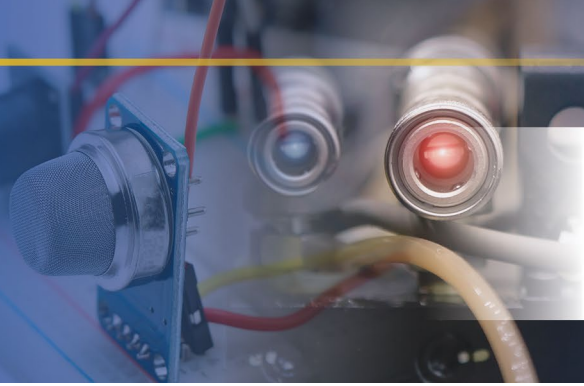
UPISC UNIVERSITY OF PITTSBURGH INFRASTRUCTURE SENSING

COLLABORATION WORKSHOP



NATIONAL ENERGY TECHNOLOGY LABORATORY

Geo-Analytics and Machine Learning for Infrastructure

Dr. Kelly Rose

Technical Director, Science-based Ai/Ml Institute (SAMI), NETL

UPISC UNIVERSITY OF PITTSBURGH INFRASTRUCTURE SENSING

COLLABORATION WORKSHOP

NATIONAL ENERGY TECHNOLOGY LABORATORY

Dr. Kelly Rose

Technical Director, Science-based Ai/Ml Institute (SAMI), NETL

Kelly Rose, PhD, is a geo-data scientist with over 20 years of service and research experience at the U.S. Department of Energy's National Energy Technology Laboratory (NETL). She is also the Technical Director for NETL's Science-based Ai/Ml Institute (SAMI). Her research focuses on developing novel science-based, data-driven methods and models for addressing energy and environmental challenges, including NETL's award-winning Energy Data eXchange® (EDX) ecosystem. Rose leads collaborative teams to deliver impactful computational data science resources and models in reusable, scalable, and reproducible formats.

Her work has been applied to many scientific and societal domains including Earth science, geoinformatics, research data management and virtualization, climate and metocean, oil spill prevention, mineral and groundwater resources, geohazards, social and environmental justice, materials innovation, infrastructure resiliency, smart cities, and smart systems. She is coauthor on more than 100 public datasets, models, tools, journal publications, and technical studies. Rose has also mentored more than fifty STEM research interns and fellows and supports additional STEM outreach activities. She holds degrees from Denison University (B.S), Virginia Tech (M.S.), and Oregon State University (PhD).

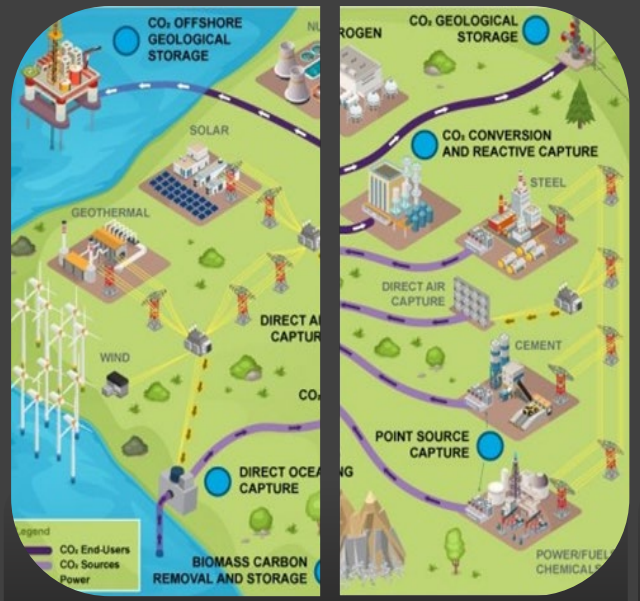
Geo-data Science and Machine Learning for Infrastructure

Innovating science-based geo-computational solutions for stakeholder needs

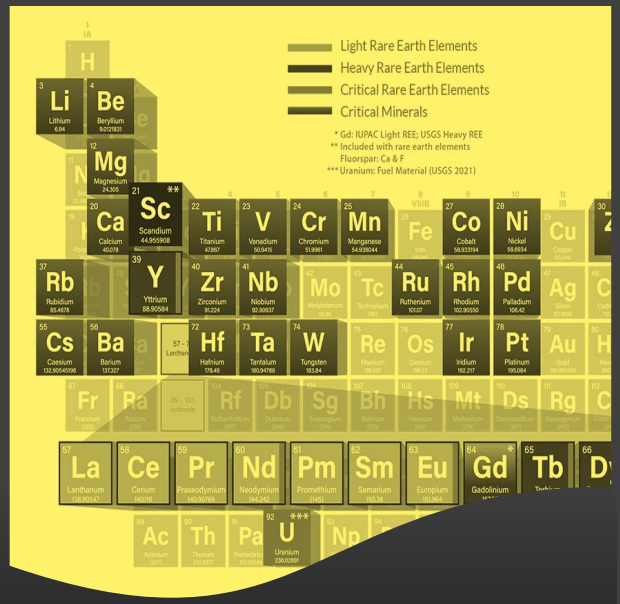
Kelly Rose, Technical Director, NETL's Science-based Ai/Ml Institute (SAMI)

NATIONAL ENERGY TECHNOLOGY LABORATORY

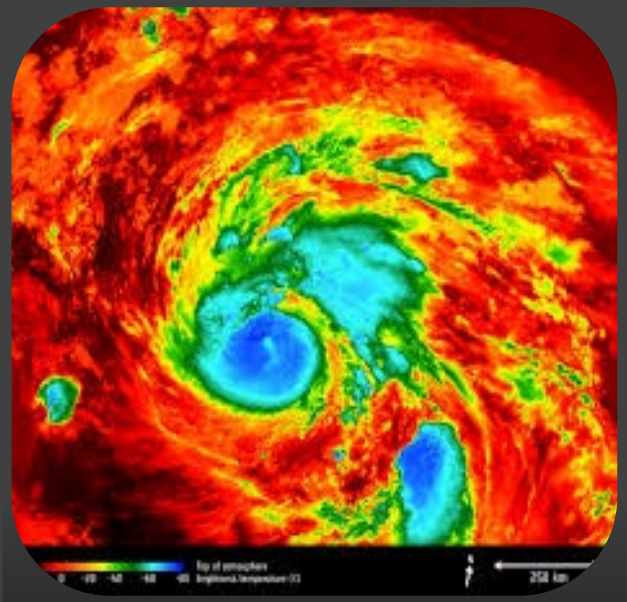
SAMI SCIENCE-BASED AI/ML INSTITUTE



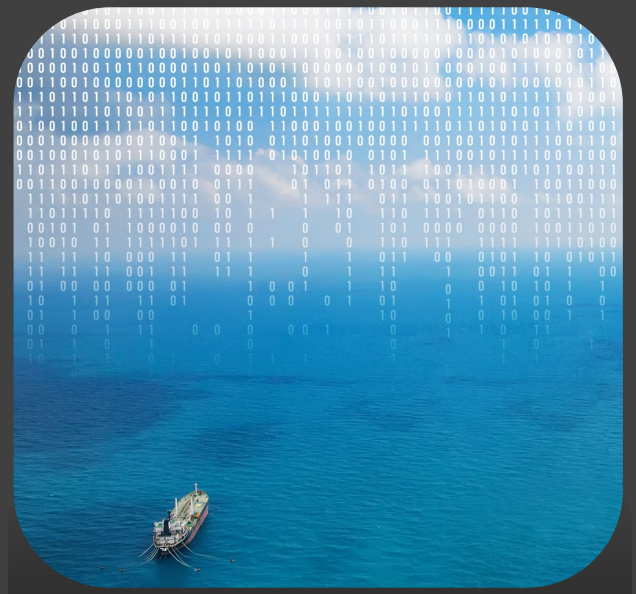
CARBON MANAGEMENT



CRITICAL MINERALS AND MATERIALS



INFRASTRUCTURE & ENVIRONMENTAL RESILIENCY



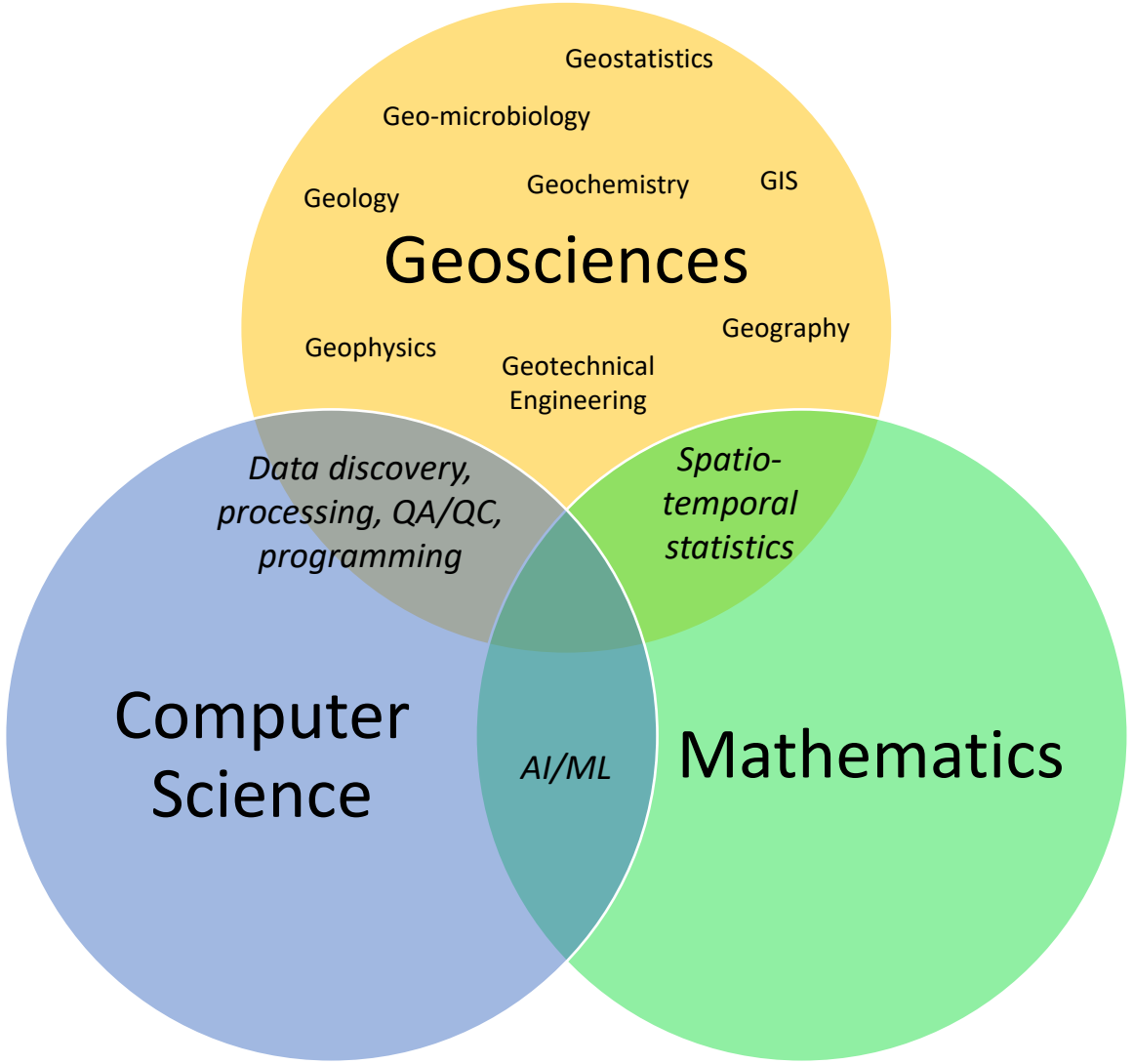
FOUNDATIONAL DATA AND A.I. SOLUTIONS

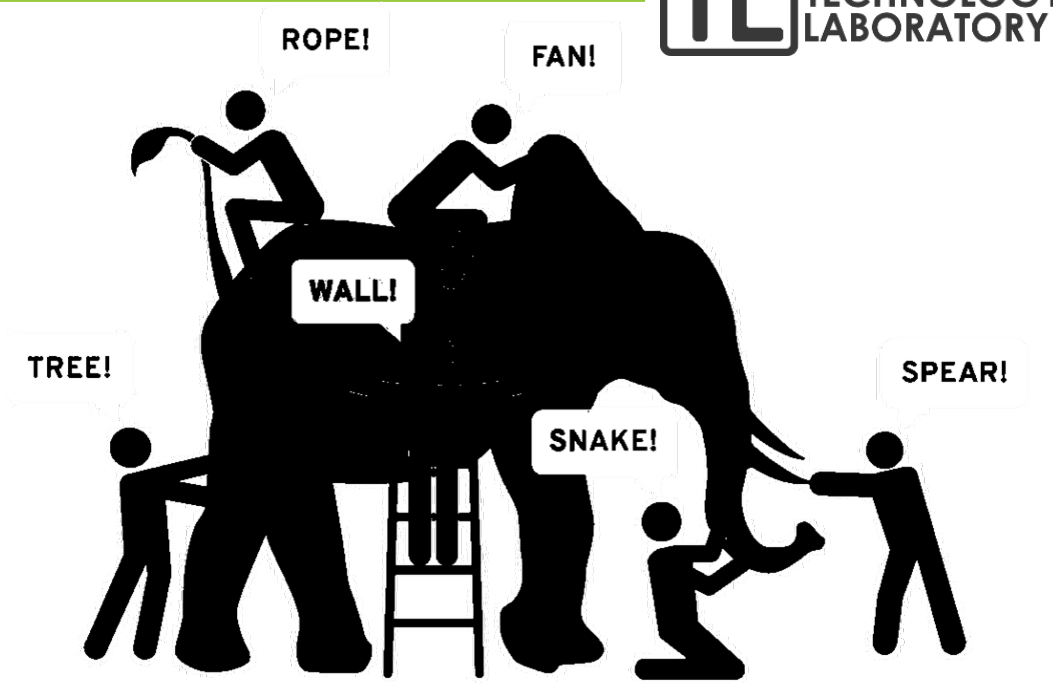
August 25, 2022

https://edx.netl.doe.gov/sami

U.S. DEPARTMENT OF ENERGY

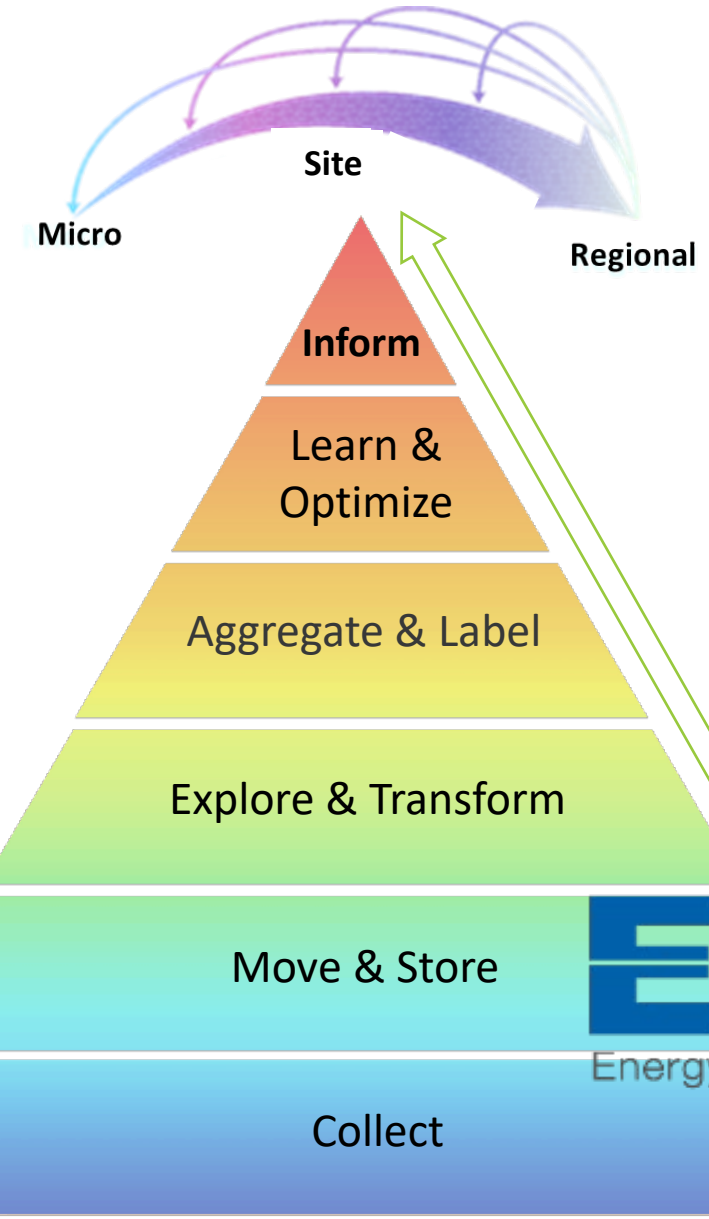
What is geo-data science?





NETL's Geo-Data Science Research

NATIONAL ENERGY TECHNOLOGY LABORATORY



Multi-scale, multi-disciplinary R&D that blends data science, advanced computing, and science-based methods to develop data, tools, models & innovate solutions for energy, human & environmental systems

EDX® Energy Data exchange

United States Department of Energy Secretary's Honor Awards 2021

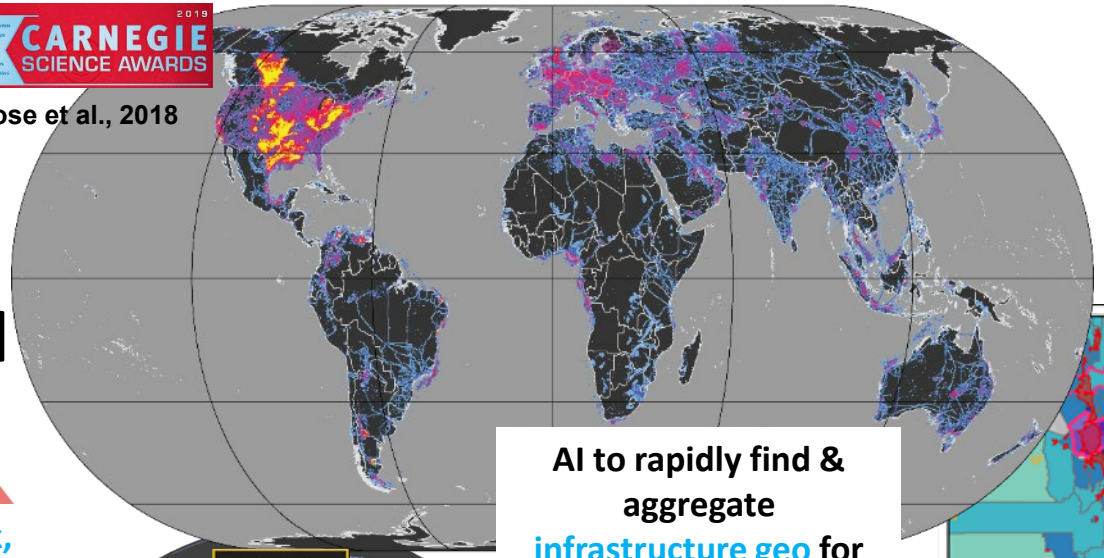
2019 CARNEGIE SCIENCE AWARDS

2022 TechConnect National Innovation Awardee

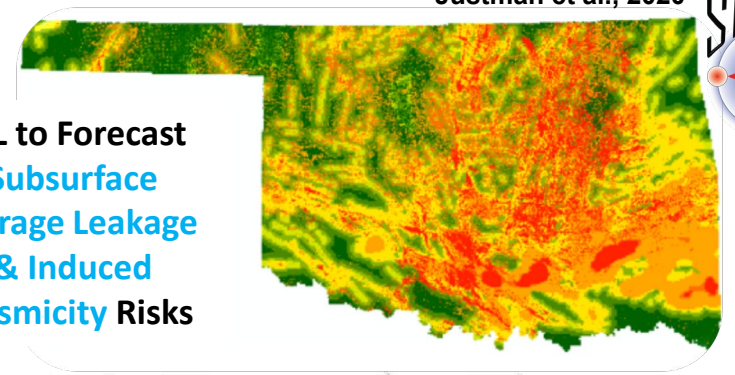
2019 R&D 100 WINNER

Addressing energy, societal & environmental needs

Rose et al., 2018

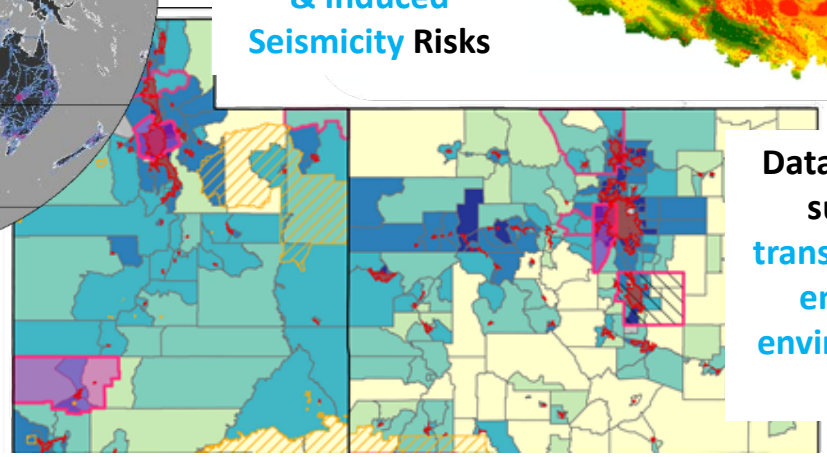


Justman et al., 2020



ML to Forecast Subsurface Storage Leakage & Induced Seismicity Risks

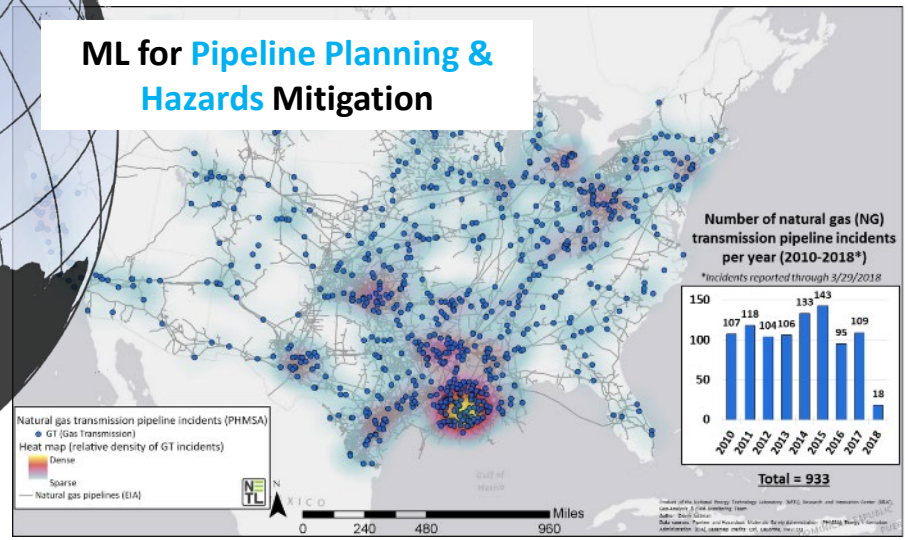
AI to rapidly find & aggregate infrastructure geo for CH4 emissions mitigation and orphan well predictions



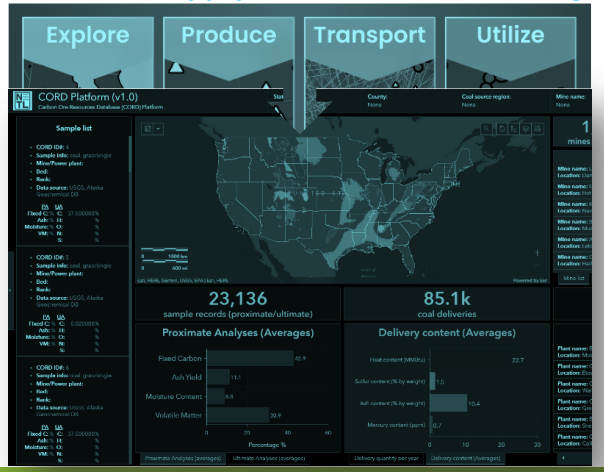
Data science & ML to support energy transitions and inform energy, social & environmental justice activities

AI & geo-data science for risk, resource & resiliency analyses

ML for Pipeline Planning & Hazards Mitigation



Big data & digitalization to support REE-CM supply chain & manufacturing



Justman, et al. (2022). **A database and framework for carbon ore resources and associated supply chain data.** *Data in Brief.* https://doi.org/10.1016/j.dib.2021.107761

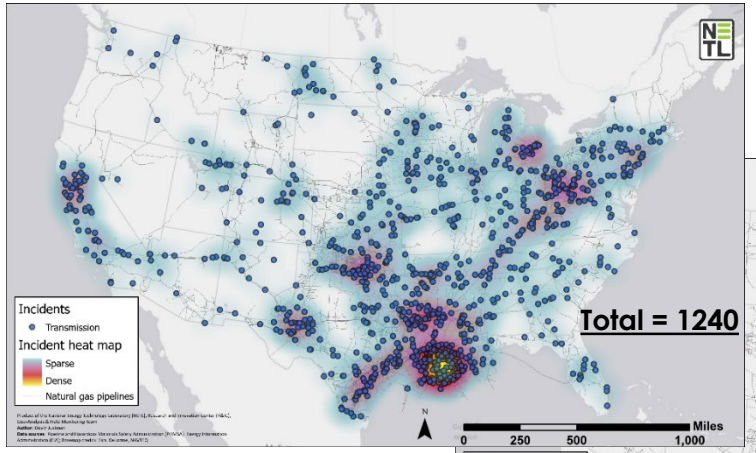
Geospatial Optimization for Pipeline Sensor Deployment

Identifying spatio-temporal trends associated with internal & external risks

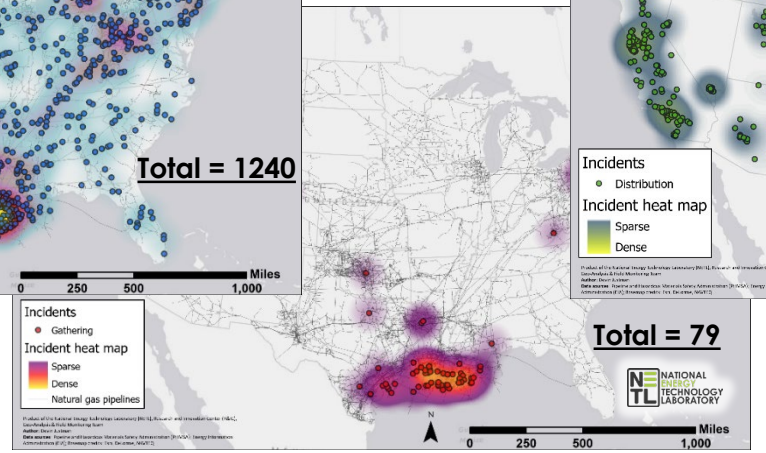
*What is the **need**?*

- **Predict** & **prevent**
  - 200 incidents/year
  - >2.5 million miles
- **~1/12,500 chance** of picking the right mile of pipeline to invest resources
- Susceptible to a **wide range of risks**
- Need to **optimize sensor placement**

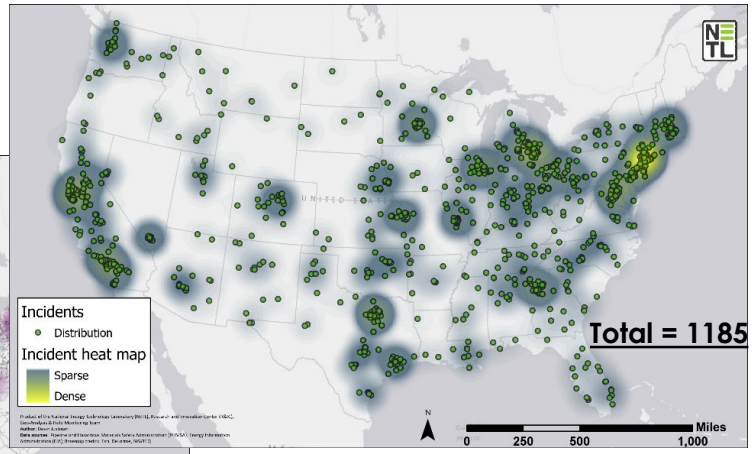
Transmission



Gathering

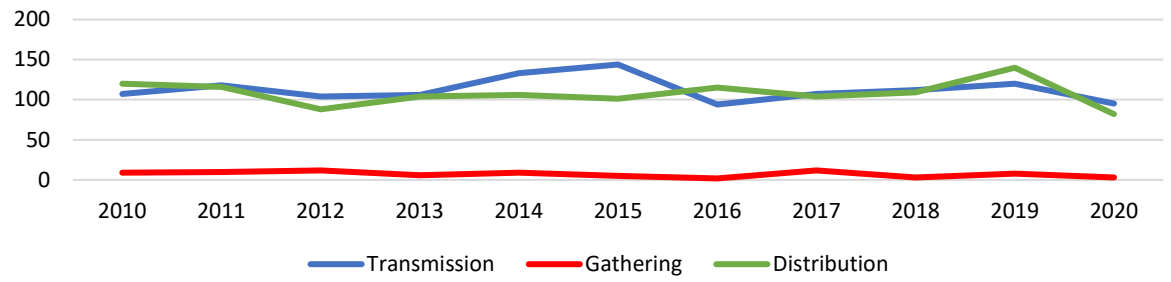


Distribution



Sources: Pipeline and Hazardous Materials Safety Administration (PHMSA) & US Energy Information Administration (EIA)

# of incidents by system per year

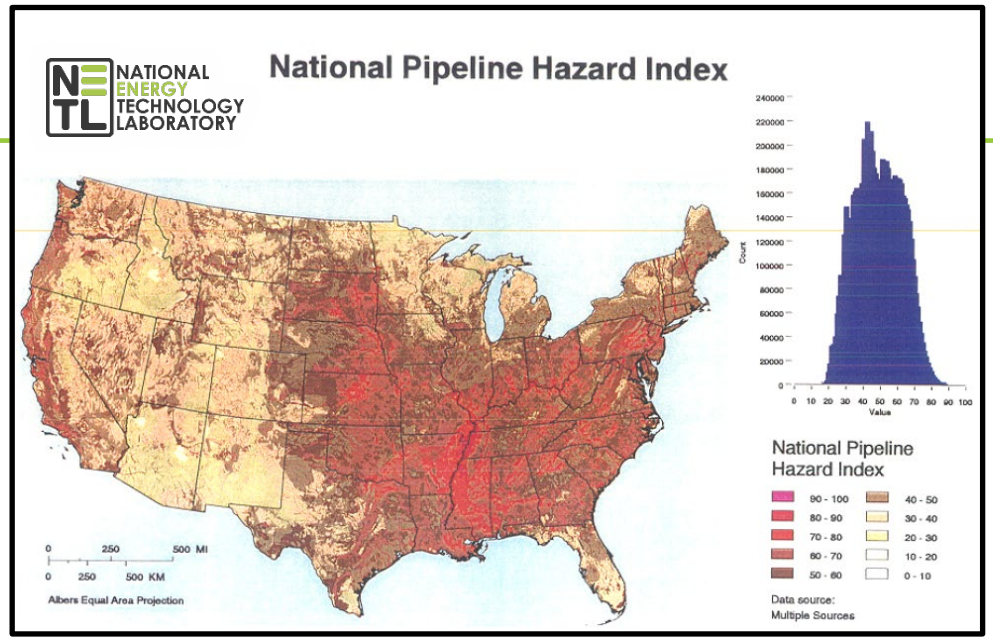


Approach & Benefits

Explored a **variety of geospatial analytics** to assess **pipeline integrity** & **failure risk**

Values Delivered

- **Identified priority areas** for **monitoring**, **maintenance**, and **improvements**
- **Optimized placement** of **advanced sensing & monitoring tools**

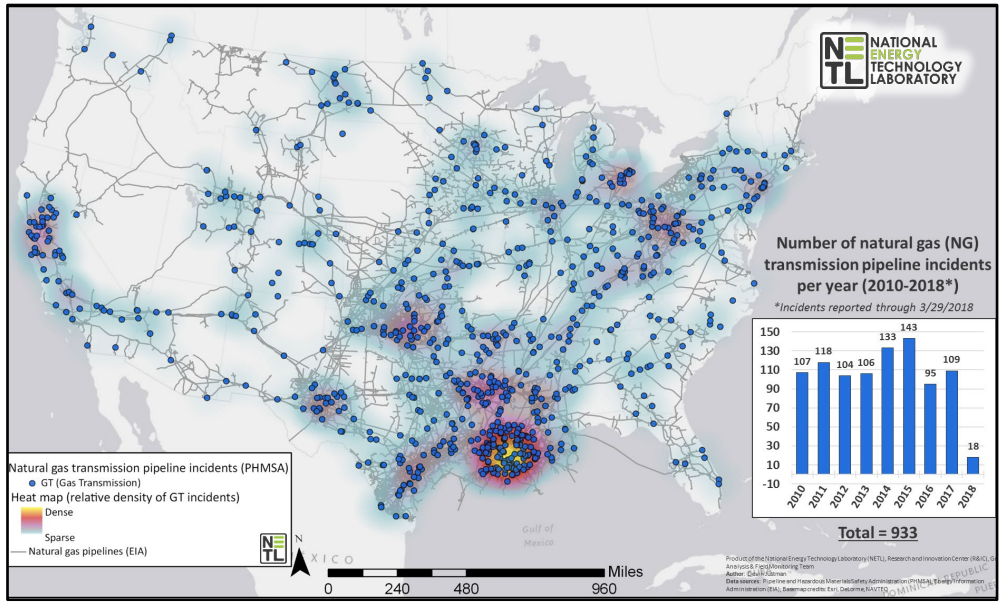


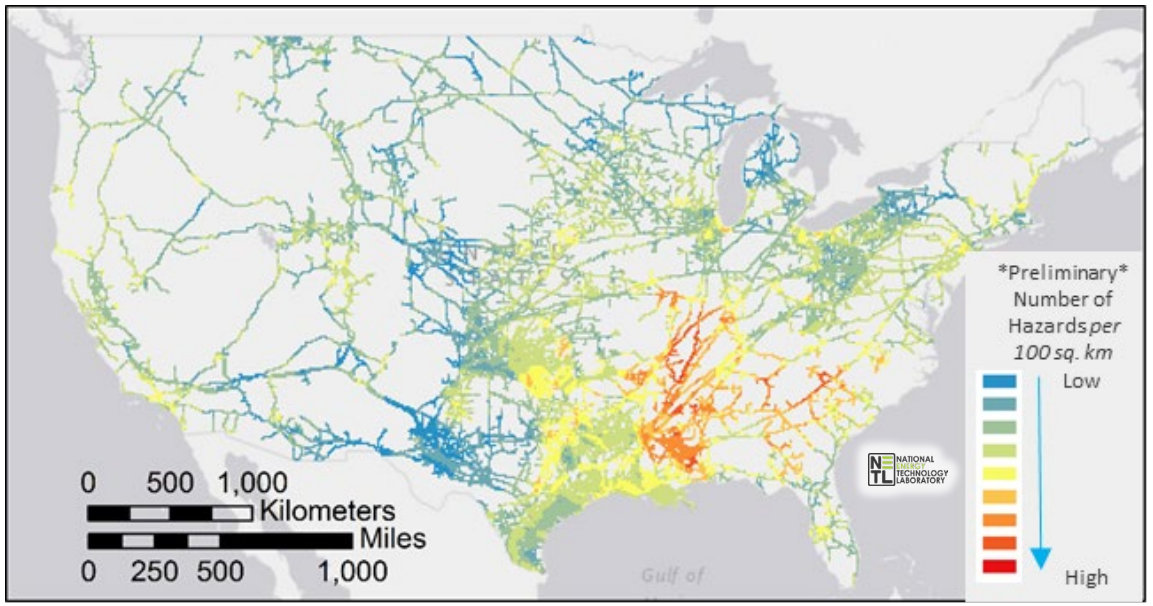
NATIONAL ENERGY TECHNOLOGY LABORATORY

Updated risk factors based on FEMA and DOT 1996 study

National Pipeline Risk Index based on natural disasters

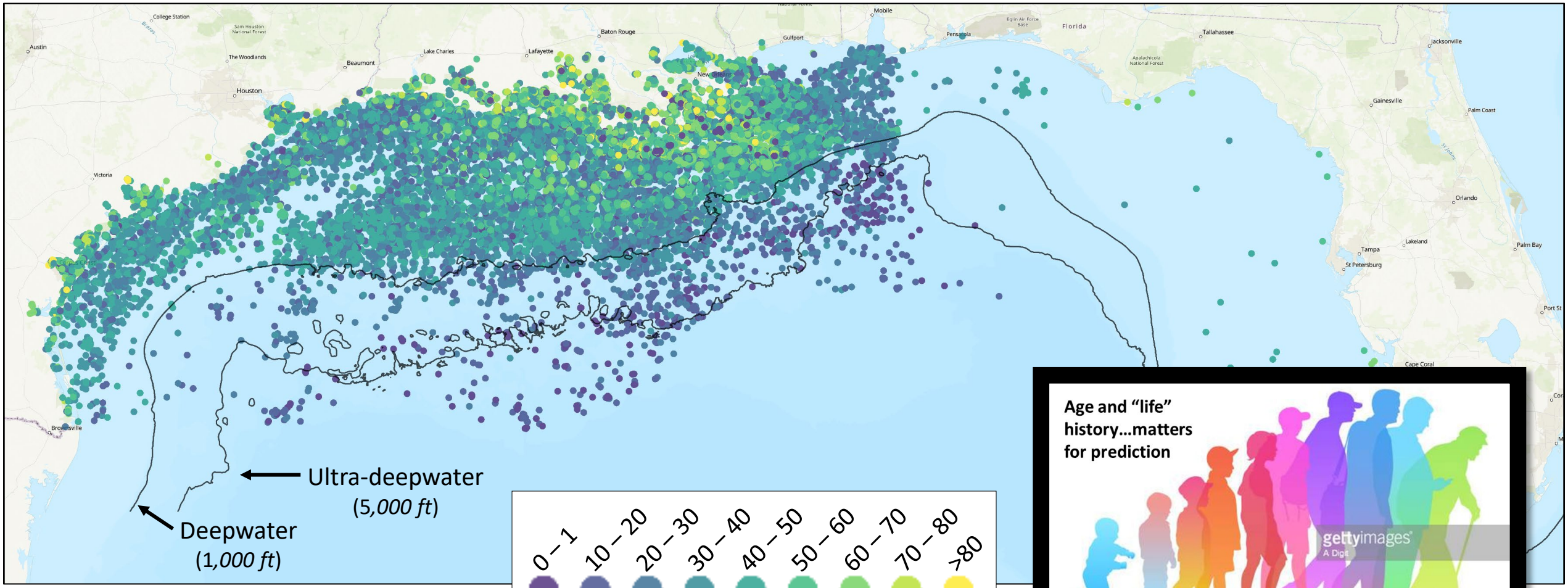
Opportunities for integration of sensing data, streaming data, and use for pipelines beyond NGI



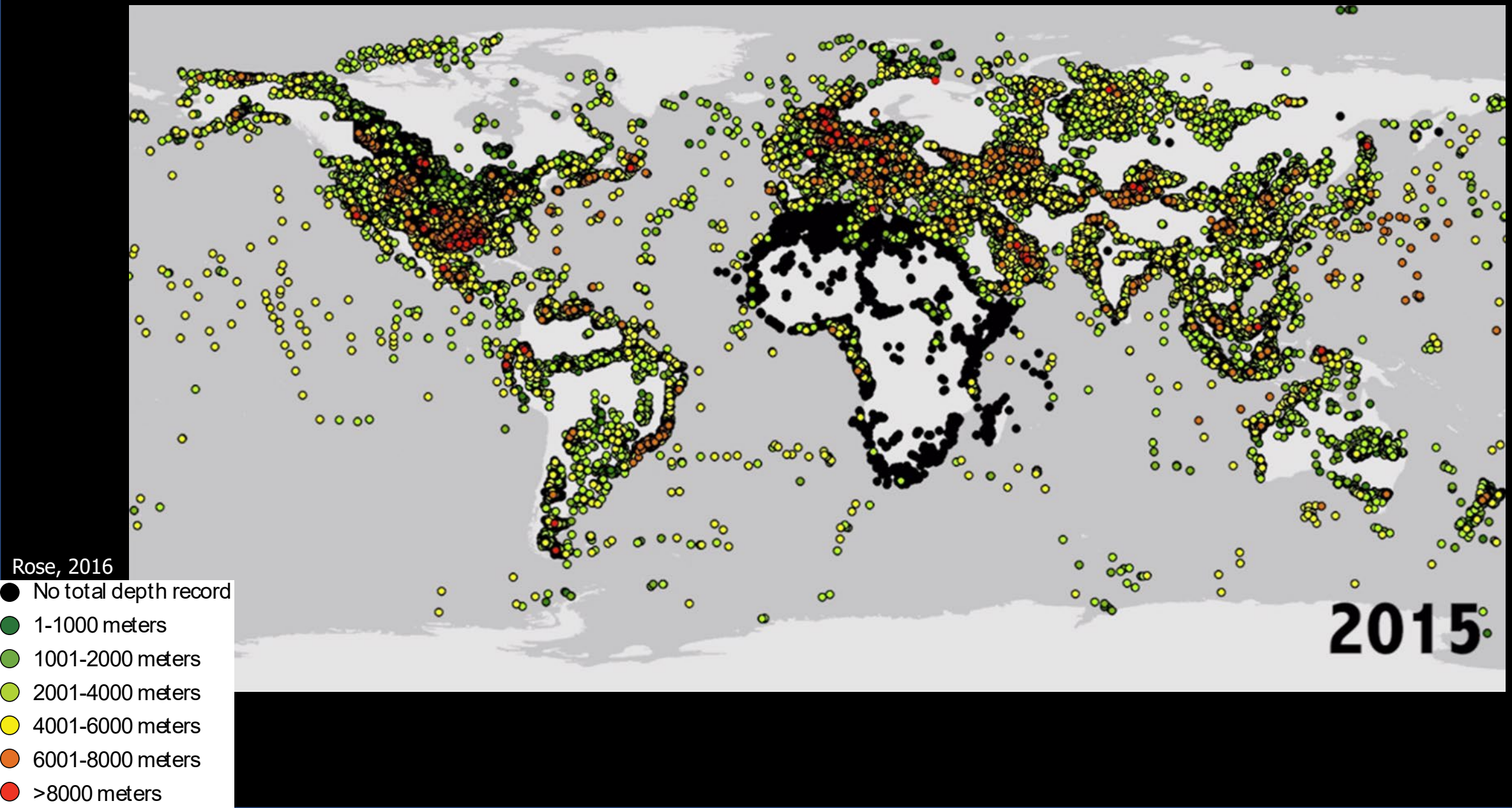


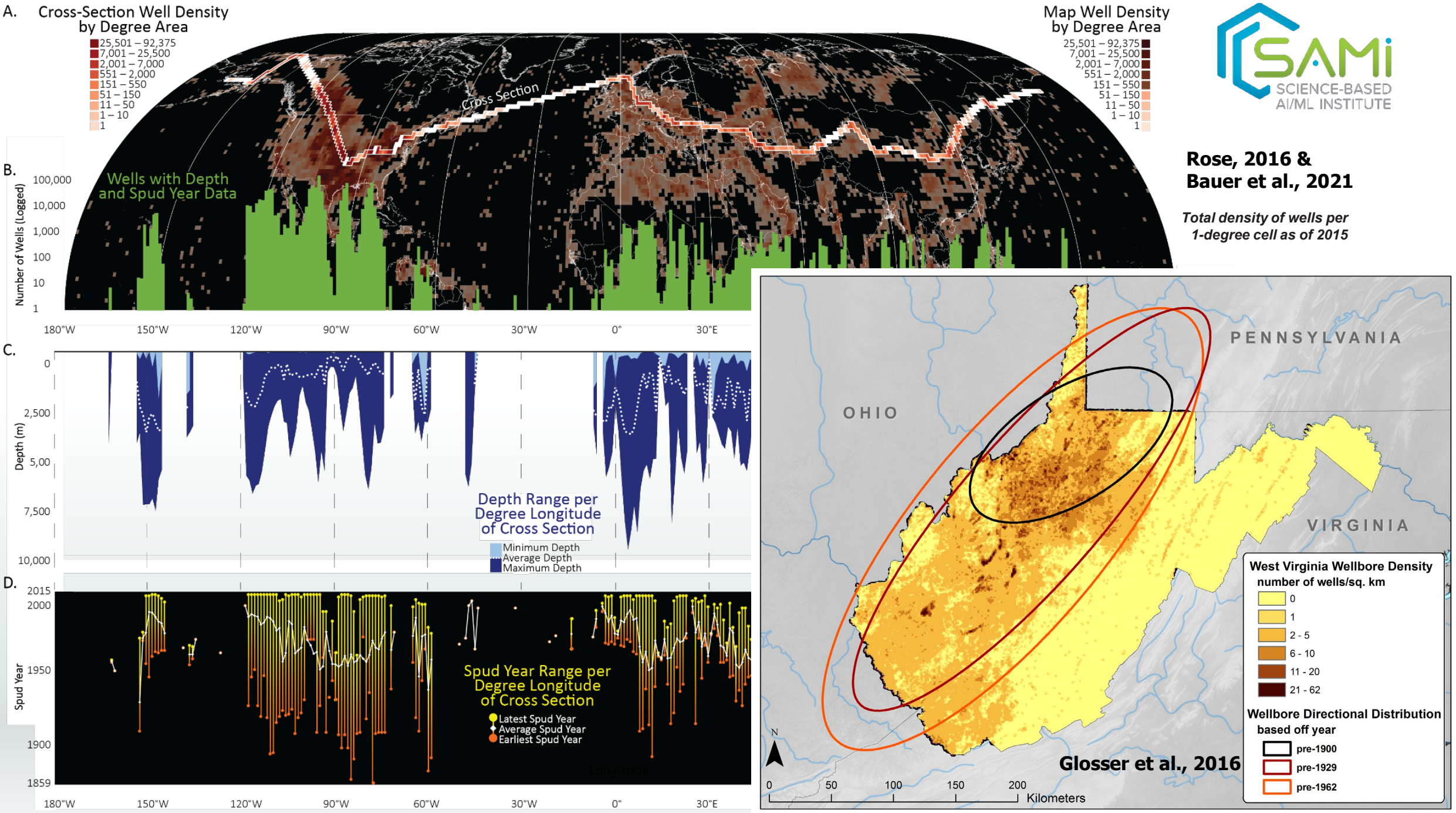
Assessing **Well Integrity – ex.** Gulf of Mexico

Ages of wells comparing today (August 1, 2022) to spud date



Data Aggregation & Analytics - 150 years of Global Deep Wells




A.
Cross-Section Well Density by Degree Area
25,501 – 92,375
7,001 – 25,500
2,001 – 7,000
551 – 2,000
151 – 550
51 – 150
11 – 50
1 – 10
1
Cross Section
Map Well Density by Degree Area
25,501 – 92,375
7,001 – 25,500
2,001 – 7,000
551 – 2,000
151 – 550
51 – 150
11 – 50
1 – 10
1
B.
Wells with Depth and Spud Year Data
Number of Wells (Logged)
100,000
10,000
1,000
100
10
1
180°W
150°W
120°W
90°W
60°W
30°W
0°
30°E
C.
Depth (m)
0
2,500
5,00
7,500
10,000
Depth Range per Degree Longitude of Cross Section
Minimum Depth
Average Depth
Maximum Depth
D.
Spud Year
2015
2000
1950
1900
1859
Spud Year Range per Degree Longitude of Cross Section
Latest Spud Year
Average Spud Year
Earliest Spud Year
Longitude
180°W
150°W
120°W
90°W
60°W
30°W
0°
30°E
PENNSYLVANIA
OHIO
VIRGINIA
West Virginia Wellbore Density
number of wells/sq. km
0
1
2 - 5
6 - 10
11 - 20
21 - 62
Wellbore Directional Distribution
based off year
pre-1900
pre-1929
pre-1962
N
0
50
100
150
200
Kilometers
Glosser et al., 2016
SAMi
SCIENCE-BASED AI/ML INSTITUTE
Rose, 2016 &
Bauer et al., 2021
Total density of wells per 1-degree cell as of 2015

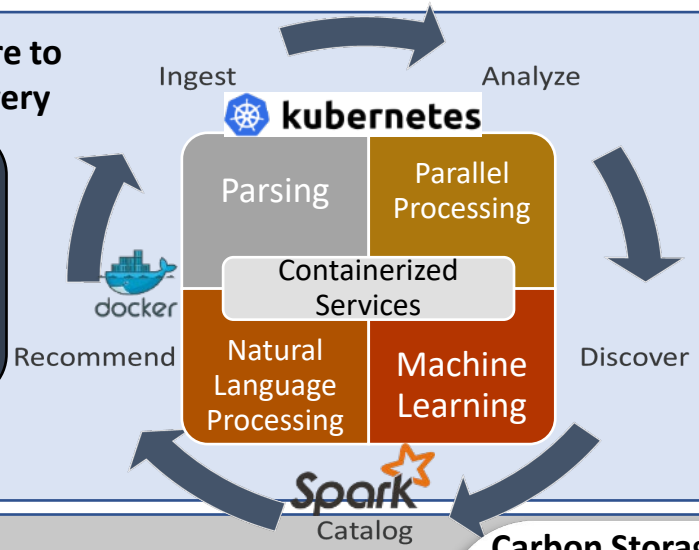
Digitalization, data management, & AI-informed data discovery

AI informed approach

Challenge: data infrastructure to AI/ML enhanced data discovery

Employing AI/ML tools to find relevant data resources

NETL SmartSearch



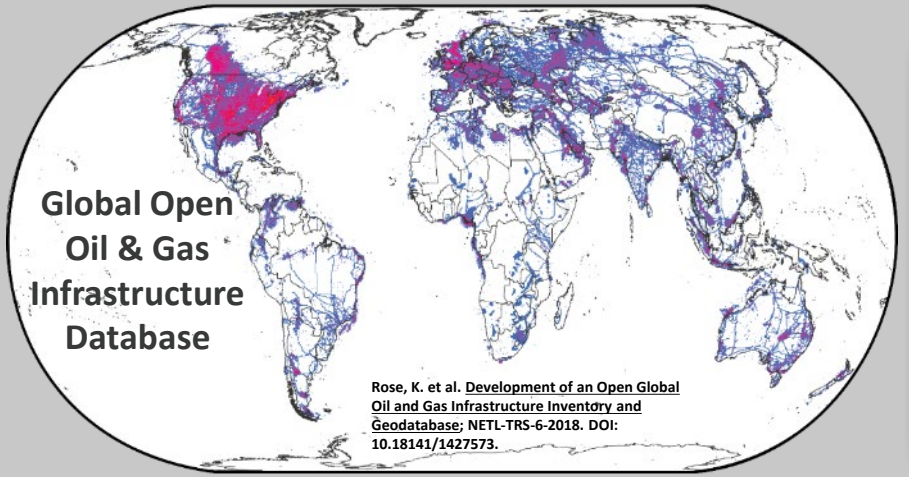
SmartSearch leverages ML+NLP to:

1) **Analyzing** content you like
2) **Finding** new content via www, local, enterprise data stores
3) Telling you **how relevant** the new data is to what you like

Opportunity:

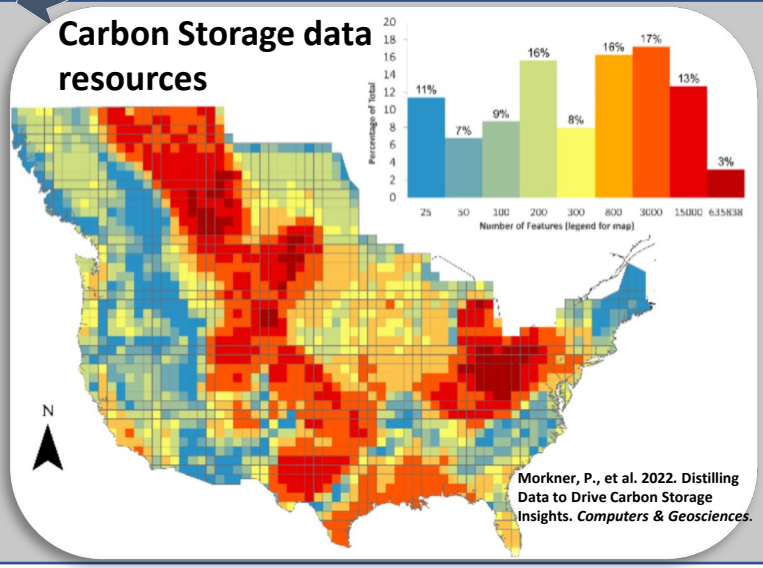
Infinitely scalable to return text, graphical, tabular, image, html, spatial, etc Driving next-generation of geo-data science R&D

Example applications to date



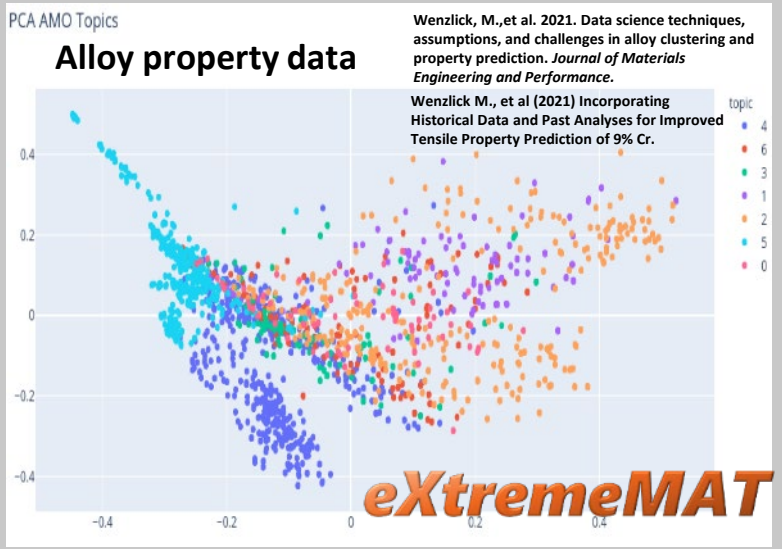
Rose, K. et al. Development of an Open Global Oil and Gas Infrastructure Inventory and Geodatabase; NETL-TRS-6-2018. DOI: 10.18141/1427573.

Carbon Storage data resources



Morkner, P., et al. 2022. Distilling Data to Drive Carbon Storage Insights. *Computers & Geosciences*.

Alloy property data



AIIM: Advanced Infrastructure Integrity Modeling

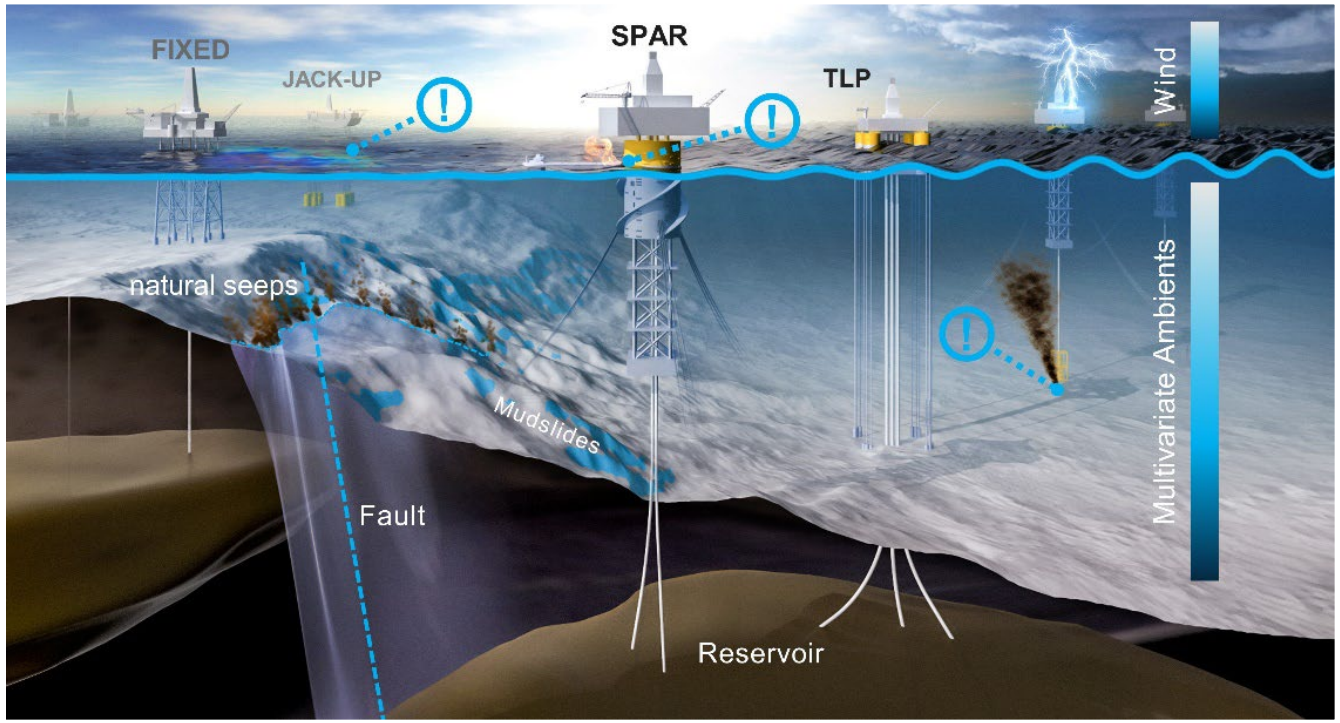
A big data, multiple AI/ML and advanced modeling framework

Assess the **current state** of **FECM infrastructure** & **risks** associated with the **extended use** or **repurposing** of **wells**, **platforms**, & **pipelines**

- Offshore operations can be hazardous
  - Environment
  - Changing climate
  - Operational wear-and-tear
- Infrastructure is aging

Dyer et al., 2021

- Published method & applications of the **AIIM framework** as applied to **platforms**
- Currently expanding **infrastructure assessments** & **analytical techniques**

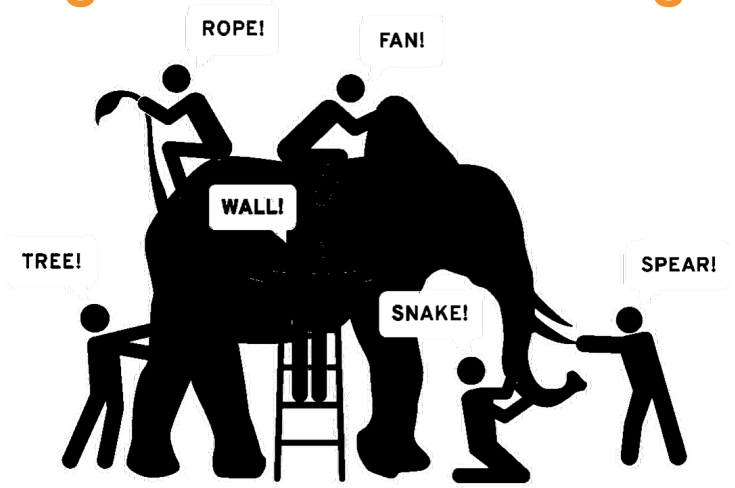


Values Delivered

- Identify infrastructure assets and liabilities
- Provide critical insights for safe reuse/repurposing strategies
- Inform operational and environmental risk prevention

Using the **Whole** to **Inform** the **Local**

Accounting for a fuller natural-engineered system

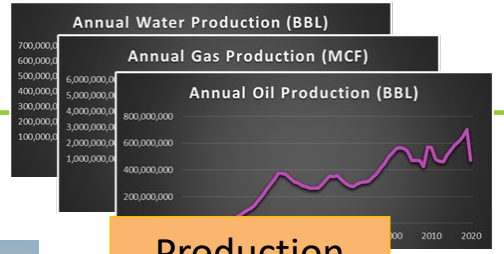


Multiple AI/ML Models **(Dyer et al., 2022)**

- Gradient Boosted Decision Trees (2 models)
- Artificial Neural Network (2 models)
- Bayesian Network

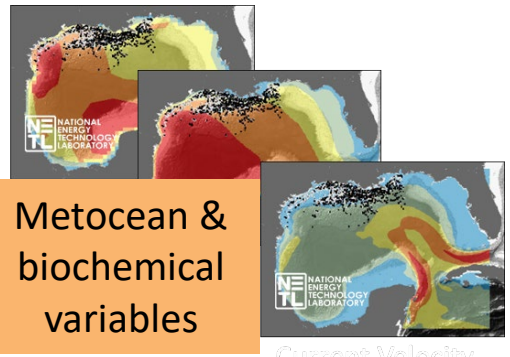
Advanced Analytics

- Geographically Weighted Regression **(Nelson et al., 2021)**
- Causality / Time series Analytics



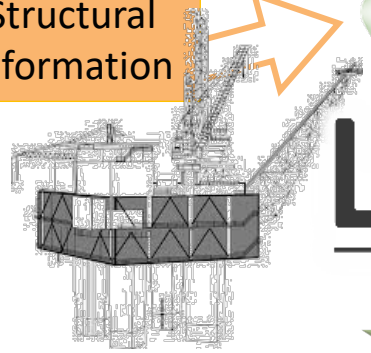
Production information

70+ years of incident reports



Metocean & biochemical variables

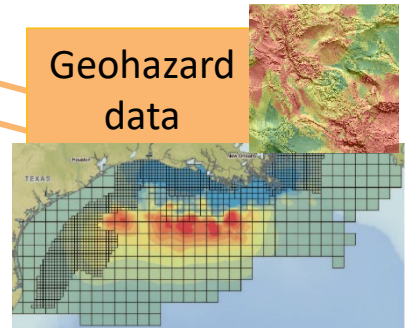
Structural information



AI/ML & Advanced Modeling

WATT The Power of AI

Geohazard data



AIIM

***Some* key findings:**

Multiple ML models capable of predicting **removal age <u>*<3 years*</u>**

Identified **significant connections** among **biochemical** variables & **incidents**

Corroborated results with **news** & **reports**

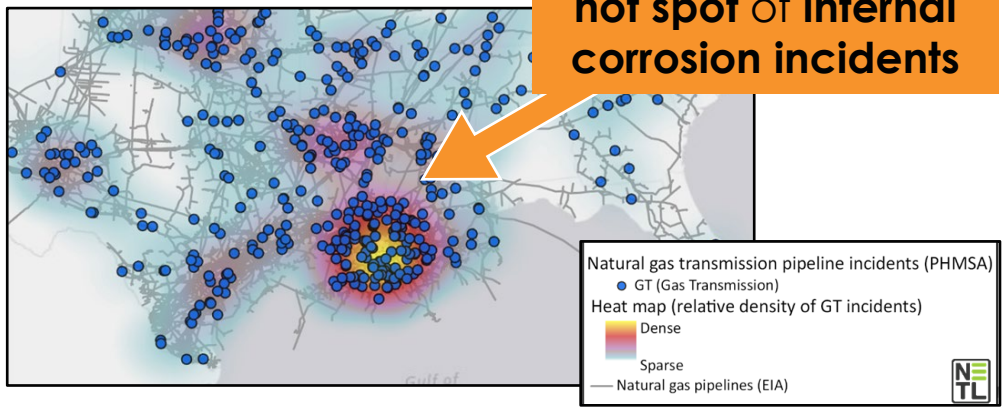
Expanding to **Pipeline** & **Well** Infrastructure

NATIONAL ENERGY TECHNOLOGY LABORATORY

Utilizing past research for today's insights & analytics

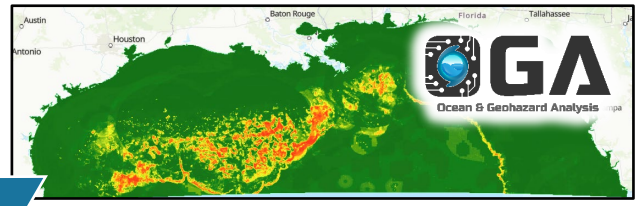
- Leveraging data & insights from **national pipeline R&D sensor placement**
- Compiled & cleaned **>30 years reported pipeline incidents**
- Processing data for **>110k wells** in federal & state waters

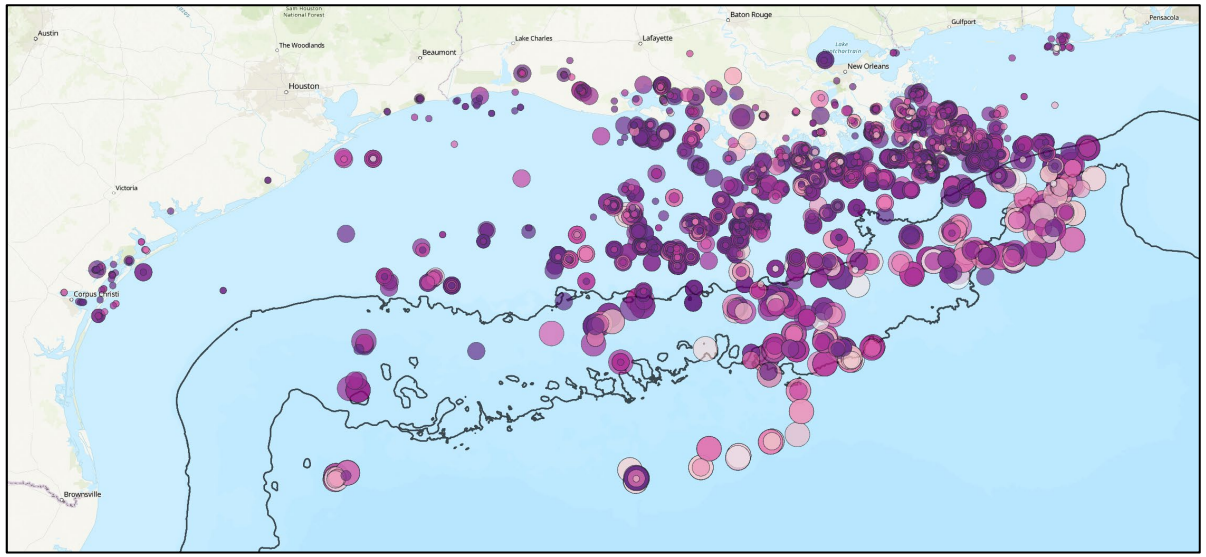
Previously identified **hot spot** of **internal corrosion incidents**



Adaptations to AIIM

- Extracting **at-depth** metocean variables
- Integrating **geohazard risk assessments**
- Evaluating **production life** and **status** of wells
  - Production timelines
  - Stats per field or area of interest
  - Amount produced or what was expected at end of life





Percent of Oil Produced at Active Wells

0 — 100

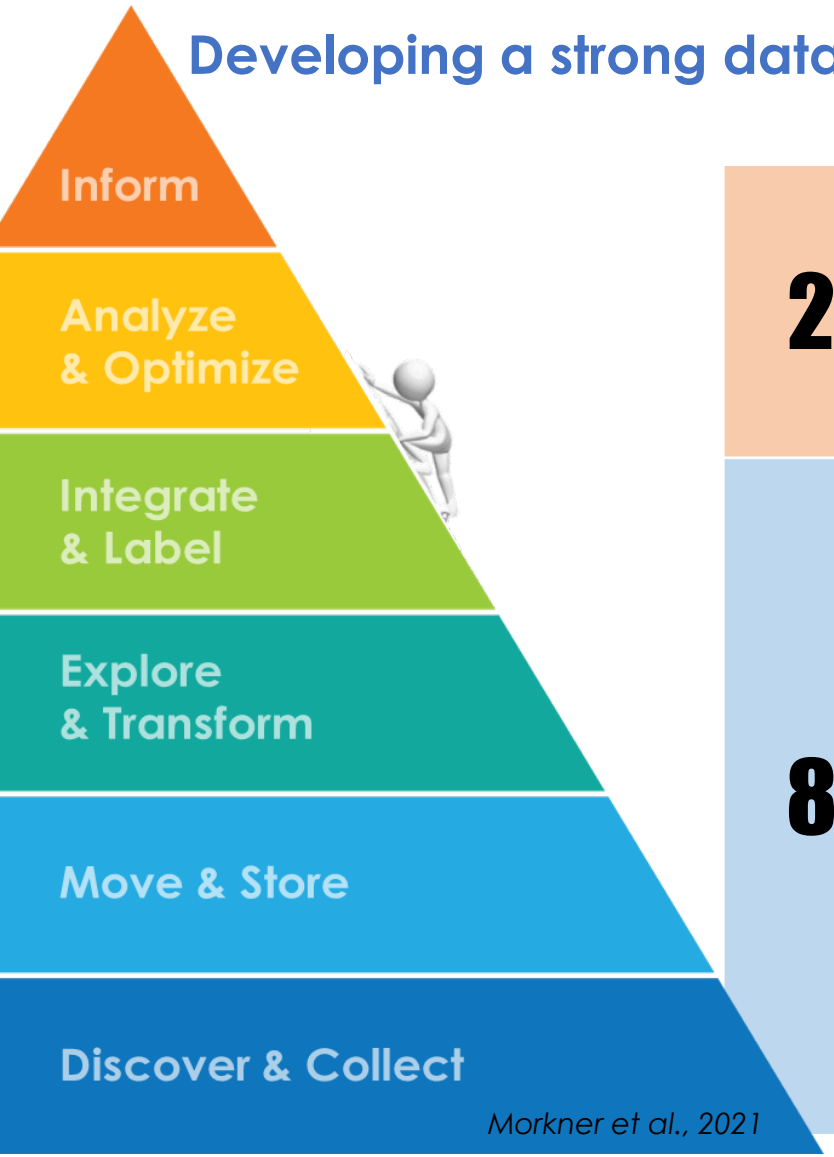
Estimated Ultimate Recovery (bbls)

<500,000 — >10,000,000

Data are the *Energy* for Analysis & Inquiry

NATIONAL ENERGY TECHNOLOGY LABORATORY

Developing a strong data foundation is key to any program/project's success



Morkner et al., 2021

20%

80%

→

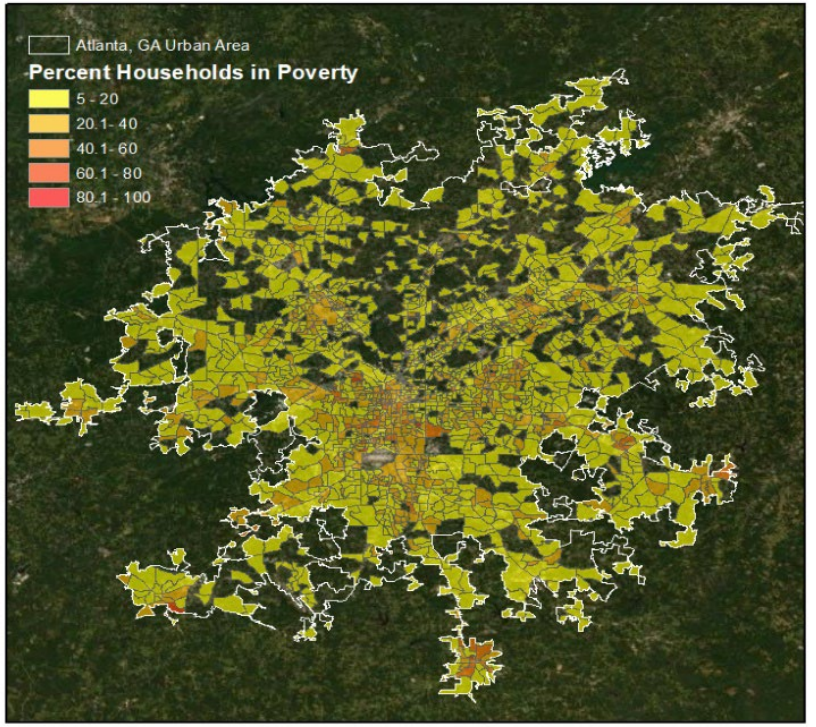
EDX has been used by other DOE Programs to reduce this data access barrier/overhead

Presently data-driven teams spend ~80% of their time addressing the bottom components of the "data pyramid"

Crowdflower 2016

High Demand for EESJ Data & Information

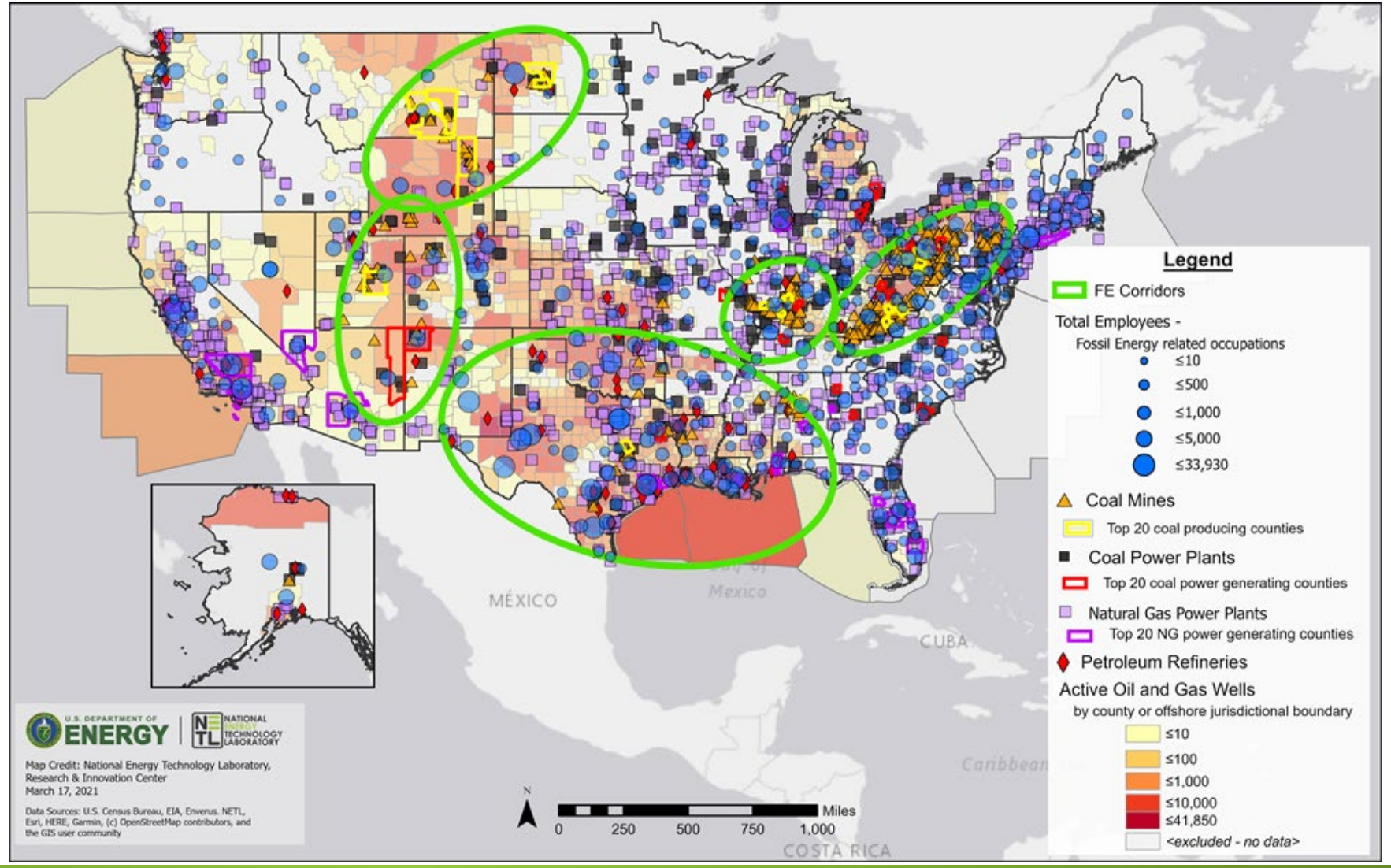
- Unprecedented opportunities for energy RDD&D
- **Energy transition and diversification** is ongoing and data-driven
- Spurred by initiatives like Justice40
- All rely on common baseline information that can help stakeholders understand and assess key energy, environmental & social burdens in relation to their activities and broader needs
- Most are strongly focused on energy infrastructure, existing & future opportunities



40% of the overall benefits of certain Federal investments—including investments in **clean energy** and **energy efficiency**; clean transit; affordable and sustainable housing; training and workforce development; the remediation and reduction of legacy pollution; and the development of clean water infrastructure—must flow to **disadvantaged communities**. (**Justice40 Initiative (EO 14008, Section 223)**)

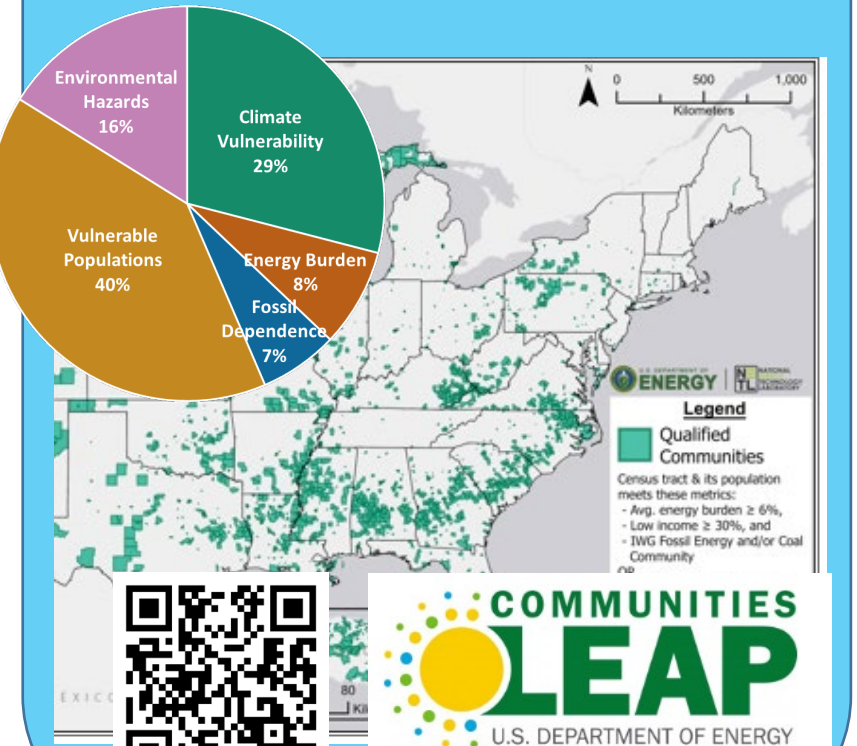
Value for Energy, Environmental & Social Justice (EESJ)

- Requires a wide range of data & information,
- At multiple spatial & temporal resolutions
- To address needs, what do you want to explore? What questions you want to evaluate?
- Role & use of these data require authoritative resources to ensure explainability & usability



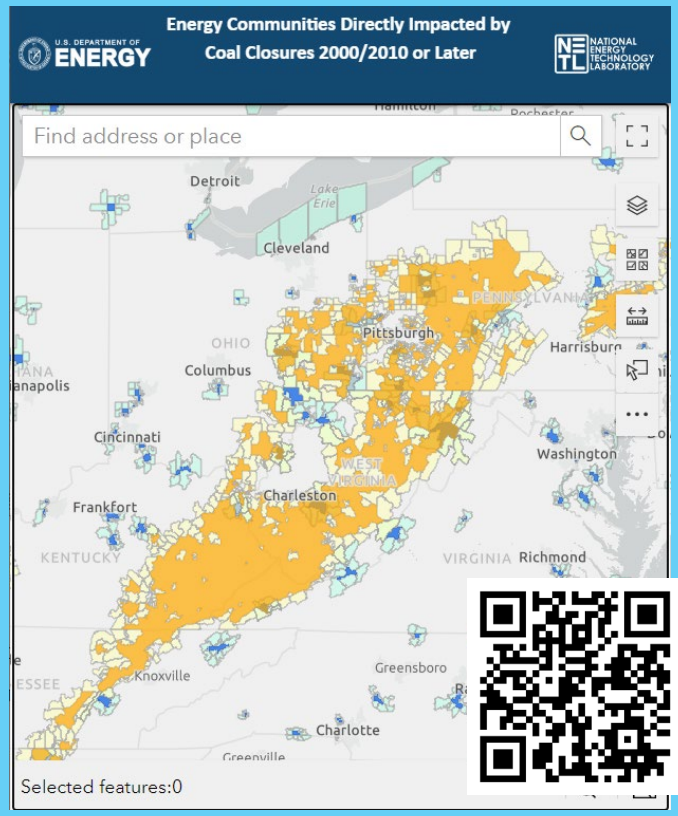
NETL's Geo-Data Science is Supporting on-going EESJ Efforts

Justice40 Initiative & Communities LEAP Pilot

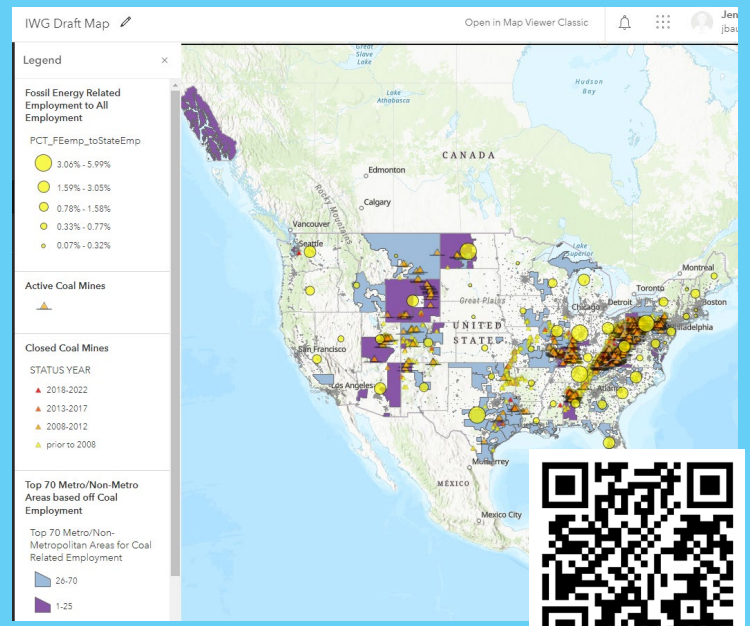


COMMUNITIES LEAP
U.S. DEPARTMENT OF ENERGY

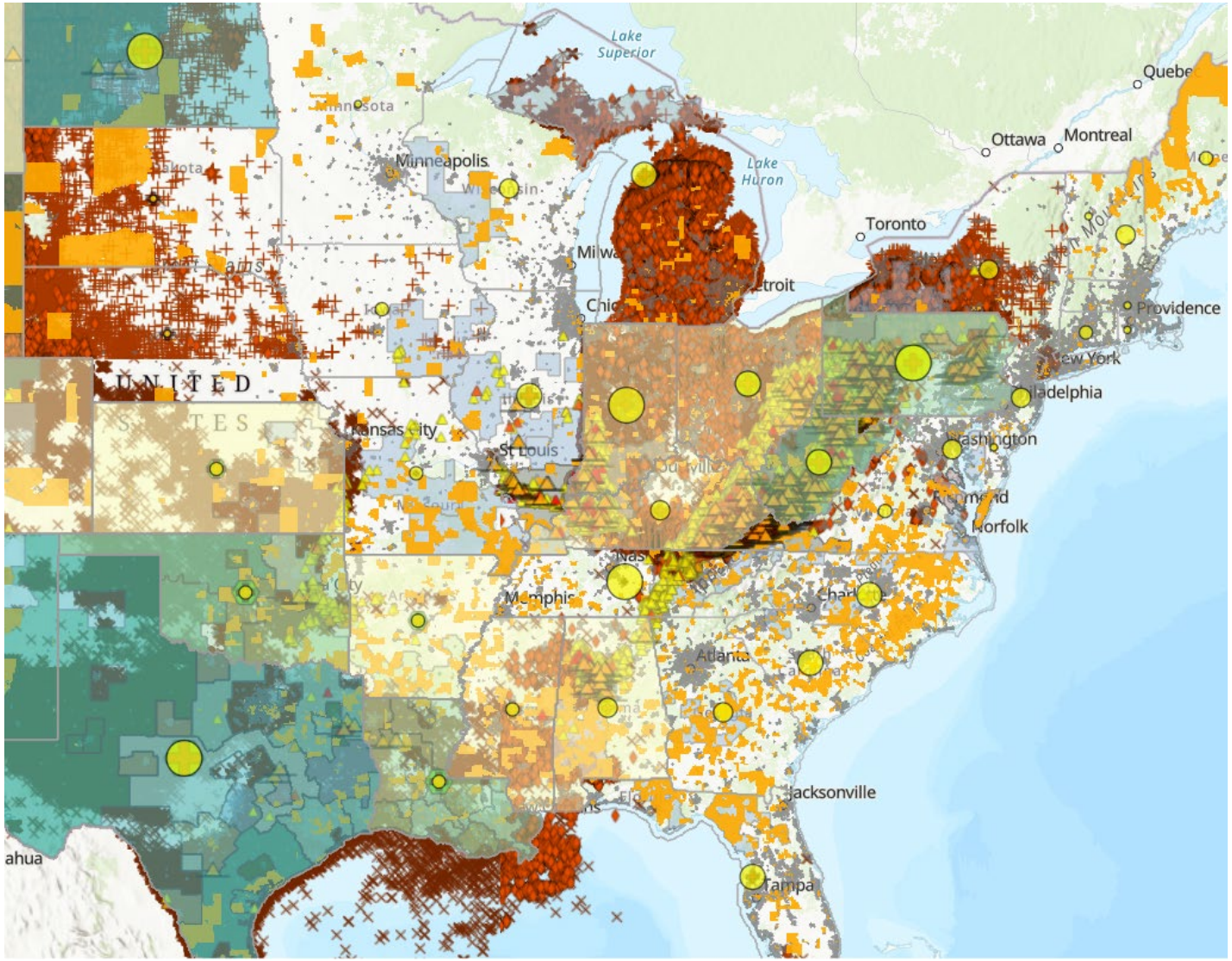
Energy Communities Directly Impacted by Coal Closures



Interagency Working Group (IWG) on Coal & Power Plant Communities and Economic Revitalization



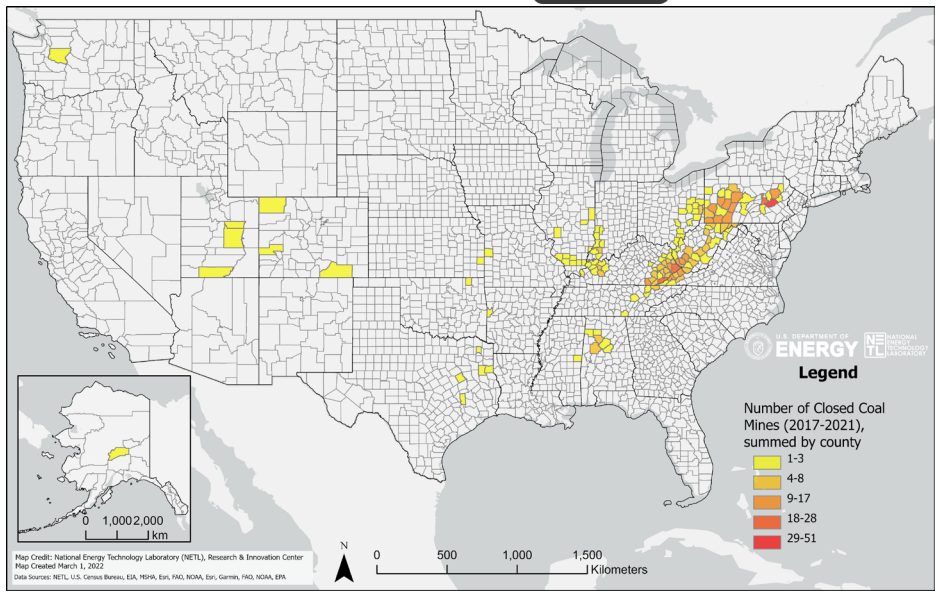
Increasing Complexity & Demand for EESJ Data

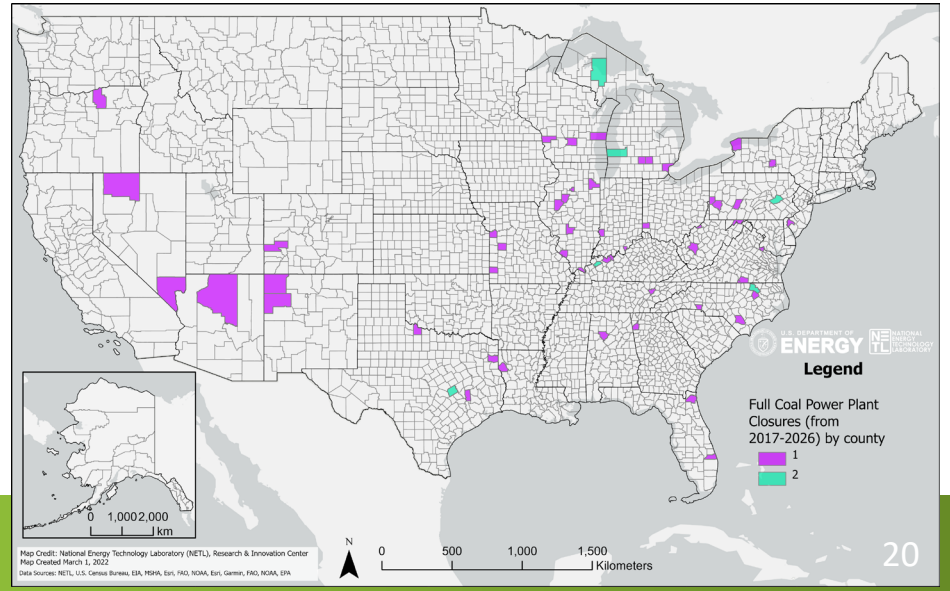


- Depending on the project, the **number of data variables & complexity of relationships** can be ridiculously overwhelming and challenging for stakeholders to effectively interpret and use
- Leveraging integrated analytics, as well as **ML & AI tools** can help simplify the data…
- But must be done smartly to reduce (or at least better characterize) underlying uncertainty and bias in these integrated analyses

Under development: Energy Transitions Atlas

- Integrate data on Communities & Energy Transitions
  - Begin analytics with Coal data, at a county scale:
    - Coal Mine production & closures
    - Coal Power Generation retirements & planned retirements
    - Pull in additional community metrics, including:
      - Tax revenue from coal
      - Urban versus Rural
      - Poverty level
      - Average travel times to/from work
      - Percent of county households considered technology limited (based off broadband internet access)
      - Average level of education
      - **And more to come…**
- Introduce criteria & weights where needed to present critical information to stakeholders

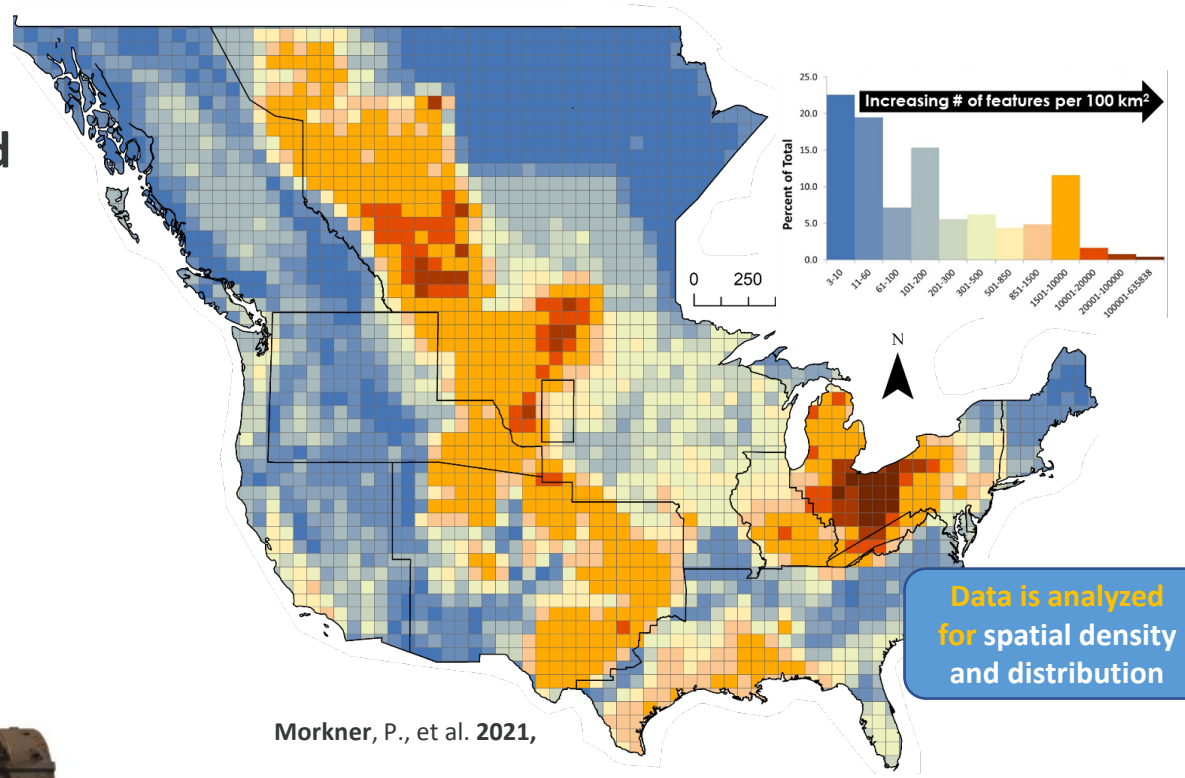




FECM has Invested in Creating a Digital Infrastructure Resource-base

Need to advance for democratized Use

- **PB of carbon storage data preserved** using the **public and private** sides of **EDX**
- Curating access to **downloadable instances** of FECM Program datasets, **models & tools**
- **Developed custom, AI/ML/NLP enhanced tools** to drive FECM data discovery and knowledge extraction
- **Enabling geospatial data and FECM web mapping for spatial data resources**



Morkner, P., et al. **2021,**

Current limitation of these data, tools, and capabilities…
*…they **still largely require the end-user to have access to the right expertise and computational resources to put them to use….***

https://edx.netl.doe.gov/about

EDX®

Energy Data eXchange

A Virtual Library and Laboratory Supporting FECM Research

Core Competencies of EDX

- AI/ML
- Big Data Compute
- Data Management
- Multi-Organizational Private Collaboration
- Public Dissemination of Data Products

these 5 core competencies support

Core Projects/Programs

- **Carbon Storage** (DisCO2ver)
- **Hydrogen** (InsigH2t)
- **Materials** (Metals, Cements, Polymers, Carbon)
- **Minerals Sustainability**
- **Infrastructure**
- **EDX Geospatial**

Ancillary Projects/Programs

- Comp Sci
- Energy Water
- Labs
- Microbiology
- Minerals Sustainability
- Sensors
- Subsurface
- System Analysis
- AND OTHERS…

EDX has fundamental capabilities that can be utilized across many research communities

NETL Research · Carbon Storage Program

SHASTA Subsurface Hydrogen Assessment, Storage, and Technology Acceleration

Natural Gas Hydrates

NRAP National Risk Assessment Partnership

SMART Offshore R&D

GEO CUBE

WATT The Power of AI

SAMI SCIENCE-BASED AI/ML INSTITUTE

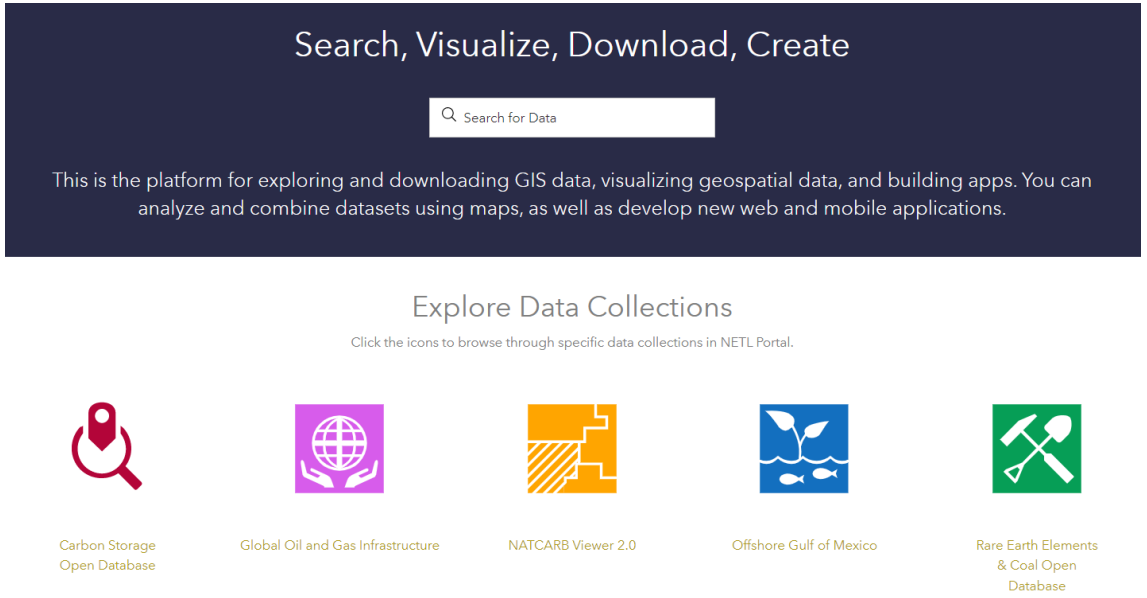
REECM

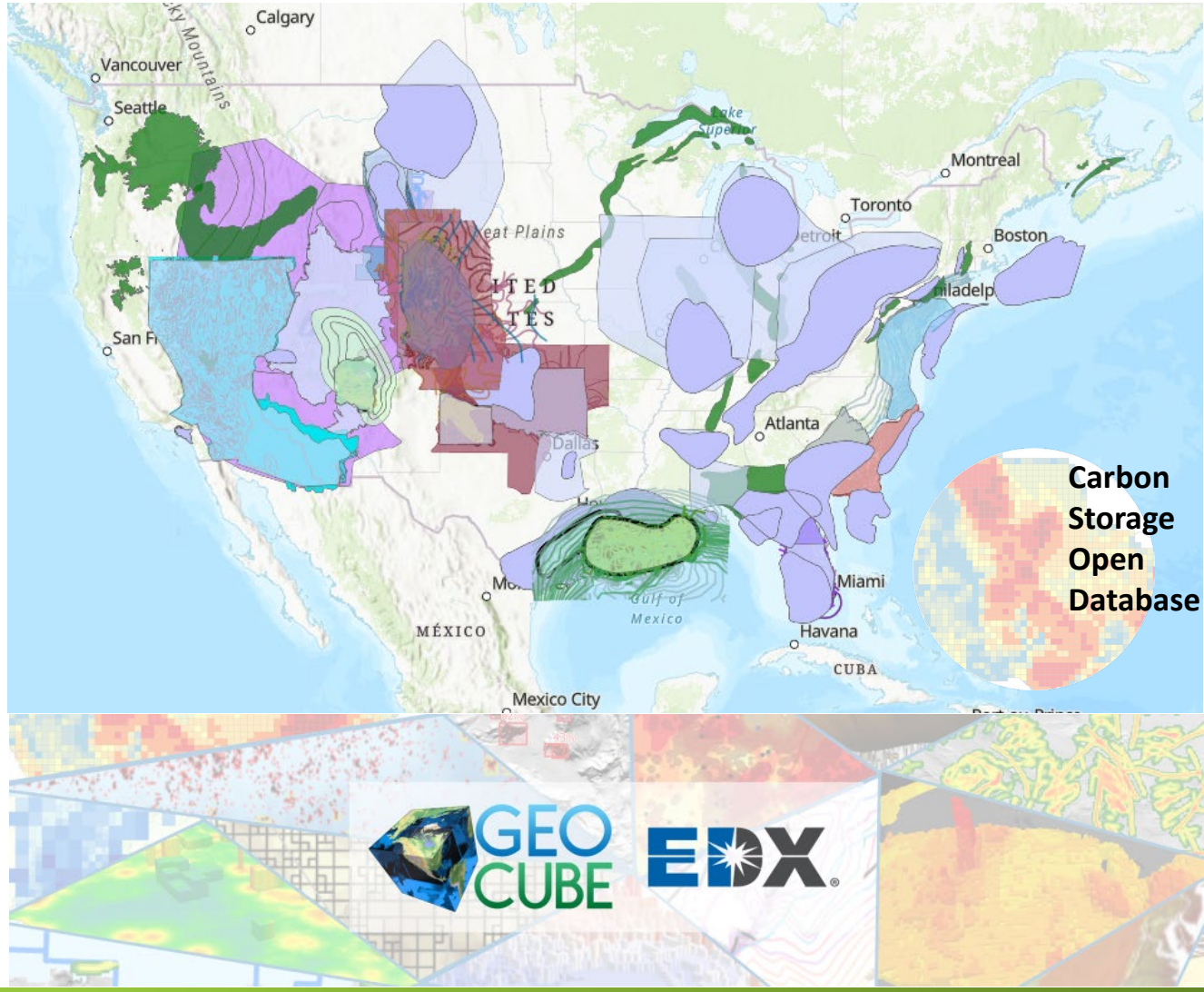
UCR Unconventional Resources

NETL SmartSearch

ETA Via FECM's Flagship Data Infrastructure Platform

- EDX's geospatial resources are available via the GeoCube tool
- Spatial datasets, maps, and capabilities for visualization, exploration, web map development, and hosting of priority geospatial data collections

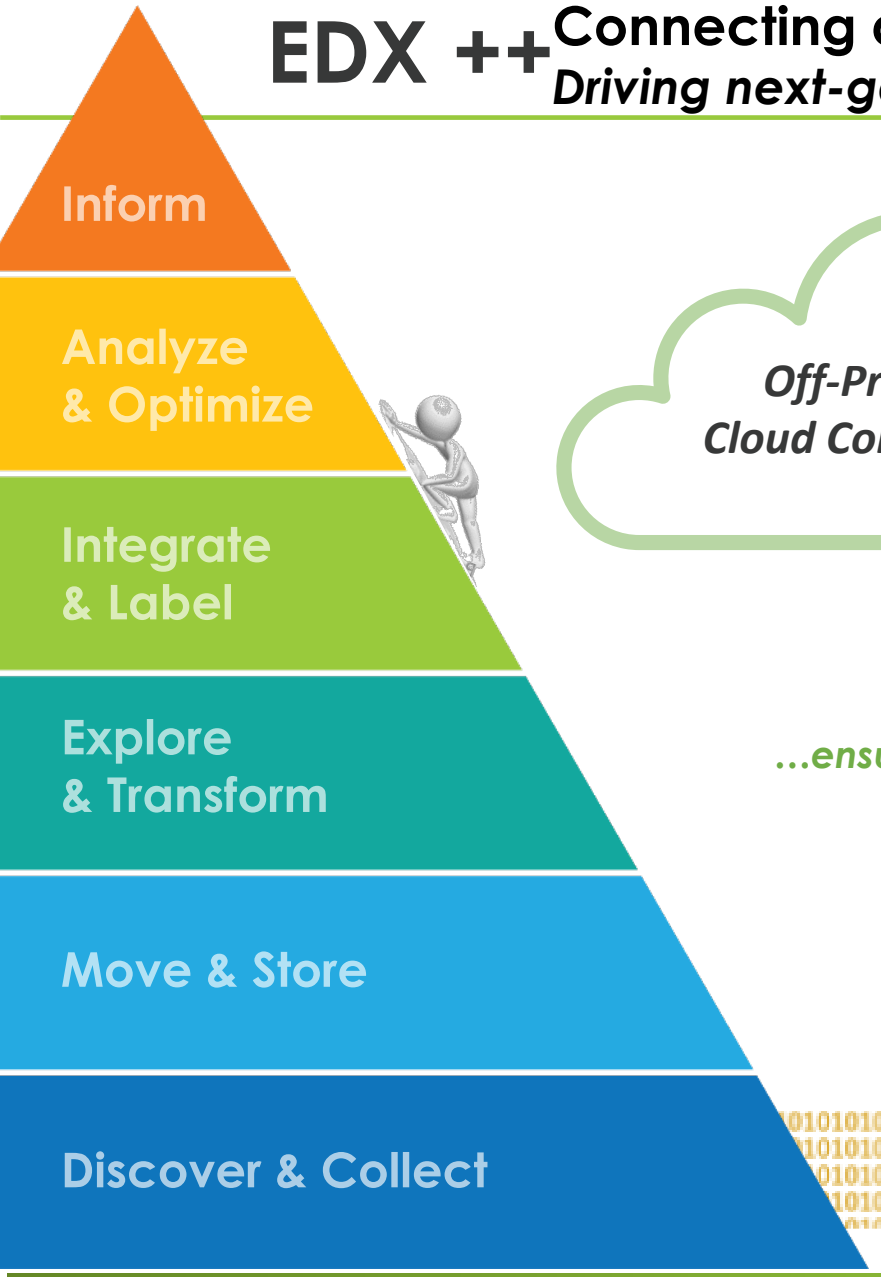




EDX ++ Connecting data to resources for analysis & computing

Driving next-gen AI R&D…scaling the pyramid

NATIONAL ENERGY TECHNOLOGY LABORATORY



Off-Prem Cloud Compute

OSTI.GOV

OpenEI Open Energy Information

Other systems & resources

NATIONAL ENERGY TECHNOLOGY LABORATORY

WSE

JOULE NETL SUPERCOMPUTER

WATT The Power of Ai

On-Prem Compute

ESnet

EDX++ FRAMEWORK

…ensuring compliance with Federal/DOE regulations

EDX® Energy Data eXchange

…ensuring preservation and access to DOE FECM knowledge and data resources

Big Data

Accelerating Commercialization and Reducing Risk

NATIONAL ENERGY TECHNOLOGY LABORATORY

Collaborative Initiatives, Addressing National and DOE Priorities

NETL Research · Carbon Storage Program · SMART · EDX Energy Data eXchange · 4CCS · NRAP National Risk Assessment Partnership

PCOR Partnership · Southern States Energy Board · Bureau of Economic Geology · GoMCarb Partnership for carbon capture & offshore geologic storage · CUSP West · Los Alamos National Laboratory · Illinois State Geological Survey Prairie Research Institute · Berkeley Lab · NETL National Energy Technology Laboratory · MRCI Midwest Regional Carbon Initiative · Pacific Northwest National Laboratory · FutureGen Alliance · Lawrence Livermore National Laboratory

Offshore R&D

NETL National Energy Technology Laboratory · U.S. Naval Research Laboratory · CSIRO · Universidad Nacional Autónoma de México · BSEE Bureau of Safety and Environmental Enforcement · CICESE · National Oceanography Centre · The Ocean Cleanup · Ocean Energy Safety Institute · BOEM Bureau of Ocean Energy Management · Pacific Northwest National Laboratory · INPE · NOAA National Oceanic and Atmospheric Administration U.S. Department of Commerce · Institut de Ciències del Mar

EDX Energy Data eXchange

NETL National Energy Technology Laboratory · Google · maximus · OSTI.GOV · esri · INL Idaho National Laboratory · aws · MATRIC · DICE · NREL Transforming ENERGY · Microsoft · EPRI Electric Power Research Institute

Unconventional REE Rare Earth Elements

NETL National Energy Technology Laboratory · U.S. Department of Energy Energy Efficiency & Renewable Energy · University of Wyoming · Office of Surface Mining Reclamation and Enforcement · USGS science for a changing world · RAMACO CARBON · WVGES Geology Underlies It All · North American Coal

Data science takes a team….

Kelly Rose, kelly.rose@netl.doe.gov
& the GAIA R&D Group,
EDX Dev/Ops, & SAMI

NATIONAL ENERGY TECHNOLOGY LABORATORY

SAMi SCIENCE-BASED AI/ML INSTITUTE

Photo, part of the GAIA R&D team "B.C."

B.C. = before covid

U.S. DEPARTMENT OF ENERGY | NATIONAL ENERGY TECHNOLOGY LABORATORY

https://edx.netl.doe.gov/sami

Citations

- Baker, D.V., Rose, K., Bauer, J., and Rager, D., 2016, **Computational Advances and Data Analytics to Reduce Subsurface Uncertainty**, ARMA 16-493, June 26-29, 2016, 16 pgs.
- Bauer, J., Justman, D., Mark-Moser, M., Romeo, L., Creason, C.G., and Rose, K., **Exploring beneath the basemap**, in Wright, D.J. and Harder, C. (Ed.), GIS for Science: Applying Mapping and Spatial Analytics: Volume 2, Redlands, CA: Esri Press, pp. 51-67, 2020, plus supplemental material.
- Dyer, et al., 2022, **Applied machine learning model comparison: Predicting offshore platform integrity with gradient boosting algorithms and neural networks**, https://doi.org/10.1016/ j.marstruc.2021.103152
- Glosser, D., Rose, K., and J. R. Bauer, 2016. **Spatio-Temporal Analysis to Constrain Uncertainty in Wellbore Datasets**: An Adaptable Analytical Approach in Support of Science-Based Decision Making. Journal of Sustainable Energy Engineering, 3(4): 299-317. Justman, et al. (2022). A database and framework **for carbon ore resources and associated supply chain data.** Data in Brief. https://doi.org/10.1016/j.dib.2021.107761
- Justman, D., Creason, C.G., Rose, K., & Bauer, J., 2020. A knowledge-data framework and **geospatial fuzzy logic-based approach to model and predict structural complexity**. Journal of Structural Geology, 104153. https://doi.org/10.1016/j.jsg.2020.104153
- Mark-Moser, M., Wingo, P., Duran, R., Dyer, A., Zaengle, D., Suhag, A., Hoover, B., Pantaleone, S., Shay, J., Bauer, J., and Rose, K. **2021. AI/ML integration for accelerated analysis and forecast of offshore hazards**. AGU Fall Meeting 2021, Dec. 13-17, New Orleans, LA/Virtual. Session: EP027 - Proven AI/ML applications in the Earth Sciences
- Morkner, P., Bauer, J., Creason, C., Sabbatino, M., Wingo, P., Greenburg, R., Walker, S., Yeates, D., and Rose, K. **Distilling data to drive carbon storage insights**, Computers & Geosciences, Volume 158, 2021, https://doi.org/10.1016/ j.cageo.2021.104945.
- Nelson, J.; Dyer, A.; Romeo, L.; Wenzlick, M.; Zaengle, D.; Duran, R.; Sabbatino, M.; Wingo, P.; Barkhurst, A.; Rose, K.; Bauer, J. **Evaluating Offshore Infrastructure Integrity**; DOE/NETL-2021/2643; NETL Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Albany, OR, 2020; p 70. doi.org/10.2172/1780656
- Nelson, J. R., Romeo, L., & Duran, R. (2021). Exploring the **Spatial Variations of Stressors Impacting Platform Removal** in the Northern Gulf of Mexico. Journal of Marine Science and Engineering, 9(11), 1223.
- Rose, K.; Bauer, J.; Baker, V.; et al., 2018, **Development of an Open Global Oil and Gas Infrastructure Inventory and Geodatabase**; NETL-TRS-6-2018; NETL Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Albany, OR, 2018; p 594; DOI: 10.18141/1427573.
- Rose, K., 2016, **Signatures in the Subsurface – Big & Small Data Approaches for the Spatio-Temporal Analysis of Geologic Properties & Uncertainty Reduction**, 162 pgs, http://hdl.handle.net/1957/59459
- Rose, K., Bauer, J.R., and Mark-Moser, M., 2020, **A systematic, science-driven approach for predicting subsurface properties,** *Interpretation, 8:1, 167-181* https://doi.org/10.1190/INT-2019-0019.1
- Wenzlick M., et al (2021) **Incorporating Historical Data and Past Analyses for Improved Tensile Property Prediction of 9% Cr.**
- Wenzlick, M.,et al. 2021. **Data science techniques, assumptions, and challenges in alloy clustering and property prediction.** Journal of Materials Engineering and Performance.