

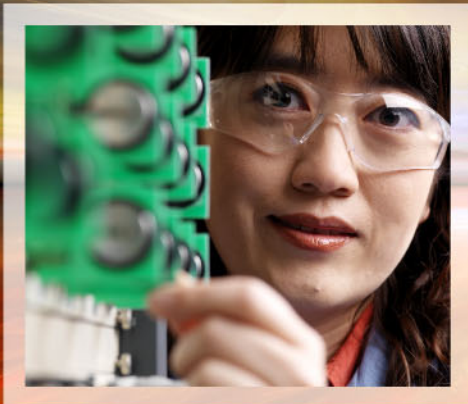


*IEEE PES AMPS Big Data Subcommittee  
Webinar  
October 10, 2018*

## Big Data Access, Analytics and Sense-Making



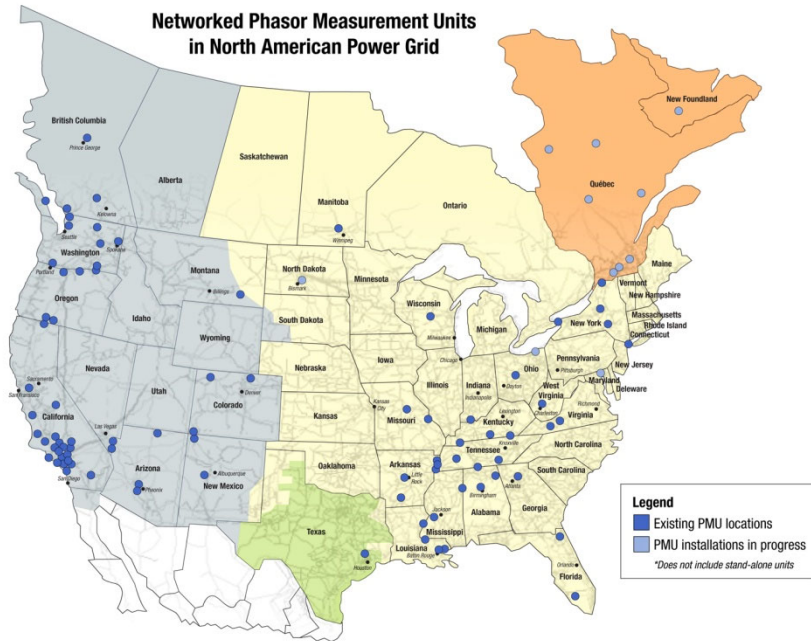
**Zhenyu (Henry) Huang, Ph.D., P.E., F.IEEE**  
Laboratory Fellow/Technical Group Manager  
Pacific Northwest National Laboratory (PNNL)



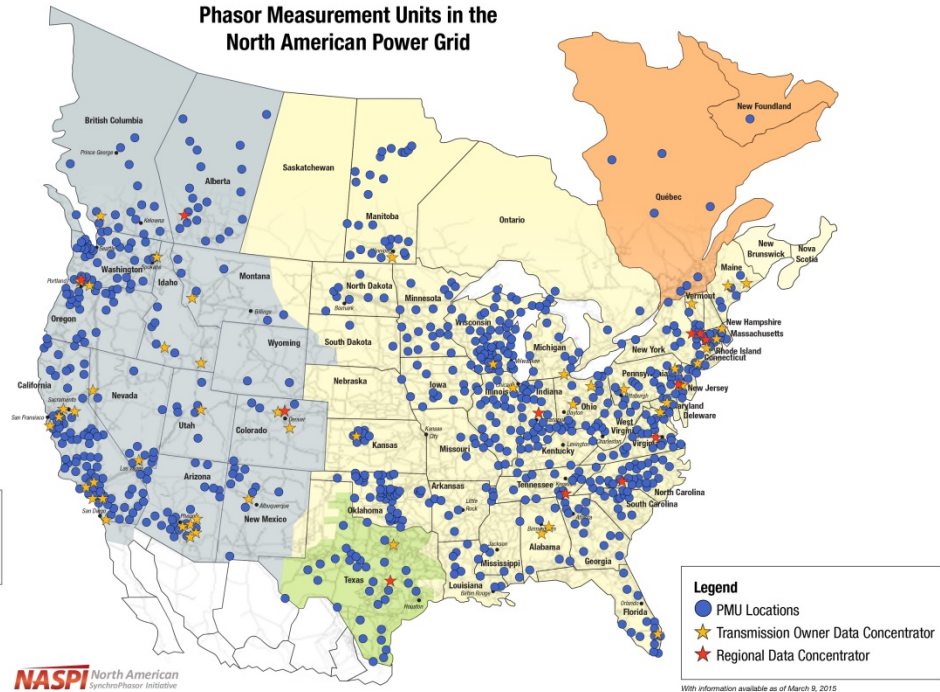
**DISCOVERY**  
*in action*

# Deployment of a vast new phasor network is generating unprecedented real-time data

**April 2007**



**March 2015**

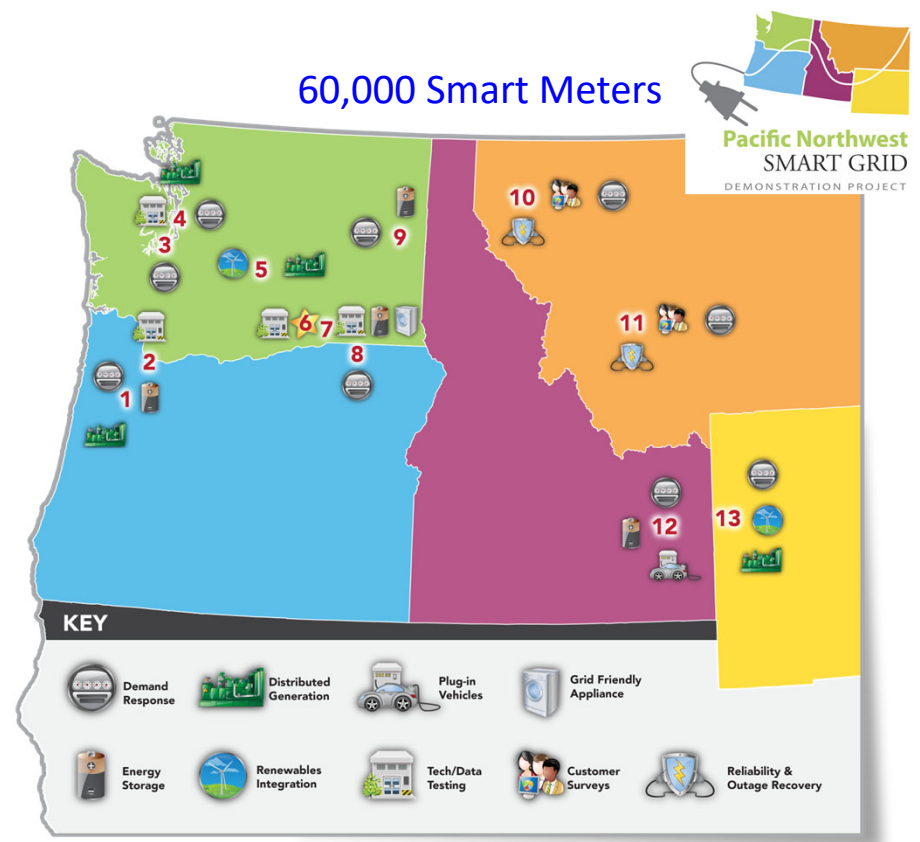
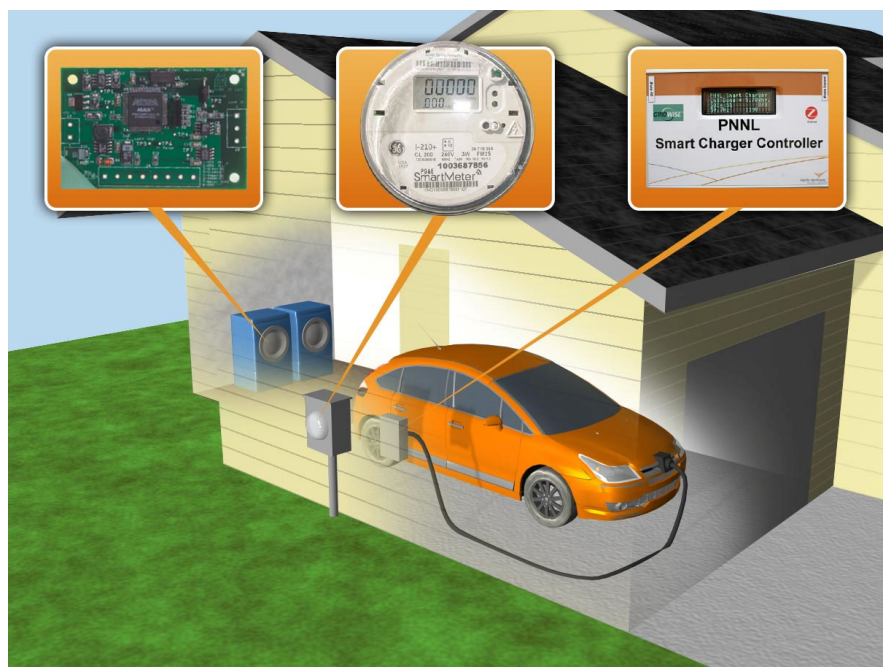


**NASPI** North American SynchroPhasor Initiative

	Today – SCADA data	Emerging – phasor data	Improvement
<b>Variety</b>	voltage + current	+ phase angle, ...	more information
<b>Velocity</b>	1 sample / 4 seconds	30-120 samples / second	~200x faster
<b>Volume</b>	8 terabytes / year	1.5 petabytes / year	~200x more data
<b>Veracity</b>	unseen ms-oscillations	oscillations seen at 10ms	greater accuracy



# Smart devices and 2-way communication offer new opportunities, greater complexity



Number of homes	100	1k+	10k+	100k+	500k+	1 Million
Compressed data size	2.5 GB	38.5 GB	366.3 GB	2.9 TB	13.6 TB	27.3 TB

# More diverse data add to the complexity

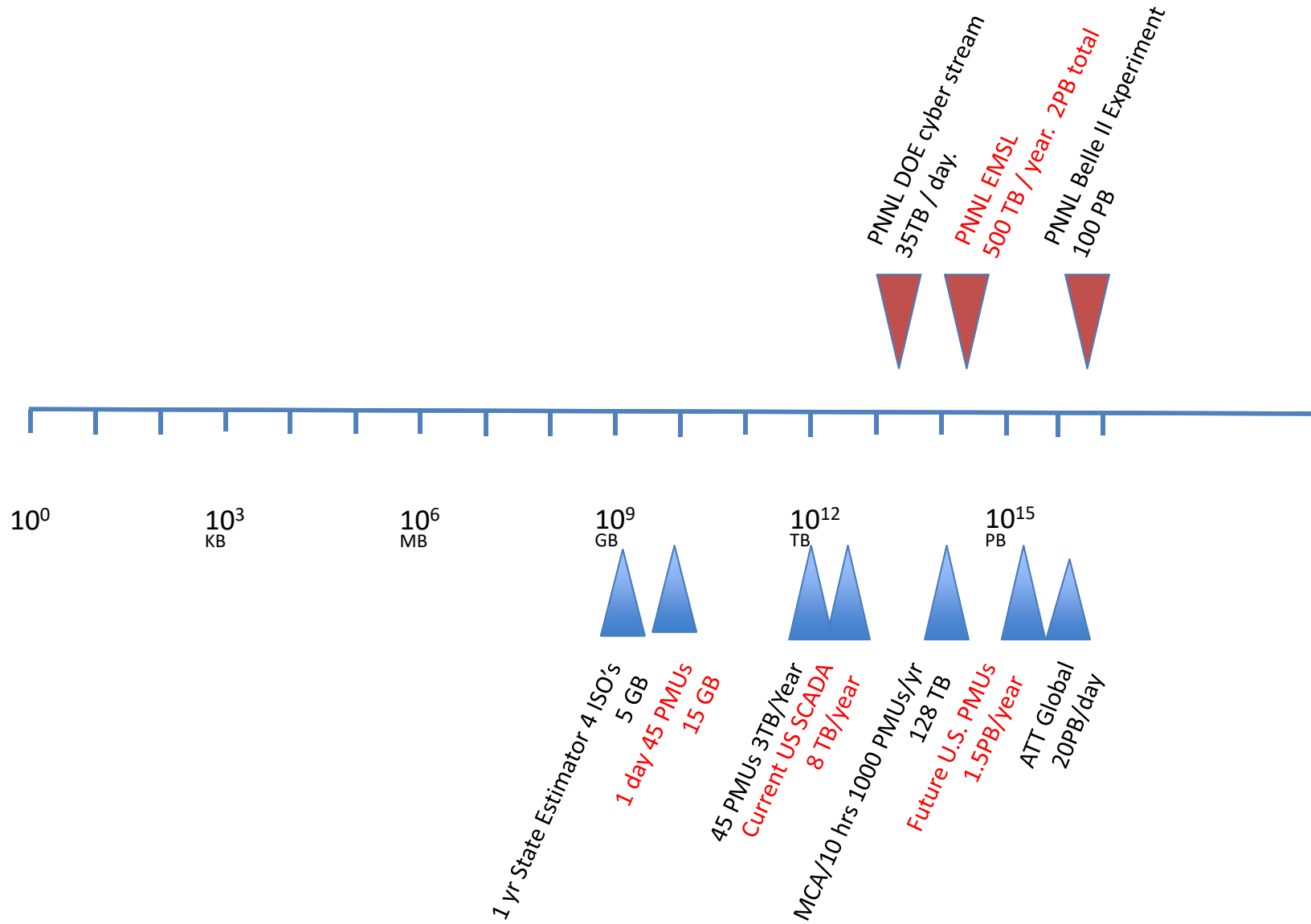
- ▶ Weather/climate data  
(e.g. PNNL ARM\* Data)
  - 300 instruments, 2000 data streams 24/7
  - 500 GB/day rising to multiple TBs/day
  - Curating 20 years' data
- ▶ Market/business data
- ▶ Cyber/communication data
- ▶ Simulated data
  - Each contingency scenario generates 0.5M bytes data, adding up to TB scale



\*ARM: Atmospheric Radiation Measurement

Contingency Analysis	Number of scenarios	Serial computing on 1 processor	Parallel computing on 512 processors	Parallel computing on 10,000 processors
WECC N-1 (full)	20,000	4 hours	~30 seconds 469x speed up	
WECC N-2 (partial)	153,600	26 hours	~3 minutes 492x speed up	~12 seconds 7877x speed up

# Data volume comparison: grid vs. big science data



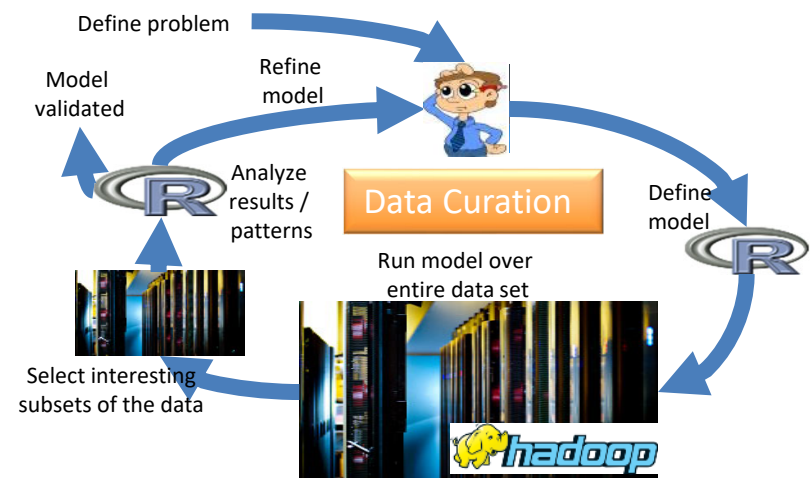
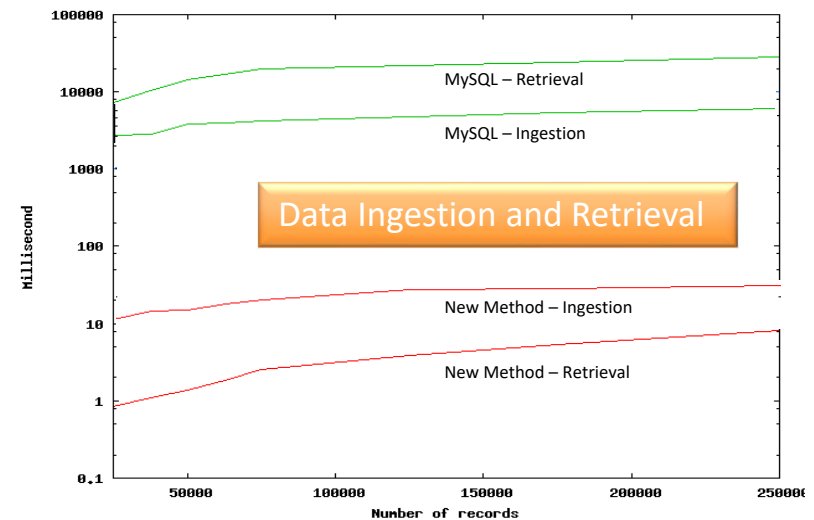
# Real-time data ingestion, retrieval, curation from a distributed sensor network

## ► Requirements

- Cyber-secure sensor network
- Data provenance and privacy
- Real-time processing

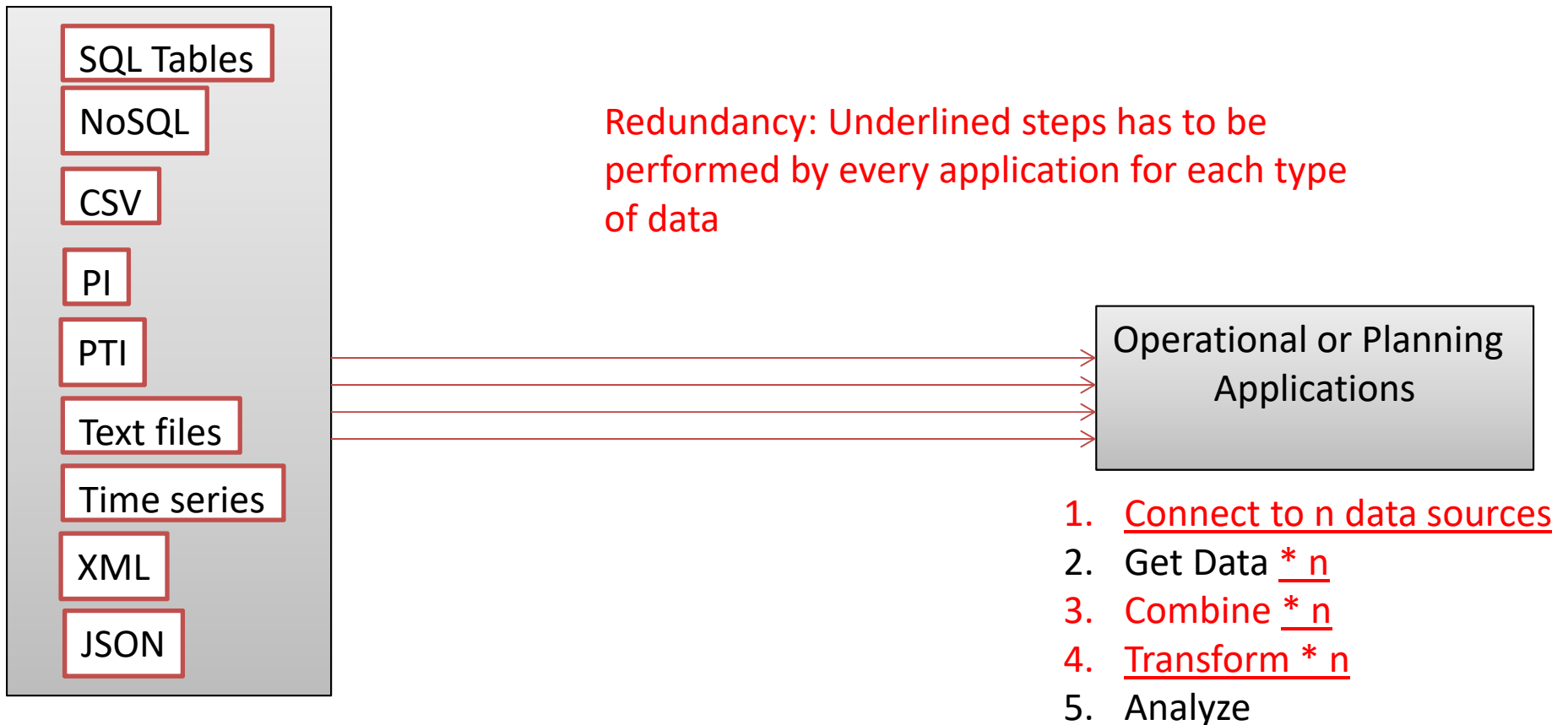
## ► Solution: scalable, flexible middleware and R/Hadoop statistical analysis capabilities

- Data ingestion is  $10^3$  times faster than MySQL
- Linearly scales to many nodes
- Data curation cleans data and detect events with confidence in real-time



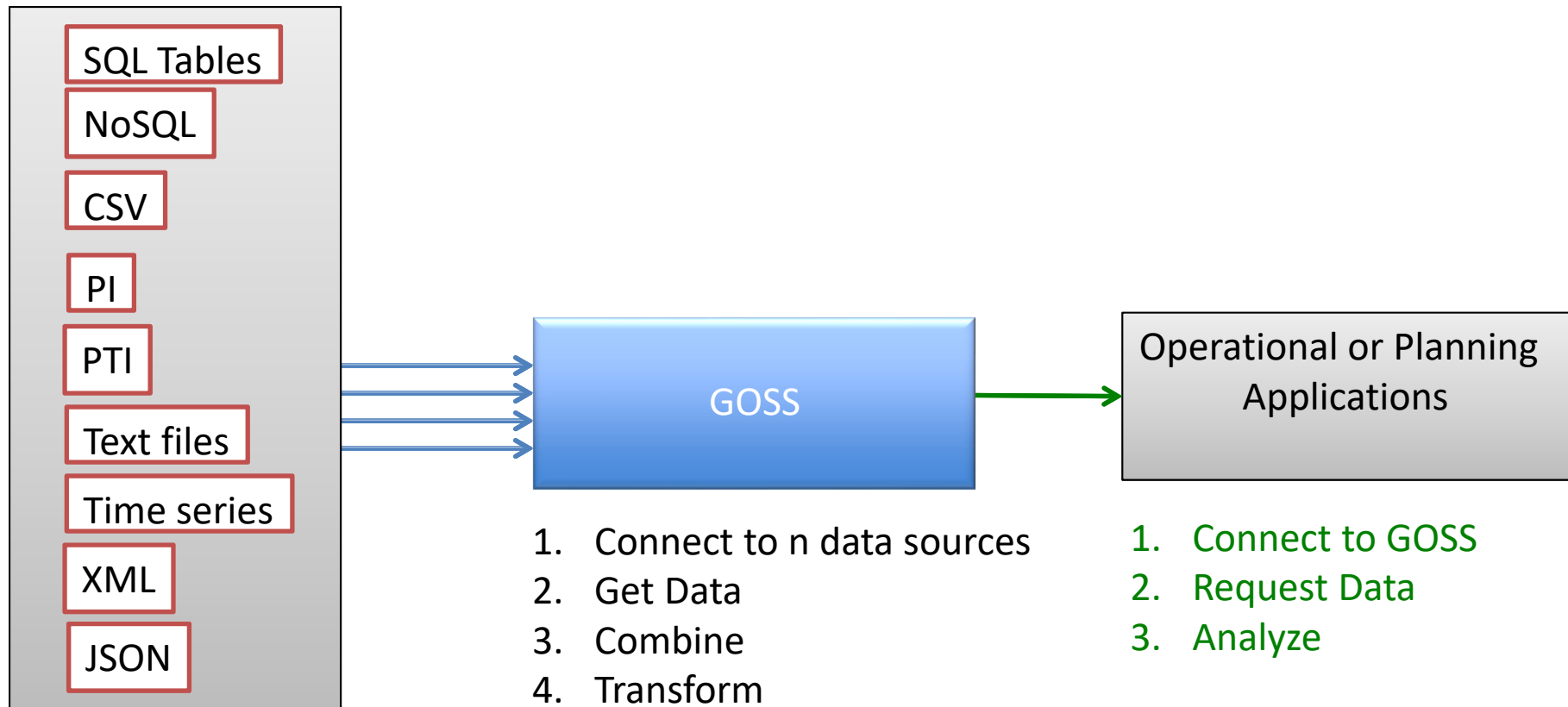
# Making data accessible is a big challenge

- ▶ Organizing and converting data to application specific formats



# Making data accessible is a big challenge

- ▶ Organizing and converting data to application specific formats

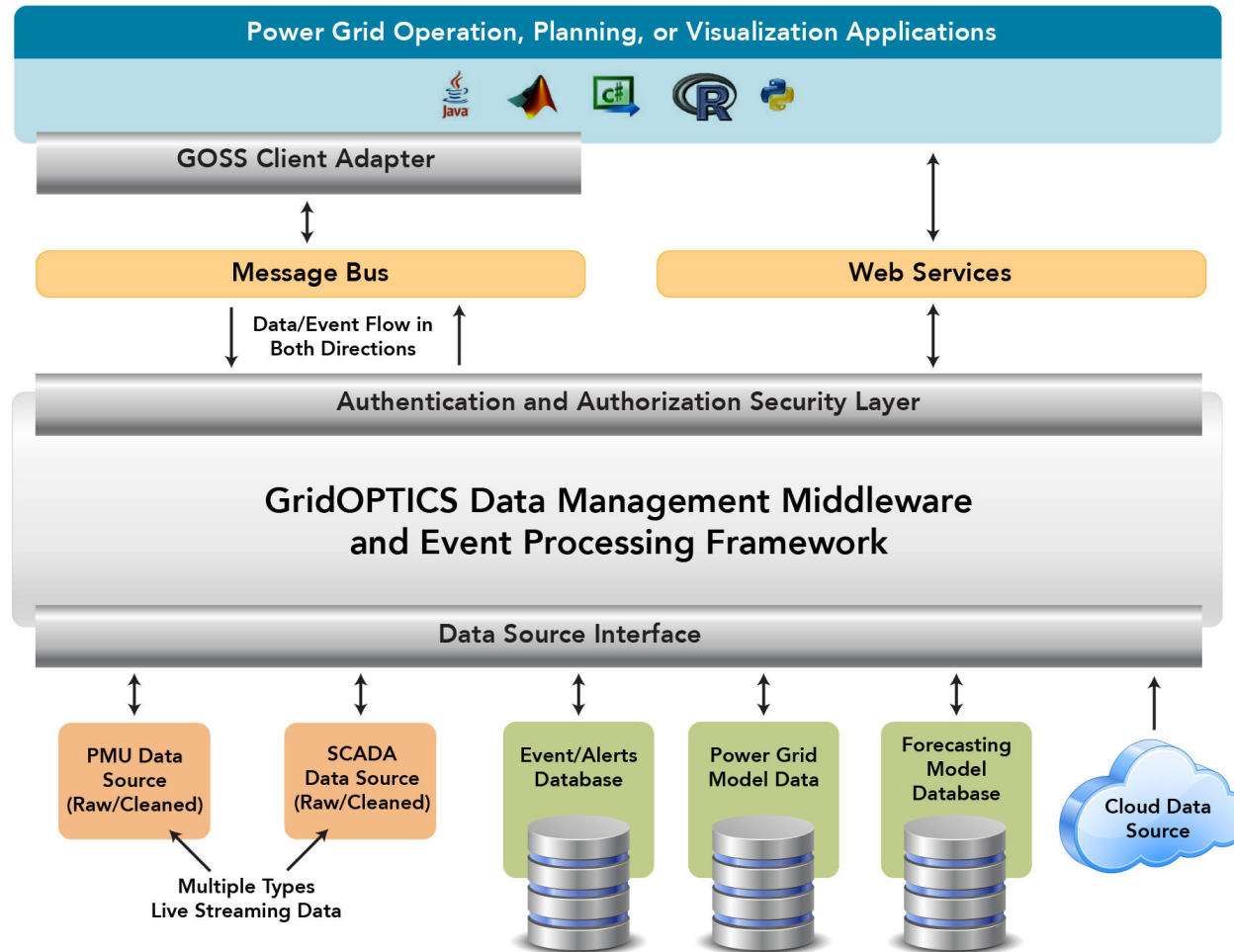


GOSS = GridOPTICS Software System, <https://github.com/GridOPTICS/GOSS>

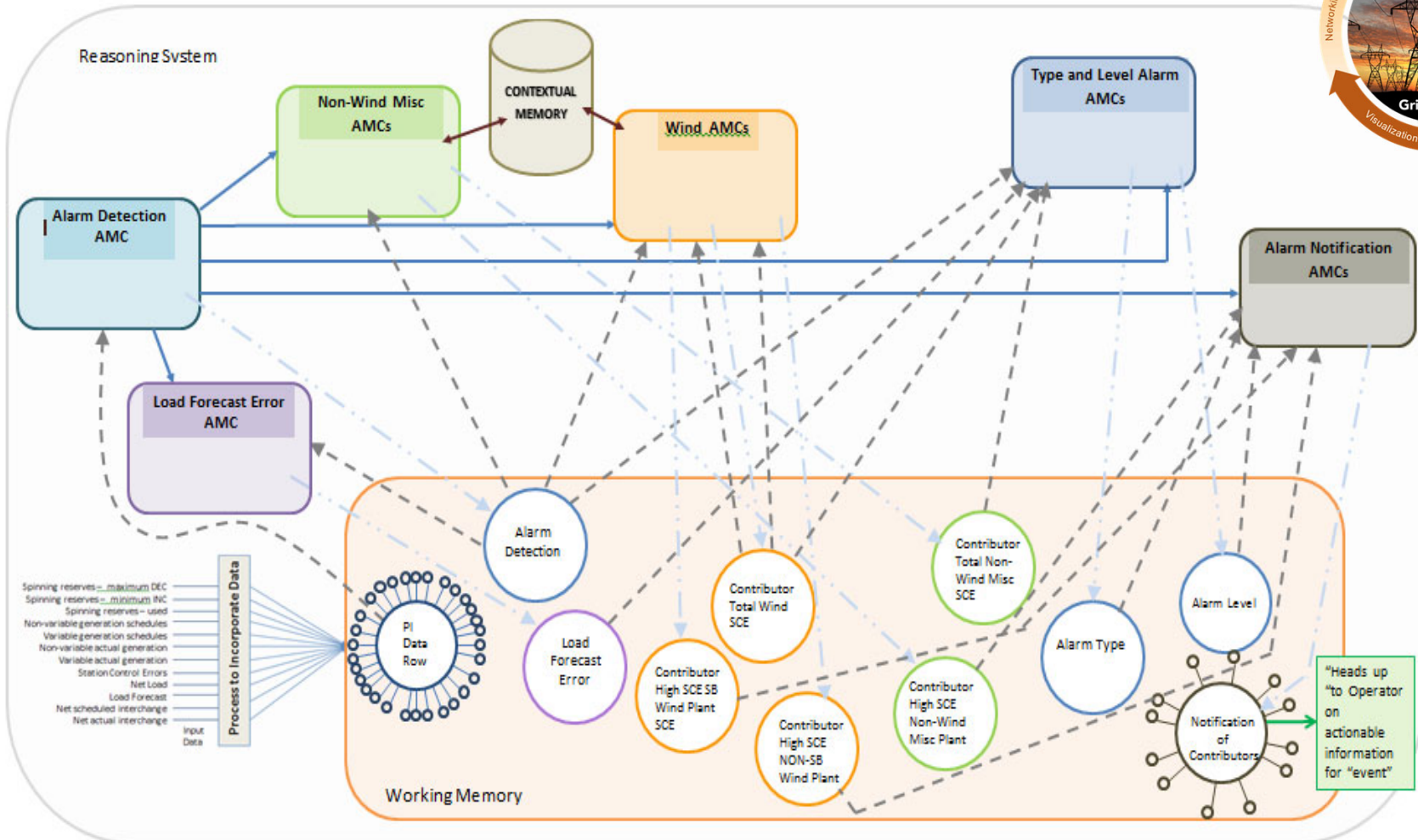
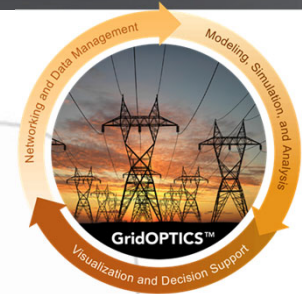


# GOSS™: link data to applications

<https://github.com/GridOPTICS/GOSS>

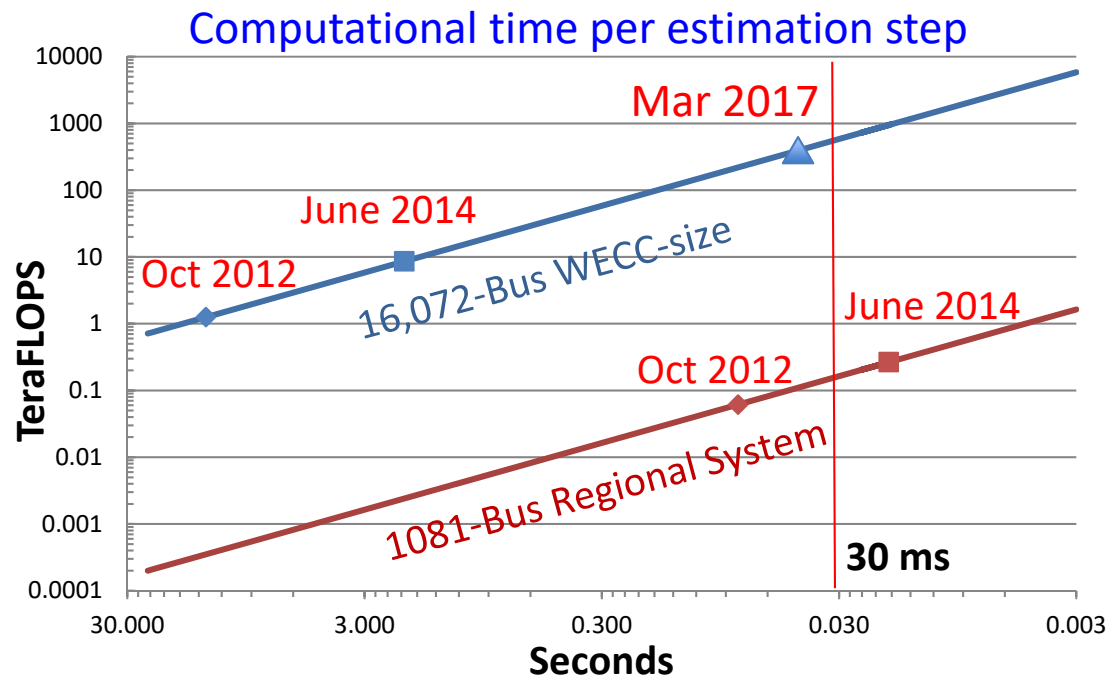


# Analytical challenges in multi-domain data-driven reasoning

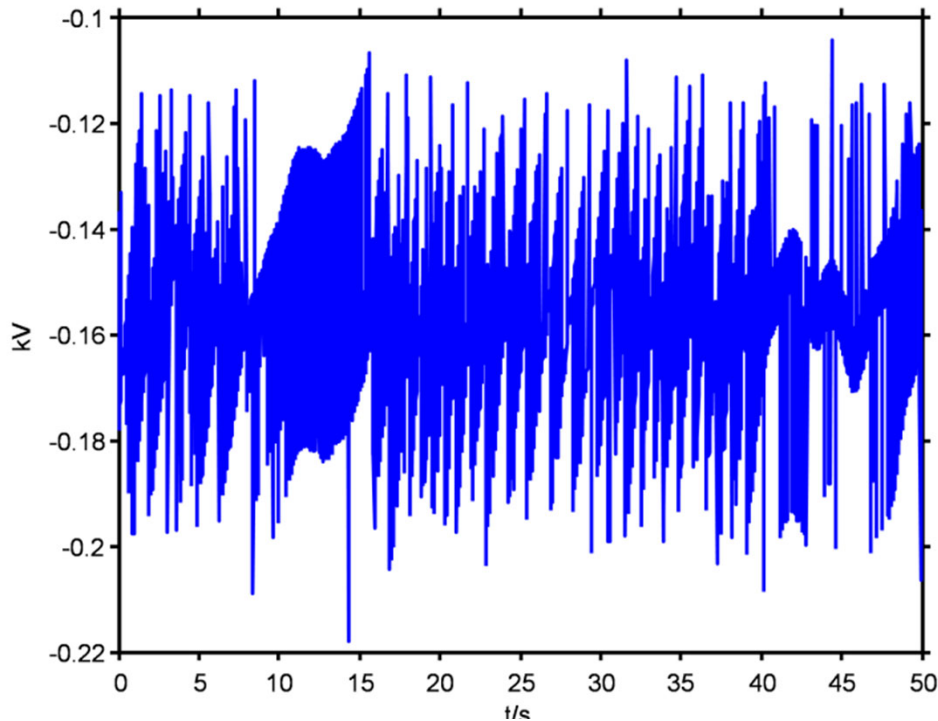


# Computational challenges in keeping up with data cycles

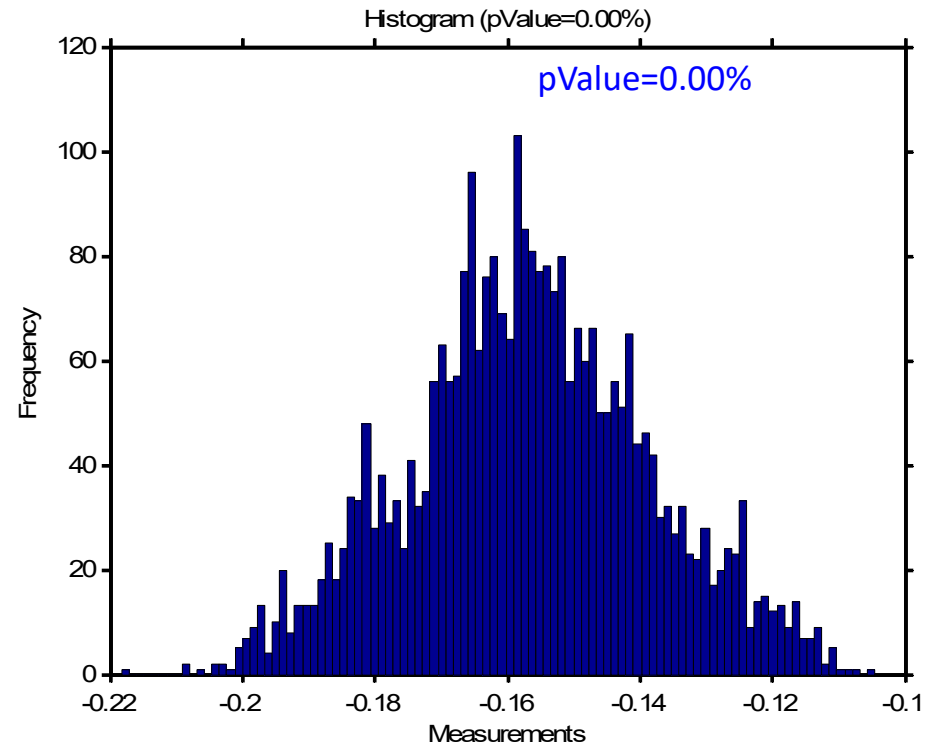
- ▶ **Dynamic state estimation:** computational performance achieved 30ms target for regional systems (1000s buses).
- ▶ WECC-size system (16,000 buses) – a 100 TF problem with 48ms performance on 3,000 cores. Remaining bottleneck is communication. Memory bandwidth advancements expect to meet the 30ms target.



# Mathematical challenges in handling non-Gaussian noise in power grid measurement



Noise extracted from PMU



Noise property analysis

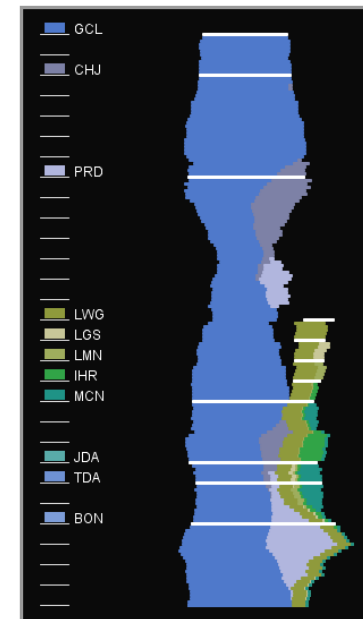
- ▶ Current data applications assume Gaussian noise.
- ▶ New mathematics needs to be developed and adapted for handling non-Gaussian noise, such as Particle Filters, Gaussian Mixture Methods.

# Advanced visualization for improving hydro state awareness (Hydromap)

- ▶ Modernize displays for hydro planning and operations
- ▶ Develop new, novel visualization techniques and paradigms for analyzing dynamic data
- ▶ Develop modular framework for deploying and integrating new data visualizations

-1	-4	163	0	966	906	0	0	807	0	0	0	DCH	2	0	
				Sum:	15493	14920	3046	2661	8953	2268	2085	1718	BLW	9	0
2	2			Plant	Gen	FB	OO	QS	QT	TW	%QS	%QS	GLN	16	0
2	3			GCL	3479	1289.0	139.3	0.0	139.3	963.5	0.0	0.0	GSP	10	0
1	1			CHJ	1238	952.6	94.5	0.0	94.4	781.4	0.0	0.0	JMS	15	-2
4	5			PRD			126.1						LOS	49	1
-3	-2			LWG	180	733.6	44.9	18.7	25.9	634.4	41.6	44.6	REA	4	0
-1	0			LGS	171	633.5	36.3	10.9	25.0	539.4	29.9	30.1	CMP	20	8
0	1			LMN	126	537.3	36.8	17.5	17.1	437.6	47.5	33.4	KN3	38	
32	33			IHR	62	437.7	34.6	25.0	9.2	340.7	72.2	80.6	STL	84	10
				EXCH	3583								VNC	24	8
				BCL	-202								WCV	80	
				CUM	3550								Other Hydro	-3	
				MCN	397	339.5	157.3	78.4	74.2	264.8	49.8	39.8	Other Thermal	13	
				JDA	645	263.1	123.2	36.6	85.0	159.6	29.6	29.6	CAWind	306	
				TDA	389	158.6	115.4	46.2	63.1	76.6	40.0	39.3			
				BON	158	74.5	121.4	74.8	34.1	12.3	61.6	57.2			

*Current data display in need of modernization*

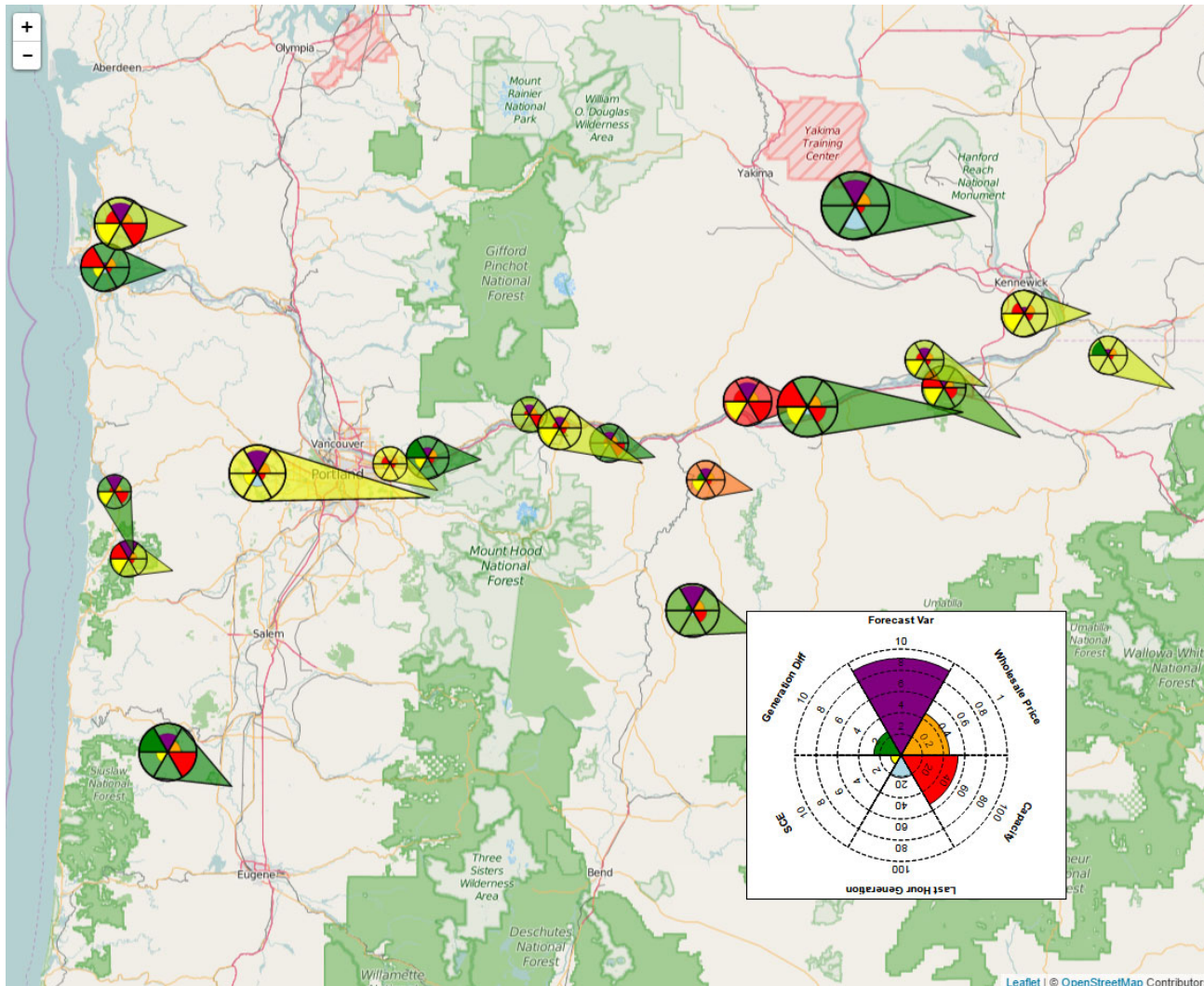


*Interactive hydro map as a new visualization paradigm*





# Multi-dimensional wind visualization (Glyphs)



Wind Visualization/Wind Forecast Visualization

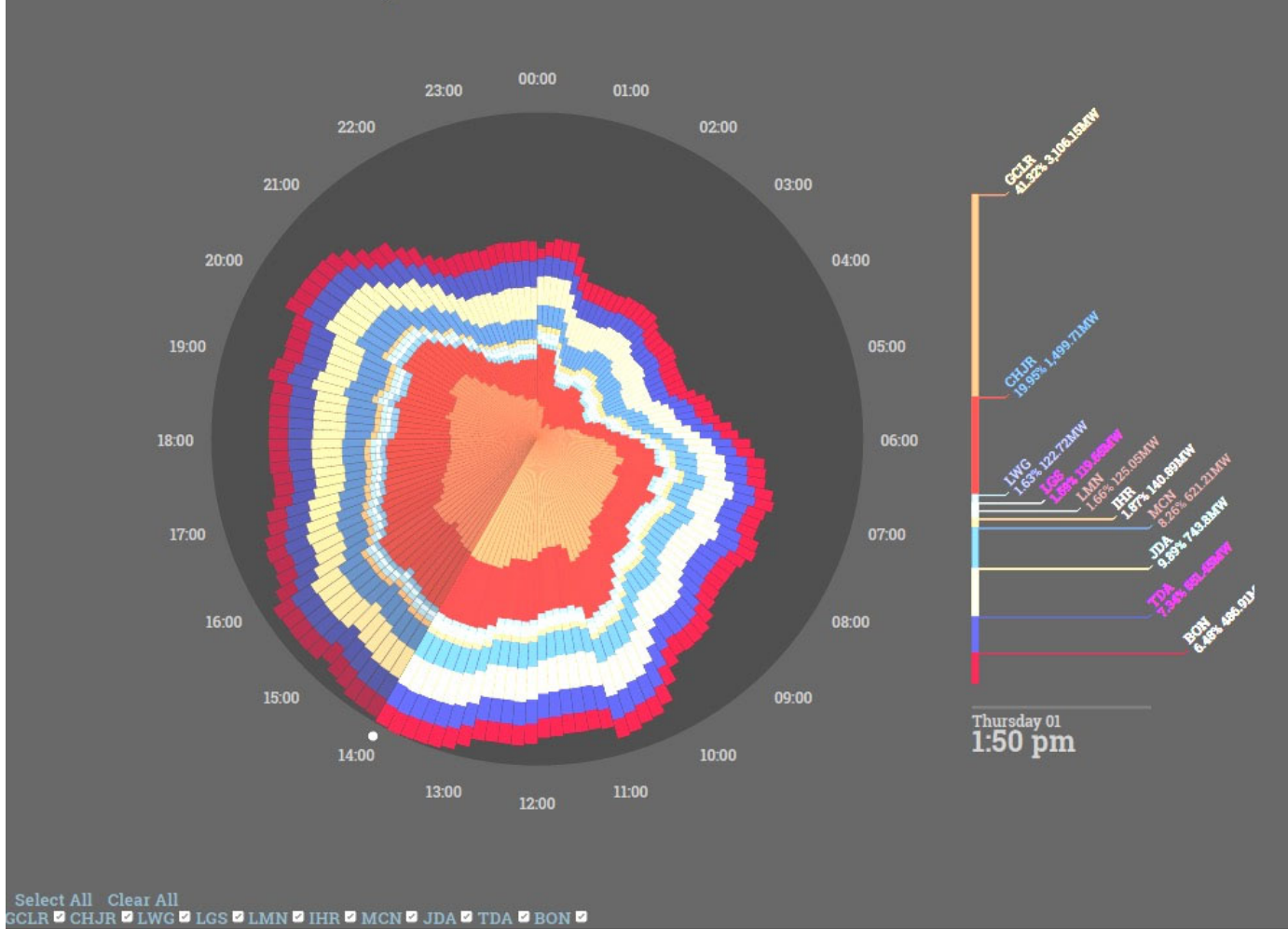
- ▶ Wind visualization showing multi-dimensional data in glyphs
  - Wind speed (length of tail)
  - Wind direction (angle of tail)
  - Generation (size of head)
  - Uncertainty (color of tail)
  - Forecast variability
  - Wholesale price
  - Capacity
  - Last hour generation
  - SCE error code
  - Generation difference from forecast

# Historical hydro view using radial visualization

Pacific Northwest  
NATIONAL LABORATORY

Proudly Operated by **Battelle** Since 1965

## Electric Generation from Hydro-electric dams - Latest 24 hours

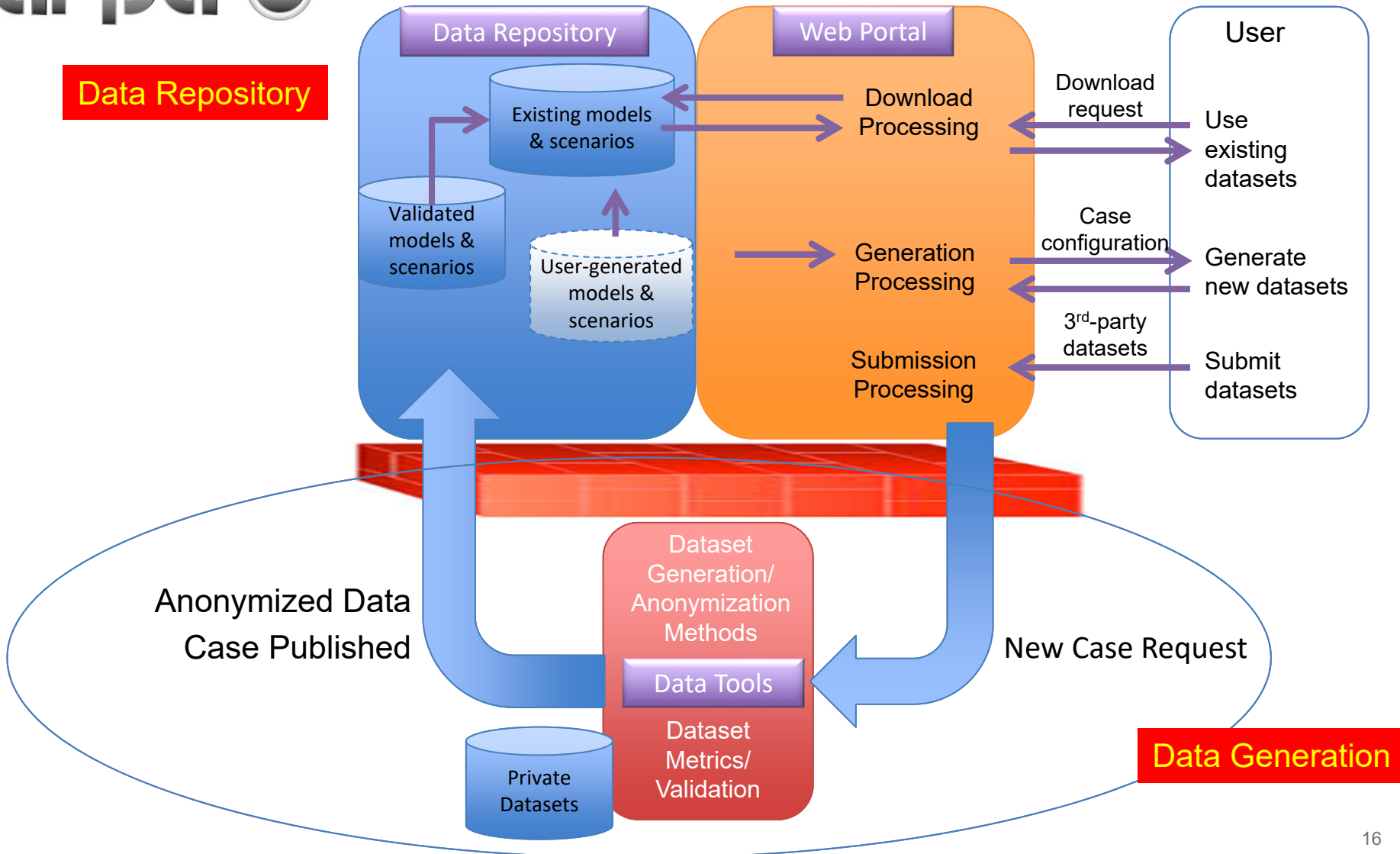


- ▶ Compares hydropower generation across different projects along Columbia and Snake rivers
- ▶ Alternative view shows generation across different sources such as hydropower, nuclear, renewables, and miscellaneous sources

# Data repository for public hosting



Data Repository



# DR POWER: ARPA-E Data Repository

<https://egriddata.org>

The screenshot shows a web browser window with the URL [egriddata.org](https://egriddata.org). The page features the DR POWER logo, a search bar, and a navigation menu. The search results section displays 57 results, with the top result being 'GO Competition Phase 0 Infeasible 179-bus'. This result includes a warning message: '\*Warning!\* The authors have discovered some serious known errors in the files and they are working on replacing them. They humbly apologize for the inconvenience. This dataset is composed of 10 scenarios, only 3 known to be feasible, based on...'. A 'zip' download button is visible below the warning. The second result is 'GO Competition Phase 0 Feas179-bus'. The left sidebar contains filters for Content Types, Topics, Tags, Format, Publisher, Author, and License.

Content Types  
Dataset  
Topics  
Transmission  
Distribution (1)  
Economics and Markets (1)  
Tags  
Format  
Publisher  
Author  
License

57 results

Search  
Date changed  
Descending  
Apply  
Reset

GO Competition Phase 0 Infeasible 179-bus  
Grid Optimization (GO) Competition  
Transmission  
\*Warning!\* The authors have discovered some serious known errors in the files and they are working on replacing them. They humbly apologize for the inconvenience. This dataset is composed of 10 scenarios, only 3 known to be feasible, based on...  
zip

GO Competition Phase 0 Feas179-bus  
Grid Optimization (GO) Competition



# ARPA-E Grid Optimization Competition (open now)

<https://gocompetition.energy.gov>



The screenshot shows a web browser window with the URL [gocompetition.energy.gov](https://gocompetition.energy.gov). The page features a dark blue header with the ARPA-E logo and the text "GRID OPTIMIZATION (GO) COMPETITION". A navigation menu includes links for Home, Background, References, Competitions, FAQs, Forum, News, and Definitions. The main content area has a large banner image of power lines with the text "GRID OPTIMIZATION COMPETITION" overlaid. Below the banner is a blue box with the text "\$4 million in prizes for better power grid optimization!". The page also includes a paragraph of introductory text, a list of links for "Now Available" (Problem Formulation, Scoring Information, Rules), and a footer note about opening the URL in a new tab.

Log In | Register

Home Background References Competitions FAQs Forum News Definitions

## GRID OPTIMIZATION COMPETITION

**\$4 million in prizes for better power grid optimization!**

The U.S. Department of Energy Advanced Research Projects Agency-Energy (**ARPA-E**) is challenging the international research and industrial communities to develop innovative power system optimization and control algorithms for the future power grid.

**Entrants** will develop transformational ways to solve power flow optimization problems across a wide range of

Open "<https://gocompetition.energy.gov>" in a new tab ral challenge underlying all grid planning and operations tools.

**Now Available**

- [Problem Formulation](#)
- [Scoring Information](#)
- [Rules](#)



- ▶ Grid data complexity is increasing with big volumes, diverse types, and various attributes.
- ▶ Such complexity poses significant challenges in data access, transformation, analytics, sense making.
- ▶ Math, computing and visualization technologies need to be developed to meet these challenges.
  - GOSS as an example big data platform.
  - Multi-domain data reasoning and high performance computing.
  - Modular visualization for information presentation.

# Acknowledgement

- ▶ PNNL Researchers: (*Data and Computing*) Poorva Sharma, Steve Elbert, Shuangshuang Jin, Bruce Palmer, George Chin; (*Power Engineering*) Yousu Chen, Mark Rice, Shaobu Wang, Karen Studarus
- ▶ Former PNNL Researchers: Ruisheng Diao, Bora Akyol, Terrence Critchlow, Ning Zhou, Ning Lu, Pengwei Du
- ▶ Funding support by:
  - PNNL Future Power Grid Initiative (FPGI)
  - DOE Office of Electricity Delivery and Energy Reliability (OE) Advanced Grid Modeling Program
  - DOE Grid Modernization Initiative
  - DOE Advanced Scientific Computing Research (ASCR) Applied Math Program
  - DOE Advanced Research Program Agency – Energy (ARPA-E)
  - Bonneville Power Administration (BPA)

# Questions?

## Further Information:

GridOPTICS: <http://gridoptics.pnnl.gov/>

GridOPTICS™ Software System (GOSS): <https://github.com/GridOPTICS/GOSS>

Interactive Visualization and Demo Center: <http://vis.pnnl.gov/>

DR POWER: ARPA-E Data Repository: <https://egriddata.org>

ARPA-E Grid Optimization Competition: <https://gocompetition.energy.gov>

**Zhenyu (Henry) Huang, Ph.D., P.E., F.IEEE**

Laboratory Fellow/Technical Group Manager

Pacific Northwest National Laboratory

[zhenyu.huang@pnnl.gov](mailto:zhenyu.huang@pnnl.gov)