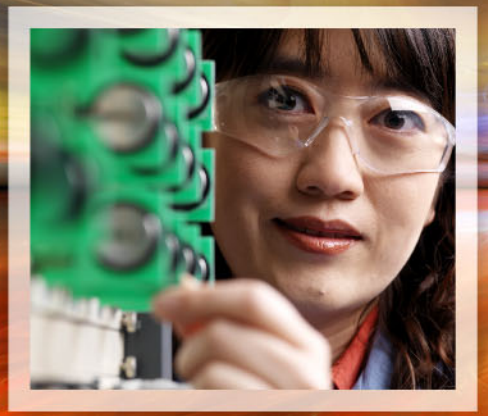


IEEE PES AMPS Big Data Subcommittee Webinar
October 10, 2018

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

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)

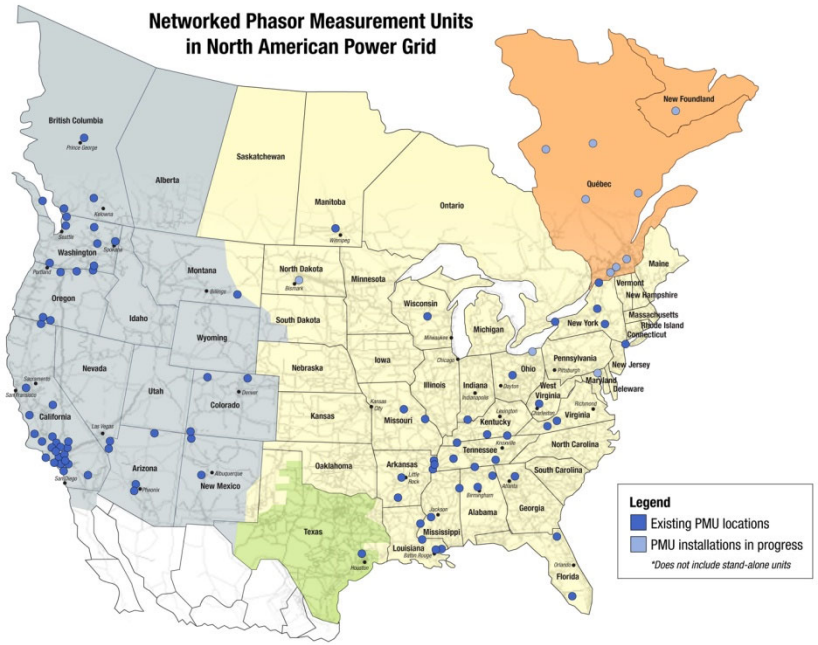
DISCOVERY *in Action*

PNNL-SA-138532

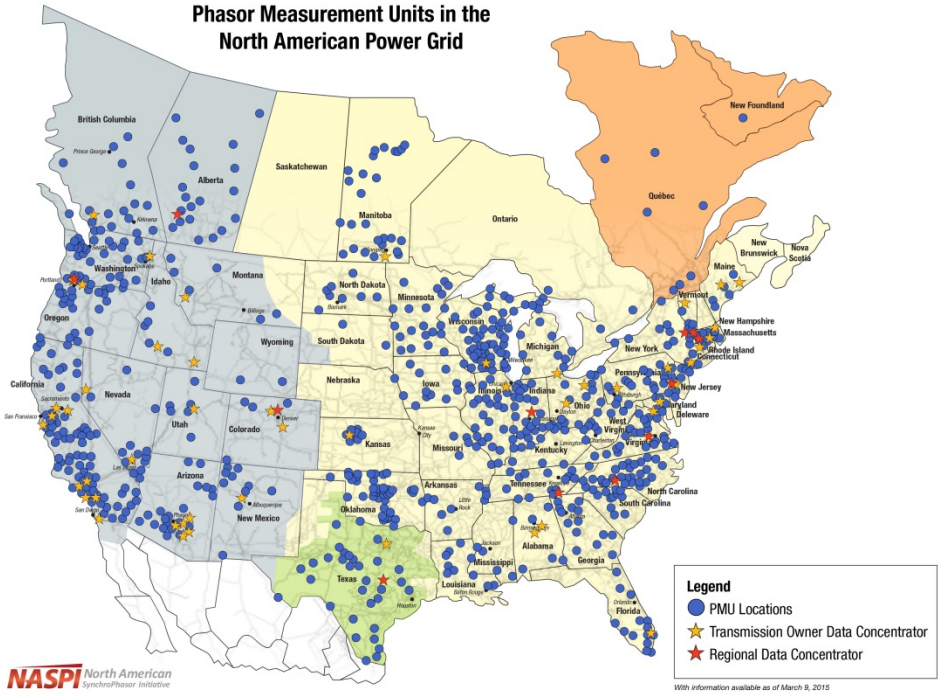
Deployment of a vast new phasor network is generating unprecedented real-time data

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

April 2007



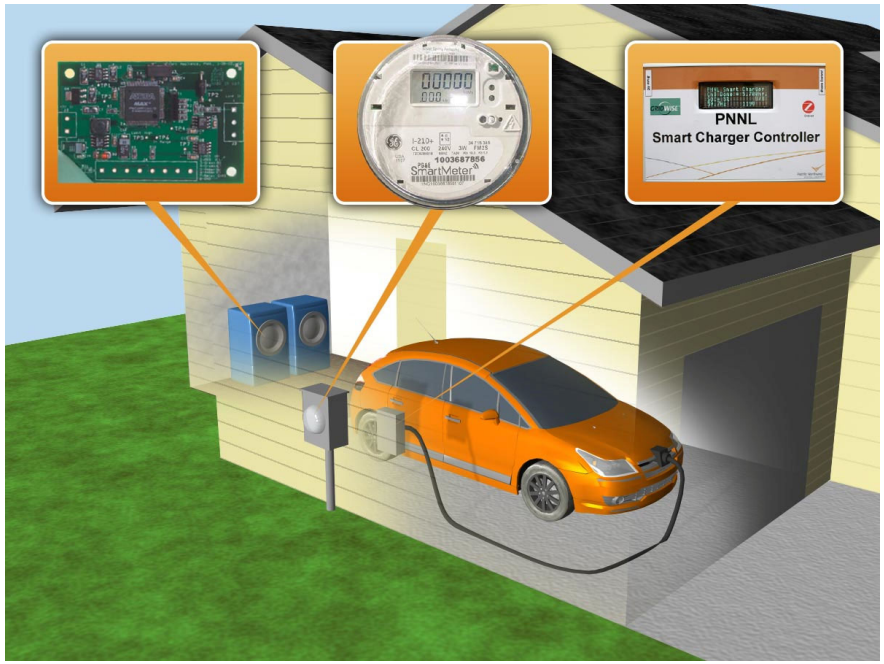
March 2015



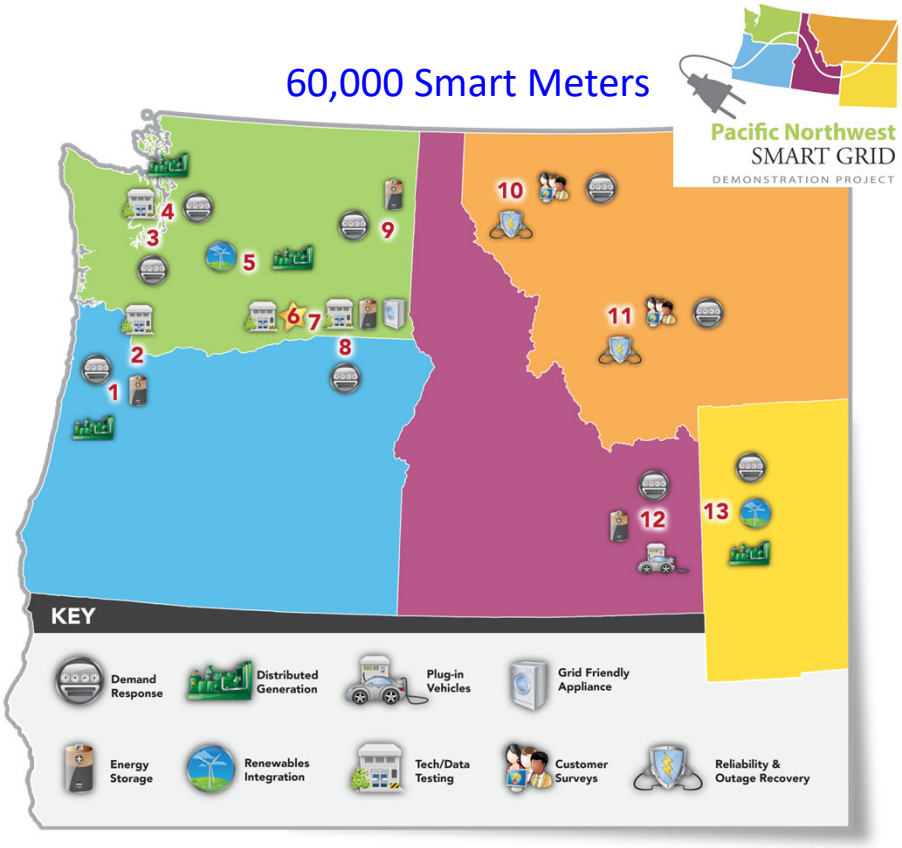
	Today – SCADA data	Emerging – phasor data	Improvement
Variety	voltage + current	+ phase angle, …	more information
Velocity	1 sample / 4 seconds	30-120 samples / second	~200x faster
Volume	8 terabytes / year	1.5 petabytes / year	~200x more data
Veracity	unseen ms-oscillations	oscillations seen at 10ms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

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

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



60,000 Smart Meters



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

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

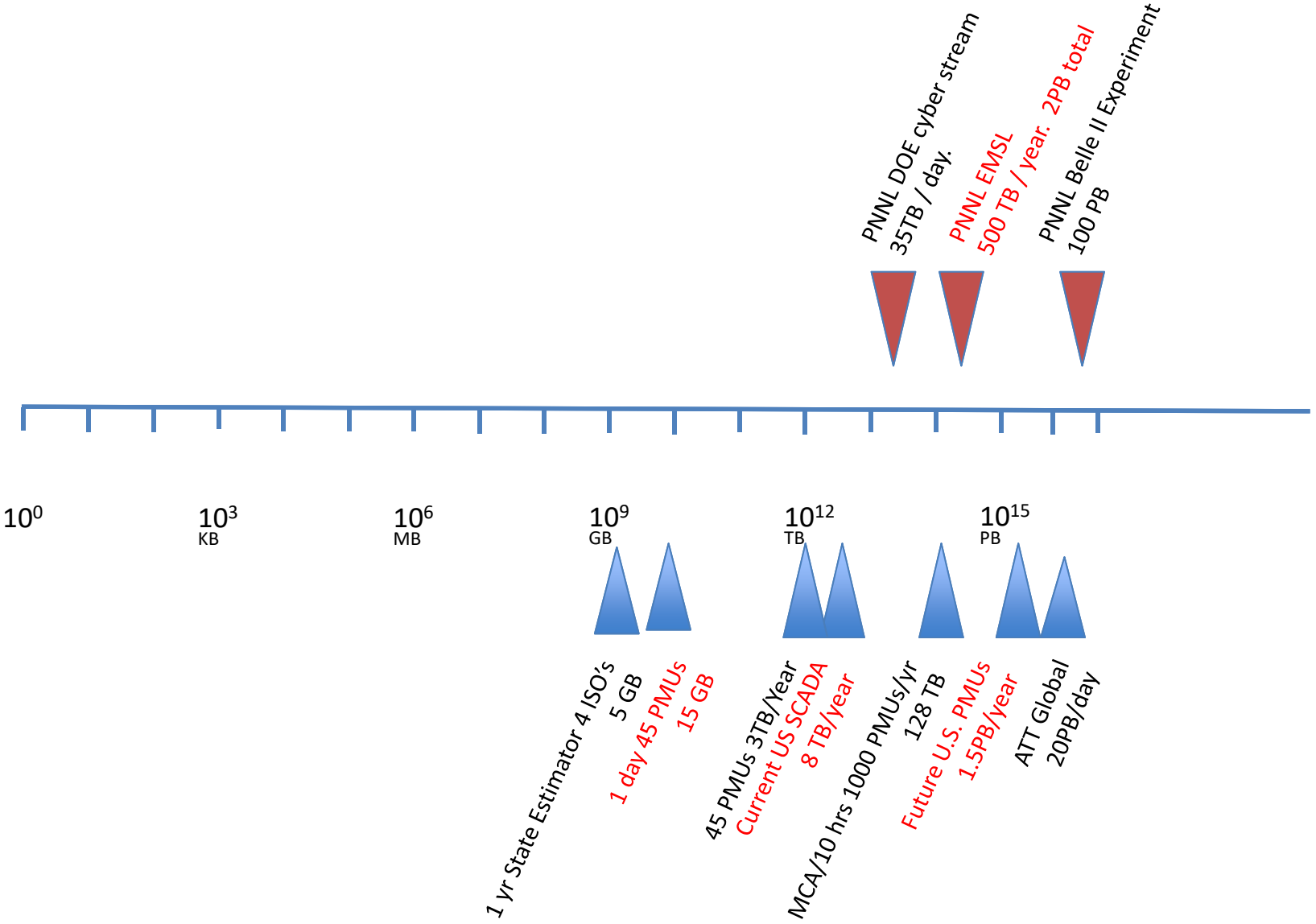
- Weather/climate data
  (e.g. PNNL ARM* Data)
  - 300 instruments, 2000 data streams 24/7
  - 500 GB/day rising to multiple TBs/day
  - Curating 20 years' data
- Market/business data
- Cyber/communication data
- Simulated data
  - Each contingency scenario generates 0.5M bytes data, adding up to TB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

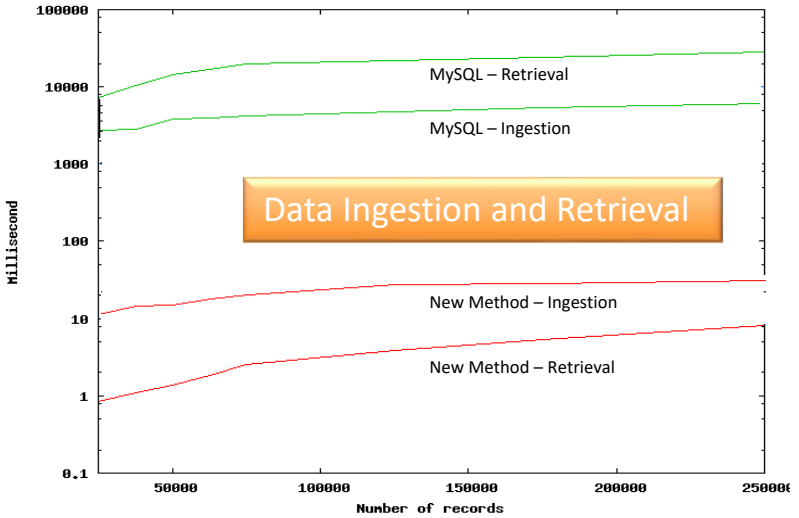
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

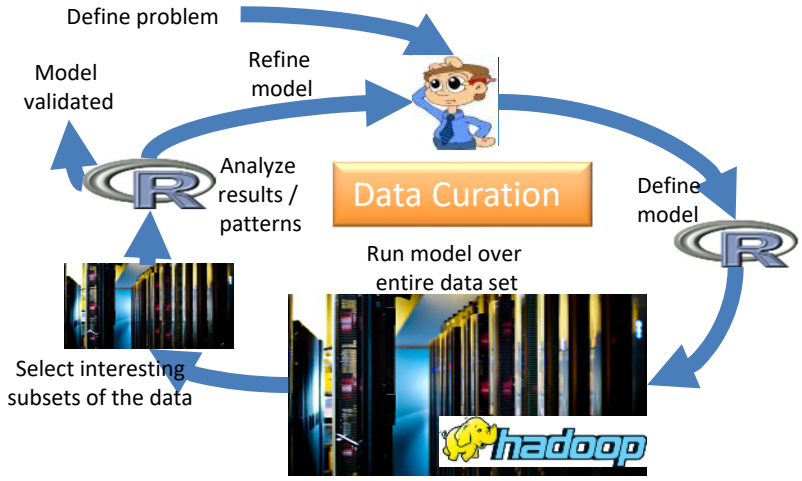


Real-time data ingestion, retrieval, curation from a distributed sensor network

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

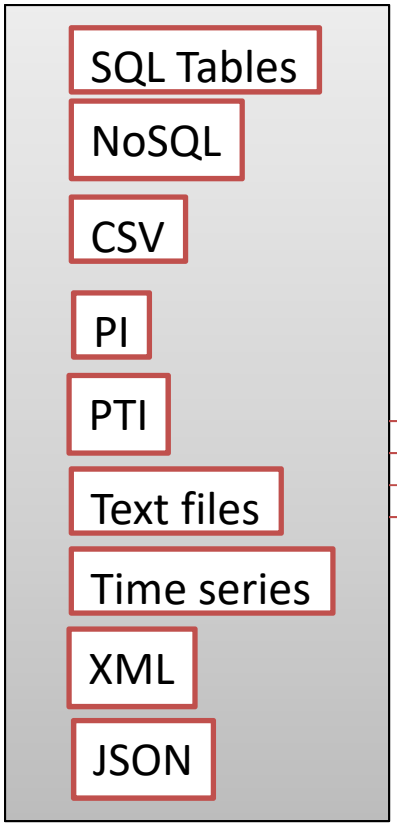
- **Requirements**
  - Cyber-secure sensor network
  - Data provenance and privacy
  - Real-time processing
- **Solution: scalable, flexible middleware and R/Hadoop statistical analysis capabilities**
  - Data ingestion is 10^3 times faster than MySQL
  - Linearly scales to many nodes
  - Data curation 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



Redundancy: Underlined steps has to be performed by every application for each type of data

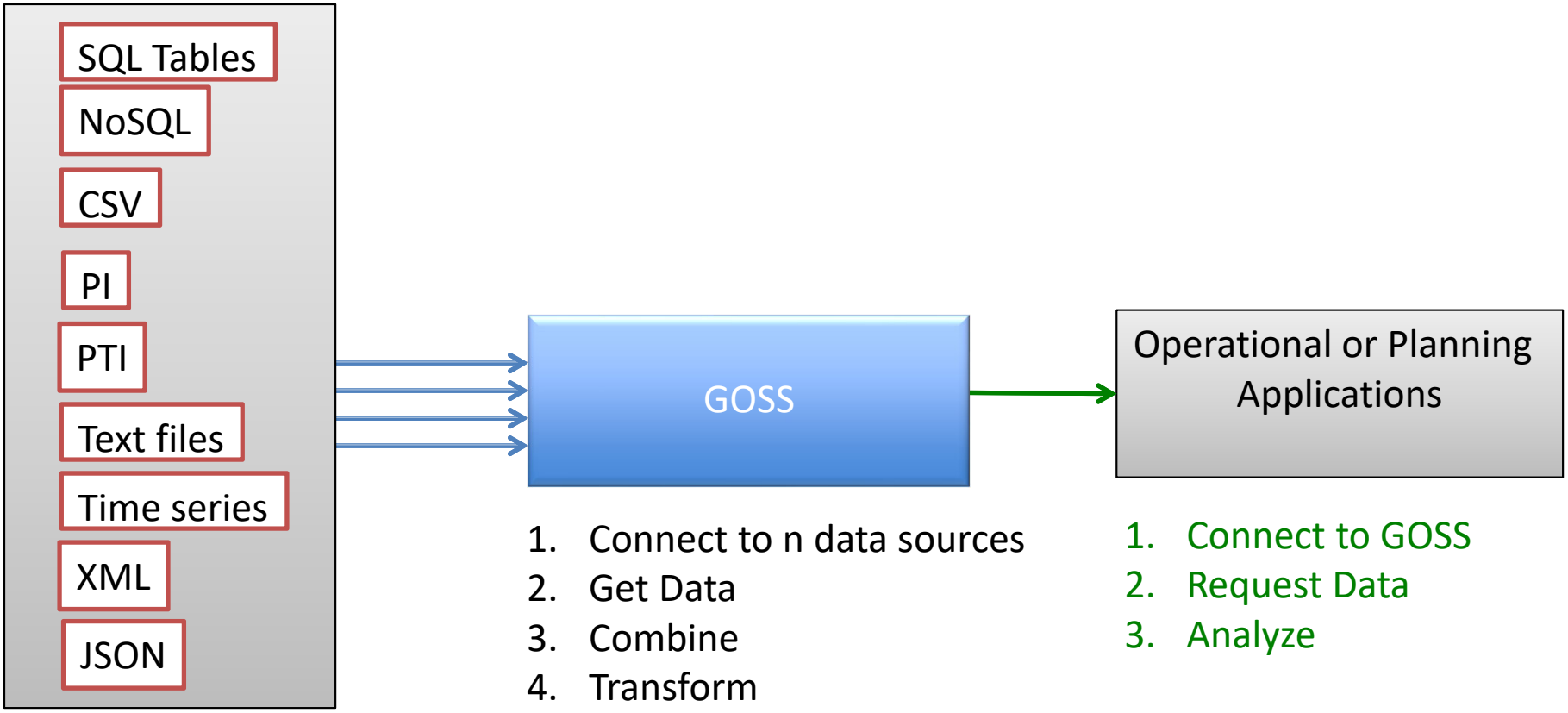
Operational or Planning Applications

1. Connect to n data sources
2. Get Data * n
3. Combine * n
4. Transform * n
5. Analyze

Making data accessible is a big challenge

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Organizing and converting data to application specific formats



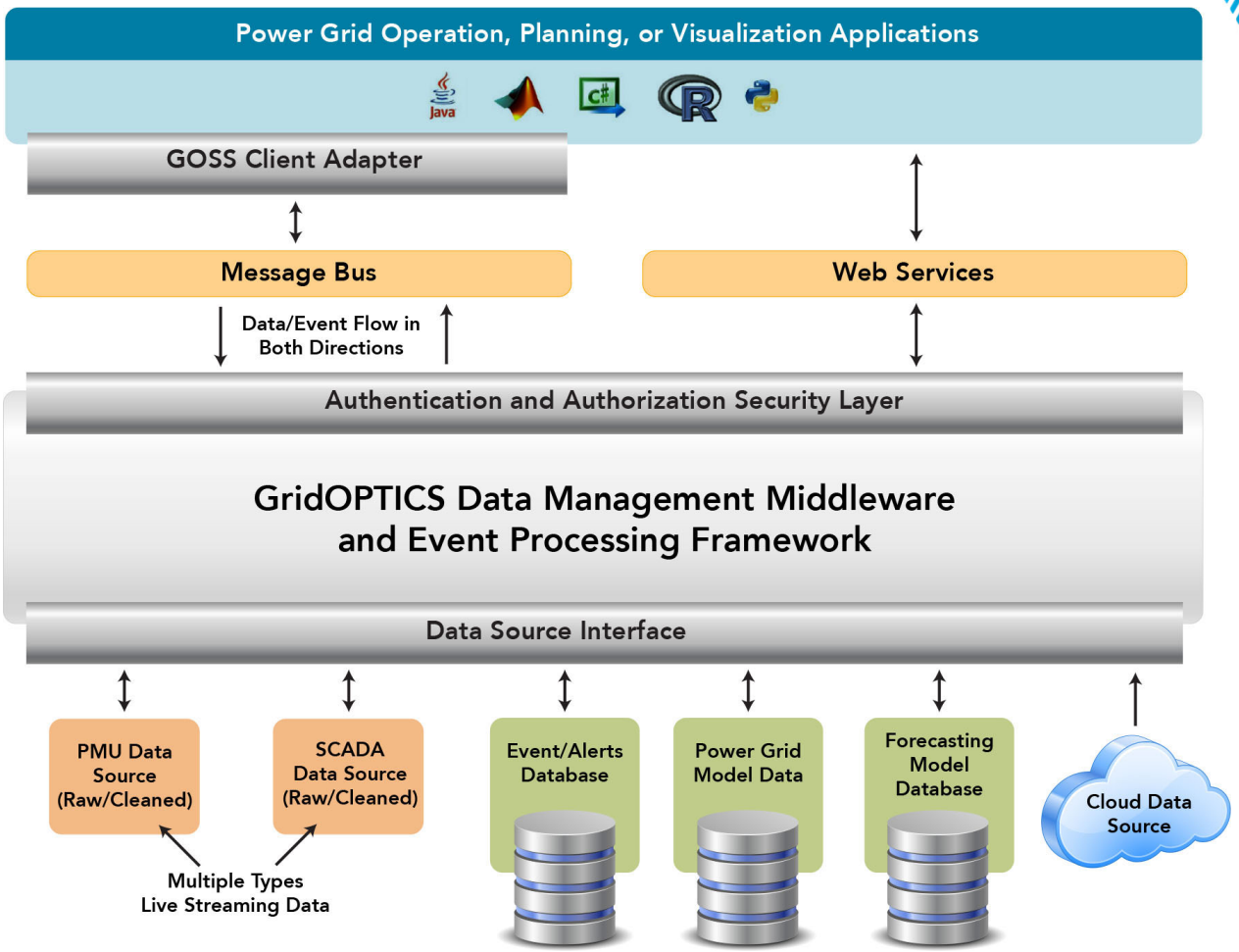
GOSS = GridOPTICS Software System, https://github.com/GridOPTICS/GOSS

GOSS™: link data to applications

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

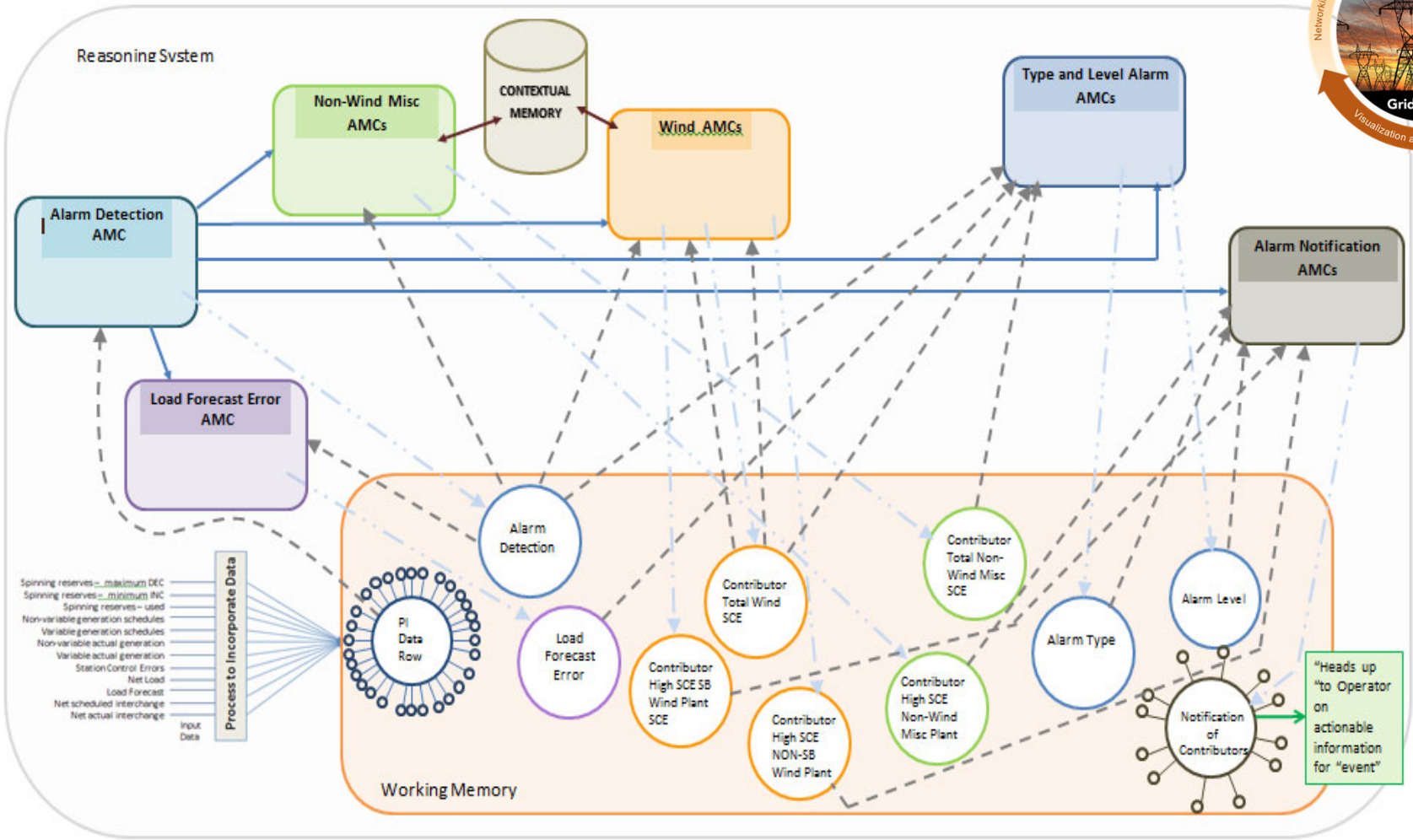
https://github.com/GridOPTICS/GOSS

GRID MODERNIZATION INITIATIVE U.S. Department of Energy



Analytical challenges in multi-domain data-driven reasoning

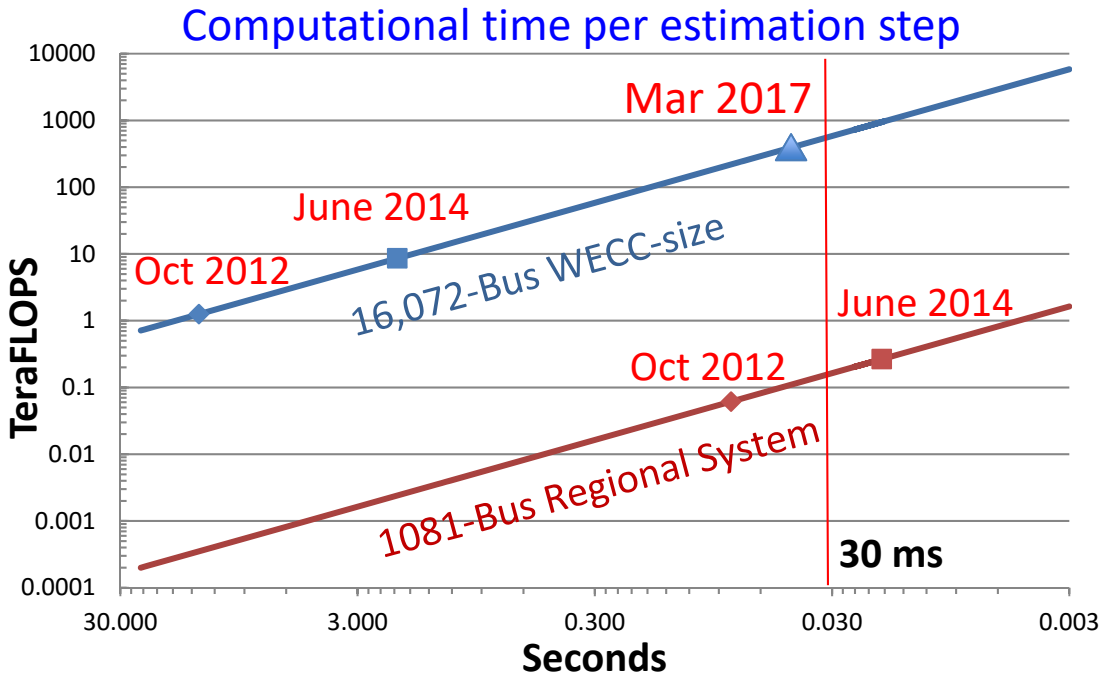
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



Computational challenges in keeping up with data cycles

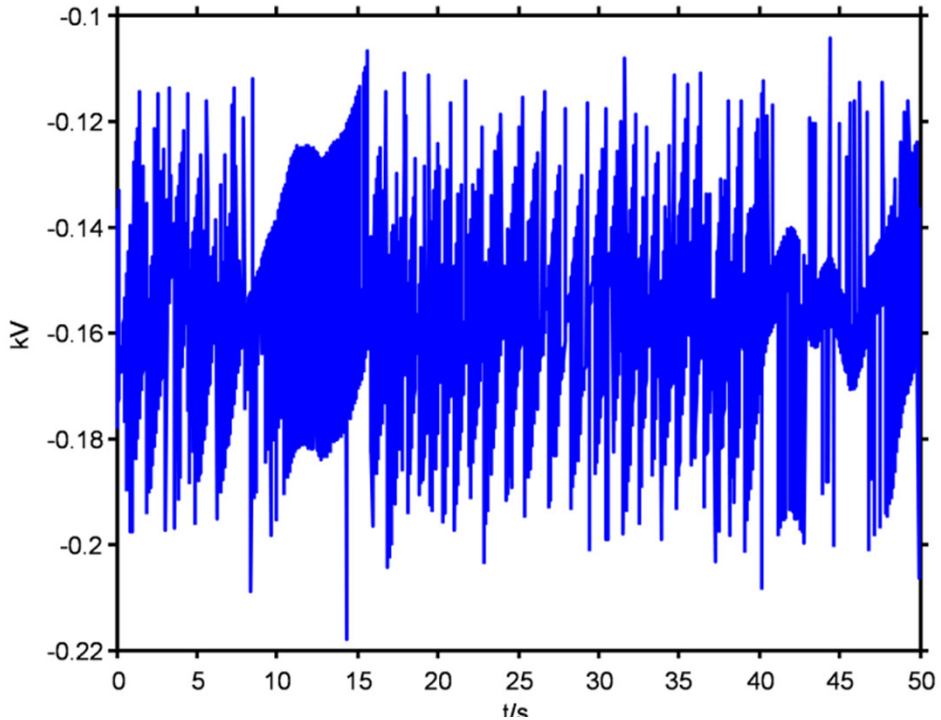
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- **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.

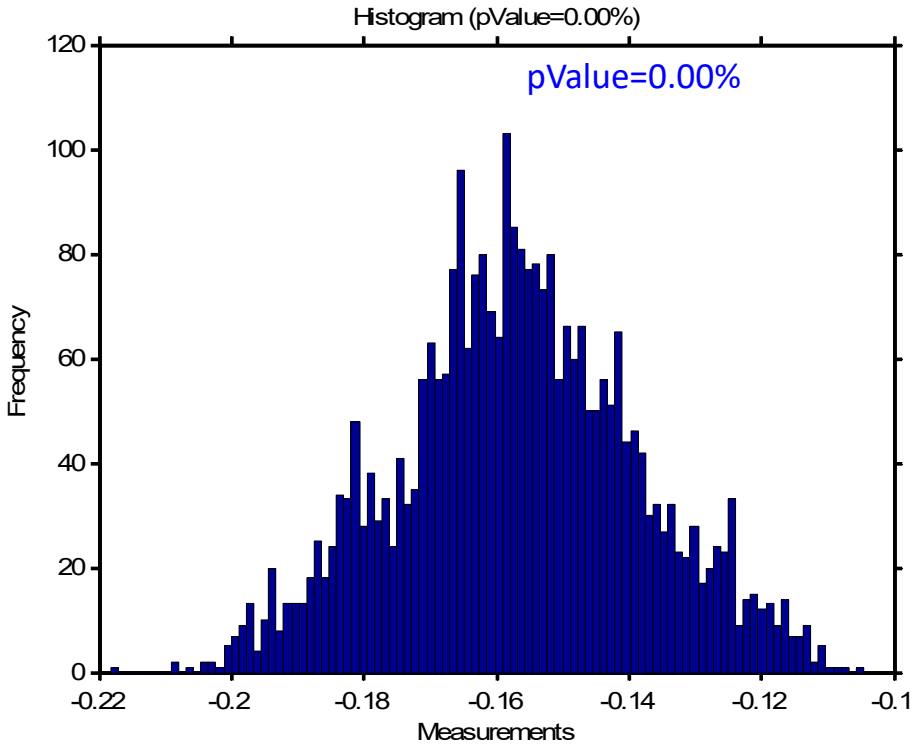


Mathematical challenges in handling non-Gaussian noise in power grid measurement

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



Noise extracted from PMU



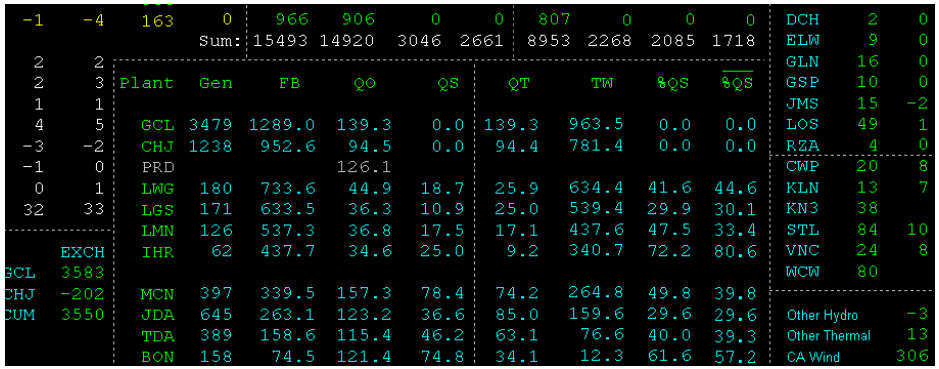
Noise property analysis

- 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)

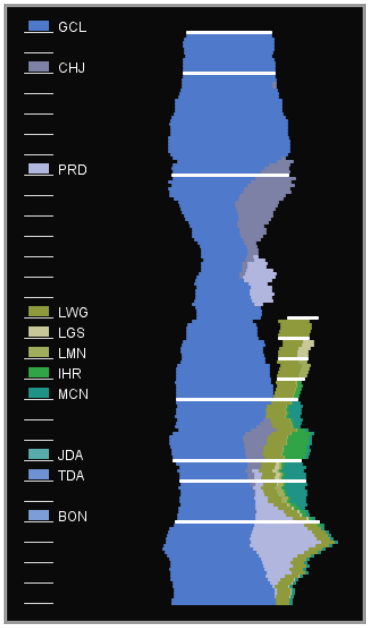
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- 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	ELW	9	0
2	2											GLN	16	0
2	3	Plant	Gen	FB	QO	QS	QT	TW	%QS	%QS		GSP	10	0
1	1											JMS	15	-2
4	5	GCL	3479	1289.0	139.3	0.0	139.3	963.5	0.0	0.0		LOS	49	1
-3	-2	CHJ	1238	952.6	94.5	0.0	94.4	781.4	0.0	0.0		RZA	4	0
-1	0	PRD			126.1							CWP	20	8
0	1	LWG	180	733.6	44.9	18.7	25.9	634.4	41.6	44.6		KLN	13	7
32	33	LGS	171	633.5	36.3	10.9	25.0	539.4	29.9	30.1		KN3	38	
		LMN	126	537.3	36.8	17.5	17.1	437.6	47.5	33.4		STL	84	10
	EXCH	IHR	62	437.7	34.6	25.0	9.2	340.7	72.2	80.6		VNC	24	8
GCL	3583											WCW	80	
CHJ	-202	MCN	397	339.5	157.3	78.4	74.2	264.8	49.8	39.8				
CUM	3550	JDA	645	263.1	123.2	36.6	85.0	159.6	29.6	29.6		Other Hydro		-3
		TDA	389	158.6	115.4	46.2	63.1	76.6	40.0	39.3		Other Thermal		13
		BON	158	74.5	121.4	74.8	34.1	12.3	61.6	57.2		CA Wind		306

Current data display in need of modernization

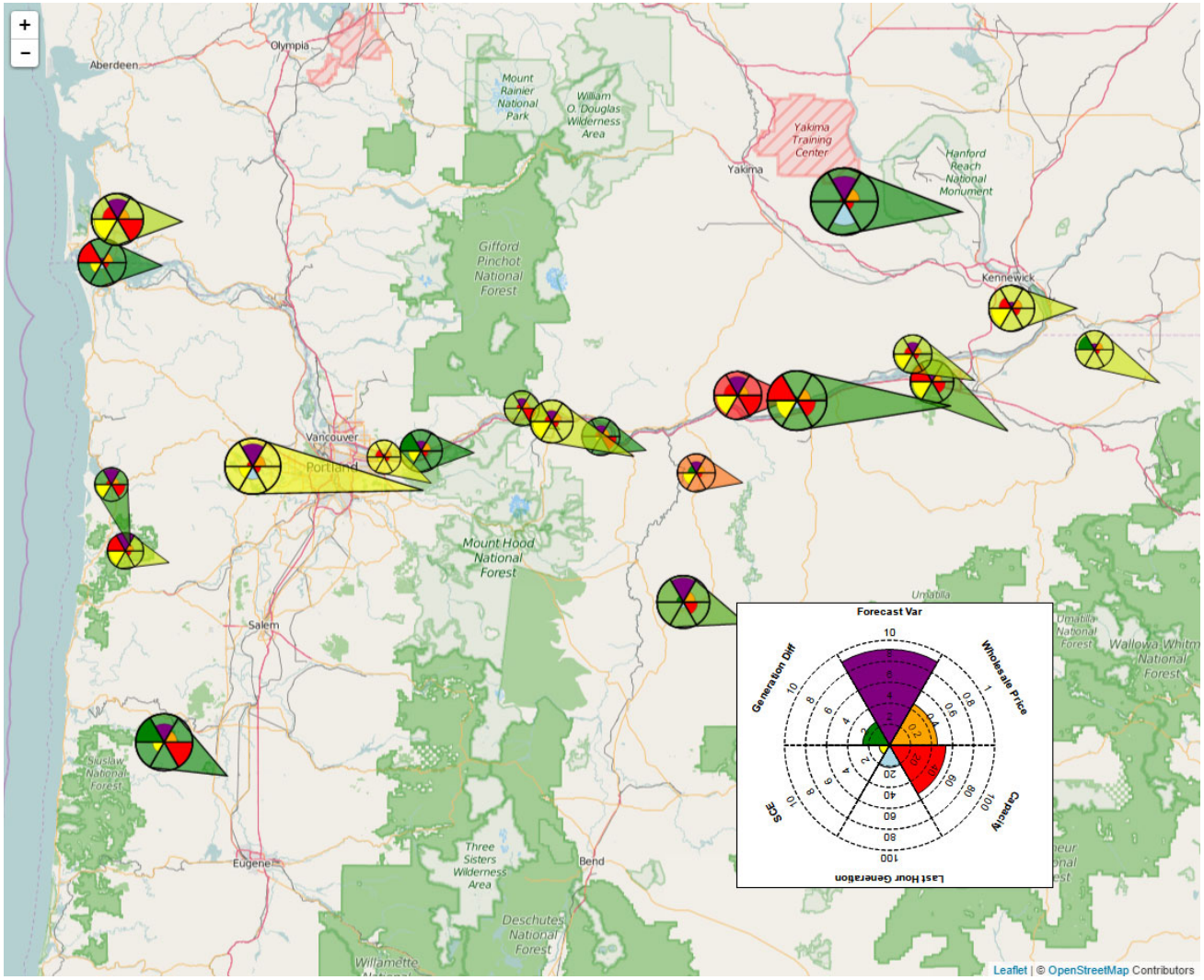


Interactive hydro map as a new visualization paradigm

BONNEVILLE POWER ADMINISTRATION

Multi-dimensional wind visualization (Glyphs)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

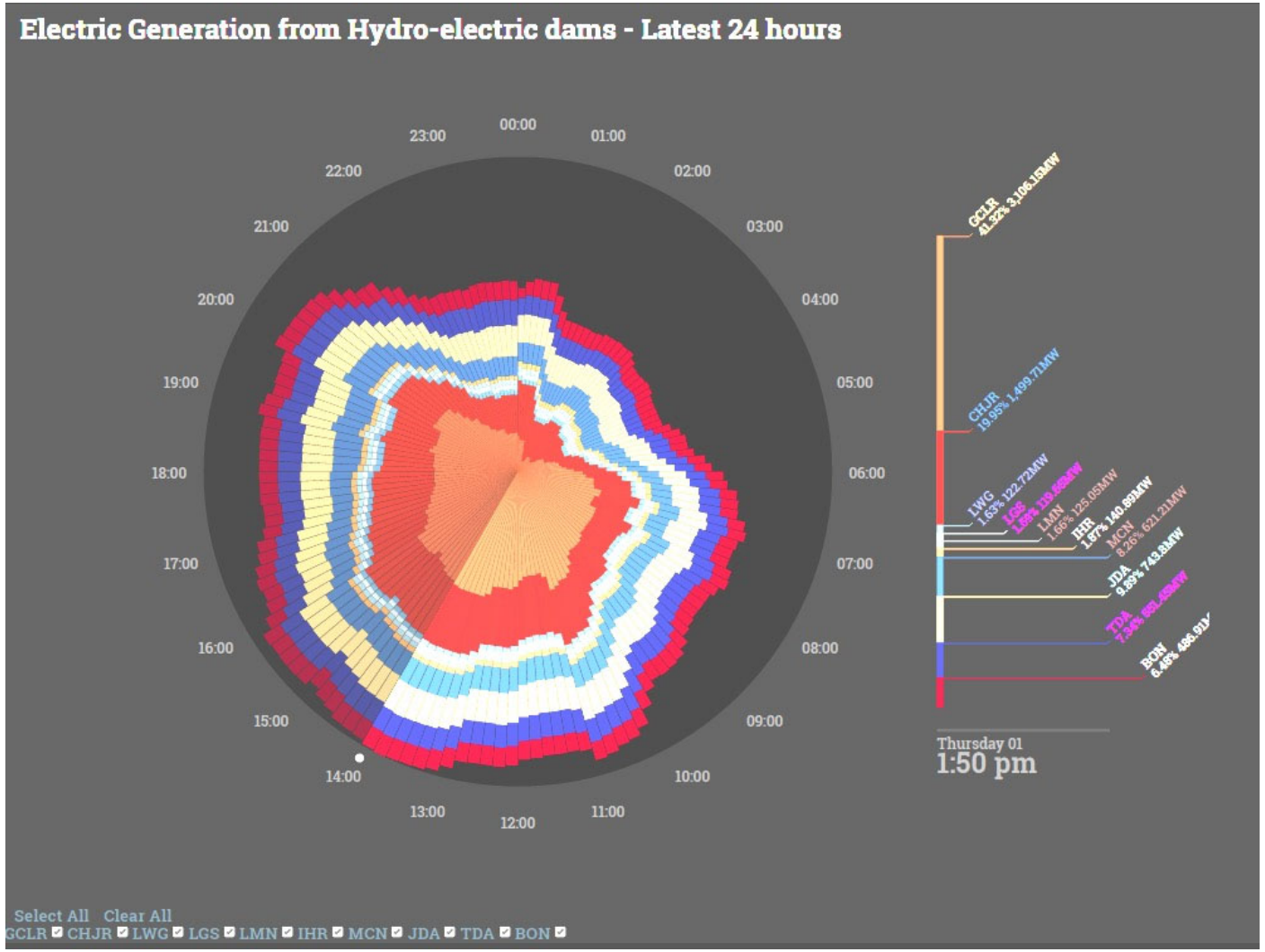


Wind Visualization/Wind Forecast Visualization

- 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



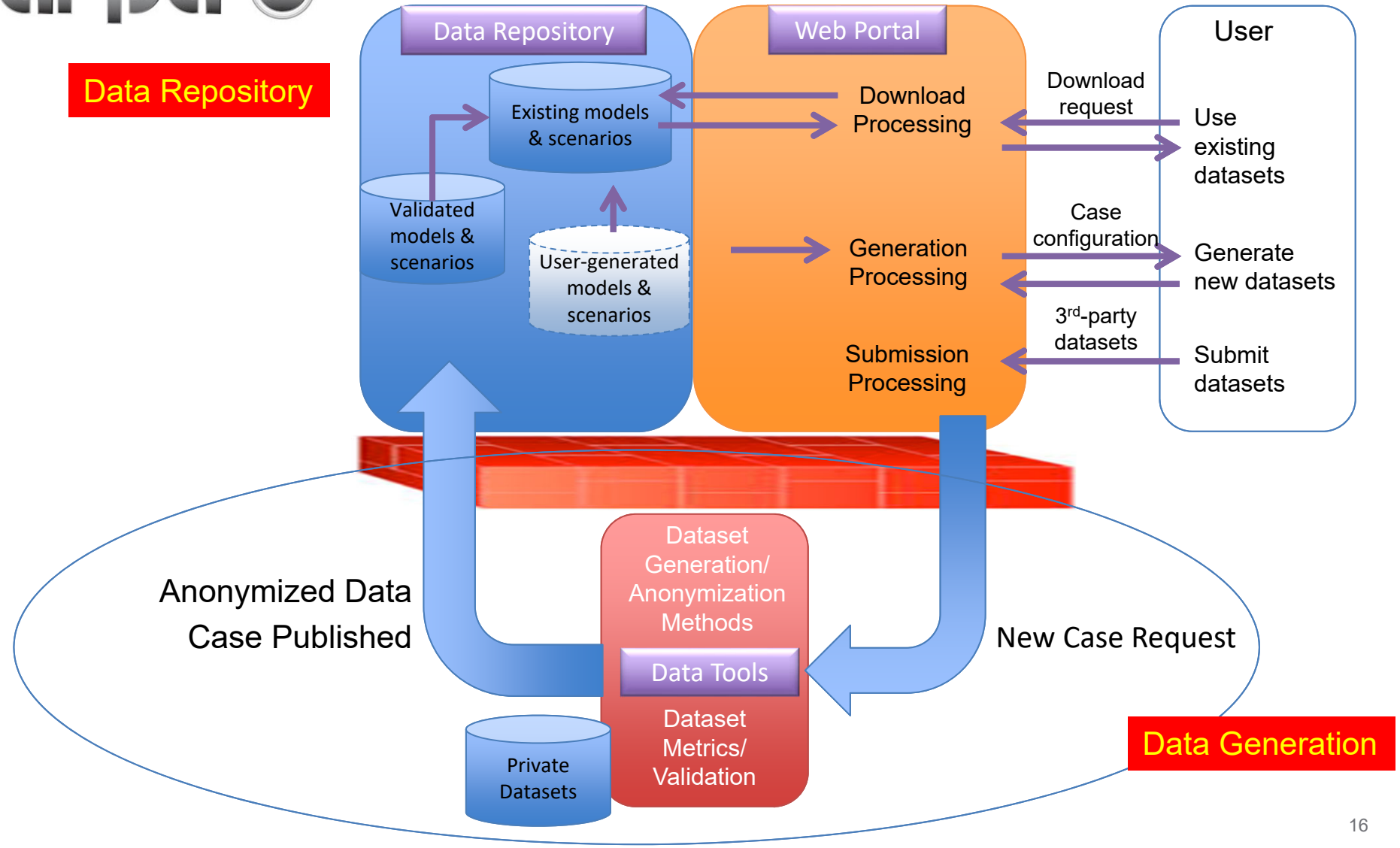
- Compares hydropower generation across different projects along Columbia and Snake rivers
- Alternative view shows generation across different sources such as hydropower, nuclear, renewables, and miscellaneous sources

Data repository for public hosting

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

arpa·e

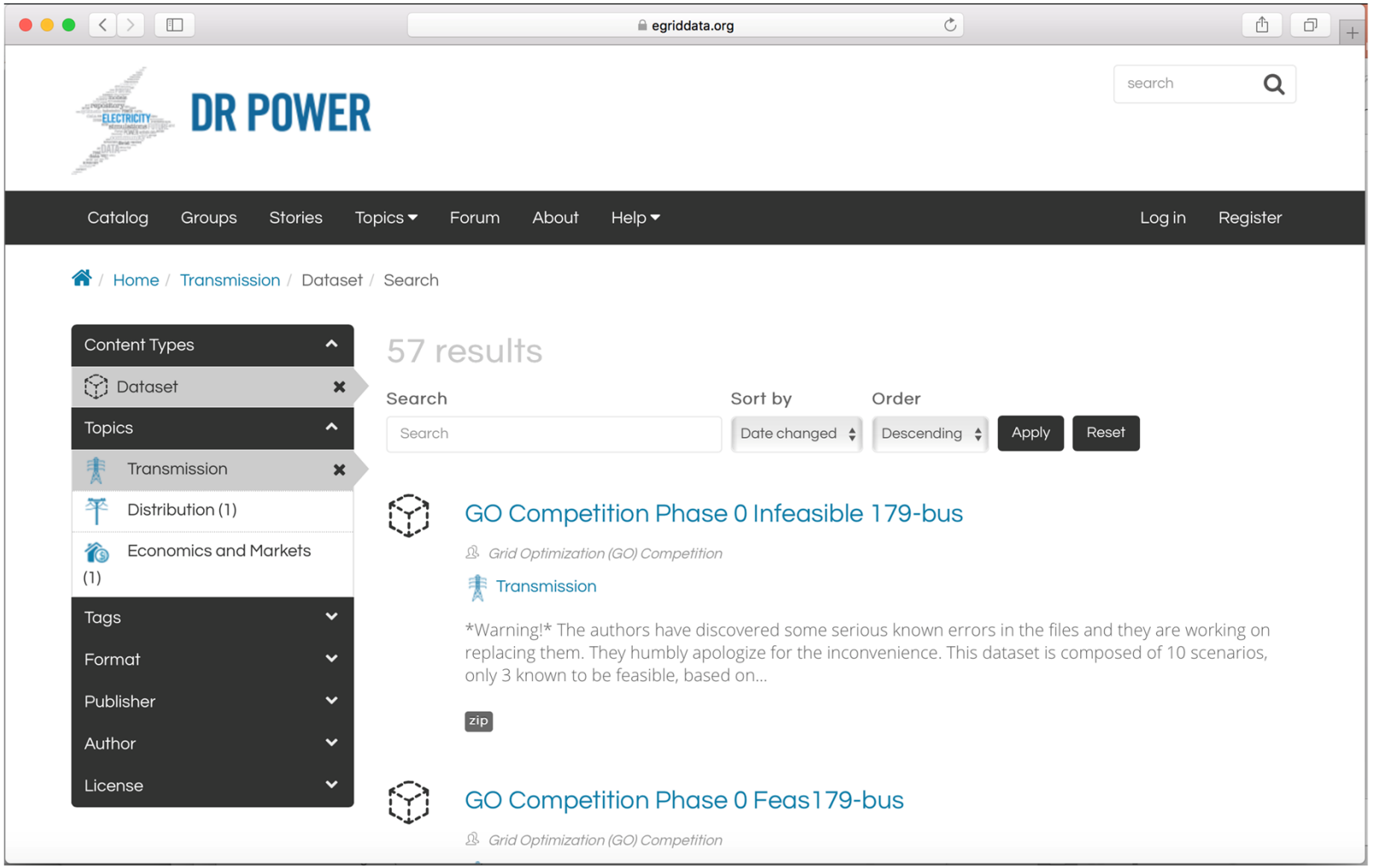
Data Repository



DR POWER: ARPA-E Data Repository

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

https://egriddata.org



ARPA-E Grid Optimization Competition (open now)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

https://gocompetition.energy.gov

Summary

- 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?

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

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

Pacific Northwest NATIONAL LABORATORY