A Self-Sustaining, Self-Perpetuating Web-Scraping Application for Crude Oil Railroad Route Information

Shih-Miao Chin

Chieh (Ross) Wang

Ho-Ling Hwang

Hyeonsup Lim

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

**U.S. DEPARTMENT OF ENERGY**

Outline

- Background and motivation

- Crowdsourcing approach: scraping crude-by-rail location information from photo sharing platforms

- Risk assessment

- Discussions and moving forward

OAK RIDGE
National Laboratory

# Motivation: Examining Crude-by-rail – A Crowdsourcing Approach

- Technological advances in horizontal drilling and hydraulic fracturing contributed to the shale oil and gas exploration boom

- Since 2010, crude-by-rail traffic has been emerged and increased drastically with the growing production of shale oil

- Information on crude oil production, consumption, and mode of transportation is needed to formulate remedial measures to counter negative impacts induced by shale oil exploration and subsequent transportation logistics

OAK RIDGE
National Laboratory

Transportation of Crude Oil

- Mode shares for transporting shale oil from the Williston Basin (Bakken):
  - Share of rail increased consistently from 2010 to 2013
  - Rail continued to hold over 50% of market shares in 2014 & 2015

- Crude by Rail
  - Majority of crude by rail originated in the Midwest (PADD 2)
  - Despite decreased in 2016 and 2017, oil by rail increased in 2018 and 2019
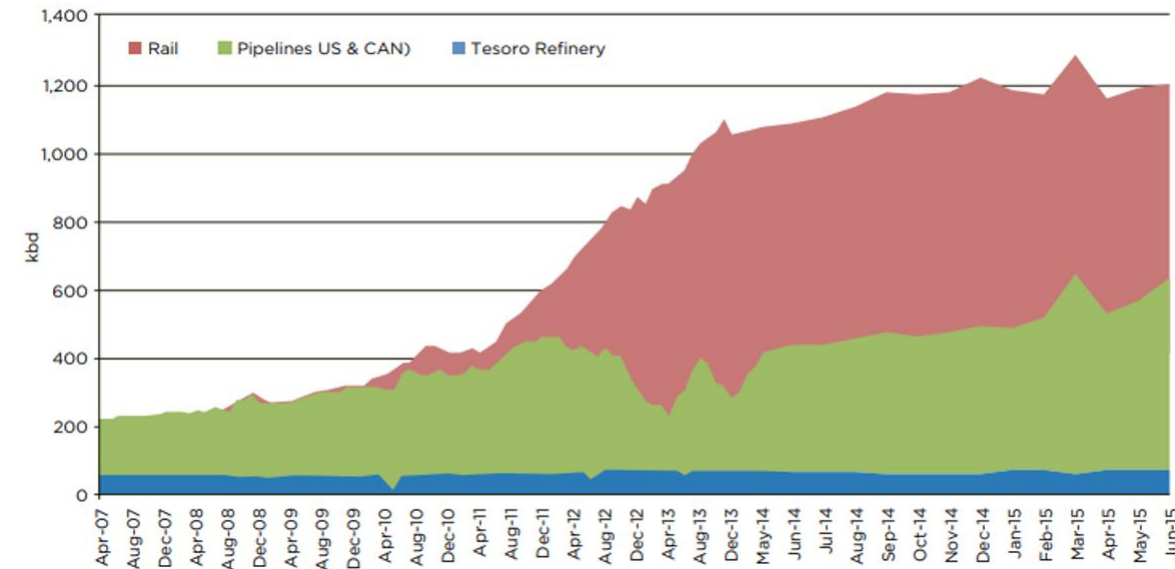
Transport of Bakken Crude (Source: North Dakota Pipeline Authority, 2015)
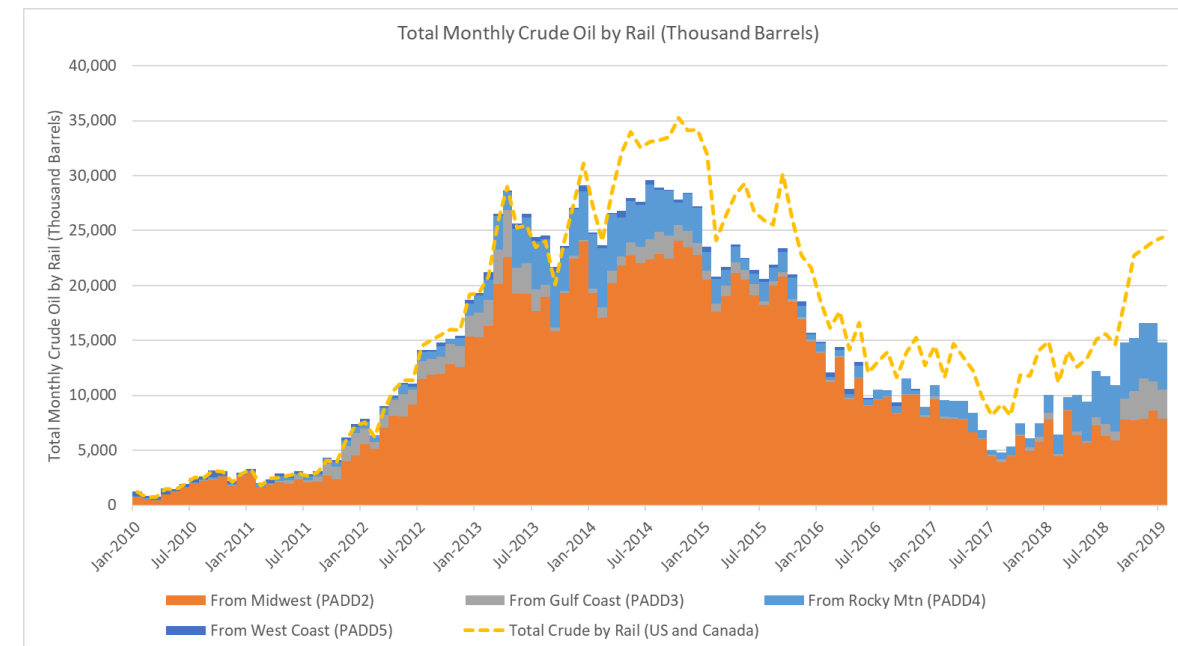
Crude by Rail among PADDs (Source: EIA, 2019)

OAK RIDGE
National Laboratory

# Unlike Other HAZMAT Transportations...

- Railroad companies are regulated by Surface Transportation Board

- Oil trains are not subject to the same strict routing requirements

- Oil trains are permitted to pass directly through cities (negative impact: undesirable consequence happened)

- In the absence of more protective regulations, people began to take matters in their own hands

Source: NTSB
Casselton, ND 2013

Lac-Megantic, Quebec 2013

Source: https://earthjustice.org/
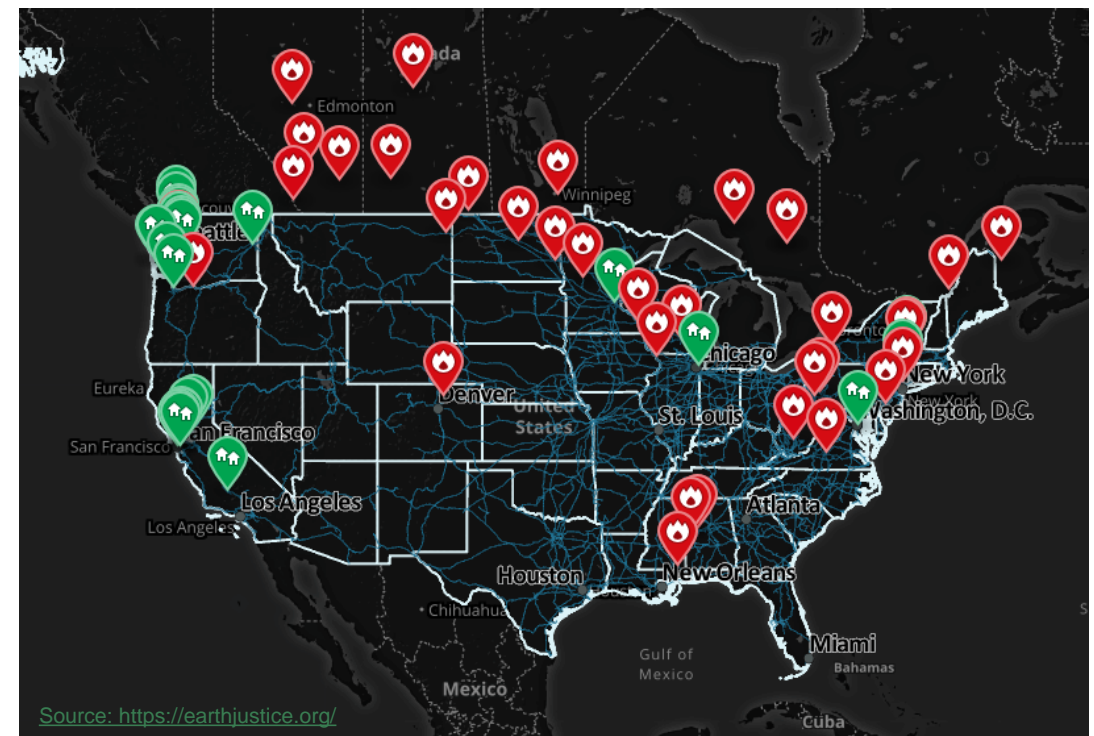
Oil Train Disasters Map since 2012 (last updated: Oct 2016)

**OAK RIDGE**
National Laboratory

Population at Risk

- LandScan data developed by ORNL was used (2011)

- Global population distribution by day and night at 30m x 30m resolution

- Buffer zones 0.5, 1, and 2 miles

Confirmed routes —

Inferred routes —

Distance to railroad track	Day			Night		
	Confirmed routes	Inferred routes	All routes	Confirmed routes	Inferred routes	All routes
<=0.5 mile	659,677	1,001,865	1,193,057	68,677	96,248	112,255
<=1 mile	938,545	1,555,130	1,758,430	107,419	136,256	156,153
<=2 miles	1,640,369	2,002,451	2,194,060	149,539	181,013	199,711

OAK RIDGE
National Laboratory

6

# Analysis of Hazardous Materials Incidents
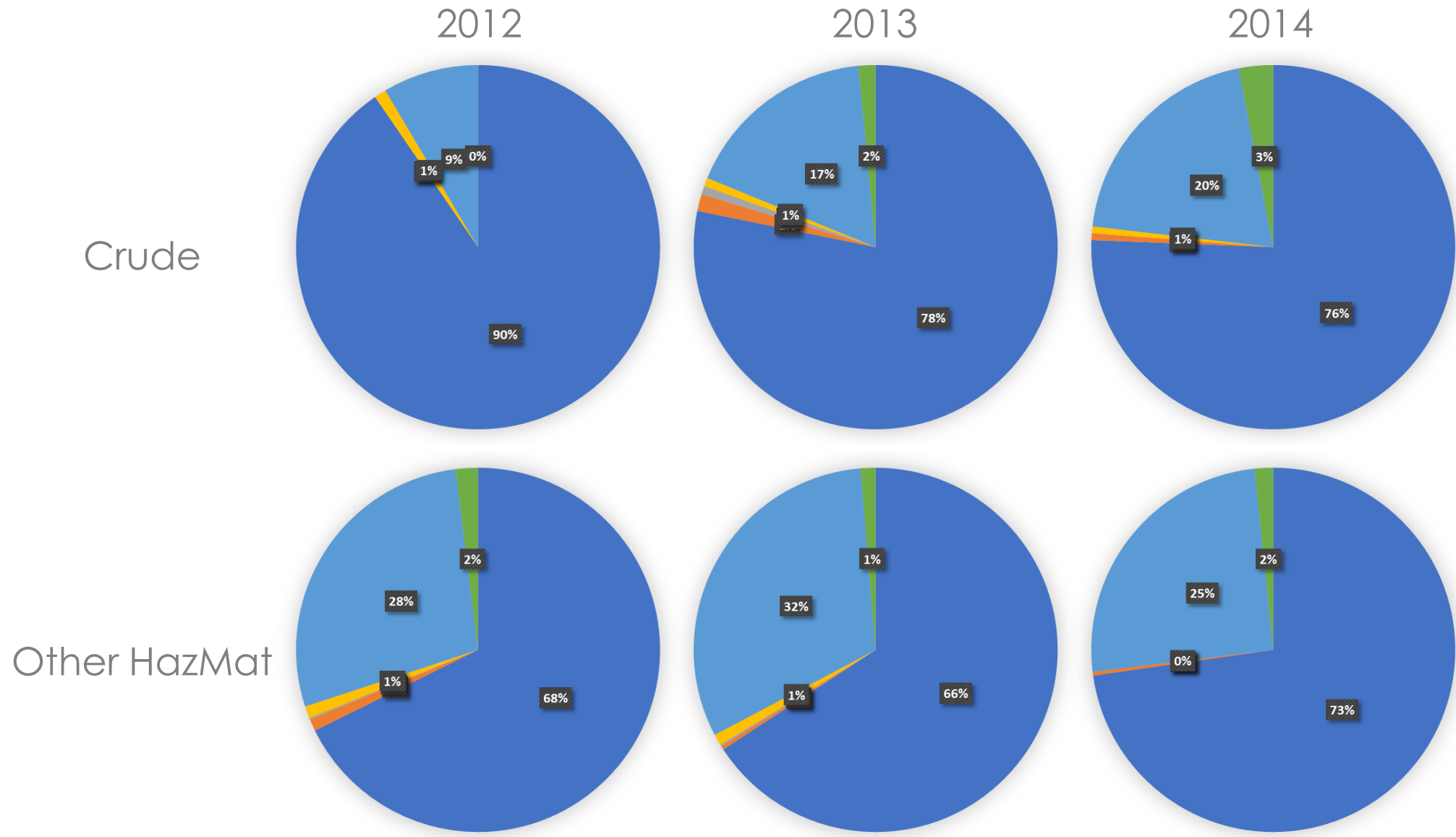
Source: PHMSA

Importance of Risk Communication

- Crude-by-rail operation information is critical for formulating remedial measures to counter any negative impacts.

- Emergency preparedness official need information to help residents of affected communities to
  - Understand the processes of oil train risk assessment and management;
  - Form scientifically-valid perceptions of the likely crude oil disasters; and
  - Participate in making decisions about how risk should be managed.

- There is a need to compile a crude oil-train flow database that can be used seamless with associated energy commodity origin-destination flows and logistic information provided by the CFS/FAF.

OAK RIDGE
National Laboratory

Crowdsourcing Approach: Photo Sharing Social Platforms

- Railway enthusiast, or railway buff, of many ages can be found worldwide - often combine their interests with other hobbies, especially photography & videography;

- These pictures taken are shared on social media platforms (publicly or privately) and many of them are

  – With detailed information on the picture's exchangeable image file format (EXIF)

  – Well described and tagged (text categories of the photo content)

  – Geo-tagged (GPS location taken)

Source: flickr.com

Source: railpictures.net

OAK RIDGE National Laboratory
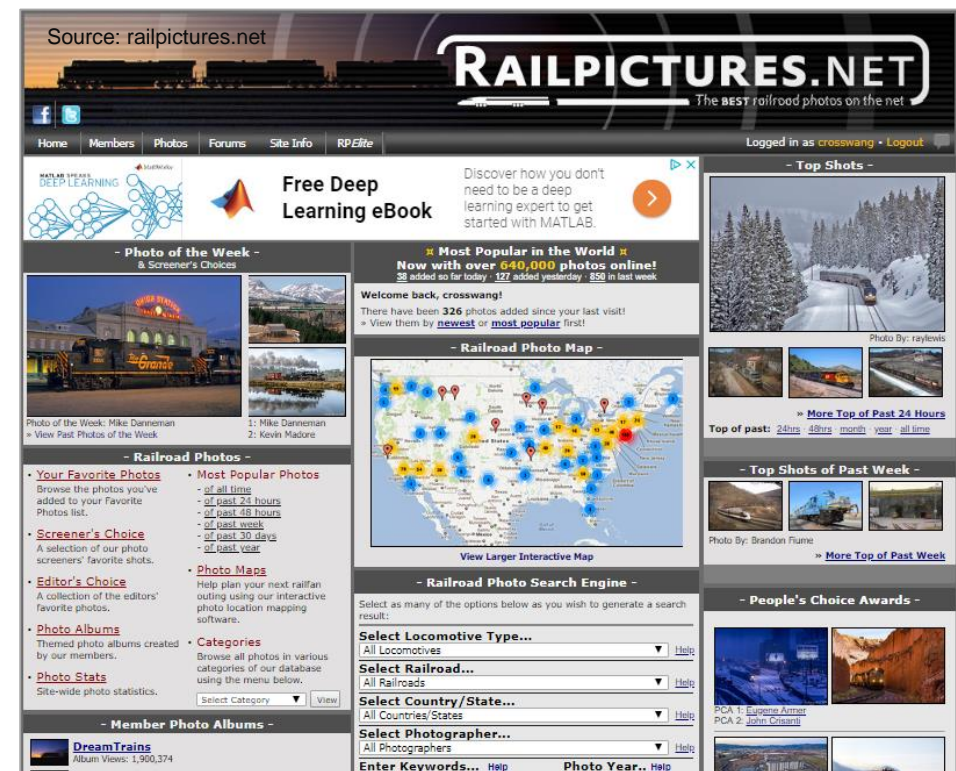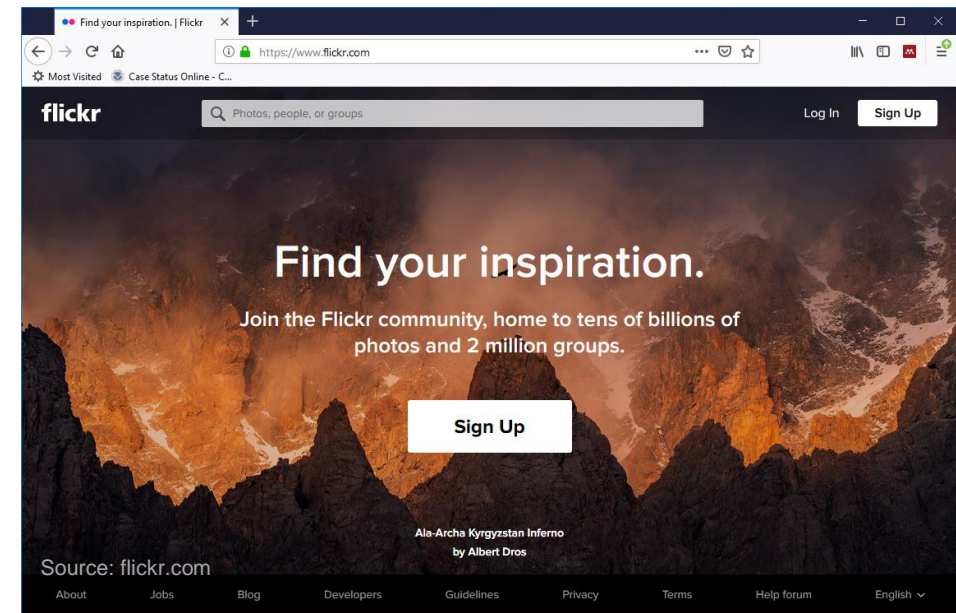
Web-scraping Crude-by-Rail Photo Information Using Flickr's API

- Flickr provides an application programming interface (API) for developers to develop applications using Flickr's services

- A **flickr.photos.search** function allows developers to search for pictures using tags and other features and get formatted search results

flickr.photos.search

Arguments

Name	Required	Send	Value
api_key	required	☐	
user_id	optional	☐	
tags	optional	☑	crude oil
tag_mode	optional	☐	
text	optional	☐	
min_upload_date	optional	☐	
max_upload_date	optional	☐	
min_taken_date	optional	☐	

▼<rsp stat="ok">
  ▼<photos page="1" pages="6" perpage="100" total="509">
    <photo id="47450352902" owner="141169384@N03" secret="b4c03e12e4" server="7891" farm="8" title="Minerva Atlantica" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="40537159853" owner="141169384@N03" secret="a9875304a7" server="7808" farm="8" title="Minerva Atlantica" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="33626467648" owner="141169384@N03" secret="c5f45183cf" server="7911" farm="8" title="Minerva Atlantica" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="47450350812" owner="141169384@N03" secret="da77da9c8d" server="7922" farm="8" title="Minerva Atlantica" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="40537157293" owner="141169384@N03" secret="c97c4bd6a7" server="7804" farm="8" title="Minerva Atlantica" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="47400154222" owner="138382841@N08" secret="a1bfcb9f18" server="7820" farm="8" title="482453752" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46490262164" owner="141169384@N03" secret="3cfecd4c82" server="7903" farm="8" title="Cap Charles" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="47161612702" owner="141169384@N03" secret="3ca3d19161" server="7886" farm="8" title="Cap Charles" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="32271458207" owner="141169384@N03" secret="2f18f75f0b" server="7814" farm="8" title="Cap Charles" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="32271457027" owner="141169384@N03" secret="a8ea4ae50d" server="7835" farm="8" title="Cap Charles" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="33338126118" owner="141169384@N03" secret="000287bf5c" server="7813" farm="8" title="Cap Charles" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46490248204" owner="141169384@N03" secret="b8c4f01699" server="7882" farm="8" title="Cap Charles" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="32271453887" owner="141169384@N03" secret="3d7713b1dd" server="7918" farm="8" title="Cap Charles" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="32262924797" owner="49410893@N04" secret="1739542a9e" server="7921" farm="8" title="crude flowing through the gates." ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="47140216981" owner="151501393@N07" secret="8ecc9b719c" server="7830" farm="8" title="Temblor / Witter Field project is situated in the San Joaquin Valley © Arctic Hunter Energy Inc. (TSX-V: AHU)" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="47085662362" owner="164178813@N06" secret="42dfeb3a1f" server="7818" farm="8" title="Woman refueling gas on a rainy day." ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="33262527048" owner="164178813@N06" secret="b62320066a" server="7814" farm="8" title="Woman refueling gas on a rainy day." ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="47044540491" owner="141169384@N03" secret="8b0093234a" server="7814" farm="8" title="Advantage Avenue" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="33168951798" owner="141169384@N03" secret="6045d327f7" server="7873" farm="8" title="Advantage Avenue" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="40079348403" owner="141169384@N03" secret="77b85047b6" server="7808" farm="8" title="Advantage Avenue" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46991448982" owner="141169384@N03" secret="8c2e6a6362" server="7921" farm="8" title="Advantage Avenue" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46991447122" owner="141169384@N03" secret="ea02783d8a" server="7860" farm="8" title="Advantage Avenue" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46991446392" owner="141169384@N03" secret="1a0253c363" server="7898" farm="8" title="Advantage Avenue" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46232709754" owner="45950591@N07" secret="0162747149" server="4837" farm="5" title="-DSC_5967 c." ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46174346404" owner="151501393@N07" secret="ebb151354d" server="4876" farm="5" title="Gas processing facility is located 7 miles from the Witter field discovery © Arctic Hunter Energy Inc. (TSX-V: AHU)" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46847156422" owner="151501393@N07" secret="a70f4e1eff" server="7912" farm="8" title="HM-5 and HM-12 wells have now had their hydrocarbon bearing zones successfully perforated © Arctic Hunter Energy Inc. (TSX-V: AHU)" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="33023789588" owner="151501393@N07" secret="b10e99fc20" server="7861" farm="8" title="NEWS RELEASE: January 11, 2019 © Arctic Hunter Energy Inc. (TSX-V: AHU)" ispublic="1" isfriend="0" isfamily="0"/>
    <photo id="46174341424" owner="151501393@N07" secret="c79f5da360" server="7855" farm="8" title="Targeting light sweet (38°API) oil formation © Arctic Hunter Energy Inc. (TSX-

Web-scraping Crude-by-Rail Photo Information by Parsing Through Webpage Sources

- All pictures on Railpictures.net are rail-related pictures

- It allows users to search for pictures using keywords

- Search results can be returned as thumbnails, each with a hyperlink to the photo and additional information

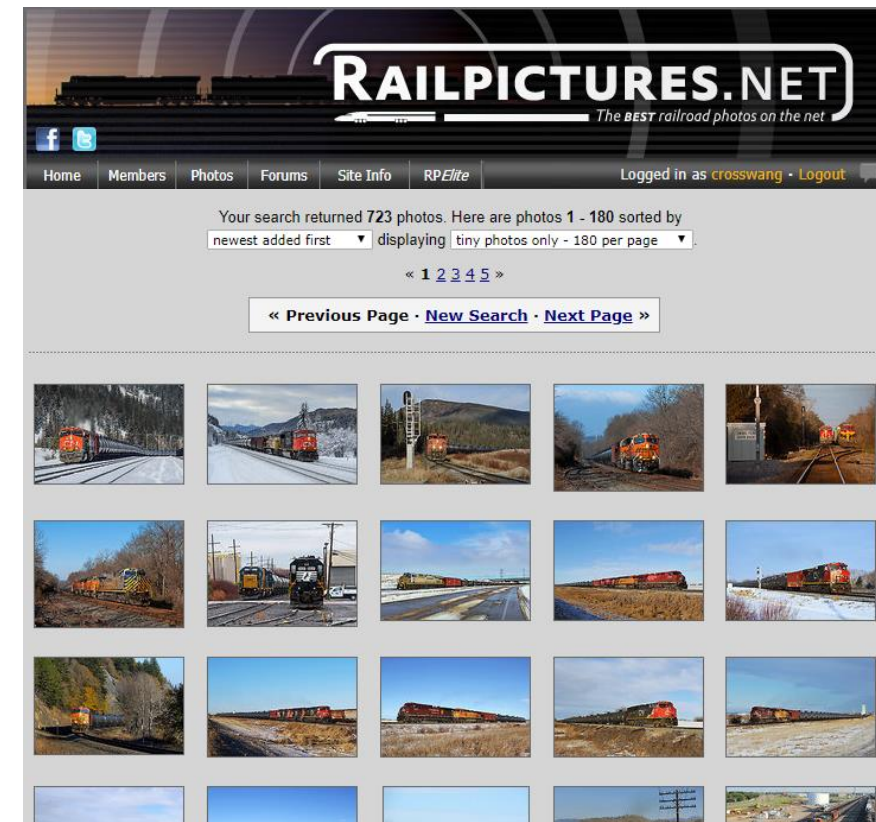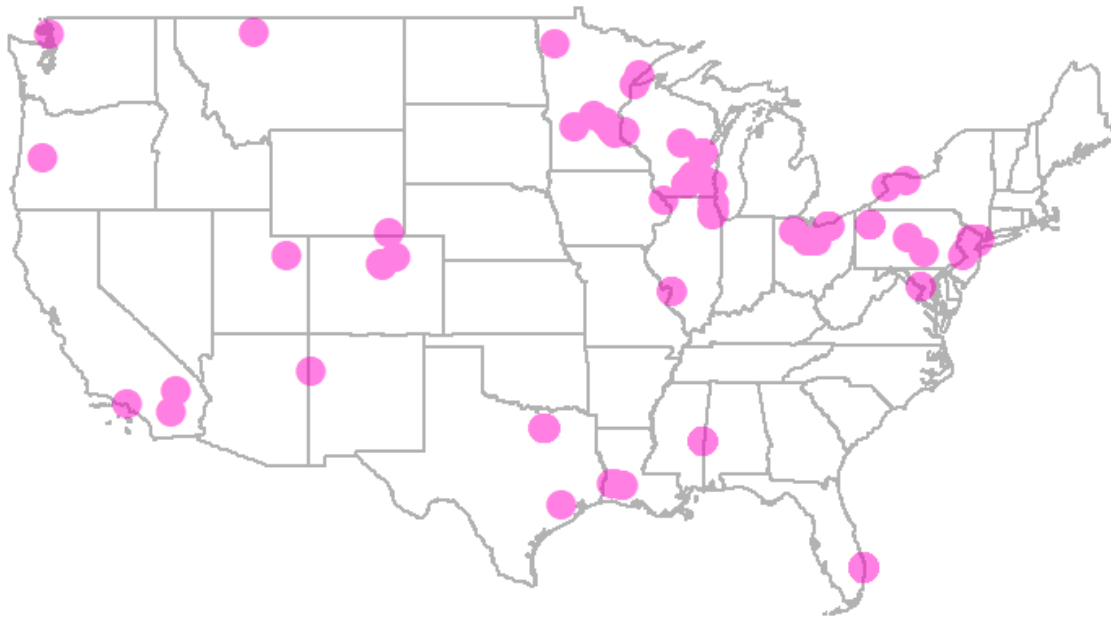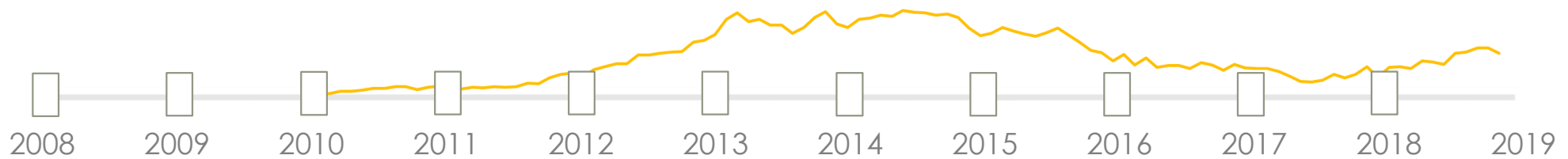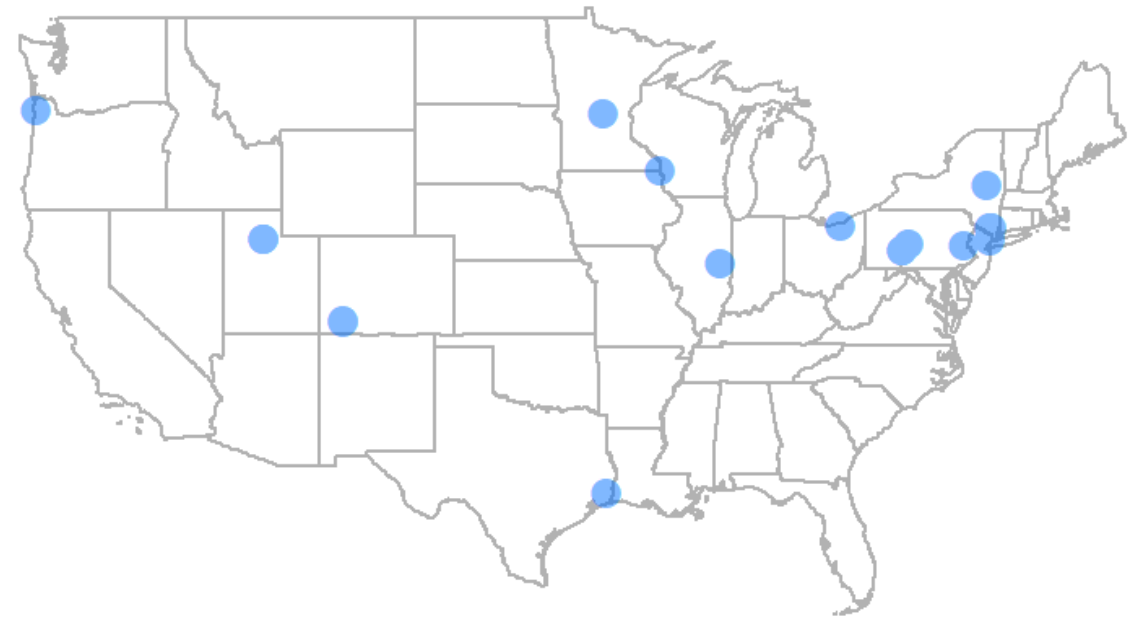- Crude-by-rail routes can be identified by "parsing" through the webpage source code of the search results and individual photo information table.

OAK RIDGE
National Laboratory

Crude-by-Rail Maps based on Photo Locations

Flickr

RailPictures

2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019

OAK RIDGE
National Laboratory

Crude-by-Rail Maps based on Photo Locations (cont'd)

Flickr   +   RailPictures

2008   2009   2010   2011   2012   2013   2014   2015   2016   2017   2018   2019

OAK RIDGE
National Laboratory
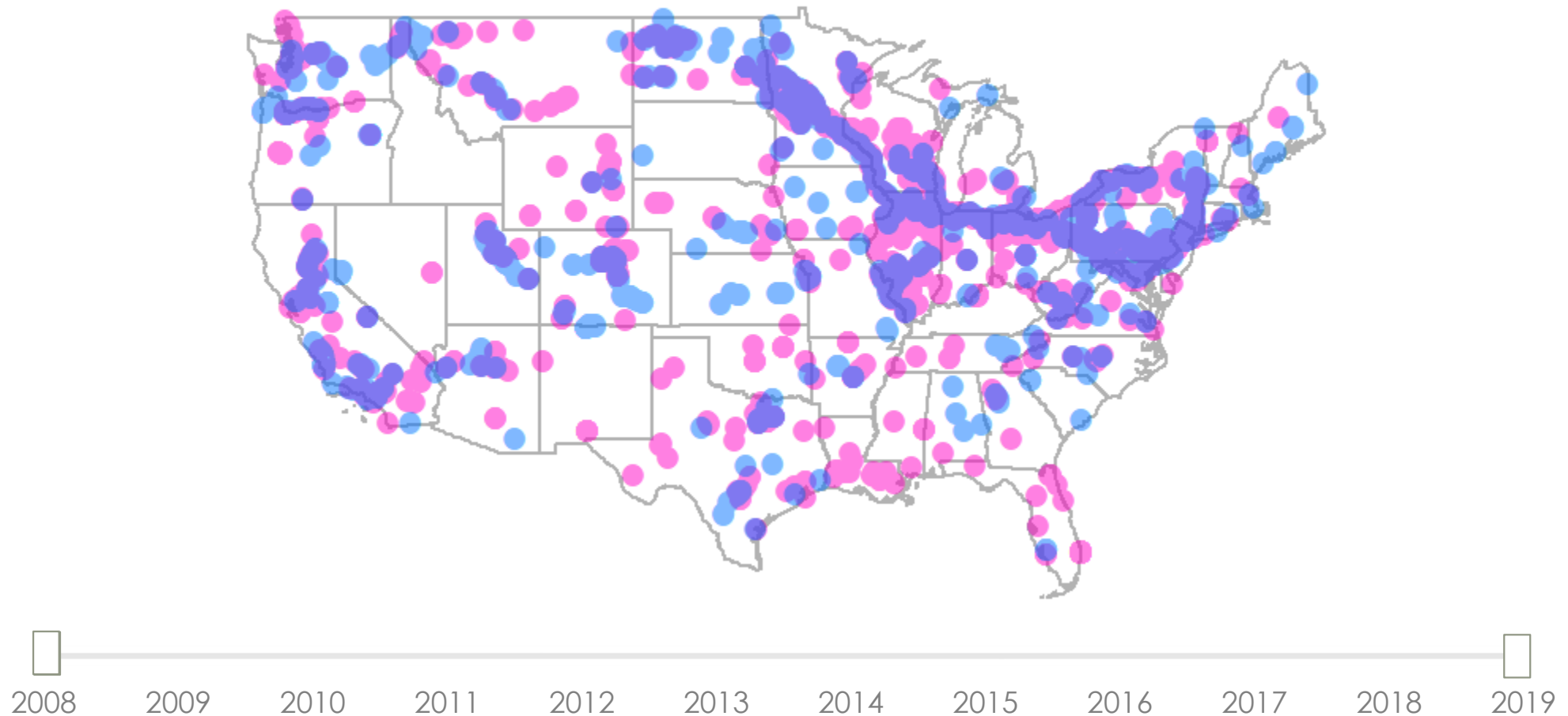
Risk Assessment: Likelihood

$$Risk = Likelihood \times Severity$$

Crude-by-Rail Map (WSJ)

Crude-by-Rail Terminals (EIA)

+

+

US Main Railroads

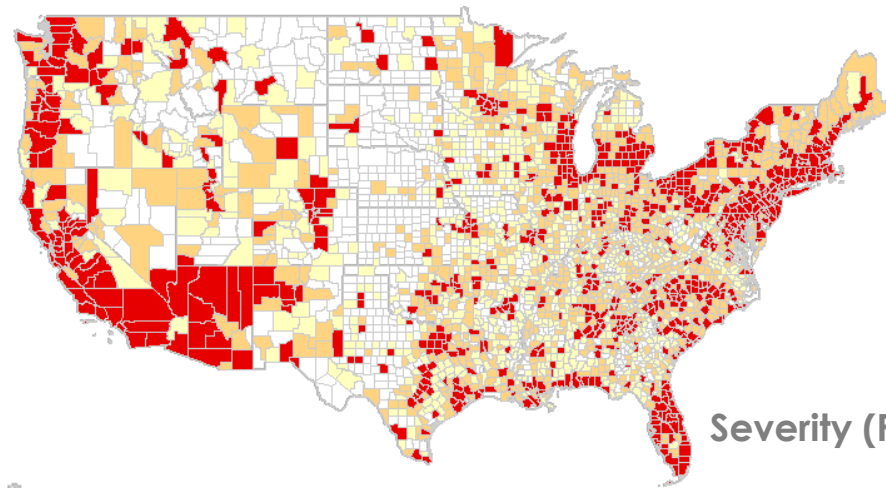Crowd-sourced Crude-by-Rail Locations

+

=

Likelihood

OAK RIDGE
National Laboratory
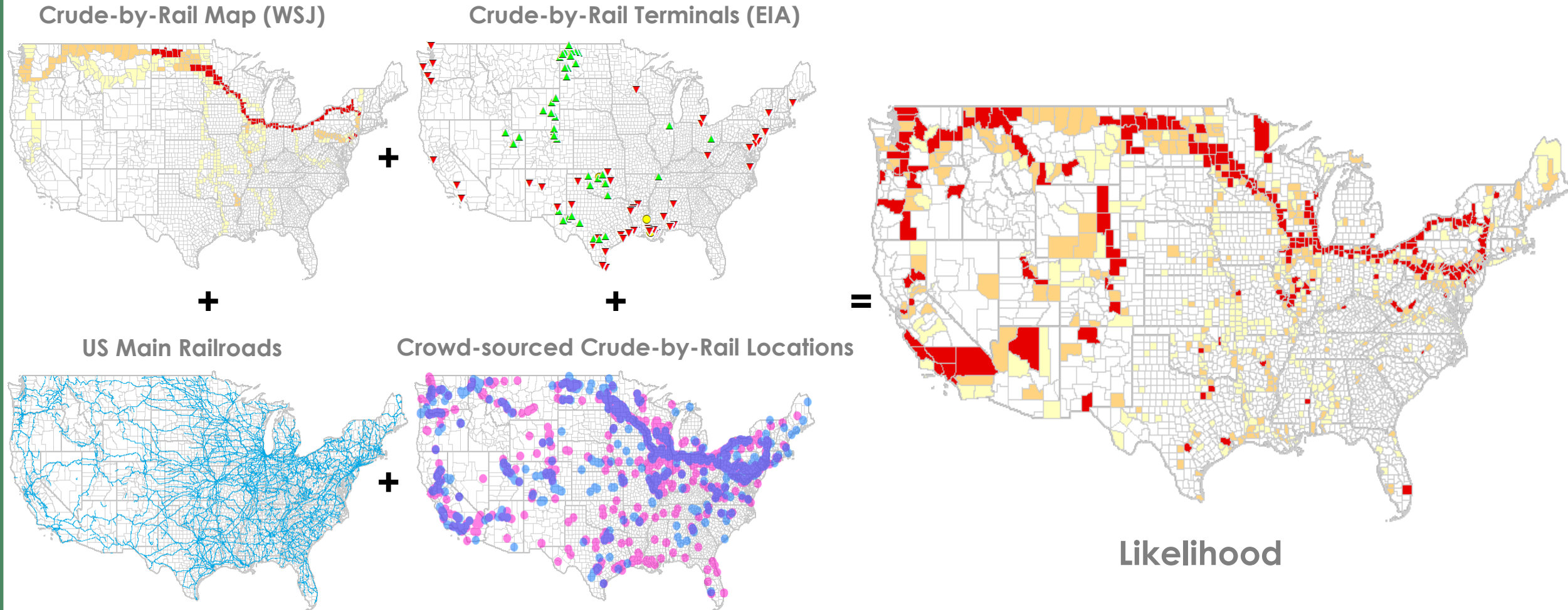
Risk Assessment (cont'd)

$$Risk = Likelihood \times Severity$$

Likelihood

X

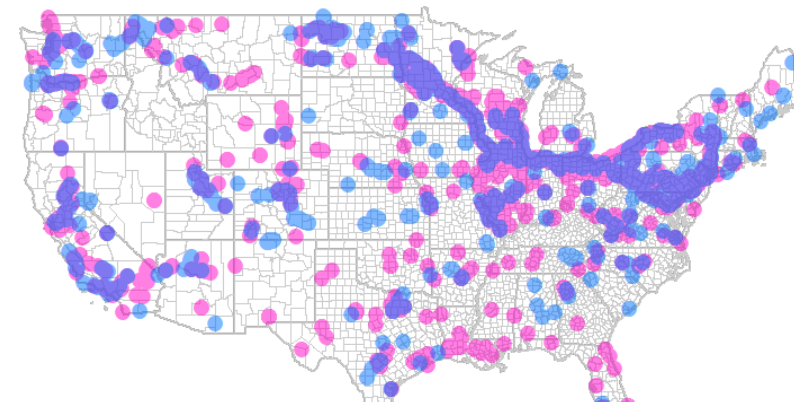**Severity (Population)**
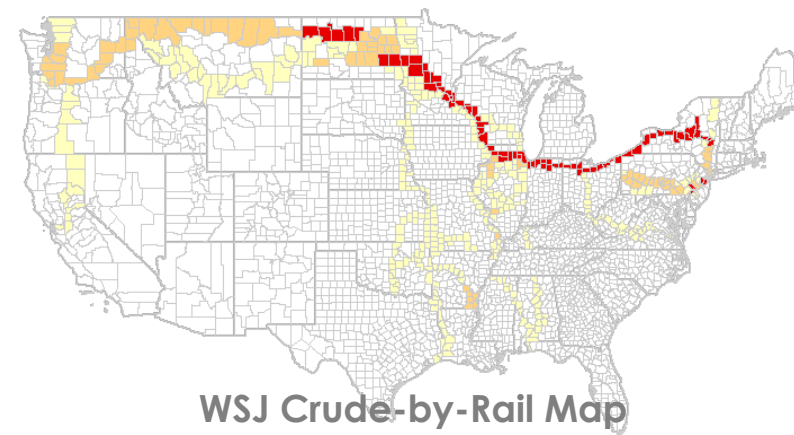
=

OAK RIDGE
National Laboratory
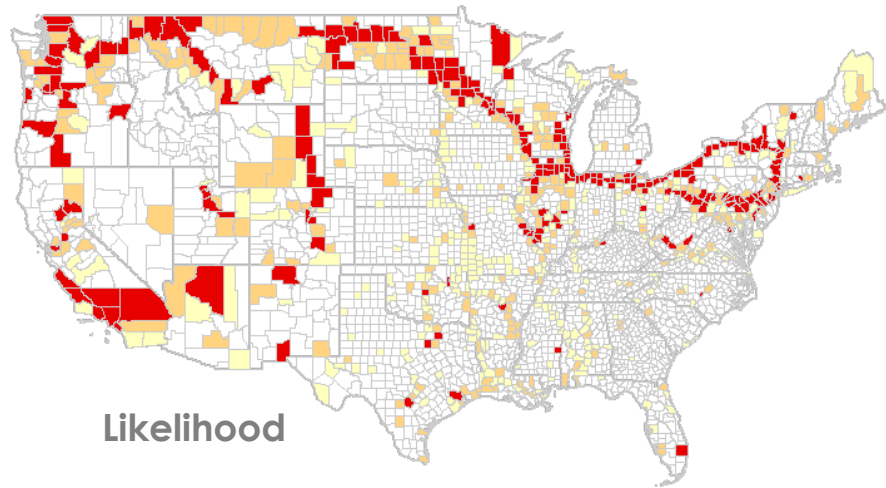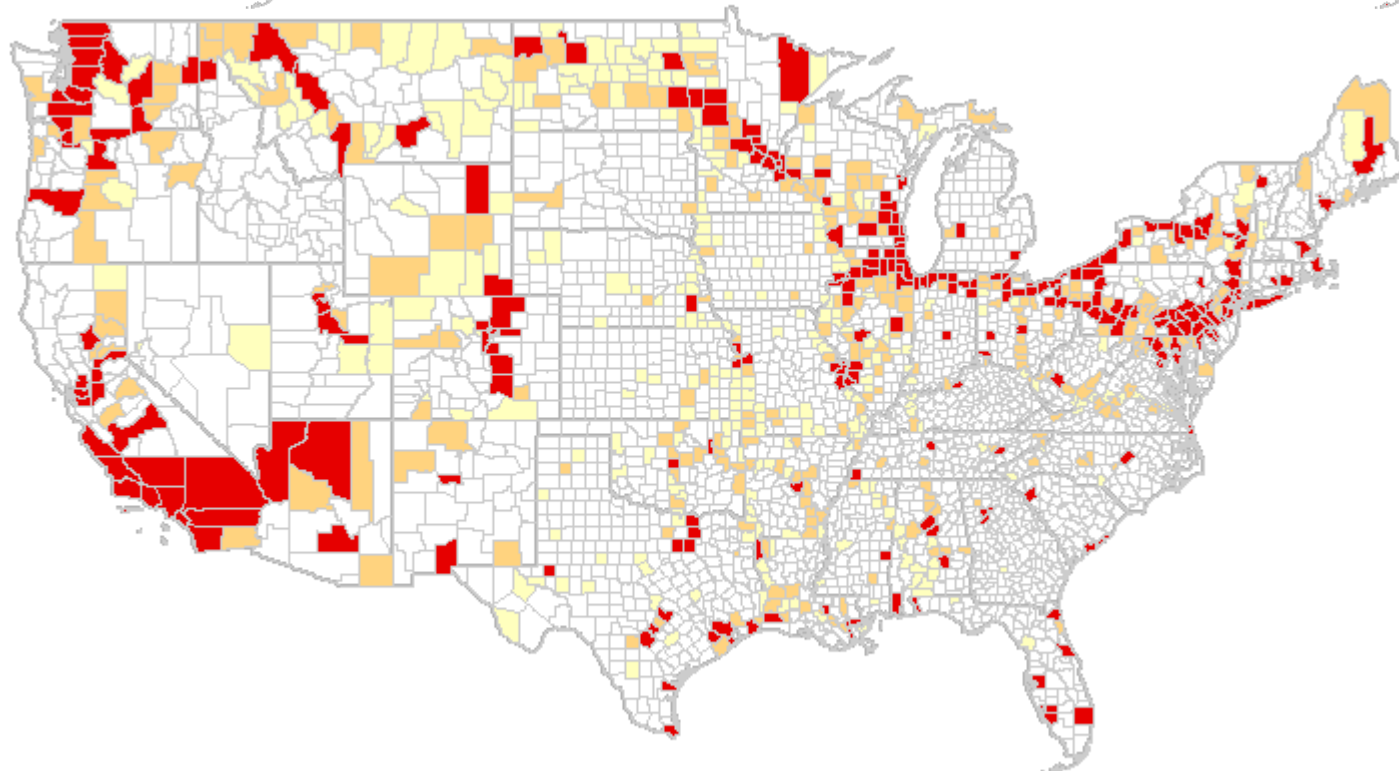
Discussions

- A self-sustaining, self-perpetuating web-scraping application

- Inferred crude-by-rail with no existing data

- Fully automated process was the focus

- Risk-based assessment for informed decision making and planning

Crowd-sourced Crude-by-Rail Locations

WSJ Crude-by-Rail Map

Crude-by-Rail Risk Map

OAK RIDGE
National Laboratory
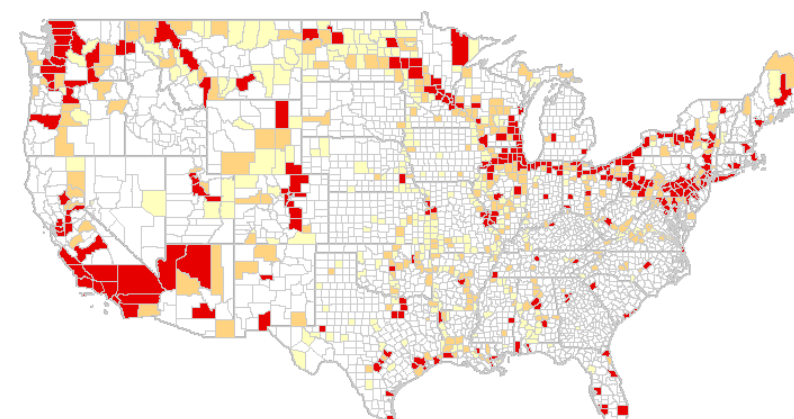
Moving Forward

- Develop automatic quality control process to verify crowdsourcing data collected by the web-scrapers

- Explore and integrate other data sources

- Finer geographic zones (e.g., census track, LandScan)

- Analyze crude-by-rail travel pattern and trends

- Formulate emergency/evacuation/communication plans

OAK RIDGE
National Laboratory

~ Thank you ~

Chieh (Ross) Wang

Energy and Transportation Science Division

Oak Ridge National Laboratory

cwang@ornl.gov

865-341-1343