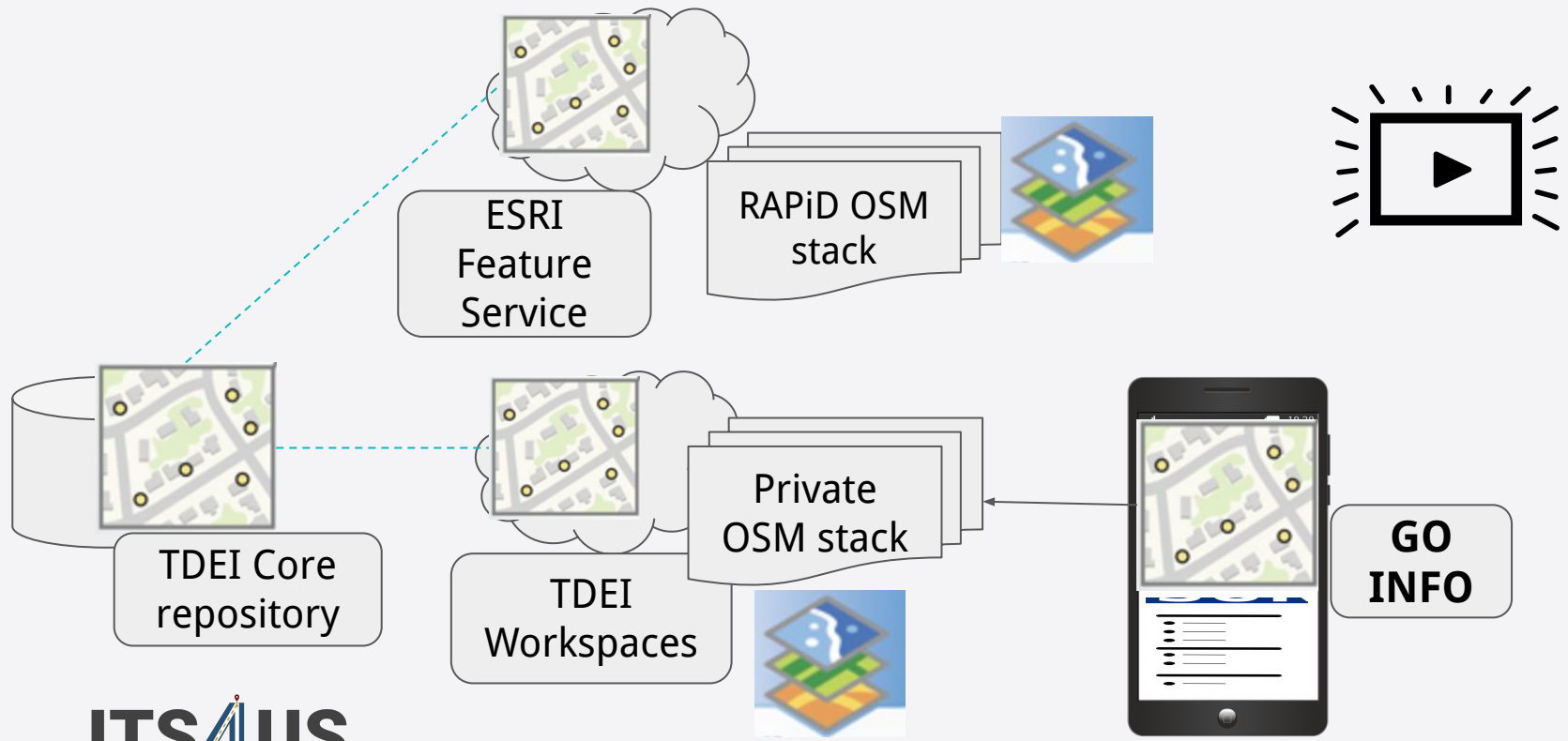
Practical questions from consumers

RAPiD EDITOR

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

Effective transportation analytics

- Leverage the **OpenSidewalks standardized pedestrian network**

- **Create mobility profiles** that describe pedestrian preferences for steepness, crosswalks, & raised curbs.

- **Generate walksheds** for a particular mobility profile and compare it against the street network.

- Calculate the **fraction of space reachable** on the **sidewalk network** versus the **street network** for a given pedestrian profile.

COMPLETE TRIP
ITS4US

Effective transportation analytics:

Pedestrian network "sidewalkscore"

COMPLETE TRIP
ITS4US

High walkscore (green background)
Low sidewalkscore (red network)
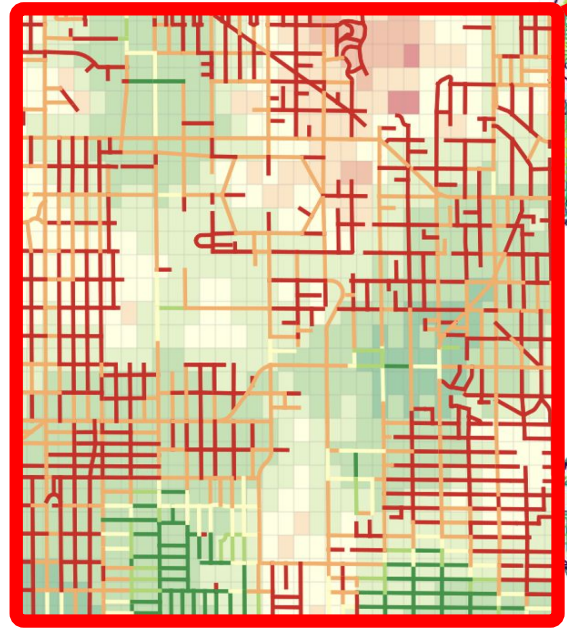
Why? No accessibility features; aging built environment; freeway
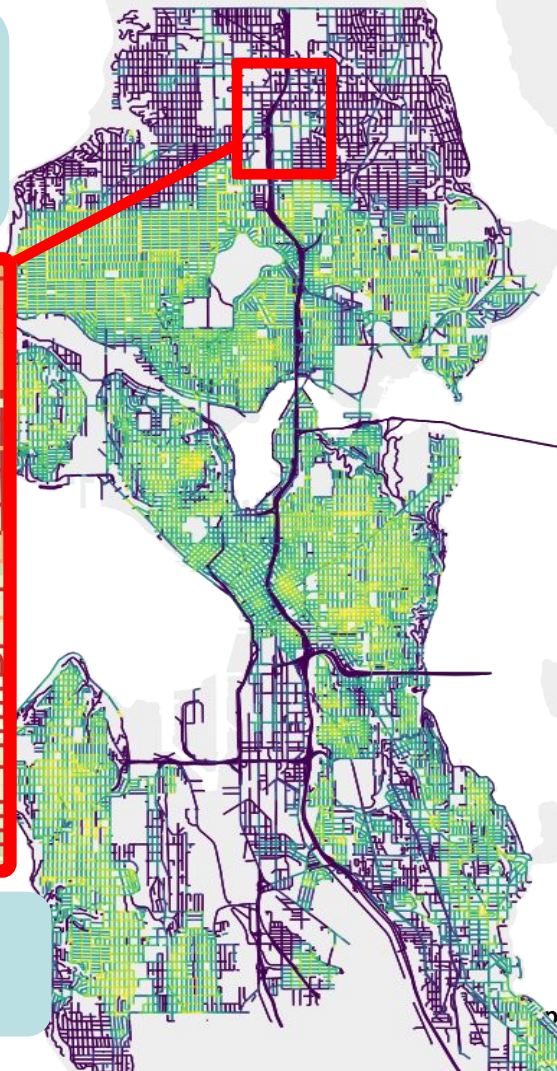
SidewalkScore interpretation:
1: On par with street network

0: much worse than street network

sidewalkscore_walk
1
0.75
0.50
0.25
0

Walkshed for unconstrained pedestrian

Effective transportation analytics:
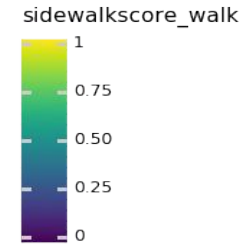
Pedestrian network "sidewalkscore"

High walkscore
	(green background)
Decent sidewalkscore
	(green/red network)

Why? Recently developed; significant improvements
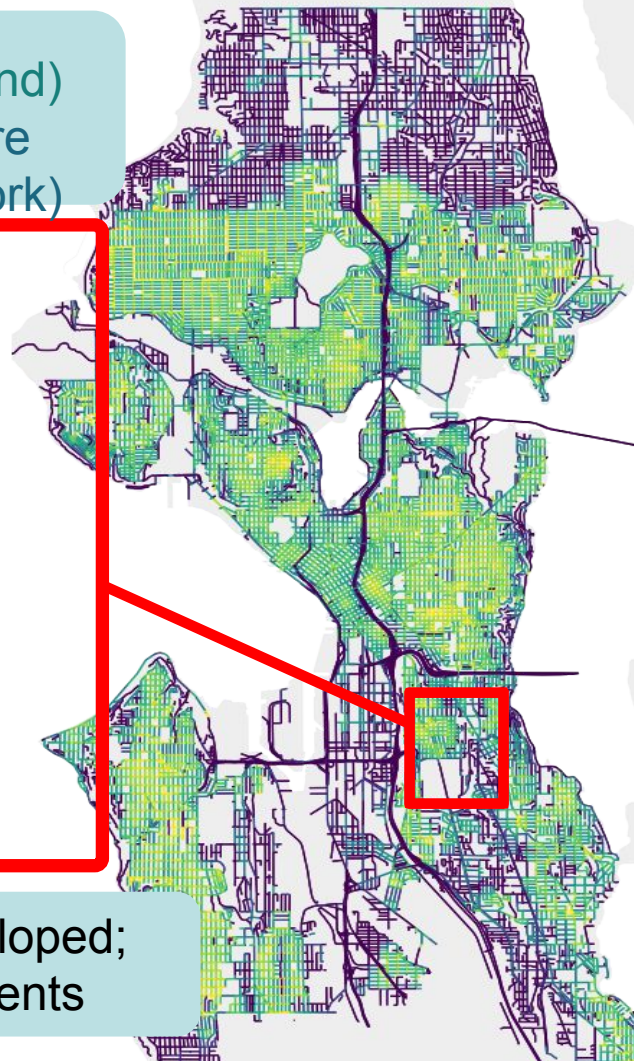
SidewalkScore interpretation:
1: On par with street network

0: much worse than street network

sidewalkscore_walk
1
0.75
0.50
0.25
0

Walkshed for unconstrained pedestrian

COMPLETE TRIP
ITS4US

Department of Transportation Program Office
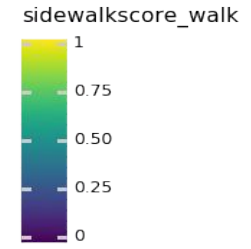
Where does necessary information come from?

Data Integration

Trained mappers

(High quality, but difficult to scale)

Automated Digital Inference

(e.g., satellite imagery, autonomous vehicles)

Legacy data from previous efforts

(may be out of date, usually special purpose, may be encumbered, sparse)

Opportunistic Passive Observation
(e.g., crowdsourced video, GPS traces. Expensive; coupled to collection and difficult   to integrate)

COMPLETE TRIP
ITS4US

**U.S. Department of Transportation
ITS Joint Program Office**

Where does necessary information come from?
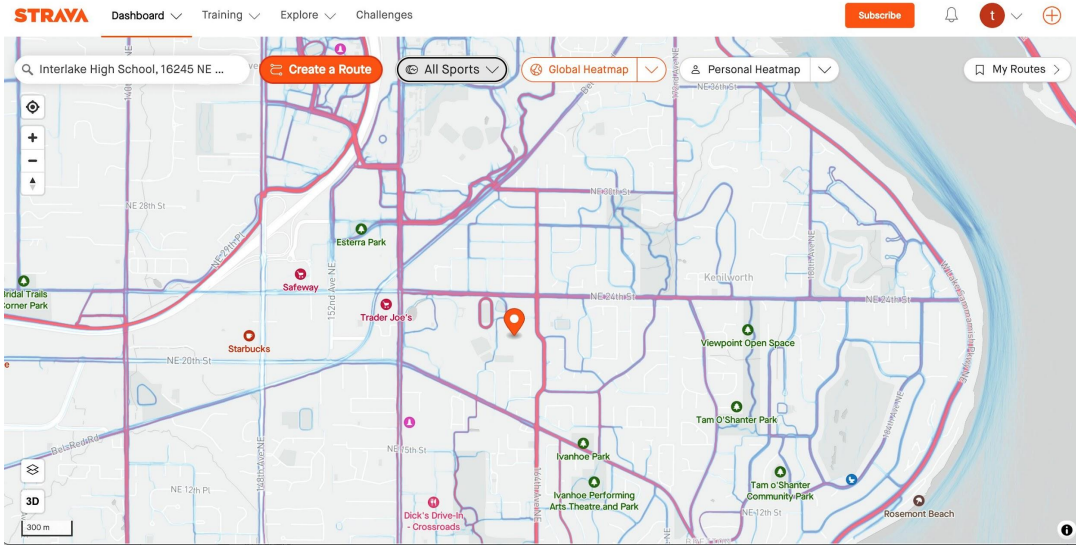
Data Integration

Example: STRAVA Road and Pedestrian Networks
(standardized, biased towards fitness enthusiasts)

Where does necessary information come from?
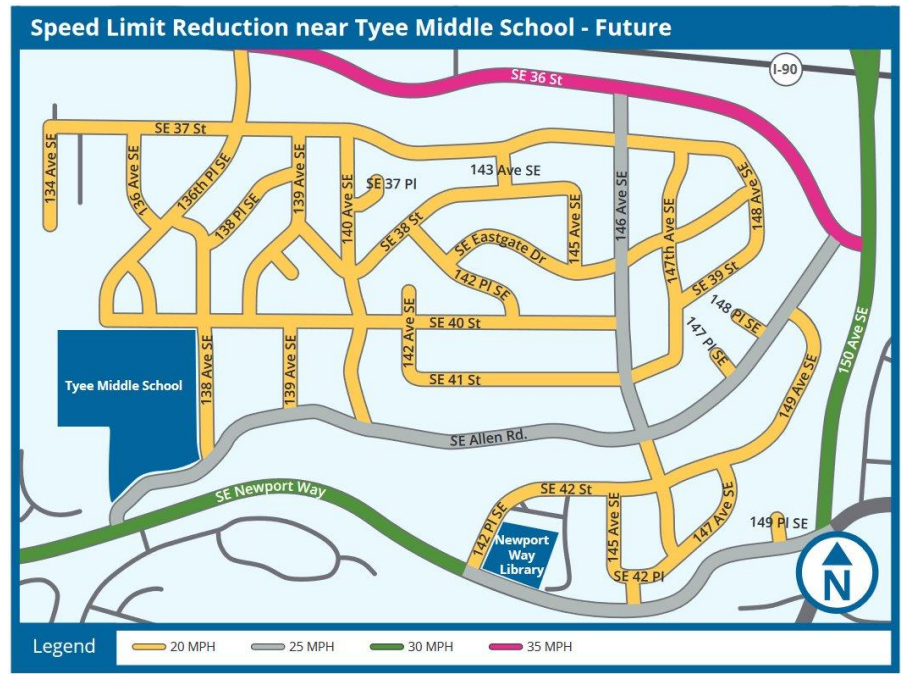
Data Integration

Neighborhood streets highlighted in yellow will have their speed limits reduced from 25 to 20 miles per hour. (Credit: City of Bellevue)
(attached to roadway, not pedestrian routes)

Where does necessary information come from?
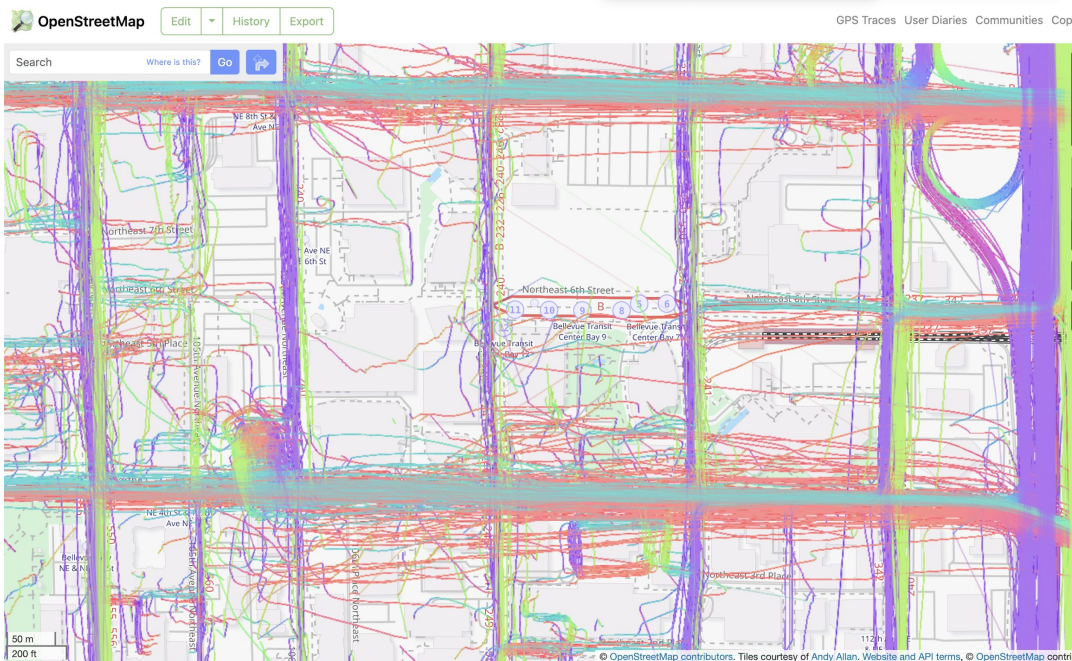
Data Integration

GPS Traces from Open Street Maps. (Credit: OSM)
(attached to roadway, not pedestrian routes)

COMPLETE TRIP
ITS4US

Where does necessary information come from?

Data Integration

Gehl Analysis considers walkability from the perspective of resident pedestrians: 12 criteria in 3 categories: protection, comfort, and enjoyment.

GEHL ANALYSIS SCORES IN NORTHEAST BELLEVUE

Legend
- Neighborhood Boundary
- Gehl Analysis Sites

Gehl Analysis Score

SOPHIA NELSON

COMPLETE TRIP
ITS4US

Where does necessary information come from?

Data Integration

Crowdsourced footpaths from Open Street Maps (Credit: OSM)
(pedestrian routes, no accessibility info)

Where does necessary information come from?
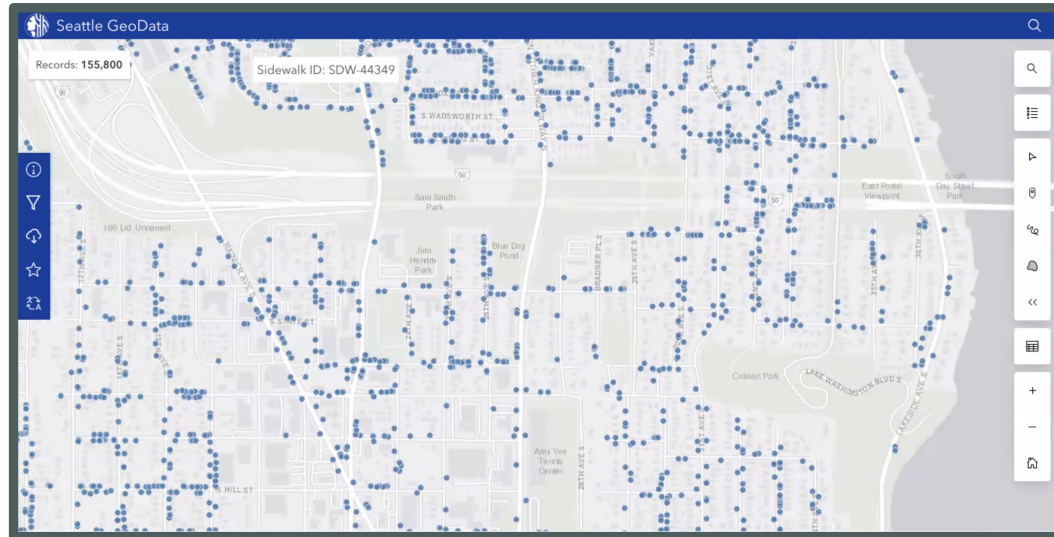
Locations of Surface Disruptions (City of Seattle)
(raw locations, not associated with pedestrian routes)

Where does necessary information come from?

Data Integration

Read Models: Integrations of Multiple Datasets to support Routing and Analytics (Credit: AccessMap)

Figure x: Using the new cost field to visualize sidewalk issues in the walksheds. Only sidewalks with observations included in the cost function are highlighted.
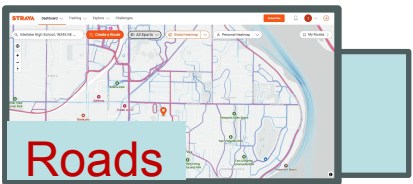
Data Integration

*The value of a data repository scales quadratically with the number of datasets it holds due to network effects*

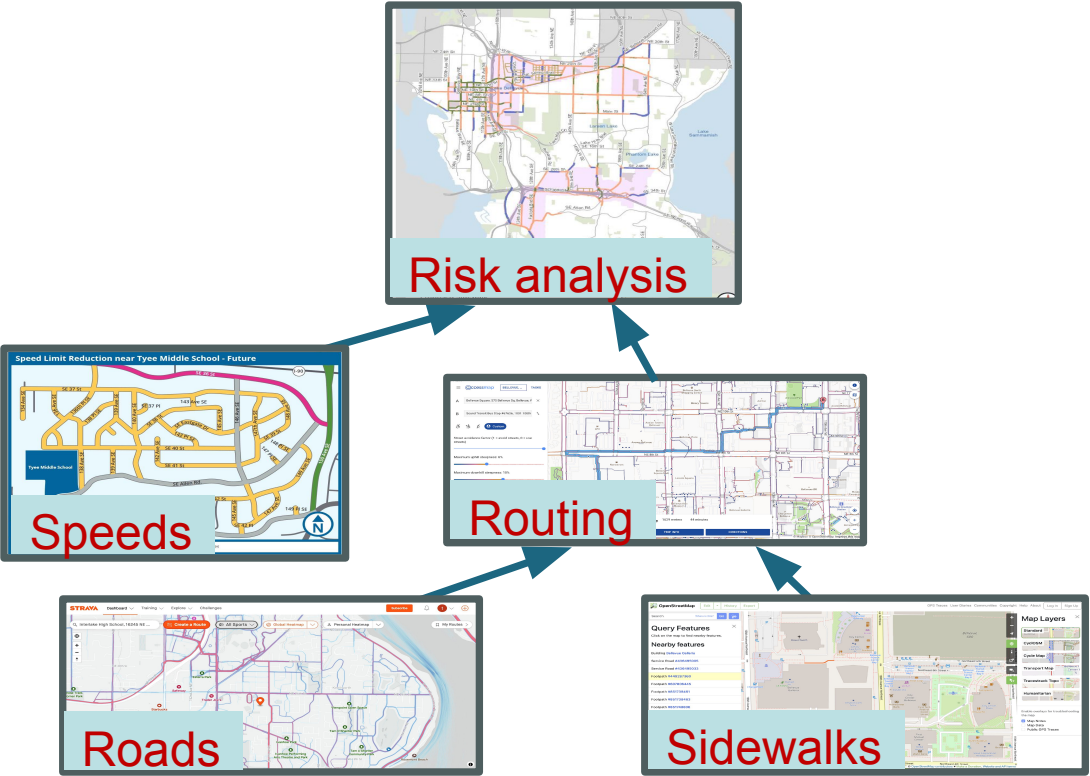# Integrating Heterogeneous Data Enables Analytics

Data Integration

Roads

Sidewalks

COMPLETE TRIP
ITS4US

Integrating Heterogeneous Data Enables Analytics

Data Integration

Routing

Roads

Sidewalks

COMPLETE TRIP
ITS4US

Integrating Heterogeneous Data Enables Analytics

Data Integration

Risk analysis

Speeds

Routing

Roads

Sidewalks

Integrating Heterogeneous Data Enables Analytics

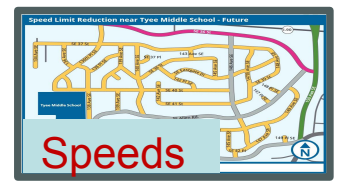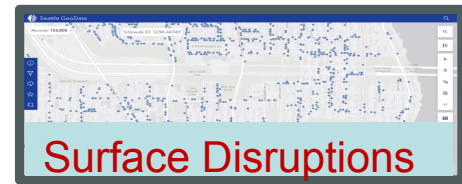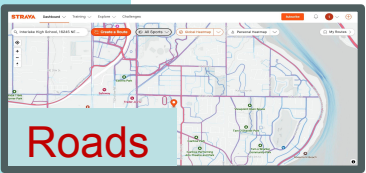**Data Integration**

Risk analysis

Routing

Cost analysis
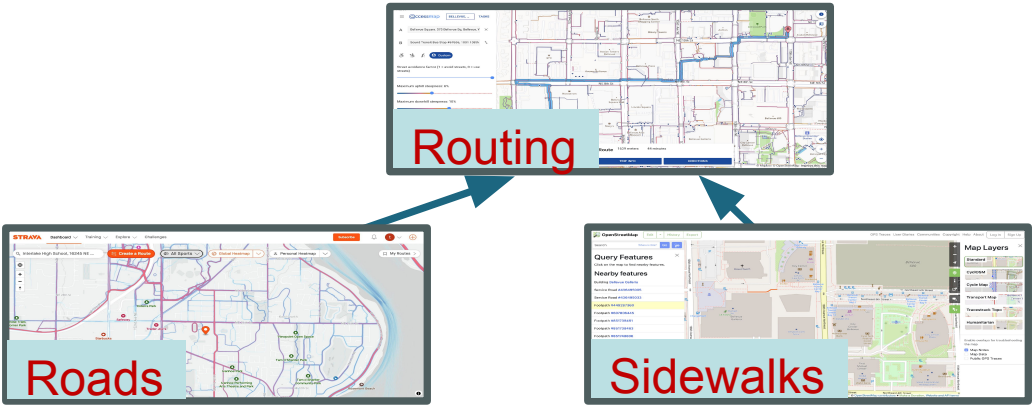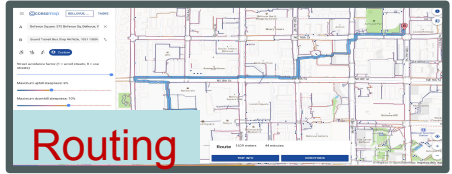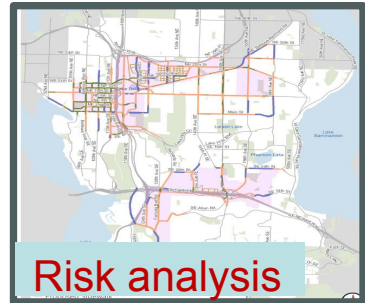
Specialized Read Models

ccessmap

Python: computational geometry + metadata harmonization +

+ Customized, per-customer

- One-off, relies on in-house engineering talent, coupled to specific applications like AccessMap

Roads

Sidewalks

Surface Disruptions

Speeds

COMPLETE TRIP
ITS4US

Integrating Heterogeneous Data Enables Analytics

Data Integration

TDEIQuery

Abstract and generalize common patterns for server-side integration

Risk analysis

Cost analysis

Speeds

Routing

Surface Disruptions

Roads

Sidewalks
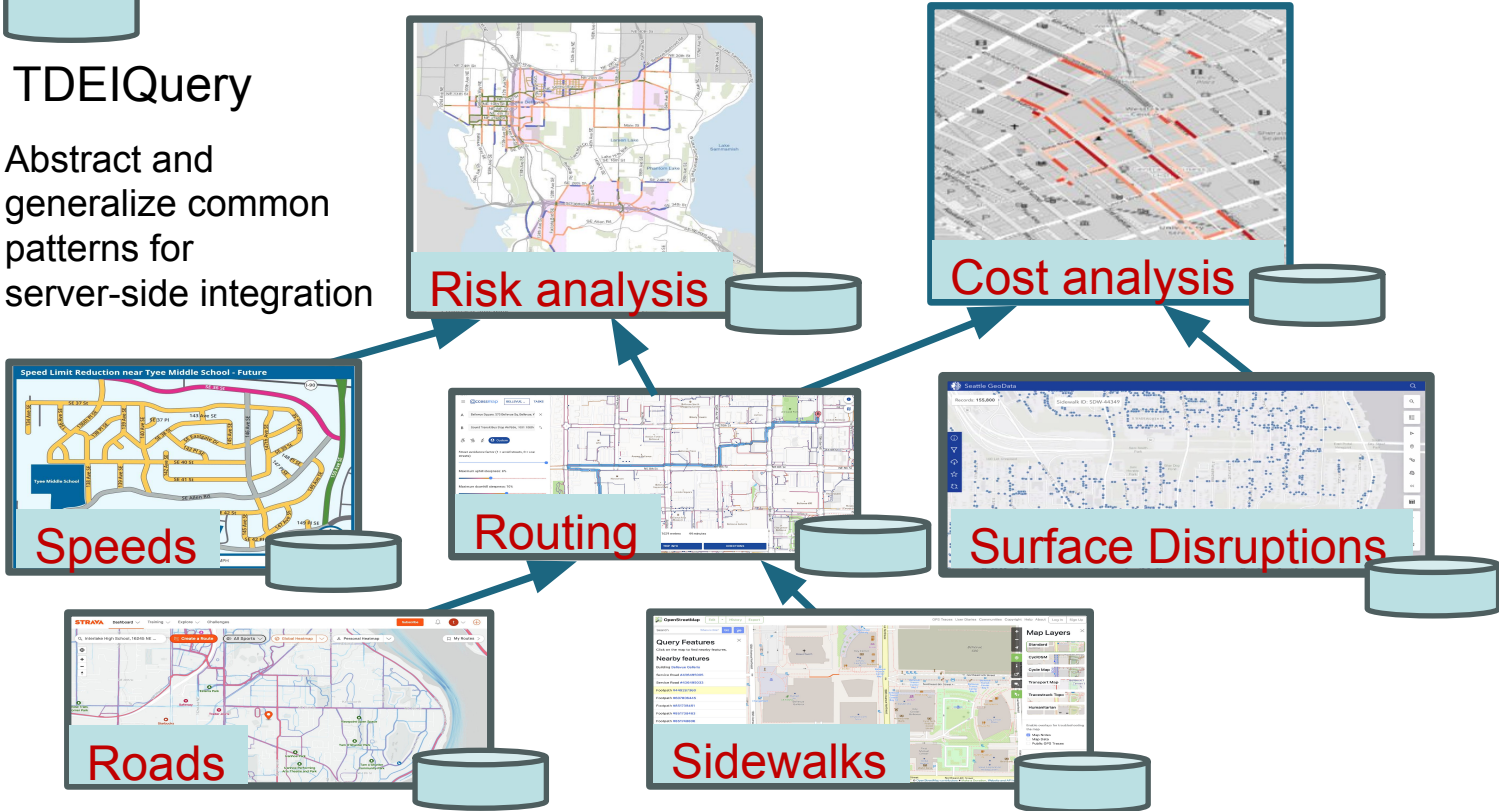
20+ Use cases for joins, integrations, conflation, unions, analytics, computational geometry, imputations, etc.,

Exploit common patterns to generalize and expose through self-serve APIs backed by spatiotemporal joins in scalable query languages

**U.S. Department of Transportation
ITS Joint Program Office**

COMPLETE TRIP
ITS4US