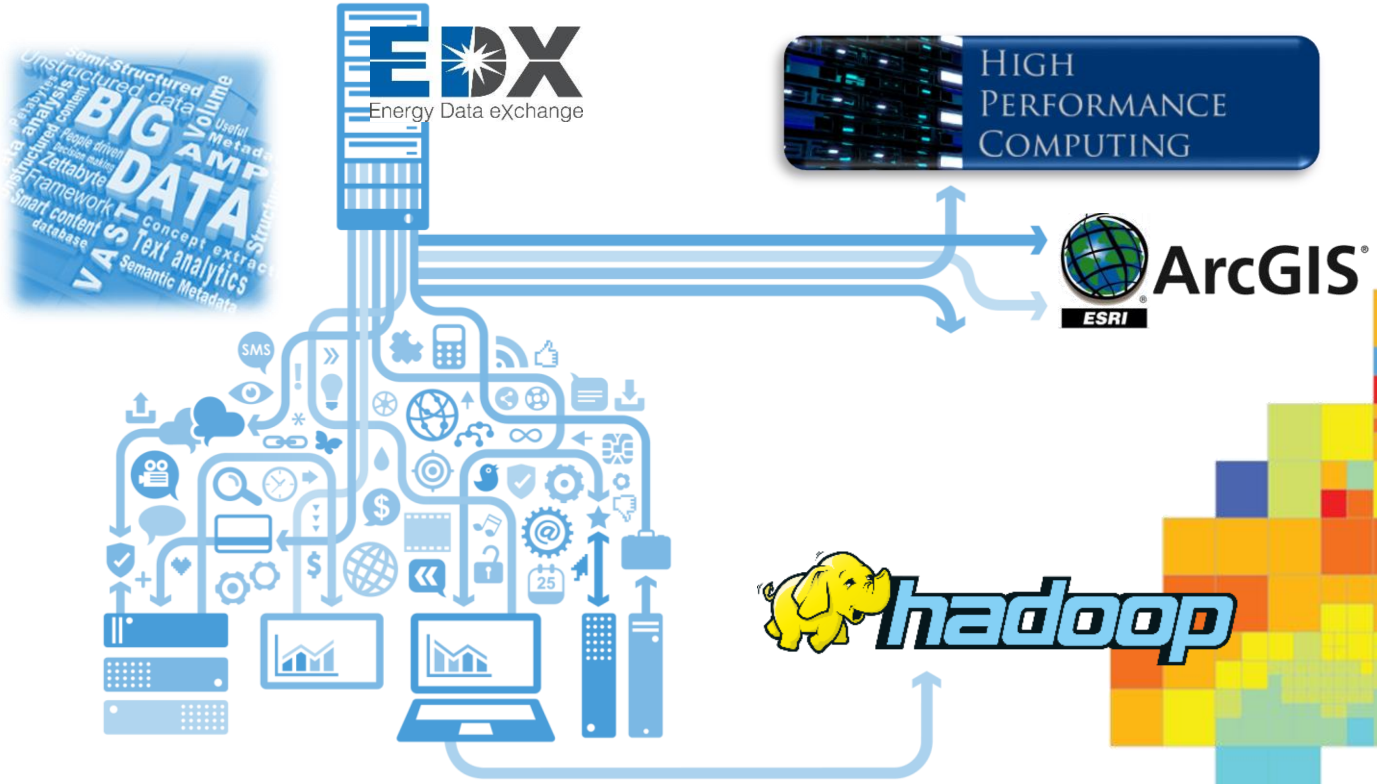
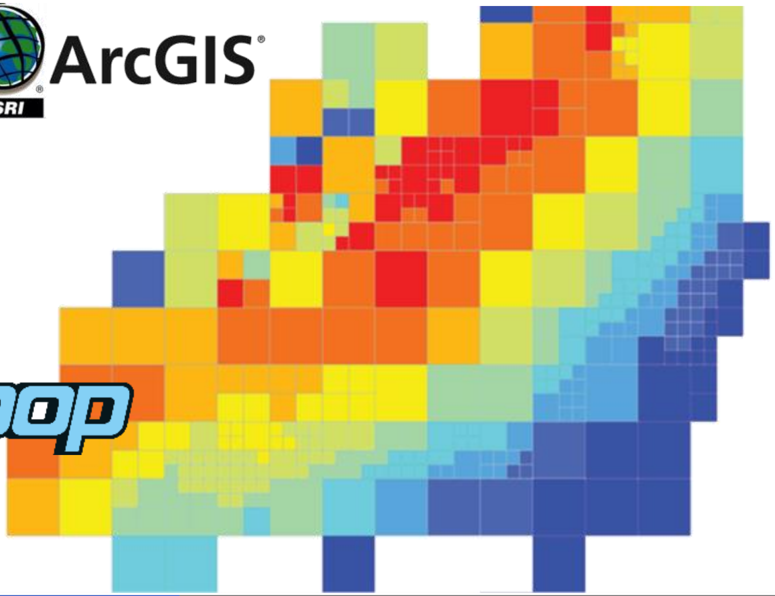
Shaking our foundation –

Innovating subsurface analysis for induced seismicity risk prediction using big data, geoscience and novel spatio-temporal methods





U.S. Department of Energy, National Energy Technology Laboratory
Mastering the Subsurface Through Technology, Innovation and Collaboration:
Carbon Storage and Oil and Natural Gas Technologies Review Meeting
August 16-18, 2016

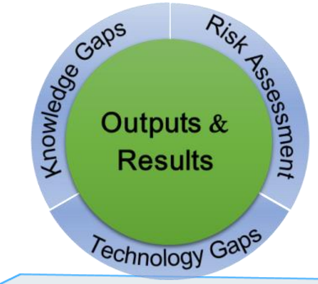
Presented by: Kelly Rose, Geology/Geospatial Researcher

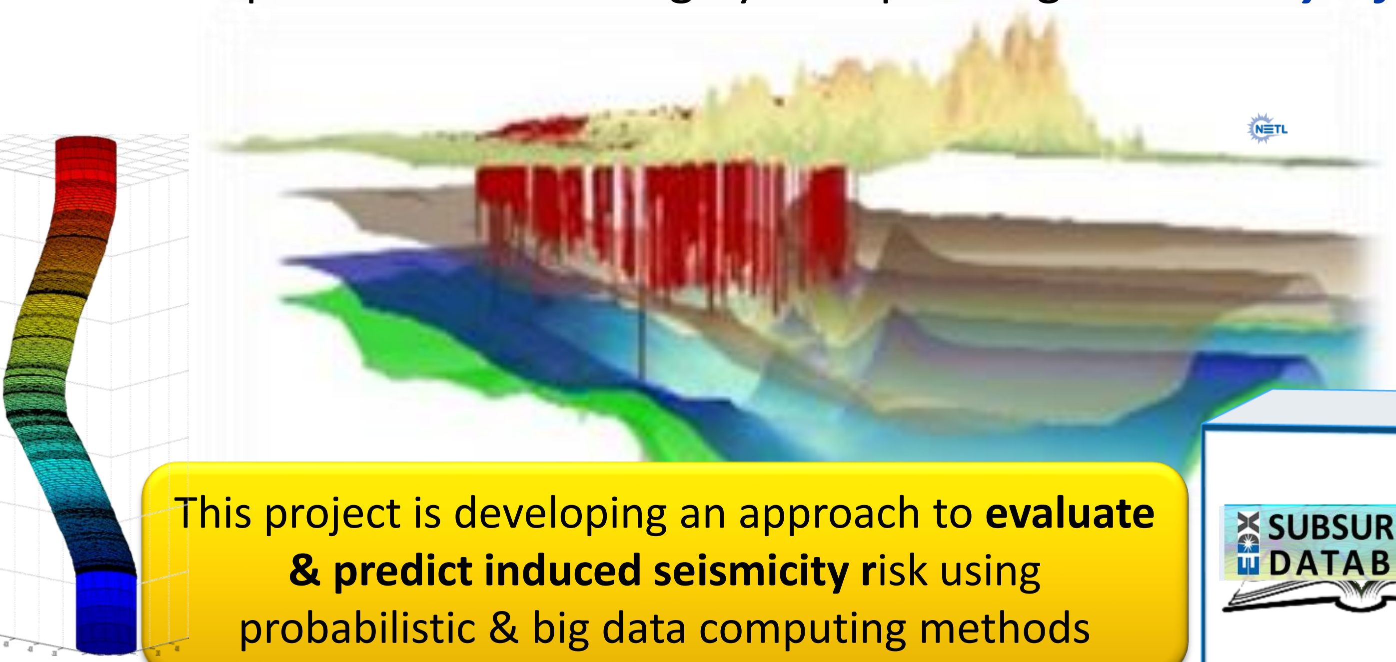
NETL Argonne Los Alamos Sandia National Laboratories AECOM MATRIC

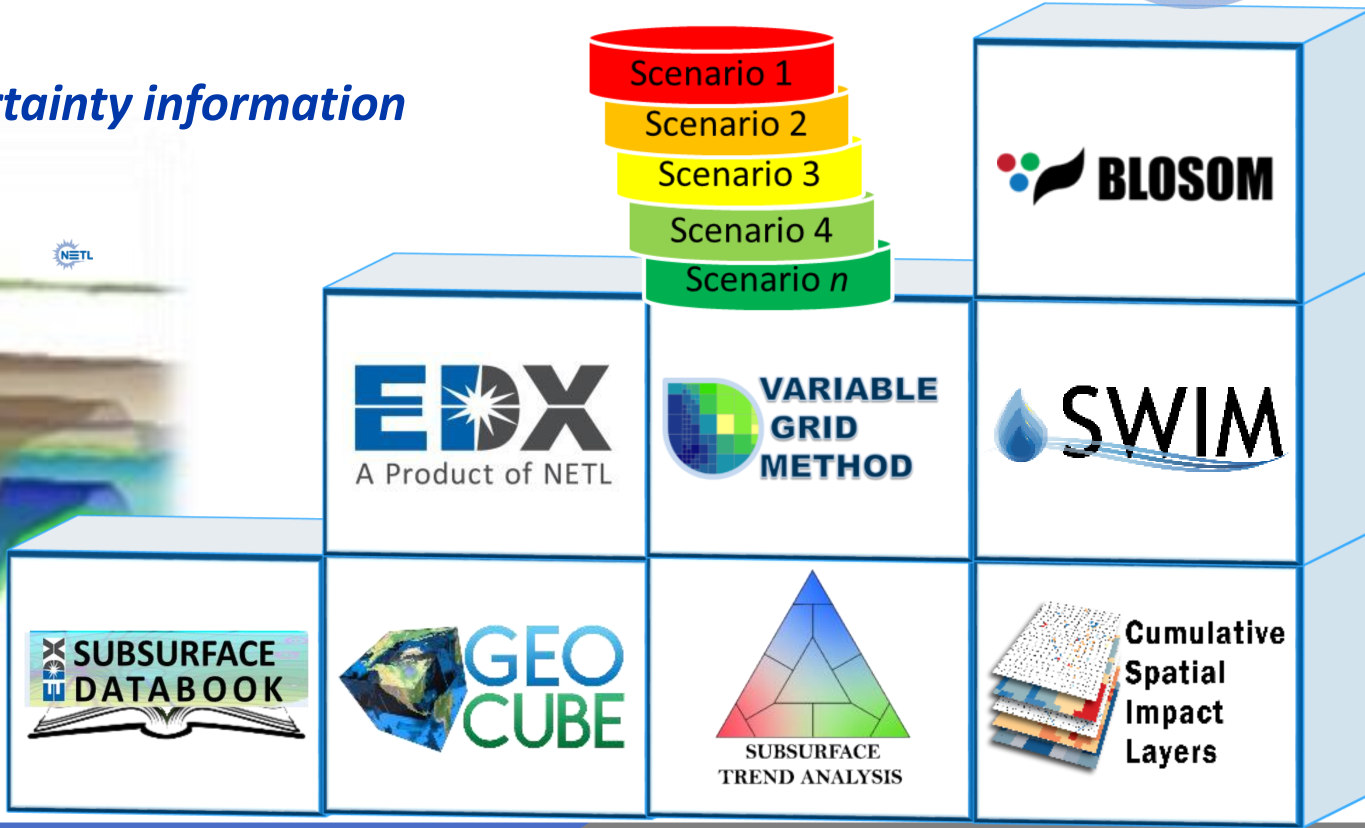
This Project aligns to DOE goals

To constrain & understand effects of subsurface engineering:

- At ***different scales and using different disciplines***
- To evaluate a range of potential ***environmental, social, and economic*** variables
- To evaluate various ***scenarios***, for risk reduction, resource evaluation, & improved efficiency
- To highlight ***knowledge and/or technology gaps***
- To improve understanding by incorporating ***uncertainty information***



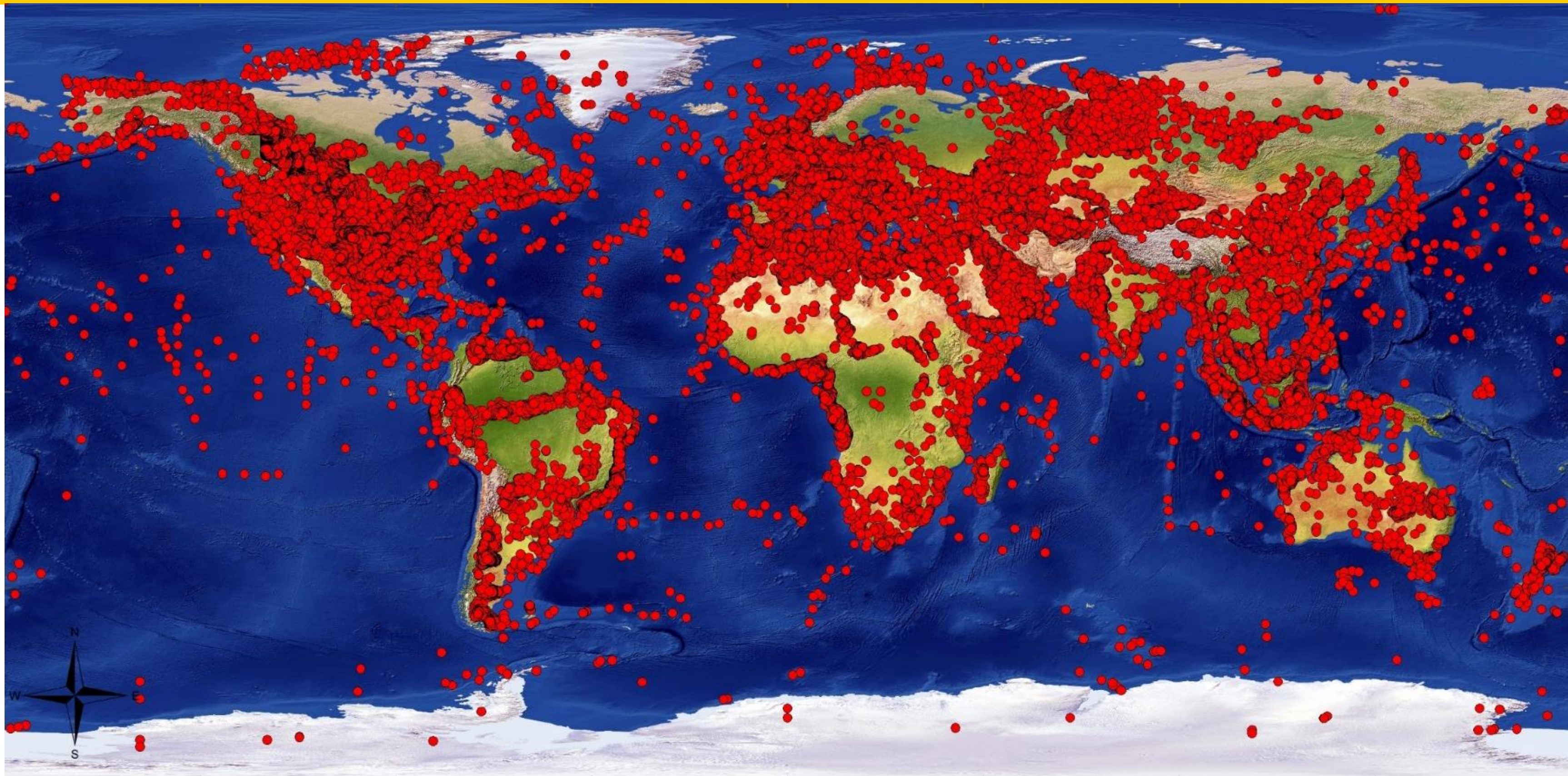




This project is developing an approach to **evaluate & predict induced seismicity r**isk using probabilistic & big data computing methods

Challenge – How do we predict subsurface behaviors?

- **Given there is now a global subsurface dataset, how can we use this to better understand the subsurface?**
  - Prediction of specific properties and subsurface features
  - Use to predict how our interactions relate to future resource
  - Improve our ability to interact with the subsurface (resource production, CO_2 storage, geohazard prediction, etc)

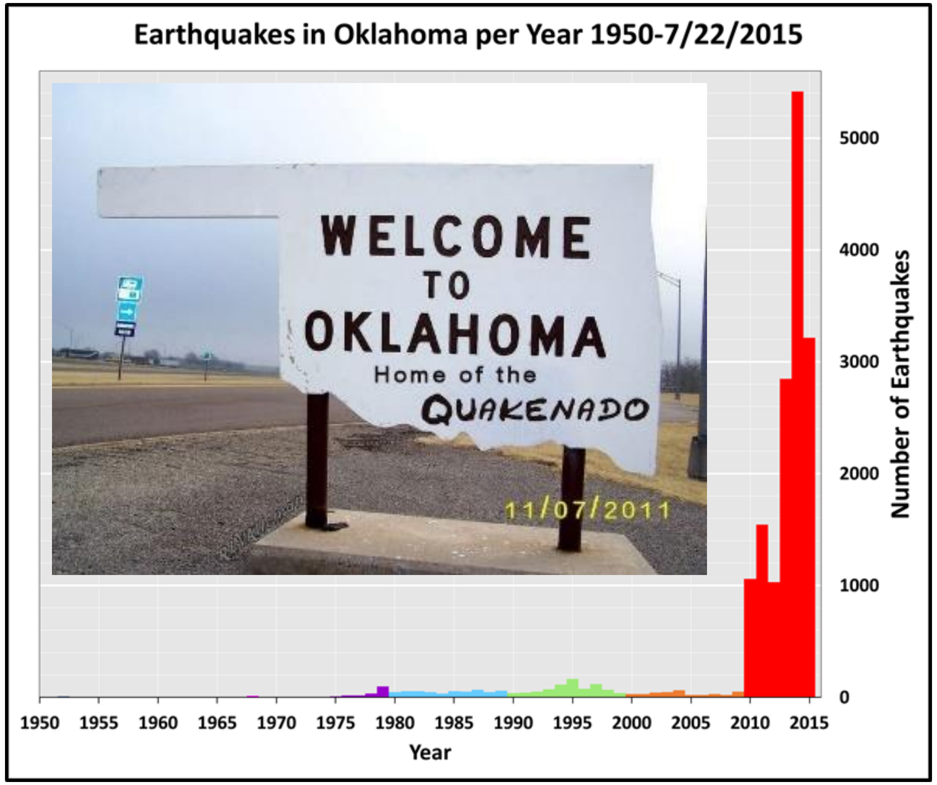


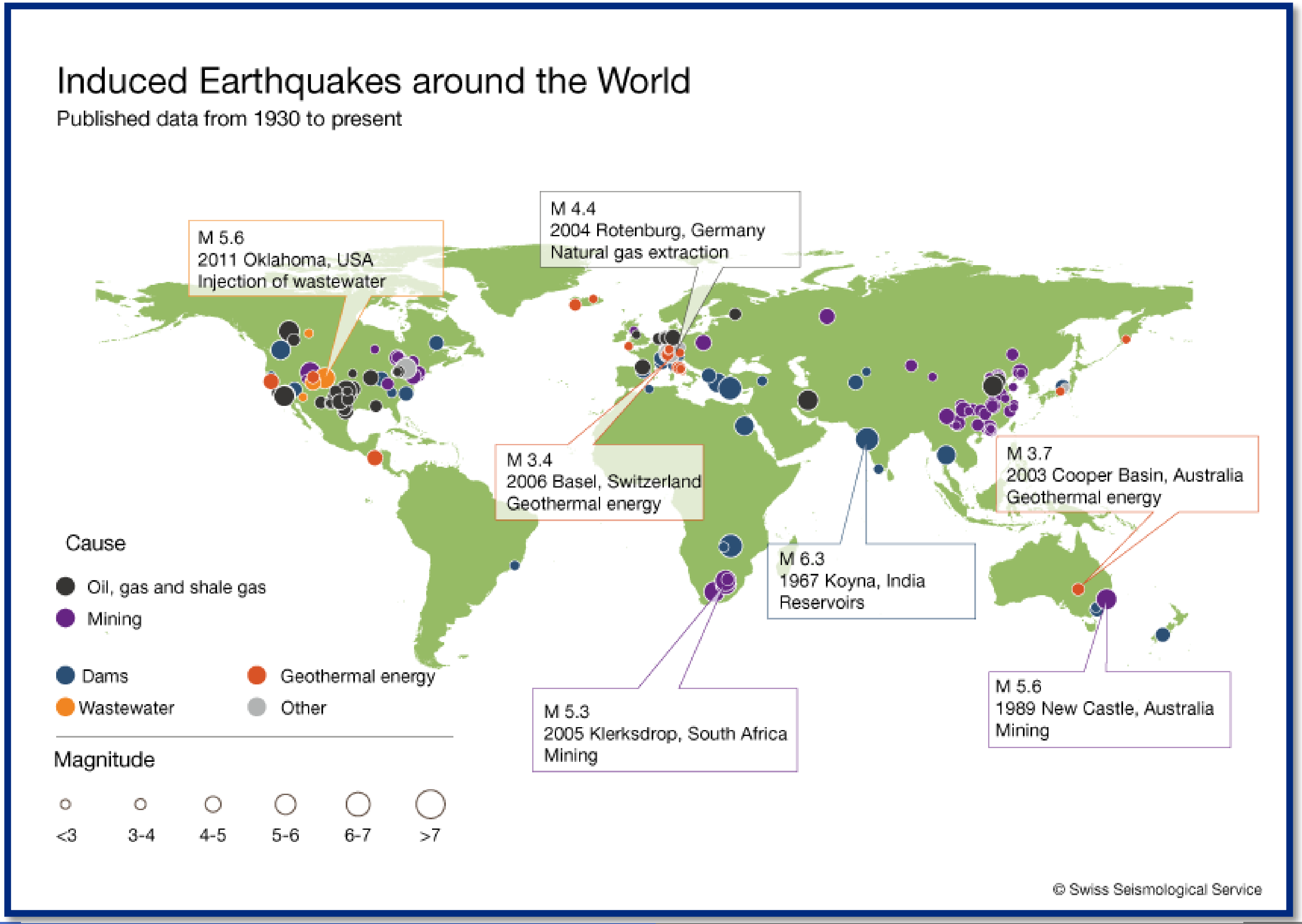
Rose, K., 2016, Signatures in the Subsurface – Big & Small Data Approaches for the Spatio-Temporal Analysis of Geologic Properties & Uncertainty Reduction, 162 pgs, http://hdl.handle.net/1957/59459

- **With more than 6 million wells worldwide, to what degree is the subsurface perturbed?**
  - What does that mean for the future?

In particular, induced seismicity is shaking things up

Human engineering of the deep subsurface is causing manmade aka induced seismic events in unprecedented #'s and locations





This project seeks to address the need for rapid, repeat evaluations

This need was spotlighted by a 2016 USGS report

USGS science for a changing world

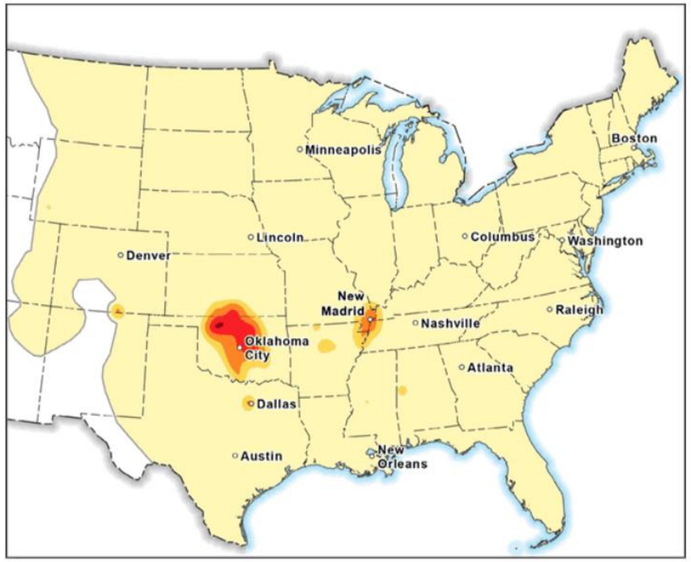
Earthquake Hazards Program

- Earthquakes
- Hazards
- Data & Products
- Learn
- Monitoring
- Research

Search...

Search

2016 One-Year Model



Map showing chance of damage from an earthquake in the Central and Eastern United States during 2016. Percent chances are represented as follows: pale yellow, less than 1 percent; dark yellow, 1 to 2 percent; orange, 2 to 5 percent; red, 5 to 10 percent; dark red, 10 to 12 percent.

The U.S. Geological Survey (USGS) has produced a 1-year seismic hazard forecast for 2016 for the Central and Eastern United States (CEUS) that includes contributions from both induced and natural earthquakes. The model assumes that earthquake rates calculated from several different time windows will remain relatively stationary and

USGS Open-File Report

2016 One-Year Seismic Hazard Forecast Induced and Natural Earthquakes - OFR

USGS News Release

USGS Science Feature

Maps for Media

- Forecast for Damage from Natural and Induced Earthquakes in 2016 (% Chance of Damage)
- Forecast for Ground Shaking Intensity from Natural and Induced Earthquakes in 2016 (Modified Mercalli Intensity)
- Map of 21 Areas Impacted by Induced Earthquakes (Wells and Seismicity)

Scientific Data

- Data
- Catalogs
- Source Code

See also Induced Seismicity Research.

Mark Zoback, a geophysicist at Stanford University who studies induced earthquakes, says: **"It's important to recognize the risk that these maps point out, but … that risk is going to change depending on what's happening on the ground."**

http://www.npr.org/sections/thetwo-way/2016/03/2[illegible]/472232829/u-s-geology-maps-reveal-areas-vulnerable-to-man-made-[illegible]uakes

PBS NEWSHOUR

TOPICS › SCIENCE

Drilling-induced earthquakes may endanger millions in 2016, USGS says

BY ANNA KUCHMENT, SCIENTIFIC AMERICAN *March 28, 2016 at 5:43 PM EDT*

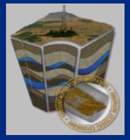
EMAIL

A billboard is seen on the side of the road set up by the law firm representing plaintiffs in a lawsuit against two oil companies near Greenbrier, Arkansas, August 6, 2013. At the time, more than a dozen homeowners in central Arkansas were suing two oil companies in federal court, claiming the disposal of wastewater from hydraulic fracturing, or fracking, triggered a swarm of more than 1,000 earthquakes in 2010 and 2011 that damaged their property. Photo by Jim Young/REUTERS

To Cathy Wallace, the earthquakes that have been rattling her tidy suburban home in Dallas feel like underground thunderstorms. First comes a distant roar, then a boom and a jolt. Her house shakes and the windows shudder. Framed prints on the walls clatter and tilt. A heavy glass vase tips over with a crash.

The worst moments are the ones between the rumble and the impact. "Every time it happens you know it's going to hit, but you don't know how severe it's going to be," she says. "Is this going to be a bigger one? Is this the part where my house falls down? It's scary. It's very scary."

Project Objectives – 2014 to 2016



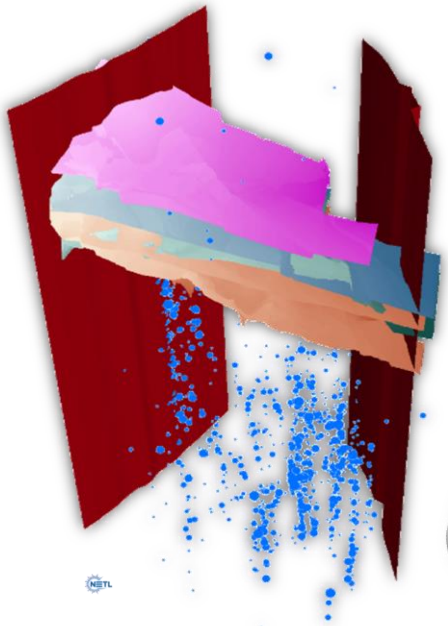
At any given point in space and time can we assess the probability of triggering an induced seismic event…?

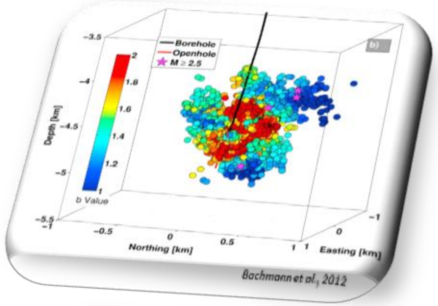
…can we integrate scientific knowledge with multiple heterogenous datastreams …

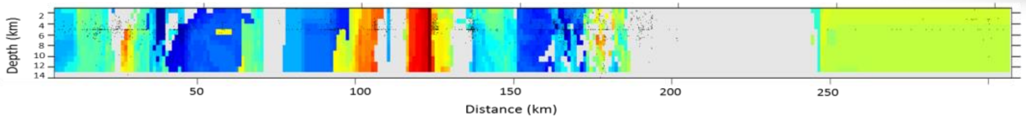
…leverage state-of-the-art computational and algorithmic resources to assess correlation and causation … ?

… use big data-geoscience computing with probabilistic methods … ?

… transform how we assess probability of future induced seismicity?





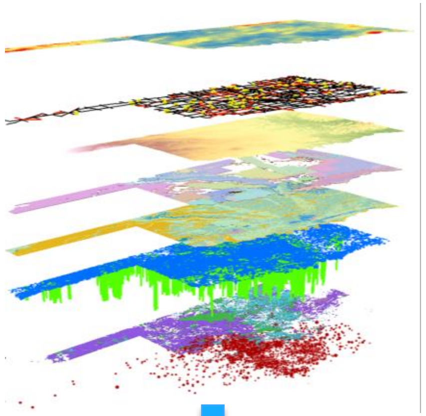


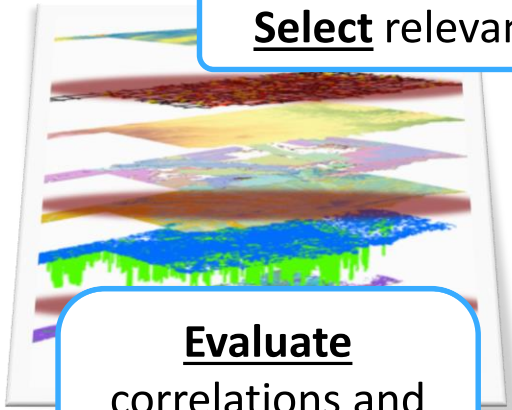
Project Goals

Goal: Address potential for causing induced seismicity events through application of computing techniques for **data mining, discovery, integration and analysis**

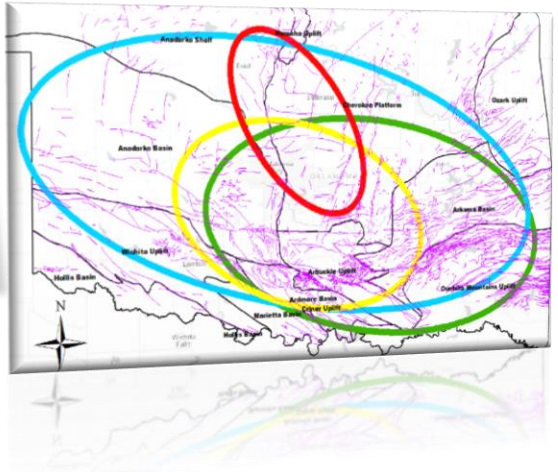
Ultimate product: Produce a platform with data, workflow, and tools that support a spatio-temporal assessments to highlights regions with an increased likelihood of induced seismicity

Access & Combine data and tools to…

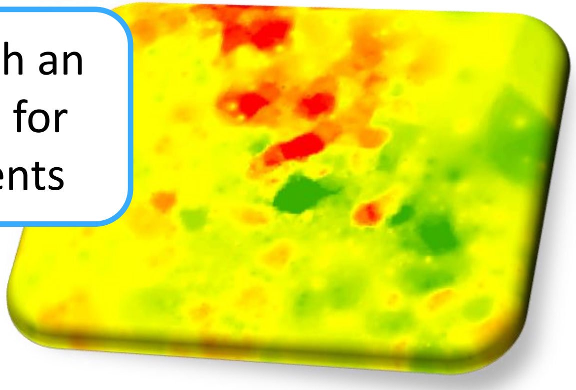




Select relevant datasets



Evaluate correlations and spatio-temporal trends



Highlight regions with an increased likelihood for induced seismic events

Organization Chart

Task 1 – Geoscience computing advances

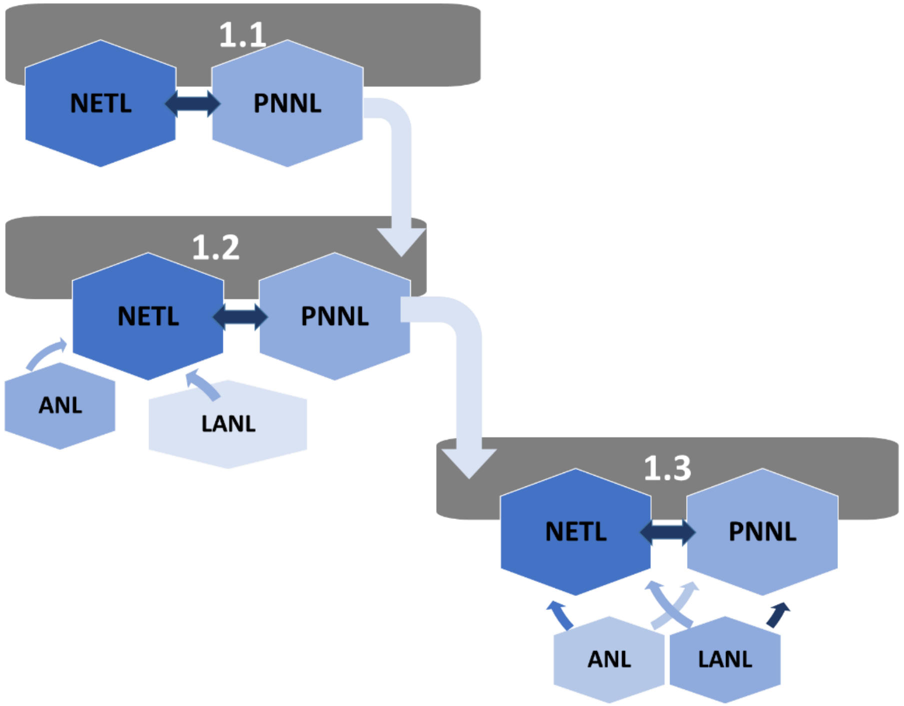
NETL (EDX team), LANL, & PNNL (Velo team)

ANL (in-kind, big data computing)

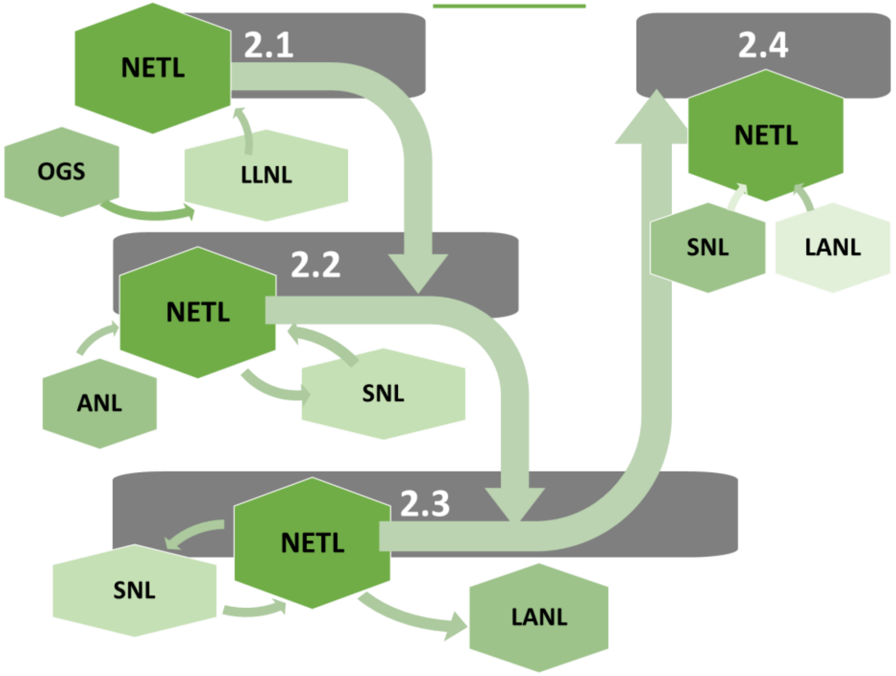
Task 2, Development of probabilistic approaches for induced seismicity

(LANL, LLNL, NETL, OGS, SNL)

Task 1



Task 2



What is "big data"

Big datasets tend to be **unstructured, distributed and complex**

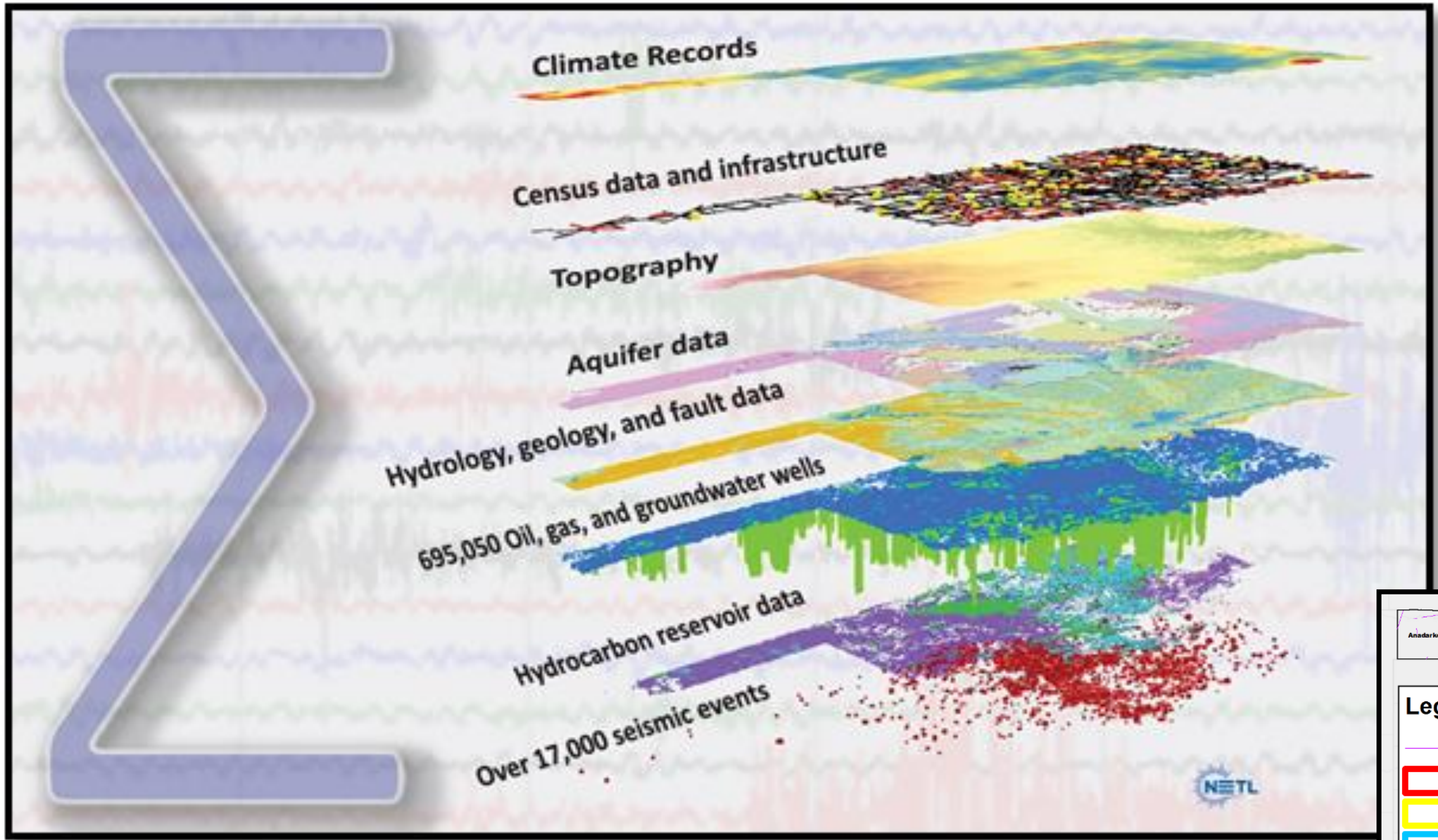
Big data can be defined by:

- The **volume** of information that systems must ingest, process and disseminate
- The **velocity** at which information grows or disappears
- The **variety** in the diversity of data sources and formats

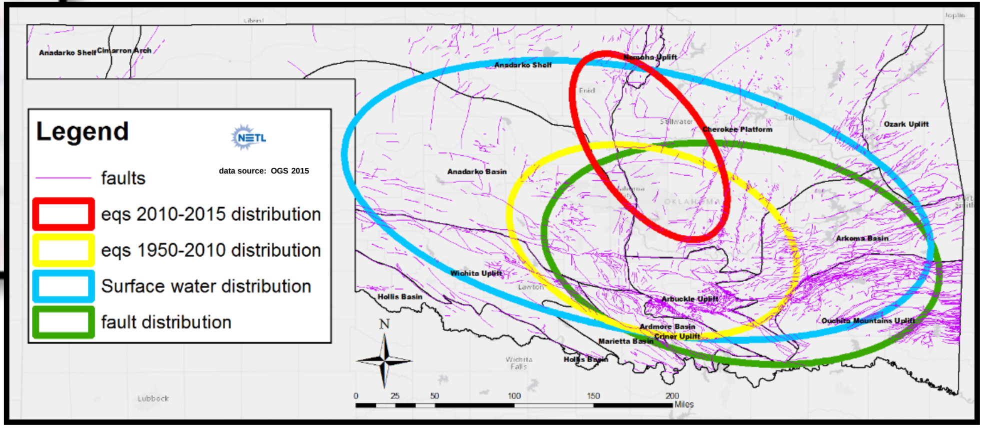
What is big data computing?

- **Combination of hardware & software technologies that make it possible to realize value from "Big Datasets"**
- **HPC vs BDC**
  - Traditional **HPC** systems are focused on performing calculations at fast speeds
  - **BDC** is focused on computing to sift through huge amounts of big datasets
  - **HPC** systems usually cost $1000's of k
  - **BDC** can operate on range of hardware, including inexpensive ($10's of k) clusters optimized for distributed, in-memory, iterative processing for analytics, query, and data mining
- **Both HPC and BDC can harness cloud server farms or add additional physical nodes**

How Are We Evaluating Induced Seismicity Probabilistically with Big Data?



- Developing tools and approaches to **manage multiple heterogeneous datasets**
- **Developing a beta probabilistic approach** that can be utilized to assess potential for induced seismicity impacts through big data analyses
- Developing approaches to **reduce uncertainty** and constrain subsurface trends
- Improving **joint analysis of multiple datasets**, using “Big Data” mining and integration techniques



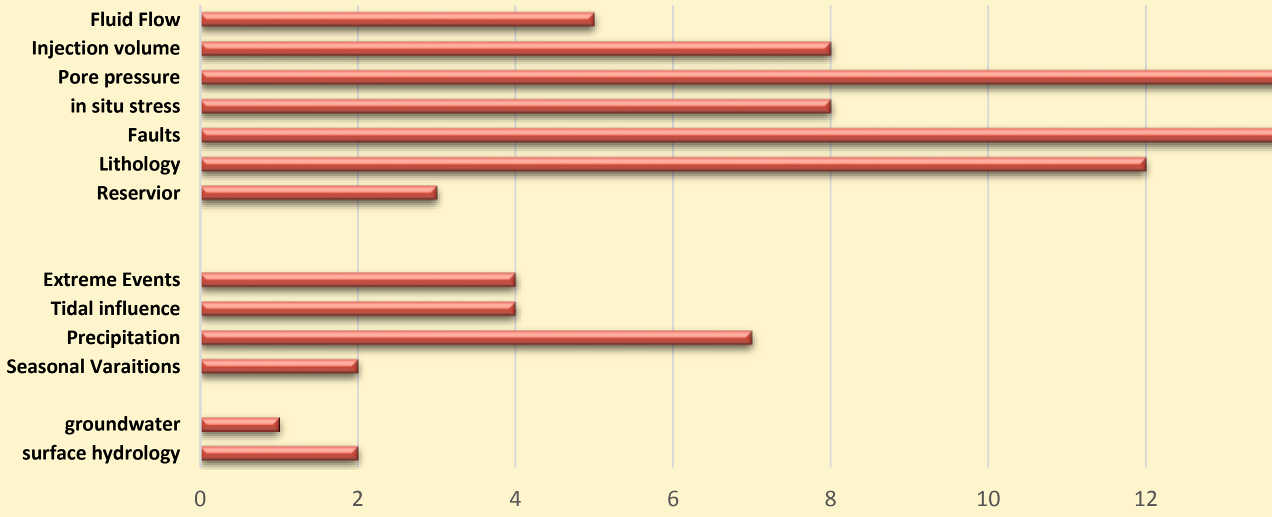
Example analysis: Spatio-Temporal Earthquake Distributions Vs. Fault and Surface Water Distributions

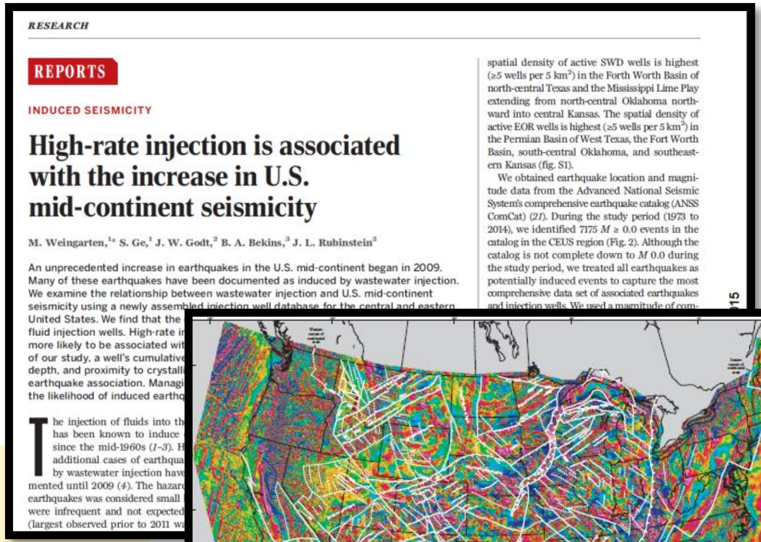
Approach & Results – Need to understand the causes

Reviewed 100's of peer-reviewed articles and other references to identify potential causes or correlation factors for both natural and induced seismicity

- Focus on hypothesis driven variables & methods

Identified variables affecting induced and natural seismicity




RESEARCH

REPORTS

INDUCED SEISMICITY

High-rate injection is associated with the increase in U.S. mid-continent seismicity

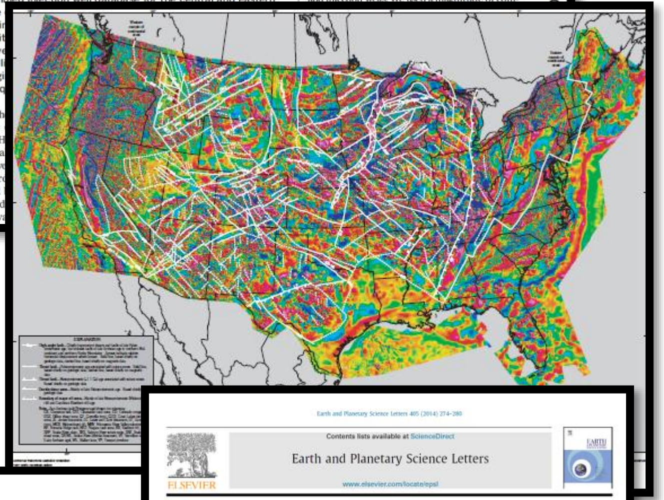
M. Weingarten, S. Ge, J. W. Godt, B. A. Bekins, J. L. Rubinstein

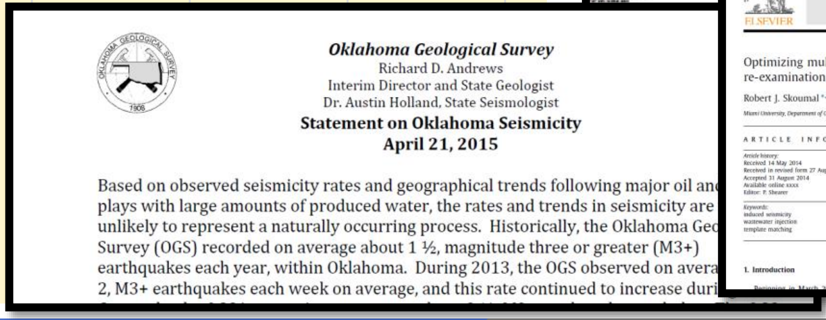
An unprecedented increase in earthquakes in the U.S. mid-continent began in 2009. Many of these earthquakes have been documented as induced by wastewater injection. We examine the relationship between wastewater injection and U.S. mid-continent seismicity using a newly assembled injection well database for the central and eastern United States. We find that the ... fluid injection wells. High-rate in... more likely to be associated wit... of our study, a well's cumulative... depth, and proximity to crystall... earthquake association. Managi... the likelihood of induced earthq...

The injection of fluids into th... has been known to induce ... since the mid-1960s (1–3). H... additional cases of earthqua... by wastewater injection have... mented until 2009 (4). The hazard... earthquakes was considered small... were infrequent and not expected... (largest observed prior to 2011 wa...

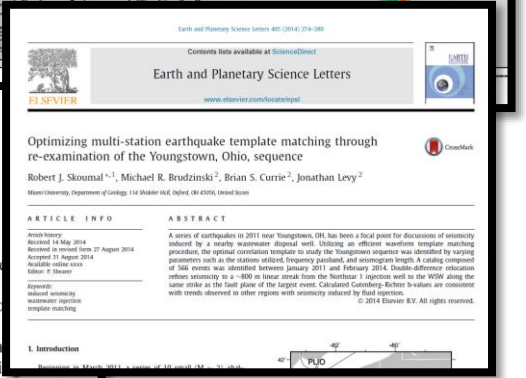
spatial density of active SWD wells is highest (≥5 wells per 5 km²) in the Forth Worth Basin of north-central Texas and the Mississippi Lime Play extending from north-central Oklahoma northward into central Kansas. The spatial density of active EOR wells is highest (≥5 wells per 5 km²) in the Permian Basin of West Texas, the Fort Worth Basin, south-central Oklahoma, and southeastern Kansas (fig. S1).

We obtained earthquake location and magnitude data from the Advanced National Seismic System's comprehensive earthquake catalog (ANSS ComCat) (21). During the study period (1973 to 2014), we identified 7175 M ≥ 0.0 events in the catalog in the CEUS region (Fig. 2). Although the catalog is not complete down to M 0.0 during the study period, we treated all earthquakes as potentially induced events to capture the most comprehensive data set of associated earthquakes and injection wells. We used a magnitude of com-




Oklahoma Geological Survey
Richard D. Andrews
Interim Director and State Geologist
Dr. Austin Holland, State Seismologist
Statement on Oklahoma Seismicity
April 21, 2015

Based on observed seismicity rates and geographical trends following major oil an... plays with large amounts of produced water, the rates and trends in seismicity are unlikely to represent a naturally occurring process. Historically, the Oklahoma Geo... Survey (OGS) recorded on average about 1 ½, magnitude three or greater (M3+) earthquakes each year, within Oklahoma. During 2013, the OGS observed on avera... 2, M3+ earthquakes each week on average, and this rate continued to increase duri...


Contents lists available at ScienceDirect

Earth and Planetary Science Letters

www.elsevier.com/locate/epsl

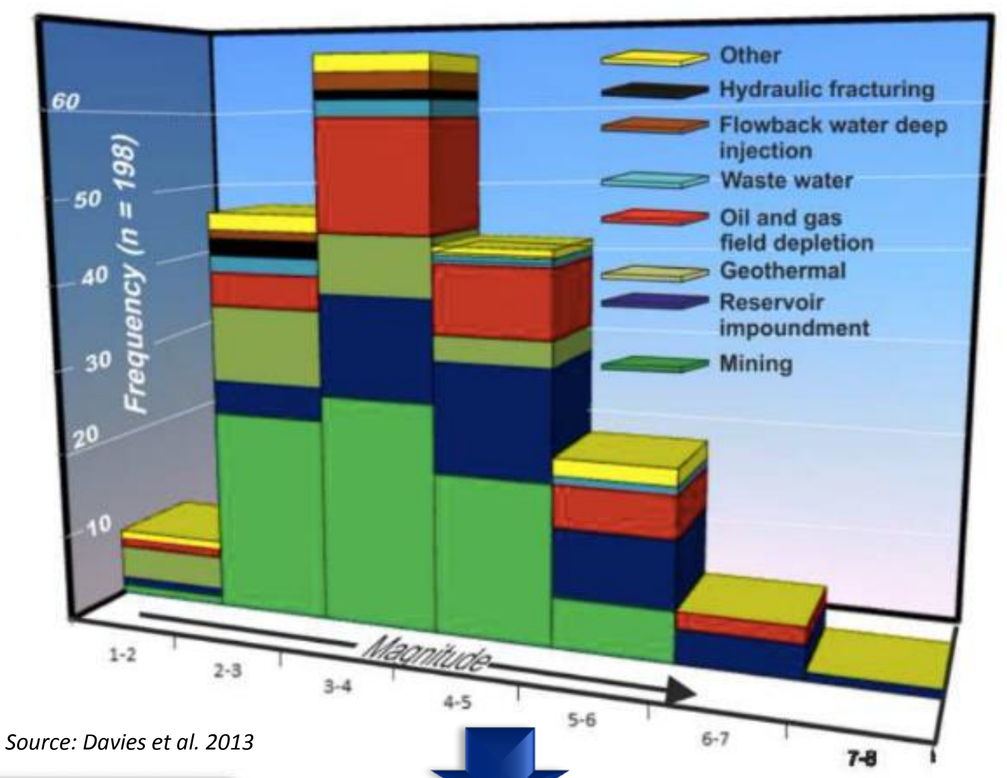
Optimizing multi-station earthquake template matching through re-examination of the Youngstown, Ohio, sequence

Robert J. Skoumal, Michael R. Brudzinski, Brian S. Currie, Jonathan Levy

Approach & Results - Induced Seismicity Causes

Which human activities are thought to trigger earthquakes?

- Wastewater injection
- Hydraulic Fracturing
- Oil and gas production
- Mining
- Geothermal energy
- Groundwater extraction
- Dammed lakes
- Large construction projects



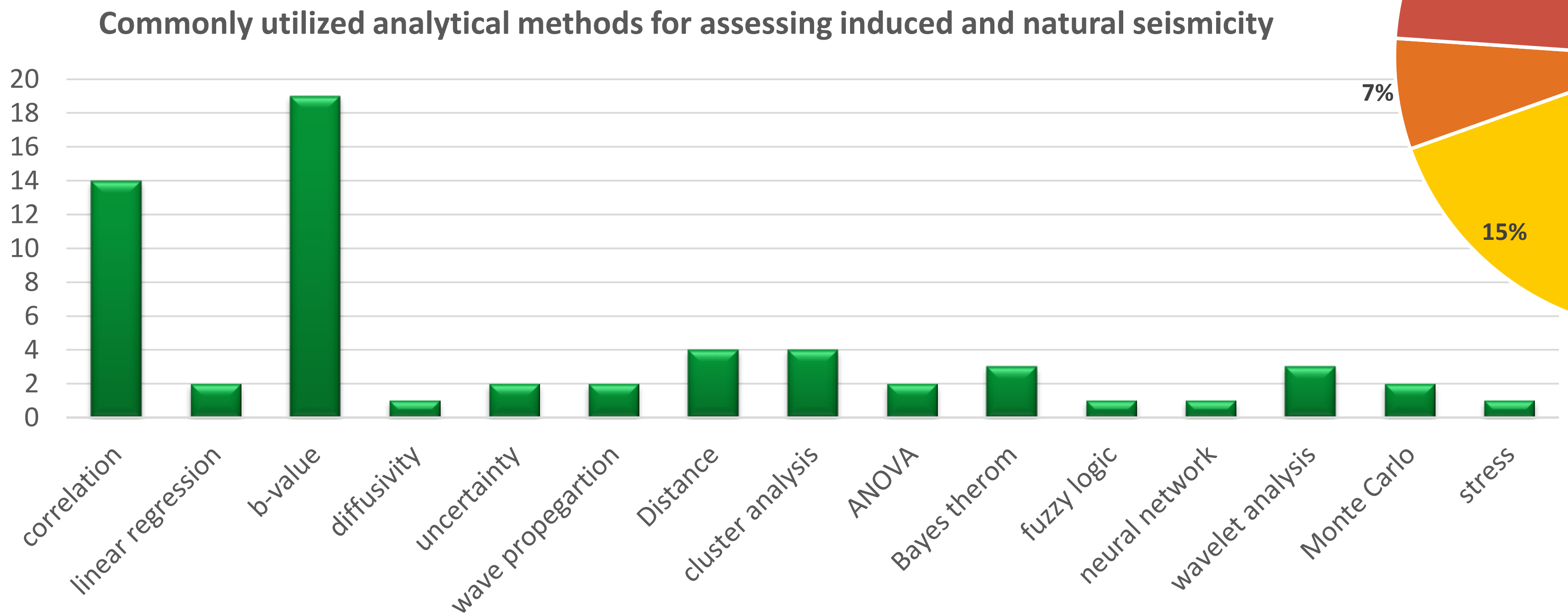
Source: Davies et al. 2013

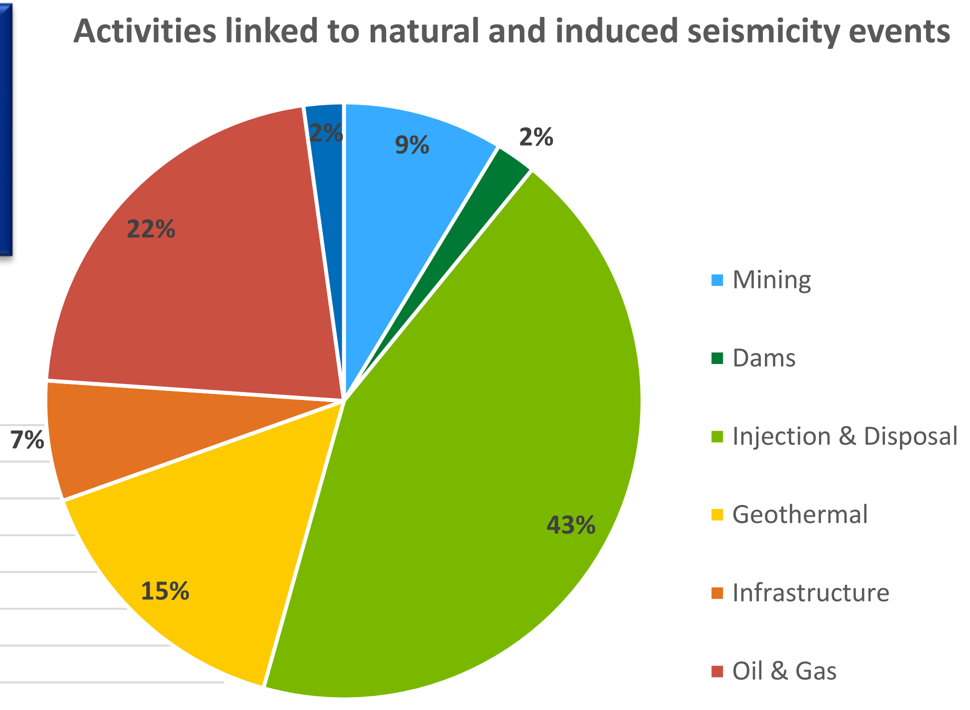
How do these activities trigger earthquakes?

- Changes in the state of stress
- Pore pressure changes
- Volume changes
- Applied forces and loads

Approach & Results - Methods for Evaluating and Analyzing Data

Identifying which method(s) are most appropriate for the data and use in the proposed probabilistic tool to assess risks of triggering induced seismicity events



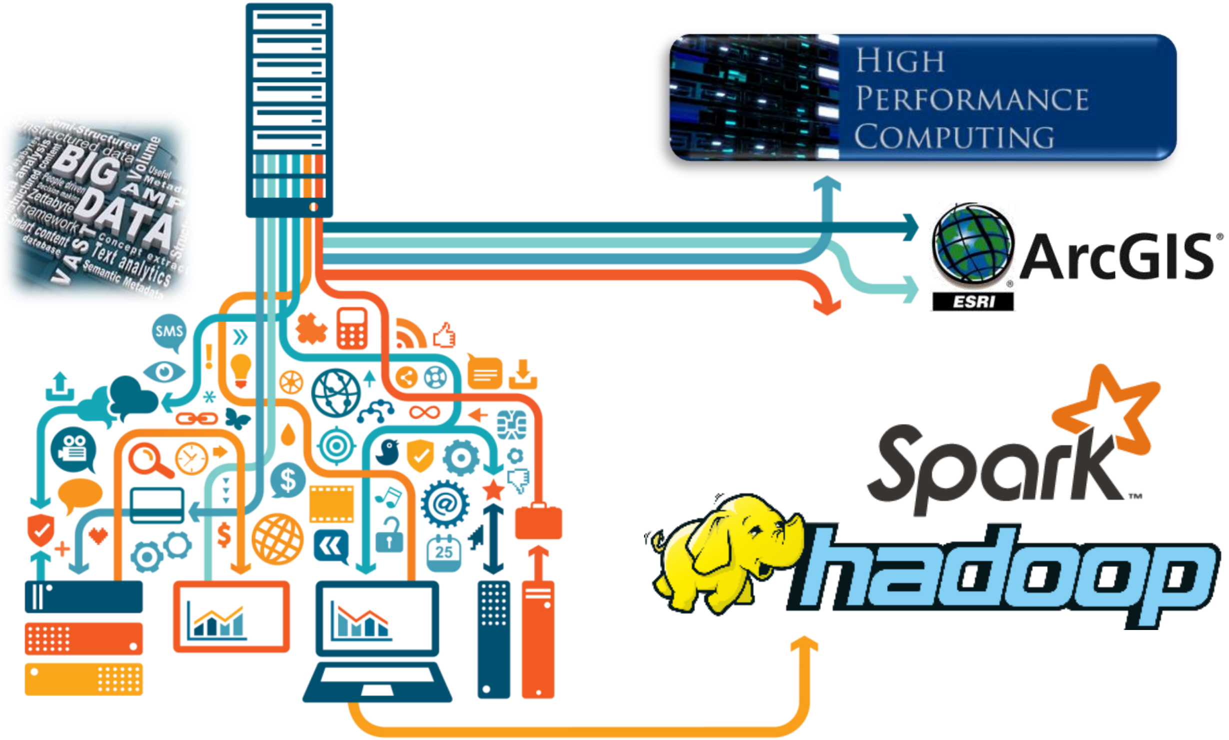


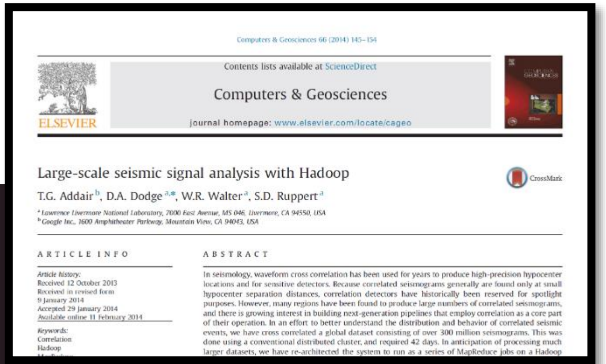
Building off Existing Knowledge

Approach - Adapting big data for geoprocessing

<u>THURSDAY, AUGUST 18, 2016</u>

1:00 PM Advances in Data Discovery, Mining, & Integration for Energy (EDX) – <u>Vic Baker</u>





Computers & Geosciences 66 (2014) 145–154
Contents lists available at ScienceDirect
Computers & Geosciences
journal homepage: www.elsevier.com/locate/cageo

Large-scale seismic signal analysis with Hadoop

T.G. Addair[b], D.A. Dodge[a,*], W.R. Walter[a], S.D. Ruppert[a]

[a] Lawrence Livermore National Laboratory, 7000 East Avenue, MS 046, Livermore, CA 94550, USA
[b] Google Inc., 1600 Amphitheater Parkway, Mountain View, CA 94043, USA

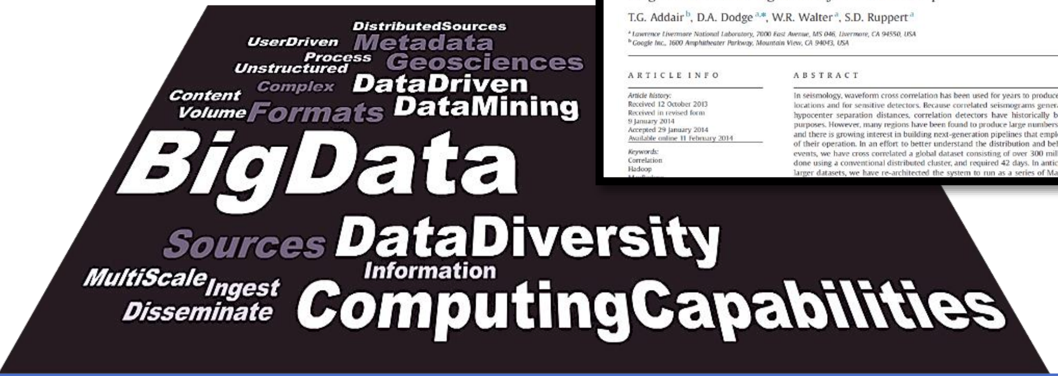
ARTICLE INFO

Article history:
Received 12 October 2013
Received in revised form
9 January 2014
Accepted 29 January 2014
Available online 11 February 2014

Keywords:
Correlation
Hadoop

ABSTRACT

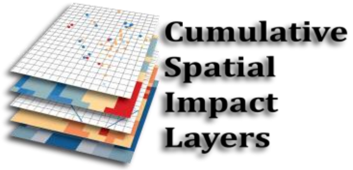
In seismology, waveform cross correlation has been used for years to produce high-precision hypocenter locations and for sensitive detectors. Because correlated seismograms generally are found only at small hypocenter separation distances, correlation detectors have historically been reserved for spotlight purposes. However, many regions have been found to produce large numbers of correlated seismograms, and there is growing interest in building next-generation pipelines that employ correlation as a core part of their operation. In an effort to better understand the distribution and behavior of correlated seismic events, we have cross correlated a global dataset consisting of over 300 million seismograms. This was done using a conventional distributed cluster, and required 42 days. In anticipation of processing much larger datasets, we have re-architected the system to run as a series of MapReduce jobs on a Hadoop



Big data + Geoprocessing will help:

- Expose geographic & temporal patterns
- Find spatial relationships
- Perform predictive modeling

Some results to date – Using NETL's CSIL for preliminary data analysis



Bauer et al., 2015

Cumulative Spatial Impact Layers (CSILs) is a spatio-temporal approach that identifies potential impacts to various socio-economic and environmental actives within a region

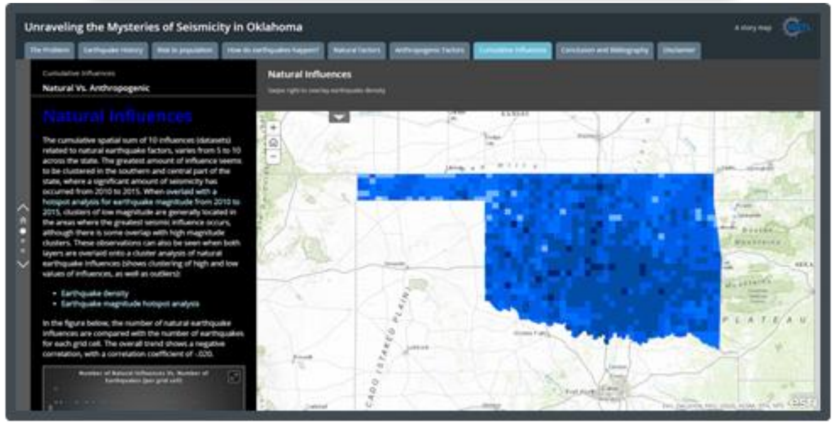
CSIL's quickly measures the total ***number of activities*** OR ***estimated "cost"*** per unit area (cell)

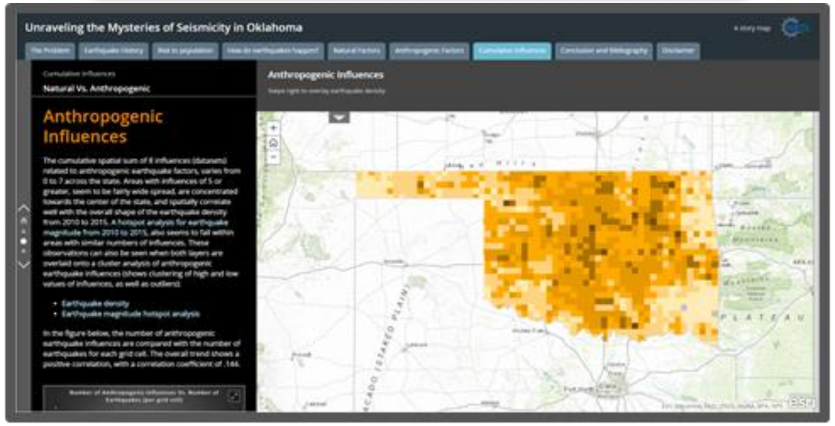
OK CSIL analysis is based on results of meta-analysis (lit review) and "seedling" data mining

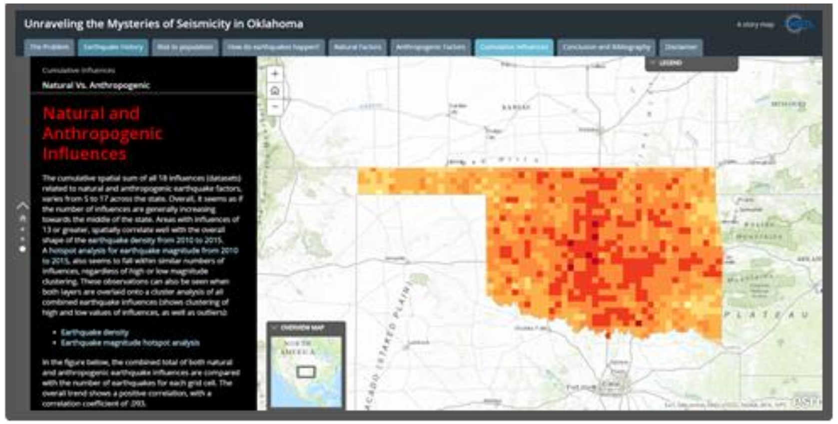
CSIL analysis of 10 key datasets relating to 5 **natural earthquake factors**

CSIL analysis 8 key datasets relating to 7 **anthropogenic earthquake factors**

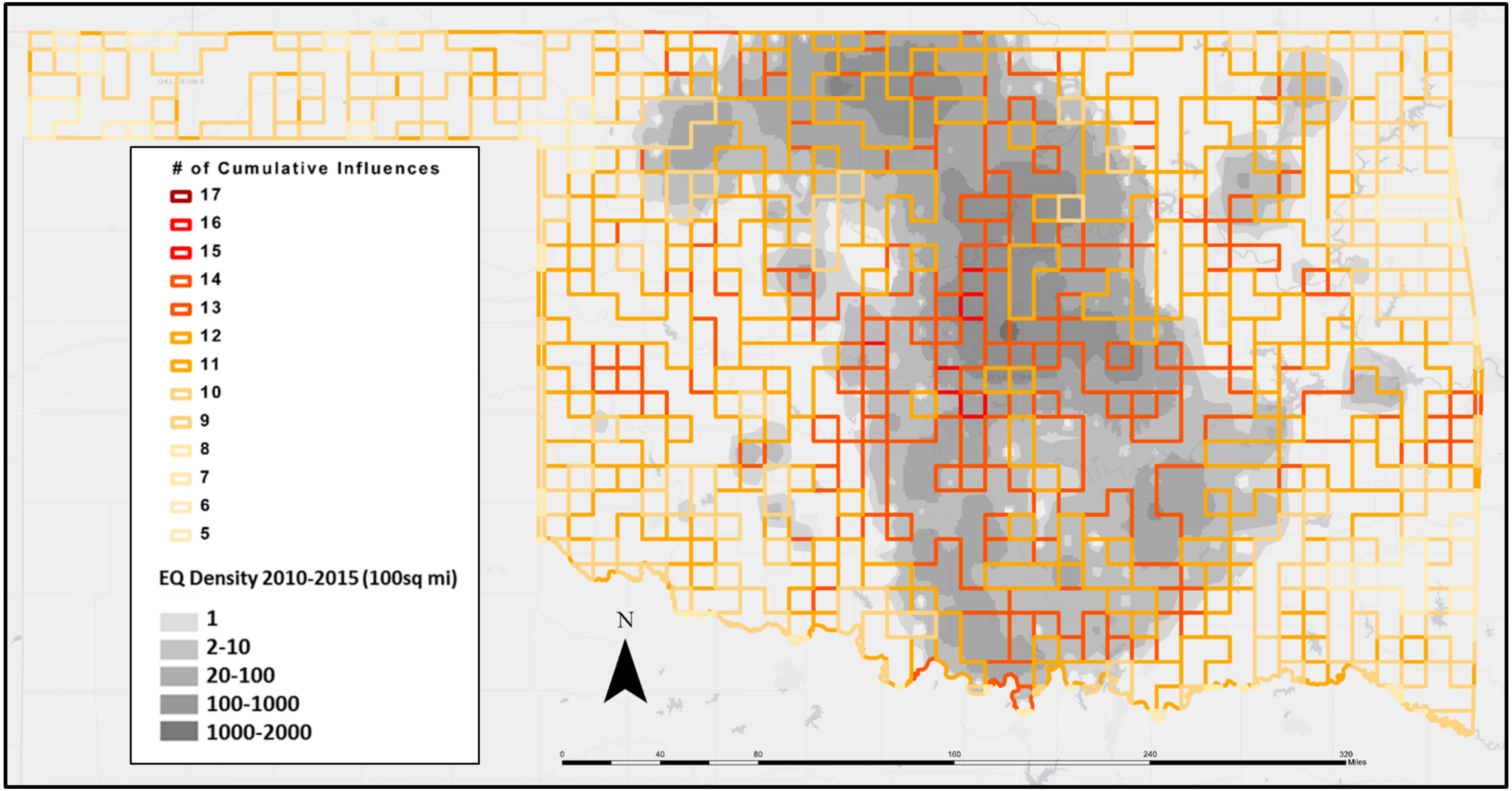
CSIL analysis of 18 key datasets related to **natural and anthropogenic earthquake factors**







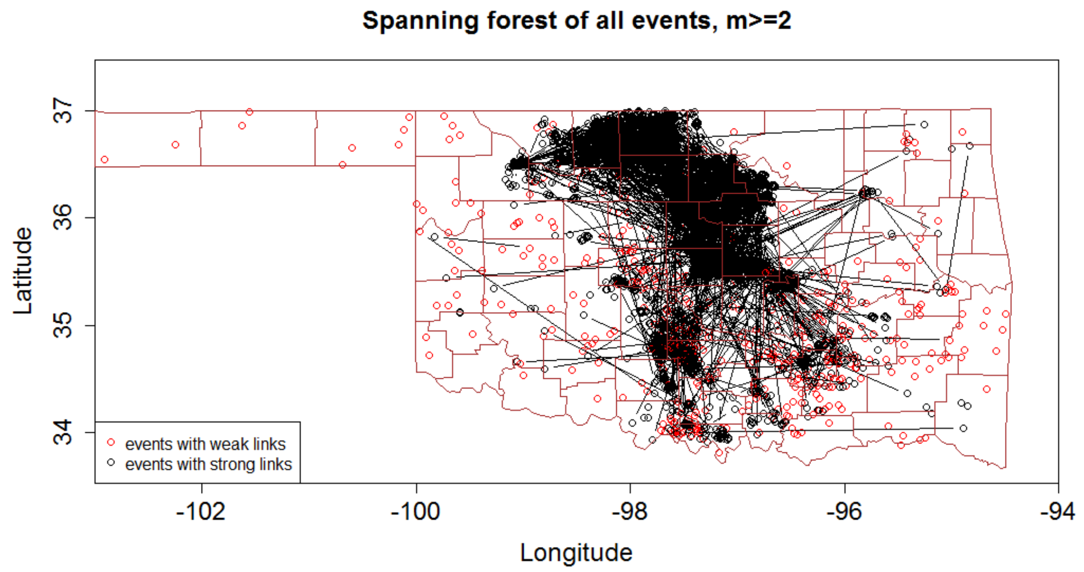
CSIL Analysis vs Earthquake Density in OK

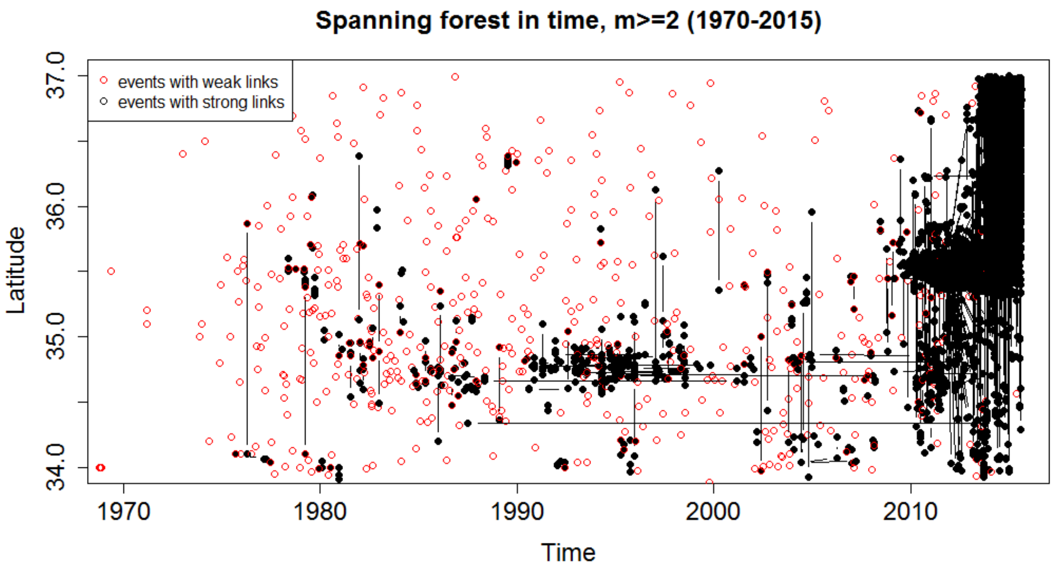


Results to date - Earthquake cluster analysis with modified NN approach

- Each earthquake j in the catalog is **connected to its nearest neighbor** (parent) i according to the NND η.
- All examined events form a single cluster: nearest-neighbor links **form a spanning network.**
- Spanning network is a **tree** (no loops).

Adapted from: Zaliapin I and Ben-Zion Y (2013), "*Earthquake clusters in southern California I: Identification and stability*", Journal of Geophysical Research: Solid Earth. Vol. 118(6), pp. 2847-2864. Wiley Online Library.





- **Weak links =** large distances between neighboring earthquake events, **strong links** = short distances
- Strong links form spanning forest (**collection of distinct trees spanning all events**)
- Two types of clusters: singles (single-event trees) and families (multi-event clusters).

Vasylkivska and Huerta, in prep

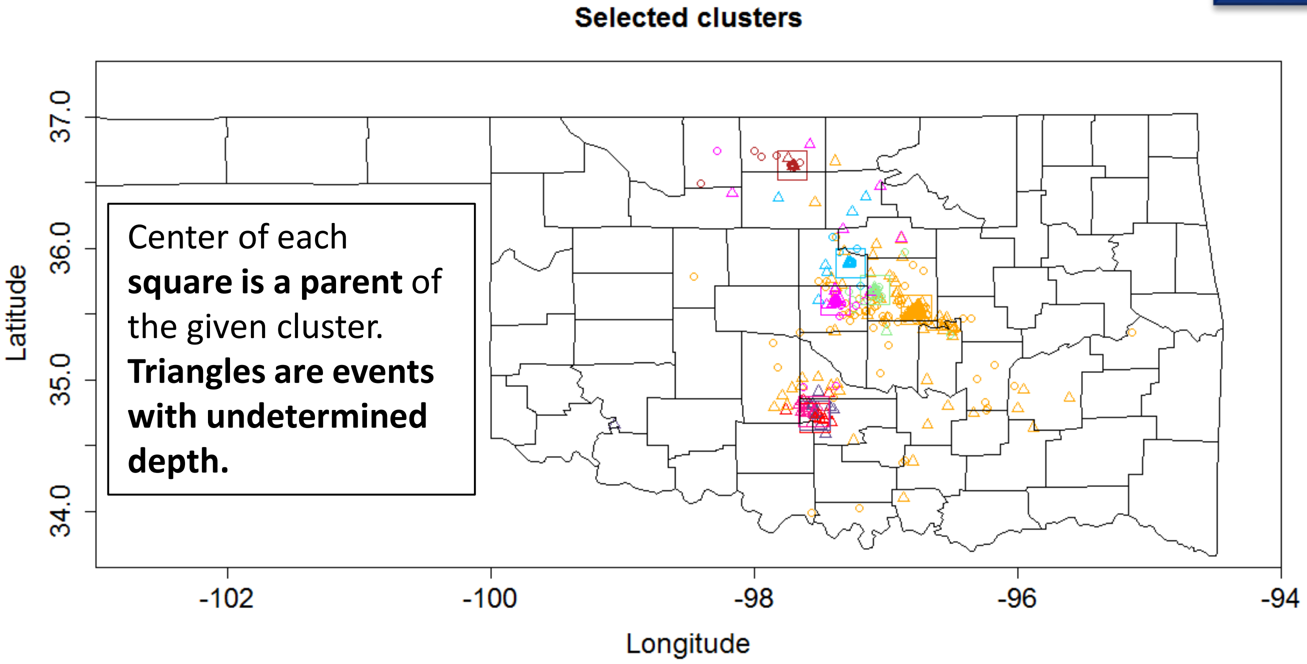
Evaluating Spatio-temporal Relations in 4D

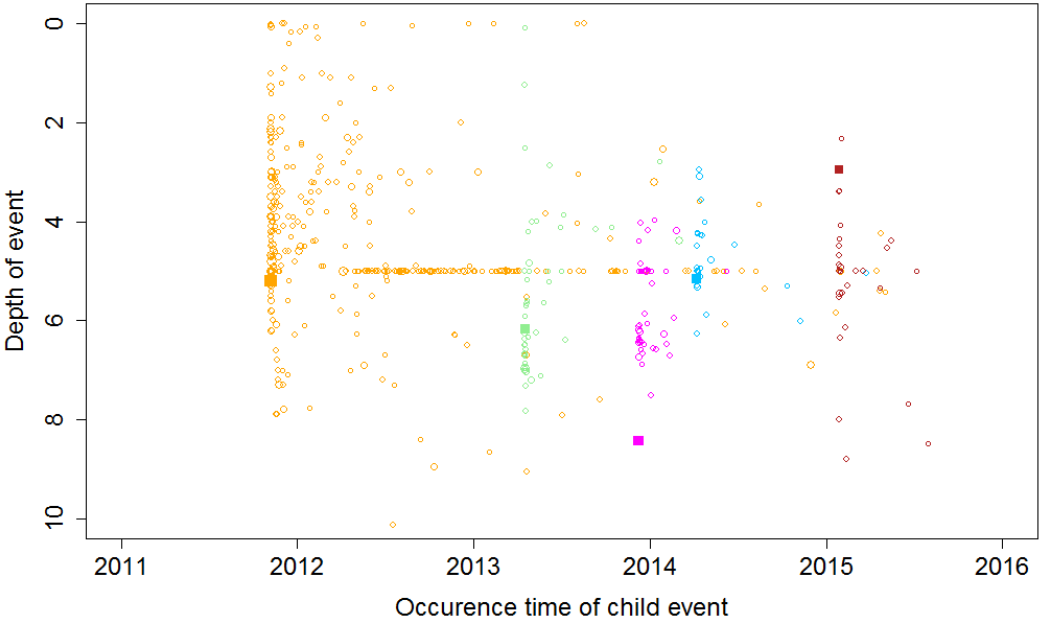
Vasylkivska and Huerta, in prep

Preliminary results in time and space (X,Y and Z)

The **largest clusters occurred in 2011-2015**.
Many events have undetermined depth of 5 km.

- Color scheme follows map.
- Many of the events in the **largest cluster have unknown depths**.
- Generally, for each cluster the **depth of the child events (empty circles) is smaller than the depth of the parent event (filled square).**





Example Results - Big Data Algorithms for Mining & Analysis

- **Search/mining algorithm, way to find, acquire and integrate open datasets from across the web**
- **NN cluster algorithm tested**
- **VGM uncertainty quant/viz algorithm**

hadoop

ArcGIS ESRI

Spark

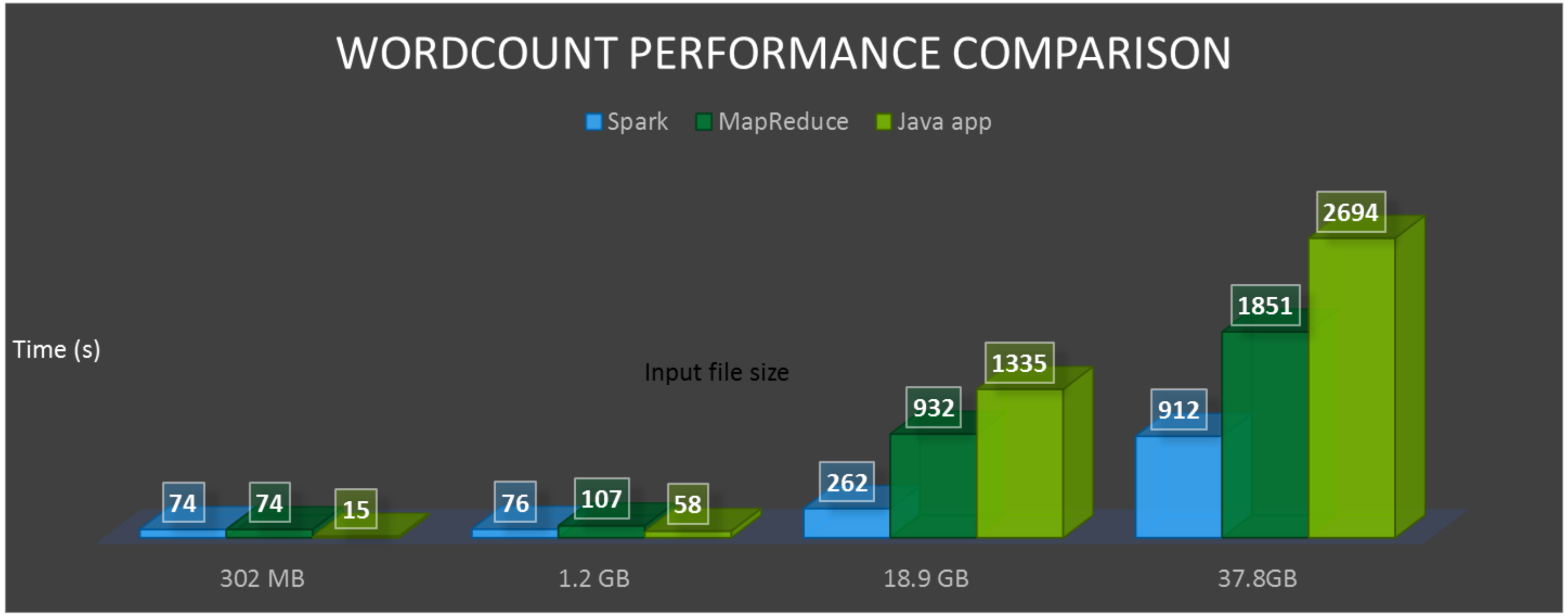
EDX Energy Data eXchange

- **Hadoop application development**
  - Hive, Java-based MapReduce 2.0, Spark, **ESRI**
- **ESRI / Hadoop development:**
  - Hive/ESRI development. Good for analysis / data mining)
  - 'Well API correlation' via Hive using spatial binning
  - Focus on Java-based MapReduce dev
- **Ongoing:**
  - **Major upgrades to EDX to improve mining, integration, analysis, and collaboration for DOE R&D community**
  - Spark application development
  - Elasticsearch add-on for **faster queries**

Example Results – big data processing time test

- **Compared execution times** for varying size data sets using **Hadoop** cluster-based **MapReduce** and **Spark** vs a stand alone, single threaded **Java** application (running on the Hadoop cluster's main node).
- **Spark's in-memory design outperform the single-threaded Java application**

Spark vs MapReduce vs Single Threaded Application



- Team succeeded in running the NN algorithm in the geoprocessing, big data cluster.
- **Time of execution went from 10 hours on desktop PC to 10 minutes**

Example Results – big data geoprocessing

Merging GIS and Big Data computing for advanced 3D/4D geospatial analysis

- **Offload intensive geometric operations** from desktop to a Hadoop cluster
- Is **highly scalable**
- **Self healing**
- The approach is **ideal for executing parallel operations on geometric operations involving many features**.

VARIABLE GRID METHOD

Hadoop-Based VGM Detailed Workflow

Well data from ArcMap

"hasM": false,
"spatialReference": {"wkid":4326},
"features": [
{
"attributes": {
"UWI__APINu": 3708520259.0,
"OR_Base_m_": 0.0,
"Surf_Lat": 41.484683,
"Salinity__": 0.0,
"WSN": 1.0,
"Surf_Lon": -80.103193,
"Brine_Dens": 0.0,
"OR_Gross_T": 0.0,
"Porosity__": 0.0,
"NET_THICKN": 0.0,
"Oriskany_T": 1190.549
},
"geometry": {
"y": 5084098.520442805,
"x": -8917047.03754837,
"spatialReference": {
"wkid": 4326,
"latestWkid": 4326
}
}
},
{
"attributes": {
"UWI__APINu": 3703920665.0,
"OR_Base_m_": 1077.9,
"Surf_Lat": 41.730652,

Example JSON from ArcMap

VGM-Step-0

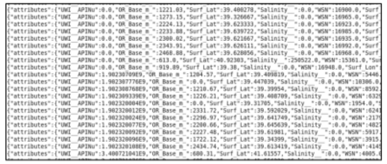
Description: Convert 'enclosed-Json' ESRI feature class into 'feature-per-row' unenclosed-Json.

Input: 'Enclosed-Json' formatted data (i.e., ORWells-wgs84.json) uploaded from ArcMap using ESRI/Hadoop toolbox tools 'Features to Json' & 'Copy to HDFS'.

Output: Processed 'Unenclosed-Json' with 'feature per row' layout suitable for Mapper.

Mapper (Setup): Create EsriFeatureClass from input file and write each feature as a row represented as unenclosed-Json.

Reducer: Aggregate Mapper output into one or more files



'Feature per row' formatted data for MapReduce

VGM-Step-1

Description: Generate multi-resolution bounding quads for input point data set (i.e., ORWells-wgs84)

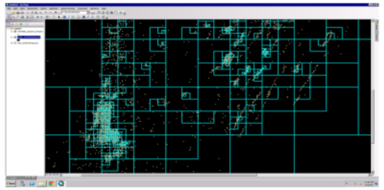
Input: vgm-step-0 output 'Unenclosed-Json' of row-per-feature representation of orwells-wgs84 data

Output: Quads of varying extents with attribution (i.e., point count, max/min/avg salinity, porosity, brine density)

Mapper (Setup): Load point features from vgm-step-0 and use to generate quadtree node extents.

Mapper: Feed mapper each row of 'unenclosed-Json' from vgm-step-0 point data and query the quadtree for all quads that contain

Reducer: Aggregate Mapper output into one or more files and store in vgm/working/output-0/.



Overlapping attributed quads (shown via ArcMap)

VGM-Step-2

Description: Generate non-overlapping topology of vgm-step-1 quads and calculate well point data per new geometries.

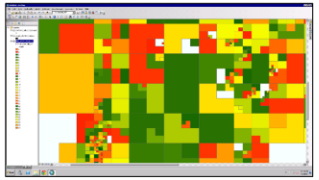
Input: Multi-resolution quads generated in vgm-step-1 AND the point data generated from vgm-step-0

Output: Non-overlapping polygons as 'unenclosed-Json' features with attribution (point count, min/max/avg porosity, etc.)

Mapper (Setup): Load vgm-step-1 output files representing attributed quads of varying resolutions to generate non-overlapping topology.

Mapper: Feed the Mapper with rows from the vgm-step-0 'unenclosed-Json' point feature data, query topology for 'point in polygon' to generate polygon's attributes, and perform geometry subtraction using ESRI Hadoop libs

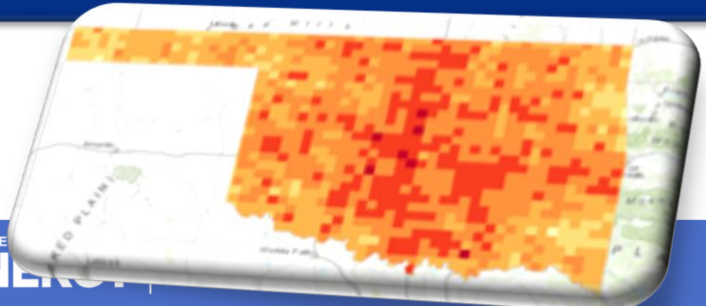
Reducer: Tally the attributes for each polygon and write attributed polygon as unenclosed-Json.

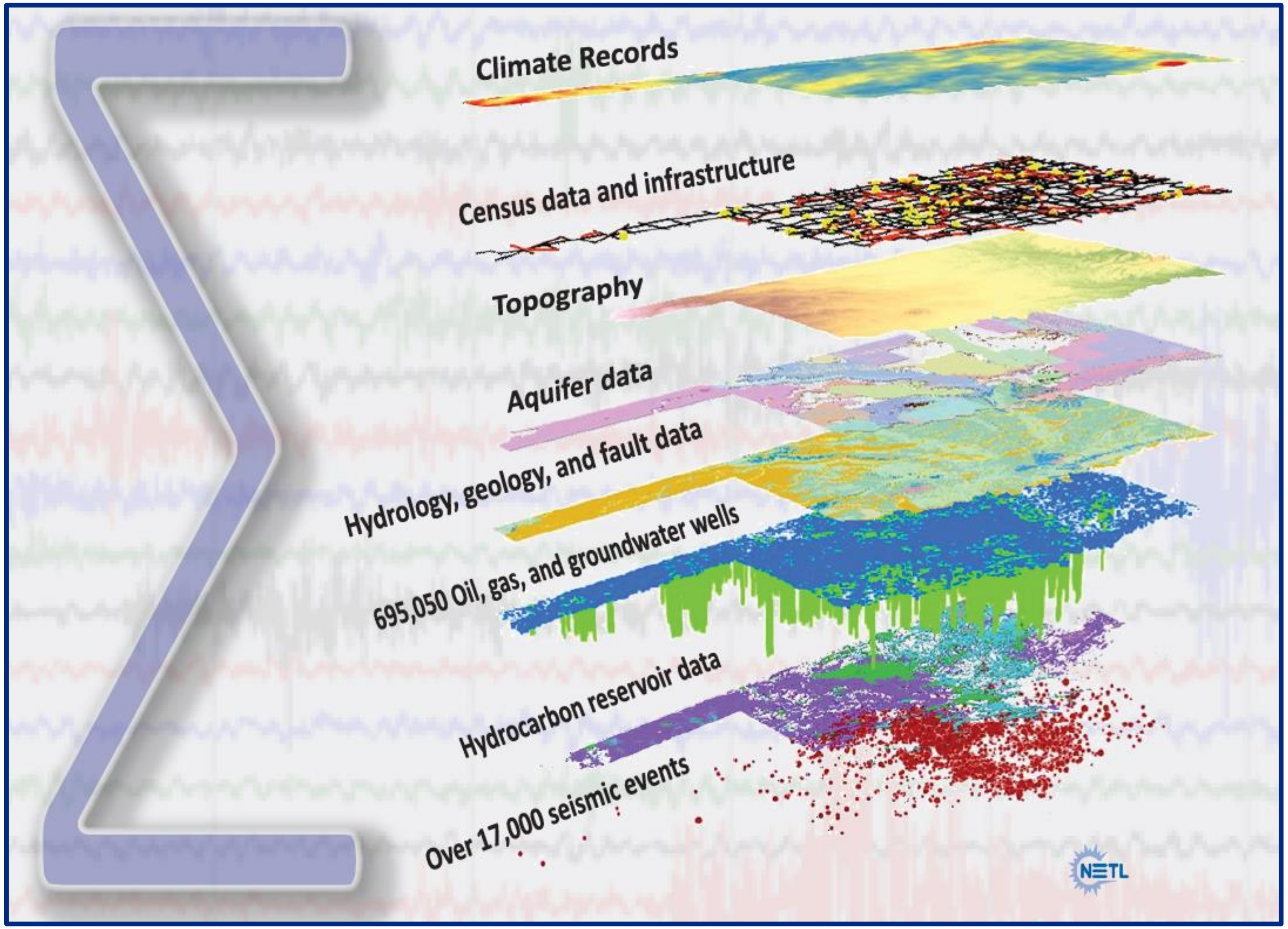


Attributed polygons for ArcMap

Tech Transfer & Accomplishments to Date

- >15 presentations & publications, more publications in prep
- Acquired & utilized millions of data records in 3D & 4D preliminary analyses
- Some analytical methods tested show promising results
- Completed review of advanced probabilistic methods, assessed strengths & weaknesses of each
- Developed novel big data mining, integration & analytical algorithms
- Integrating existing DOE computing capabilities with new BDC capabilities resulting in geoprocessing advances





Synergy Opportunities - Alignment to SubTER

- A probabilistic method to **assess potential of future induced seismicity events**
- New **multi-scale, multi-type analyses** for uncertainty reduction & spatial analysis
- Geoprocessing and **computing advances**

NRAP National Risk Assessment Partnership

UCR Unconventional Resources

Ties also to geothermal, waste disposal, unconventional, offshore, and carbon storage projects, tools, and data

Ties to pillars, elements & projects

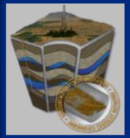
Intelligent Wellbore Systems	Subsurface Stress & Induced Seismicity	Permeability Manipulation	New Subsurface Signals
Improved well construction materials and techniques	Measurement of stress and induced seismicity	Physicochemical fluid-rock interactions	New sensing approaches
Autonomous completions for well integrity modeling	Manipulation of stress and induced seismicity	Manipulating flowpaths	Integration of multi-scale, multi-type data
New diagnostics for wellbore integrity	Relating stress manipulation and induced seismicity to permeability	Characterizing fractures, dynamics, and flows	Adaptive control processes
Remediation tools and technologies	Applied risk analysis of subsurface manipulation	Novel stimulation methods	Diagnostic signatures and critical thresholds
Fit-for-purpose drilling and completion tools (e.g. anticipative drilling, centralizers, monitoring)			
HT/HP well construction / completion technologies			

Energy Field Observatories

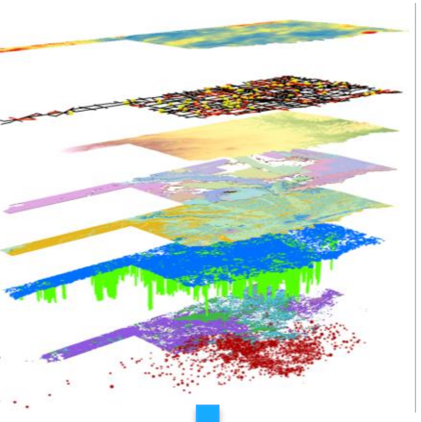
Fit For Purpose Simulation Capabilities

Anticipated next steps …

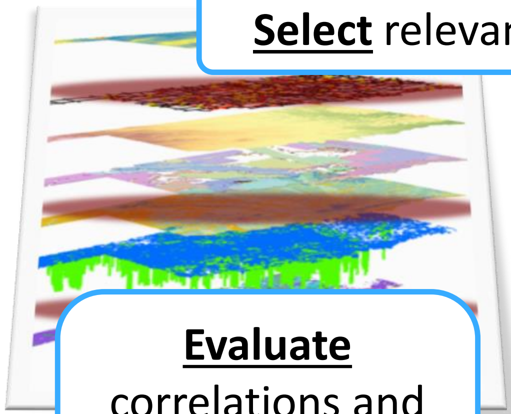
Ultimate product: Produce a platform with data, workflow, and tools to support efficient & repeatable probabilistic assessments to highlight regions with increased likelihood of induced seismicity



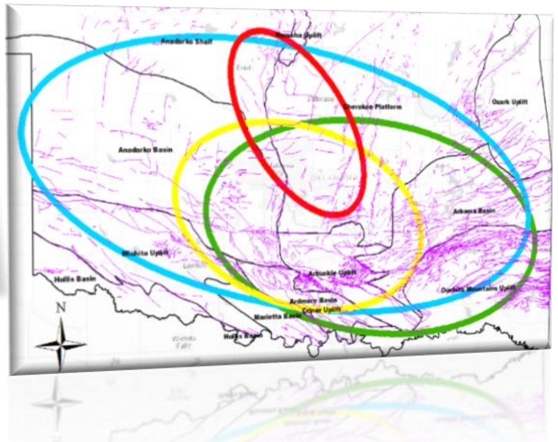
Access & Combine data and tools to…



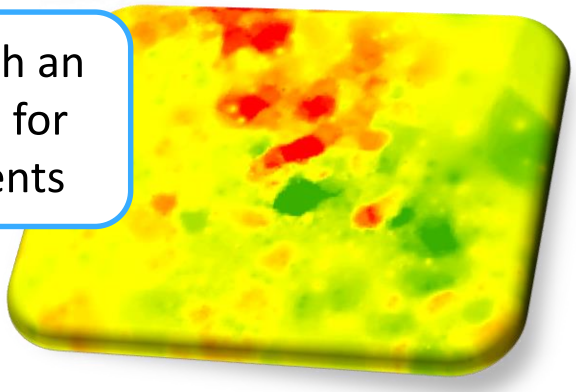
<u>Select</u> relevant datasets



<u>Evaluate</u> correlations and spatio-temporal trends



<u>Highlight</u> regions with an increased likelihood for induced seismic events



- Further develop, test & then refine **probabilistic approach**
- **Test** at different scales, other areas, e.g. Geysers
- **Validation**, address and test "why" we are seeing results we are seeing in test cases etc.
- Considerations on scale or evaluation & advanced **uncertainty analyses**
- **Computing**, continue growing and enhancing capabilities from big data computing world, HPC arena, and GIS domain
- Continue leveraging capabilities and tools from DOE efforts

Phase 1 Project Team:

1st Name	Last Name	Role	Organization
Kelly	Rose	Project PI & NETL POC	NETL
Jennifer	Bauer	Task 2 alt POC NETL	NETL
Vic	Baker	Task 1 alt POC NETL	NETL/Matric
Devin	Justman	GIS/Geologist	NETL
Veronika	Vasylkivska	Applied Math/Stats	NETL
Dennis	Donaldson	Structural Geology	NETL
Roy	Miller III	GIS/Geologist	NETL
Brian	Tost	Wellbore geophysics	NETL
Gary	Black	PNNL POC/Lead	PNNL
Chandrika	Sivaramakrishnan	Velo developer	PNNL
Zoe	Guillen	Velo developer	PNNL
Xingyuan	Chen	Velo, data assimilation etc	PNNL
Carina	Lansing	Velo developer	PNNL
Pavan	Balaji	Computer Scientist	ANL
Randy	Gentry	ANL POC/Lead	ANL
Joanne	Wendelberger	Group Leader Statistical Sciences	LANL
James	Ahrens	Data science at scale	LANL
Lawrence	Ticknor	Statistician	LANL
Divya	Banesh	Post-Doc	LANL

1st Name	Last Name	Role	Organization
Diane	Woodbridge	SNL POC/Lead computer scientist, data mining projects	SNL
Danny	Dunlavy	Tensor analysis	SNL
Dave	Cuyler	software eng. NGDS, CKAN, uncertainty	SNL
Randy	Brost	Data mining/Stats	SNL
David	Stracuzzi	Data mining/Stats	SNL
Kristina	Czuchlweski	data mining, stats, artificial intelligence	SNL
Stephen	Myers	Geophysics / Bayesloc	LLNL
Stanley	Ruppert	Geoscience HPC & Big Data	LLNL
Joe	Morris	Geophysics	LLNL
Rob	Mellors	Geophysics	LLNL
Steven	Magana-Zook	LLNL POC / Data scientist Developer	LLNL
Amberlee	Darold	OGS POC / Seismologist	OGS

For more information on the SubTER crosscut:

http://esd.lbl.gov/subter

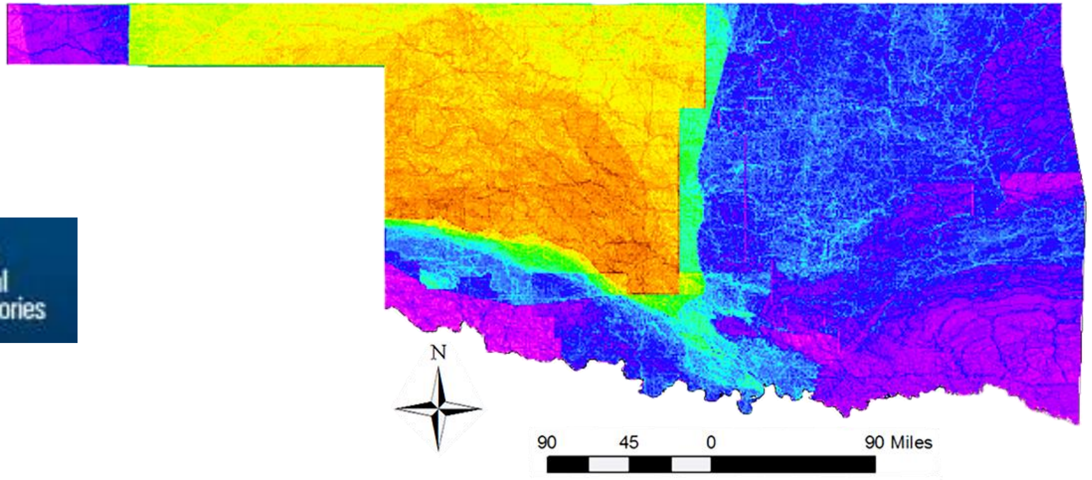
For more information on data and tools visit:

https://edx.netl.doe.gov

EDX A Product of NETL

Kelly Rose Geology & Geospatial Researcher

kelly.rose@netl.doe.gov



Acknowledgment: This technical effort was performed in support of the National Energy Technology Laboratory's ongoing research under the RES contract DE-FE0004000.

TUESDAY, AUGUST 16, 2016

- 12:40 PM Monitoring Groundwater Impacts - Christina Lopano
- 1:55 PM Multi Variate Examination of the Cause of Increasing Induced Seismicity – Kelly Rose
- 4:40 PM Exploring the Behavior of Shales as Seals and Storage Reservoirs for CO_2 – Ernest Lindner
- 5:05 PM Risk Assessment for Offshore Systems – Kelly Rose
- 5:30 PM Metal-based systems in Extreme Environments – Jeff Hawk
- 6:15 p.m. Poster Session
  - Kelly Rose - Developing a carbon storage resource assessment methodology for offshore systems
  - Doug Kauffman - Catalytic Conversion of CO2 to Ind. Chem. And eval. Of CO2 Use and Re-Use
  - Liwel Zhang - Numerical simulation of pressure and CO2 saturation above an imperfect seal as a result of CO2 injection: implications for CO2 migration detection

WEDNESDAY, AUGUST 17, 2016

- 12:30 PM MVA Field Activities – Hank Edenborn
- 1:20 PM Microseismicity –Erik Zorn
- 2:35 PM Resource Assessment – Angela Goodman
- 2:35 PM Understanding Impacts to Air Quality from Unconventional Natural Gas – Natalie Pekney
- 4:05 PM Improving Science-Base for Wellbore Integrity, Barrier Interface Performance – Nik Huerta
- 5:20 PM Wellbore Integrity and Mitigation – Barbara Kutchko

THURSDAY, AUGUST 18, 2016

- 1:00 PM Advances in Big Data Discovery, Mining, & Analysis for Energy (EDX) – Vic Baker
- 1:25 PM Methods for Locating Legacy Wells – Garrett Veloski
- 2:40 PM Reservoir Performance – Johnathan Moore
- 3:05 PM Geochemical Evolution of Hydraulically-Fractured Shales – Ale Hakala

Appendix

– These slides will not be discussed during the presentation, but are mandatory

29

Organization Chart

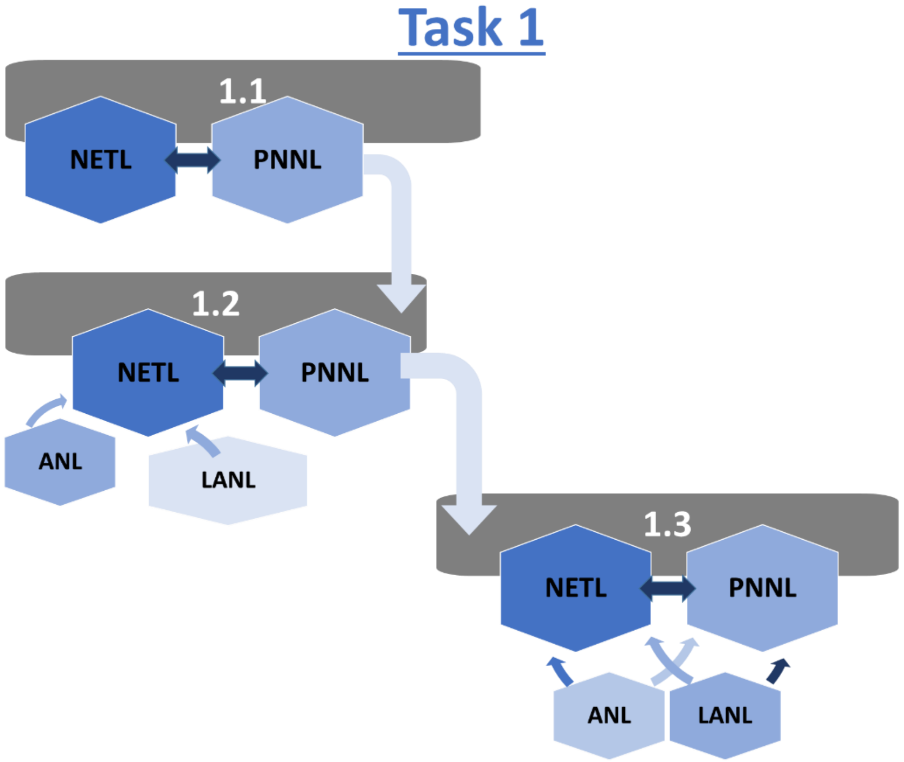
Task 1 – Geoscience computing advances

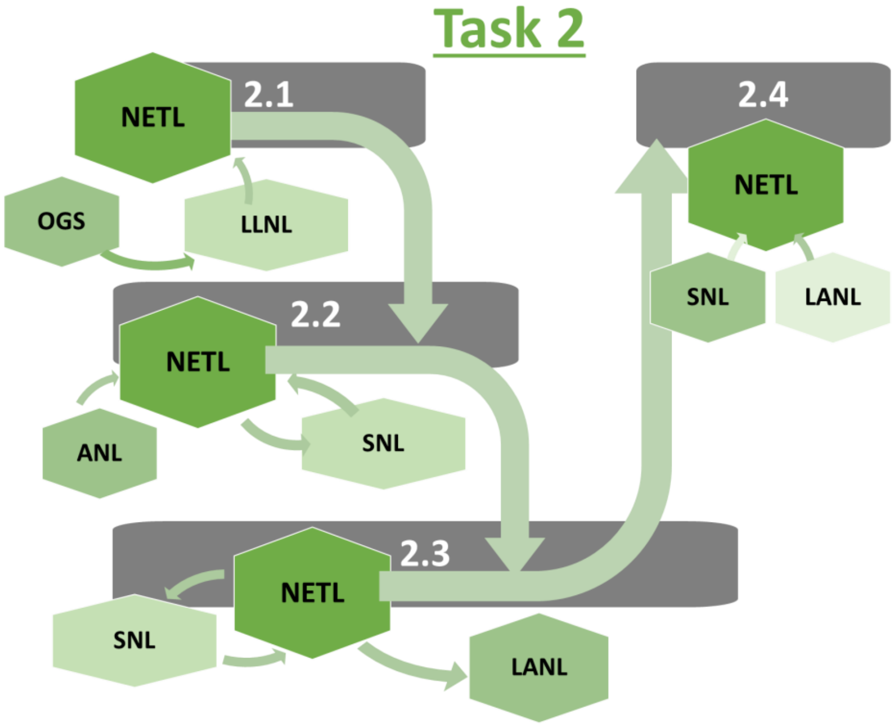
NETL (EDX team), LANL, & PNNL (Velo team)

ANL (in-kind, big data computing)

Task 2, Development of probabilistic approaches for induced seismicity

(LANL, LLNL, NETL, OGS, SNL)



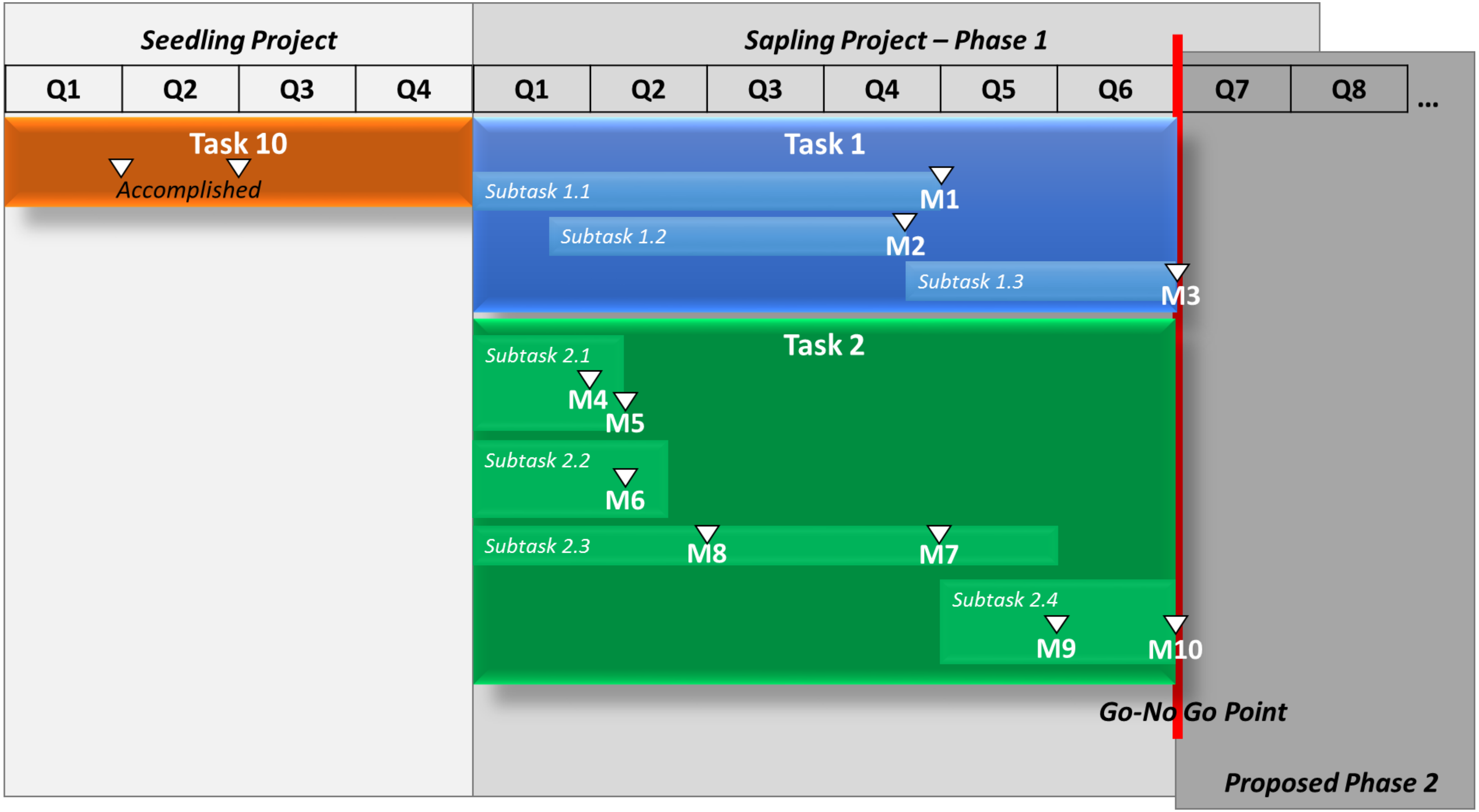


Organization Chart

Phase 1 Project Team:

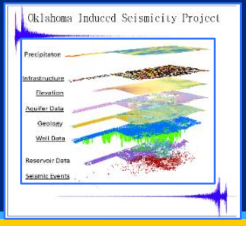
Task	Task	1st Name	Last Name	Role	Phone	Email	Organization
1	2	Kelly	Rose	Project PI & NETL POC	5419675883	kelly.rose@netl.doe.gov	NETL
1	2	Jennifer	Bauer	Task 2 alt POC NETL	5419184507	jennifer.bauer@netl.doe.gov	NETL
1		Vic	Baker	Task 1 alt POC NETL		vic.baker@matricinnovates.com	NETL/Matric
	2	Devin	Justman	GIS/Geologist	5419184561	devin.justman@netl.doe.gov	NETL
	2	Veronika	Vasylkivska	Applied Math/Stats	5419185964	veronika.vasylkivska@netl.doe.gov	NETL
	2	Dennis	Donaldson	Structural Geology	4123867344	dennis.donaldson@netl.doe.gov	NETL
	2	Roy	Miller III	GIS/Geologist	5419184561	roy.miller@netl.doe.gov	NETL
	2	Brian	Tost	Wellbore geophysics	5419184561	brian.tost@netl.doe.gov	NETL
1		Gary	Black	PNNL POC/Lead	5093752313	Gary.Black@pnnl.gov	PNNL
1		Chandrika	Sivaramakrishnan	Velo developer	5093726032	chandrika@pnnl.gov	PNNL
1		Zoe	Guillen	Velo developer	5093726857	zoe@pnnl.gov	PNNL
1		Xingyuan	Chen	Velo, data assimilation etc	509-371-7510	xingyuan.chen@pnnl.gov	PNNL
1		Carina	Lansing	Velo developer	5093752482	carina.lansing@pnnl.gov	PNNL
1		Pavan	Balaji	Computer Scientist		balaji@anl.gov	ANL
	2	Randy	Gentry	ANL POC/Lead		rgentry@anl.gov	ANL
	2	Joanne	Wendelberger	Group Leader Statistical Sciences		joanne@lanl.gov	LANL
1	2	James	Ahrens	Data science at scale		ahrens@lanl.gov	LANL
	2	Lawrence	Ticknor	Statistician		lot@lanl.gov	LANL
?	?	Divya	Banesh	Post-Doc		dbanesh@lanl.gov	LANL
	2	Diane	Woodbridge	SNL POC/Lead computer scientist, data mining projects		dwoodbr@sandia.gov	SNL
	2	Danny	Dunlavy	Tensor analysis		dmdunla@sandia.gov	SNL
		Dave	Cuyler	software eng. NGDS, CKAN, uncertainty			SNL
	2	Randy	Brost			rcbrost@sandia.gov	SNL
	2	David	Stracuzzi				SNL
	2	Kristina	Czuchlweski	data mining, stats, artificial intelligence		krczuch@sandia.gov	SNL
	2	Stephen	Myers	Geophysics / Bayesloc	9254234988	myers30@llnl.gov	LLNL
	2	Stanley	Ruppert	Geoscience HPC & Big Data	9254237552	ruppert1@llnl.gov	LLNL
	2	Joe	Morris	Geophysics	9254242263	morris50@llnl.gov	LLNL
	2	Rob	Mellors	Geophysics	9254230578	mellors1@llnl.gov	LLNL
	2	Steven	Magana-Zook	LLNL POC / Data scientist Developer	9254226007	maganazook1@llnl.gov	LLNL
	2	Amberlee	Darold	OGS POC / Seismologist	405/325-8611	adarold@ou.edu	OGS

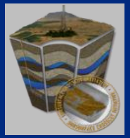
Gantt Chart



Seedling Project
Sapling Project – Phase 1
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
...
Task 10
Accomplished
Task 1
Subtask 1.1
M1
Subtask 1.2
M2
Subtask 1.3
M3
Task 2
Subtask 2.1
M4
M5
Subtask 2.2
M6
Subtask 2.3
M8
M7
Subtask 2.4
M9
M10
Go-No Go Point
Proposed Phase 2

Phase 1, Task 1 – Goals, Milestones, Schedule





Task 1 Research Leads -
Rose (PI & EDX Coordinator)
Baker (EDX Developer)
~~Gary Black~~ (Velo lead)

Task 1 – Geoscience computing advances

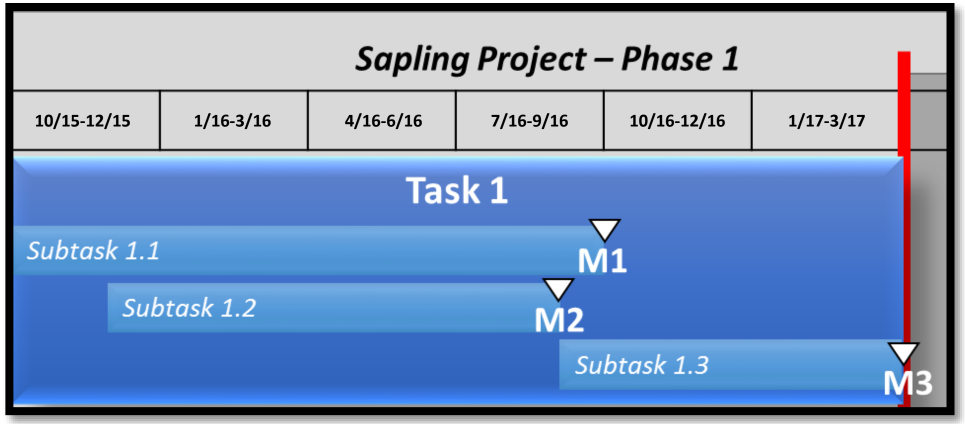
NETL (EDX team), LANL, & PNNL (Velo team)

ANL (in-kind, big data computing)

1.1 - Develop an EDX / Velo / Hadoop integration to support data gathering, mining, and analytical needs (NETL, PNNL)

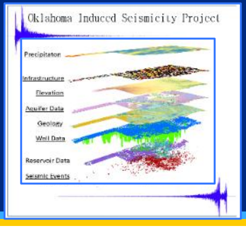
1.2 - Evaluate tools for irregular data management
(ANL, LANL, NETL, PNNL)

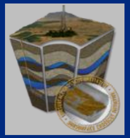
1.3 - Integrate irregular data management capabilities
(NETL, LANL, ANL, PNNL)



Phase 1, Task 1 - Milestone Summary Table						
Task No.	Task Title	MS Type	MS #	Milestone Description	Milestone Verification Process (Who, What, When, Where)	**Anticipated Date of Completion**
1.1	Develop an EDX / Velo / Hadoop integration to support data gathering, mining, and analytical needs for the induced seismicity use case	MS	1	Deploy baseline mining-fusion functionality via EDX	Quarterly report	Month 12
1.2	Evaluate ANL tools for irregular data management and fusion capabilities for integration within EDX	MS	2	Complete assessment of irregular data management options	Quarterly report	Month 11
1.3	Integrate preliminary irregular data management capabilities based on 1.2 recommendations/priorities	MS	3	Complete use case based testing of new EDX / Velo functionality	Quarterly report & EDX capabilities	Month 18

Task 2 – Goals, Milestones, Schedule





Task 2, Development of probabilistic approaches for induced seismicity (LANL, LLNL, NETL, OGS, SNL)

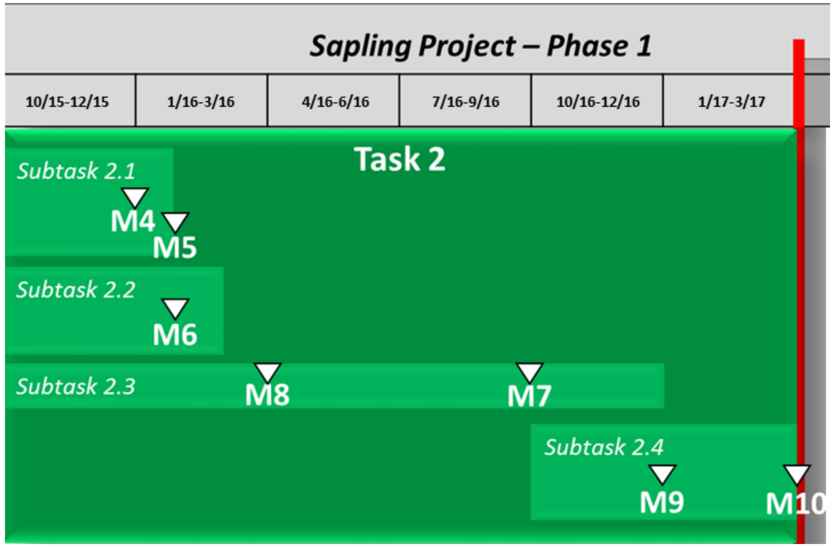
2.1 - Data processing, reduce uncertainty seismic events
(OGS, LLNL, NETL)

2.2 - Finalize data gathering & integration (
NETL, ANL, SNL)

Task 2 Research Leads -
Rose (PI)
Bauer (GIS/Geospatial)

2.3 - Probabilistic analyses and uncertainty sources
(NETL, SNL, OSU, LANL)

2.4 - Develop beta version of stochastic approach
(NETL, SNL, LANL)



Milestone Summary Table

Task #	MS #	Milestone Description	Milestone Verification	Anticipated Date of Completion
2.1	4	Complete data conversion of OK seismic records	Quarterly report	Month 3
2.1	5	Complete calculation of hypocenters for historical OK seismic data…	Quarterly report	Month 4
2.2	6	Gathering and integration of high resolution climate data and other key datasets from Southern Plains ARM facility	Quarterly report	Month 4
2.3	7	Evaluate adapting NETL's VGM to pair with LANL uncertainty propagation approaches…	Quarterly report	Month 12
2.3	8	Test spatio-temporal workflow for assessing probability of induced seismicity risk …	Quarterly report	Month 6
2.4	9	Incorporate findings from NL's stochastic analyses to develop a workflow for IS events	Quarterly report	Month 15
2.4	10	Develop beta version of a stochastic tool for IS	Quarterly report	Month 18

Bibliography

- Baker, D.V., Rose, K., Bauer, J., and Rager, D., *accepted*, Computational Advances and Data Analytics to Reduce Subsurface Uncertainty, *ARMA 16-493*, June 26-29, 2016, 16 pgs.
- Baker, V., Rose, K., Bauer, J., and D. Rager. Computational advances and data analytics to reduce subsurface uncertainty, American Rock Mechanics Association (ARMA) Symposium, Houston, TX, June 26-29, 2016. http://armasymposium.org/
- Baker, D.V. "Vic", Bauer, Jennifer, Rose, Kelly. Big Data Geo-Analytical Tool Development for Spatial Analysis Uncertainty Visualization and Quantification Needs. American Geophysical Union Fall Meeting, December 14-18, 2015. http://fallmeeting.agu.org/2015/.
- Bauer, J. R., Rose, K., Vasylkivska, V., Justman, D., Glosser, D., Miller, R., and D. V. Baker. Big data and spatio-temporal approaches for multivariate analysis, Esri User Conference, San Diego, CA, June 27-July 1, 2016. http://www.esri.com/events/user-conference
- Bauer, J., Rose, K., Romeo, L., Nelson, J., and Dick, D. Analyzing multisource and multidimensional data to evaluate spatio-temporal trends and patterns of potential impacts related to modelled offshore hazards, CoDA conference, Santa Fe, MN, March 2-4 2016. http://www.cvent.com/events/coda-2016-conference-on-data-analysis-2016/event-summary-a11ed42531524891a3ebeb626147a980.aspx
- Justman, D., Rose, K., Bauer, J., Miller, R., and Vasylkivska, V. Evaluating the Mysteries of Seismicity in Oklahoma, San Diego, CA, June 27-July 1, 2016. http://www.esri.com/events/user-conference
- Justman, D., Rose, K., Bauer, J., Miller, R., Vasylkivska, V., Romeo, L., *in review,* Communicating and Evaluating the Causes of Seismicity in Oklahoma Using ArcGIS Online Story Map Web Applications. American Geophysical Union Fall Meeting, San Francisco, CA, December 12-16, 2016. http://fallmeeting.agu.org/2016/
- Rose, K. and Bauer, J. Wrangling Big Data to herd Geoprocessing for Analysis of Engineered-Natural Systems. Quality and Productivity Research Conference (QPRC), Tempe, AZ, June 14-16, 2016. *Invited Session & Presentation*. http://www.qprc2016.com/home.html
- Rose, K., Justman, D., Bauer, J., Baker, D.V., Miller III, R., Vasylkivska, V., Guillen, Z., Wendelberger, J., and Woodbridge, D., 2016, A multi-variate examination of the cause(s) of increasing induced seismicity events using geoscience computing & big data techniques, SPE/SEG Workshop: Injection Induced Seismicity – Engineering Integration, Evaluation and Mitigation, 28 - 30 March 2016, Fort Worth, Texas, USA | Omni Fort Worth http://www.spe.org/events/16jftw/
- Rose, K., Bauer, J., M., Baker, D.V., Miller III, R., and Vasylkivska, V., 2016, Innovating beyond today's GIS, advanced tools and analysis for energy-geosystems. Esri Federal GIS Conference & Developers Summit, Washington, D.C., February 24-26, 2016. http://www.esri.com/events/federal
- Rose, K., 2016, A multi-variate geospatial approach to constrain subsurface properties and reduce uncertainty, CoDA conference, Santa Fe, MN, March 2-4 2016. http://www.cvent.com/events/coda-2016-conference-on-data-analysis-2016/event-summary-a11ed42531524891a3ebeb626147a980.aspx
- Rose, K., Bauer, J., Baker, D. V., Justman, D., Miller, R., and A. Tiwari, 2015, A Multi-variate examination of the cause(s) of increasing induced seismicity events using geoscience computing & big data techniques, Carbon Storage R&D Project Review Meeting, Pittsburgh, PA, August 18-20, 2015. http://www.netl.doe.gov/events/2015-carbon-storage-rd
- Ticknor, T., and Casleton, E., 2016, Big Data for Engineered-Natural Systems, Quality and Productivity Research Conference (QPRC), Tempe, AZ, June 14-16, 2016. *Invited Session & Presentation*. http://www.qprc2016.com/home.html
- Vasylkivska, V., Huerta, N., *in review,* Cluster analysis and fault identification from earthquake catalog. American Geophysical Union Fall Meeting, San Francisco, CA. December 12-16, 2016. http://fallmeeting.agu.org/2016/
- Vasylkivska, V., Huerta, N., Rose, K., Bauer, J., Probabilistic analysis of earthquake occurrence in Oklahoma, CoDA conference, Santa Fe, MN, March 2-4 2016. http://www.cvent.com/events/coda-2016-conference-on-data-analysis-2016/event-summary-a11ed42531524891a3ebeb626147a980.aspx