ADVANCED RESEARCH PROJECTS AGENCY – ENERGY
RESEARCH PERFORMANCE PROGRESS REPORT

Open Funding Opportunity Announcement 2018
(OPEN 2018) DE-FOA-0001858

Award: DE-AR0001104
Prime Recipient: PingThings Inc.
Project Title: A National Infrastructure for Artificial Intelligence on the Grid
(PingThings, Inc.)
Principal Investigator: CEO Sean Murphy
3016765595 (phone) sean@pingthings.io
Reporting Period: 4/1/2021 to 6/30/2021

Technical - Accomplishments and Milestone Updates

M1.1: Time series benchmark suite v1.0 released	**Due Date:** 11/11/2019	**% complete:** 100
Release of the open source time series benchmark suite that fulfills the requirements determined in T1.1 as open source software including documentation. Documentation submitted to ARPA-E for PD approval.		
Performer Input: Completed		

M1.2: Cloud provider selected	**Due Date:** 2/11/2020	**% complete:** 100
Selection of the official cloud provider for the National Infrastructure project. Both the technical evaluation, including all relevant criteria, and the estimated non-discounted operational costs of the platform running on two different cloud providers will be fully documented and published for public consumption. Documentation submitted to ARPA-E for PD approval.		
Performer Input: Though we chose AWS as the cloud service provider to host the NI4AI platform, we remain committed to working with other vendors based on customer preferences.		

M1.3: Time series compression algorithm testing completed	**Due Date:** 5/11/2020	**% complete:** 100
The effectiveness and computational performance of the various lossy and lossless compression algorithms will be tested and documented on synchrophasor and point on wave data. Open source implementations of published time series compression algorithms (lossy and lossless) will be created. Current platform compression ratio improved by at least 5%. Documentation submitted to ARPA-E for PD approval.		
Performer Input: Completed		

M1.5: Go-No/Go: Platform deployed in selected cloud provider	**Due Date:** 8/11/2020	**% complete:** 100
Platform deployed in selected cloud provider and operational, ingesting data from at least 1 external sensor continuously. Demonstration of platform to ARPA-E.		
Performer Input: The platform is available at plot.ni4ai.org. New users may register for an account at ni4ai.org.		

M1.7: Data quality assessments developed, tested, and deployed on the platform.	**Due Date:** 2/11/2021	**% complete:** 100
Development and testing of a suite of data quality assessment algorithms including at least zero data, null data, and repeating data tests and then deployment on the live operational platform. Demonstration of data quality assessment tools to ARPA-E.		
Performer Input: We completed this milestone by deploying four data quality checks which are performed on ingest: zero data, null values (or gaps), repeat values, and duplicate time stamps. Checks rely on our at-scale calculation engine DISTIL (see Milestone 1.10). We have deployed an experimental suite of API functions (under btrdb-extras) to enable users to explore data quality information about streams.		

M1.10: Platform optimization documented	**Due Date:** 8/11/2021	**% complete:** 100
Evaluation and complete documentation of platform performance as a function of cost using different tiers of cloud storage and compute optimizations. This evaluation will include the use of the benchmark suite developed for M1.1. Submission of documentation to ARPA-E for PD approval. If cost effective, the platform will be upgraded with the enhancement(s).		
Performer Input: We deployed two cost optimizations under this milestone. Compute: We have deployed a new data processing engine called DISTIL, which uses the database structure and versioning to make batch and streaming calculations easier to deploy and more performant to run. Storage: We have deployed tiered storage. Frequently used data are kept in "hot" storage, while older data are rolled into cheaper "cold" storage to reduce data retention costs.		

M1.11: ML/AI Open Source Library Integration	**Due Date:** 8/11/2021	**% complete:** 95
Two selected machine learning and/or deep learning libraries have been implemented into the platform and are available for immediate use.		
Performer Input: The two libraries we have selected are Ray and SciKitLearn (sklearn). We are currently working to integrate these into the core platform functionality.		

M1.12: Additional platform enhancement gathering complete.	**Due Date:** 2/11/2022	**% complete:** 80
Completion of survey of current users for desired platform enhancements complete with findings documented, submitted to ARPA-E for PD approval, and publicly available.		
Performer Input: We have identified a number of possible enhancements and are engaging with new, existing, and potential users to focus efforts on developing enhancements that will maximize benefits to users.		

M1.13.1: Additional platform enhancement 1 completed	**Due Date:** 2/11/2022	**% complete:** 50
Development, testing, and launch of additional platform enhancement 1 given requirements collected in T1.12. Demonstration of additional enhancements to ARPA-E.		
Performer Input: This milestone is in progress.		

M1.13.2: Additional platform enhancement 2 completed	**Due Date:** 5/11/2022	**% complete:** 20
Development, testing, and launch of additional platform enhancement 2 given requirements collected in T1.12. Demonstration of additional enhancements to ARPA-E.		
Performer Input: We are working to define the scope of this milestone. See report for more detail.		

M1.13.3: Additional platform enhancement 3 completed	**Due Date:** 8/11/2022	**% complete:** 20
Development, testing, and launch of additional platform enhancement 3 given requirements collected in T1.12. Demonstration of additional enhancements to ARPA-E.		
Performer Input: We are working to define the scope of this milestone. See report for more detail.		

M2.1.2: Previously generated sensor data set ingested and available	**Due Date:** 11/11/2020	**% complete:** 100
The first previously generated sensor data set is ingested into the platform and made broadly available to the community.		
Performer Input: Completed		

M2.2: First in vitro data is streaming to the platform.	**Due Date:** 11/11/2020	**% complete:** 100
The first in vitro data set has been streamed into the platform.		
Performer Input: Completed		

M2.3.1: Go-No/Go: Successful sensor deployment to control group	**Due Date:** 2/11/2021	**% complete:** 100
The first wave of sensors to create the wide area or national data set have been acquired, tested, and deployed with team members or trusted collaborators. Data is streaming to the platform. Demonstration of streaming data to ARPA-E.		
Performer Input: Completed		

M2.3.2: Successful sensor deployment for national coverage data set	**Due Date:** 8/11/2021	**% complete:** 50
The first 20 sensors have been acquired, shipped, and deployed to generate the wide-area data set. Data is actively streaming to the platform. Demonstration of streaming data to ARPA-E.		
Performer Input: This milestone is in progress.		

M2.3.3: Successful sensor deployment for boutique data set	**Due Date:** 5/11/2021	**% complete:** 100
At least half the sensors for the first boutique data set have been acquired and deployed for the creation of the first boutique (site-specific) data set. Data is streaming into the platform. Demonstration of streaming data to ARPA-E.		
Performer Input: This milestone includes ingestion of both streaming and historical data under a series of "proof of concept" platform deployments making the platform available to potentially high-value customers managing large sensor fleets.		

M2.4.1: Legal document collection and analysis	**Due Date:** 8/11/2020	**% complete:** 100
Collection of at least 5 separate non-disclosure agreements and data sharing agreements from appropriate organizations. Legal analysis of the 5 separate documents to identify commonalities and essential components.		
Performer Input: Complete		

M2.4.2: “Common” NDA and DSA developed and released.	**Due Date:** 2/11/2021	**% complete:** 100
N/A		
Performer Input: Reporting materials include a copy of the GridBright NDA. We have also included a sensor hosting agreement outlining terms of data ownership for sensors deployed under the NI4AI project.		

M2.4.3: Deployed and tested data escrow functionality.	**Due Date:** 8/11/2021	**% complete:** 50
Data escrow requirements implemented as a service running on top of the platform with a successful exchange of data between two separate parties.		
Performer Input: See report.		

M3.2: Industry/academia project launch announcement	**Due Date:** 11/11/2019	**% complete:** 100
Project announcement across industry, academia, and utilities. This includes the dissemination of the initial press releases via ARPA-E and direct contact to at least three dozen select individuals and organizations of interest. This also includes outreach to potential project partners.		
Performer Input: Complete		

M3.3.1: Project blog launched	**Due Date:** 11/11/2019	**% complete:** 100
Successful design, creation, and structuring of the project blog with at least one initial post.		
Performer Input: Complete		

M3.3.2: Year 1 blog content creation completed	**Due Date:** 8/11/2020	**% complete:** 100
Successfully created and posted 20 blogs over the course of the year, with roughly even spacing throughout the calendar.		
Performer Input: Complete		

M3.3.3: Year 2 blog content creation completed	**Due Date:** 8/11/2021	**% complete:** 95
Successfully created and posted 20 blogs over the course of the year, with roughly even spacing throughout the calendar.		
Performer Input: This milestone is in progress.		

M3.4: Project website launched	**Due Date:** 2/11/2020	**% complete:** 100
Project website designed, built, and launched, complete with a project overview section and a team page.		
Performer Input: Complete.		

M3.5: Platform content created	**Due Date:** 2/11/2021	**% complete:** 100
The team has created at least 3 videos totaling 90 minutes in length, a dozen different Jupyter Notebooks for the platform, and comprehensive API documentation.		
Performer Input: Videos: https://blog.ni4ai.org/post/2020-10-31-workshops/ Jupyter notebooks: https://github.com/PingThingsIO/ni4ai-notebooks API documentation: https://btrdb.readthedocs.io/en/latest/		

M3.6.3: Conference attended	**Due Date:** 2/11/2021	**% complete:** 100
N/A		
Performer Input: Attended multiple conferences both as a presenter and as the manager of a virtual booth.		

M3.6.4: Conference attended	**Due Date:** 8/11/2021	**% complete:** 100
N/A		
Performer Input: Attended multiple conferences both as a presenter and as the manager of a virtual booth.		

M3.7.3: Project-focused workshop hosted	**Due Date:** 8/11/2022	**% complete:** 100
N/A		
Performer Input: This milestone is in progress.		

M3.8.1: Data science competition or hackathon hosted.	**Due Date:** 2/11/2021	**% complete:** 50
N/A		
Performer Input: We have been engaging pilot participants to gather use cases of interest, analytical methods, and implementation requirements. We are engaging in pair programming sessions and are building additional training materials to better enable practitioners to explore questions of interest to them in their data.		

Note that all text in *blue Italics* is from the last full SOPO Attachment that was exchanged between PingThings and ARPA-E. Also, the tasks appear ahead of the Milestones for which they lead into.

Task 1 - Data Platform Deployment and Enhancement

The Project Team will deploy an instance of the advanced sensor analytics and AI platform onto one of the premier cloud service providers as a platform-as-a-service. The team will continuously add any and all necessary features to the platform to support the identified tasks in the project and will support and maintain the platform during and after the project. Task 1 proceeds in 3 phases: (1) pre-launch; (2) platform launch; and (3) post-launch enhancements. Task 1 includes the following deliverables: (1) open source time series benchmark suite and documentation; (2) documentation of cloud provider selection criteria; (3) documentation of cloud provider final selection decision and rationale; (4) documentation of the performance of lossy and lossless compression algorithms; (5) documentation describing deployment of platform to multiple cloud providers; (6) additional platform language bindings; (7) additional data ingestors for selected sensors; (8) documentation of the cloud storage tier performance/cost optimization

T1.0 - Deployment of existing platform as project demo

Due Q1, Year 1

(Pre-launch) To hit the ground running, the team will launch a demo version of the platform in PingThings' private co-located server facility. This demo platform can be shown and used immediately.

9/30/2019 - 100% - Task Complete.

The existing technology stack has been deployed as a demo version of the NI4AI platform. We opted to deploy this in the AWS cloud rather than at a co-located facility to enable a seamless transition from the demo platform to a production platform over the next few months. In addition to the existing technology deployment, we have already deployed a custom-branded portal site to tie together the community-facing aspects of the project such as the Blog and project mission statements. The site allows for federated authentication using social providers to allow for potential members of the community to rapidly gain access to the platform and existing data sets. Please visit [https://ni4ai.org/].

T1.1 - Open source time series benchmark

Due Q1, Year 1

(Pre-launch) The team will immediately develop a benchmarking suite for time series data. This will allow us to better distinguish the platform from competitors using a standardized suite of benchmarks that allows utilities to compare and contrast all phases of time series work including data ingest, reading/access, storage, and, importantly, analysis. We are working across vendors to develop such a test suite.

The benchmarks will contain two major areas: (1) traditional database benchmarks (read/write/query data) and (2) analytics benchmarks with particular read/write patterns and computational complexity reflective of traditional engineering and sensor related tasks. This task will be broken down into (a) requirements [10%], (b) implementation [50%], (c) testing [70%], (d) release [80%], and (e) documentation [100%].

12/31/2019 - 100% - Task Complete

This task has been completed. See M1.1 for more details.

M1.1 - Time series benchmark suite v1.0 released

Release of the open source time series benchmark suite that fulfills the requirements determined in T1.1 as open source software including documentation. Documentation submitted to ARPA-E for PD approval.

12/31/2019 - 100% - Milestone Complete

Version 1.0 of the benchmark suite, fulfilling the requirements of T1.1 and M1.1, was released on 11/11/2019. This initial version includes results for the current market leaders (InfluxDB, Timescale) as well as results for the NI4AI demo platform deployed in T1.0. New versions will be released periodically throughout the project as the requirements and available technology offerings change. Please see the completion report for more information

T1.2 - Platform cloud provider selection

(Pre-launch) The team will evaluate multiple cloud providers and determine the "best" provider. The criteria for provider selection will be based on technical merit, cost, willingness to negotiate discounted rates or "cloud credits," experience using the provider, and platform benchmarks. The list of potential cloud providers includes but is not necessarily limited to AWS, Google Cloud, Microsoft Azure, Digital Ocean, and Linode. Steps include (a) technical criteria development [20%], (b) cloud provider technical and cost evaluation [40%], (c) cloud provider negotiation [60%], (d) test deployment and benchmarking [90%], and (e) provider selection.

12/31/2019 - 40%

We have gone end to end with the cloud platform for which we have the most experience - Amazon - and consider this task to be iterative, so that the lessons learned evaluating the first provider can then be applied to better and more efficiently evaluate the second and so on.

	Amazon	Google	Microsoft
Technical Criteria Development	Drafted	-	-
Cloud Provider Technical and Cost Evaluation	Complete	Ongoing	Ongoing
Cloud Provider Negotiation	Ongoing	Complete	Initiated
Test Deployment and Benchmarking	Complete	Ongoing	Not Started Yet

Negotiations for Google Cloud Platform (GCP) credits have been completed and we have secured cloud credits for the project.

We are working to identify the correct individuals within Amazon, Microsoft, and IBM and others to discuss possible collaborations and potential cloud usage discounts.

03/31/2020 - 100%
This task is documented in the Milestone report.

M1.2 - Cloud provider selected
Selection of the official cloud provider for the National Infrastructure project. Both the technical evaluation, including all relevant criteria, and the estimated non-discounted operational costs of the platform running on two different cloud providers will be fully documented and published for public consumption. Documentation submitted to ARPA-E for PD approval.

12/31/2019 - 25%
We have thoroughly evaluated AWS for the project and have started to evaluate Google's and Microsoft's cloud platforms.

03/31/2020 - 100%
Please see Milestone report.

T1.3 - Enhanced time series compression algorithms
(Pre-launch) Cloud storage costs remain a key driver for operating the national infrastructure platform. As such, both lossy and lossless compression approaches will be investigated in depth with the goal of implementing state of the art lossy and lossless compression approaches for time series sensor data. This task will

- *(a) identify current state of the art for both lossy and lossless time series data compression [20%],*
- *(b) prototype 2 lossy and lossless compression schemes [40%],*
- *(c) attempt to improve given prototype schemes [60%],*
- *(d) benchmark and document compression schemes [80%], and*
- *(e) implement final scheme for platform [100%].*

12/31/2019 - 20%
We must be careful to maintain a tight scope around this task as it is highly likely that it could be repeated for each sensor type. The optimal compression approach will likely change depending on the nature of the signal. The best strategy for compressing point on wave data may be very different than the best approach for compressing voltage phase angle data or frequency data. Thus, the evaluation framework above (a-e), is applicable to each type of signal compression. We are focusing our initial work on timestamp compression, especially for higher frequency sensor data from such things as synchrophasors and continuous point on wave; half of the data volume will come from regularly spaced time stamps and there is much storage space to be saved.

03/31/2020 - 60%
We have investigated various lossy compression schemes to determine their computational efficiency and how much data they lose.

06/31/2020 - 100%
See the milestone report.

M1.3 - Time series compression algorithm testing completed
The effectiveness and computational performance of the various lossy and lossless compression algorithms will be tested and documented on synchrophasor and point on wave data. Open source implementations of published time series compression algorithms (lossy and lossless) will be created. Current platform compression ratio improved by at least 5%. Documentation submitted to ARPA-E for PD approval.

12/31/2019 - 20%
Initial progress has been made in implementing current best compression approaches. More details will be coming in the next quarterly update.

03/31/2020 - 60%
See above

06/31/2020 - 100%
See the milestone report.

T1.4 - Development of an online method to onboard a new sensor
Due Q6, Year 2
(Pre-launch) Adding a new sensor to the platform must be easy for individuals with different backgrounds and levels of experience. The team will build an online mechanism that makes onboarding a new sensor simple and fast. This process will allow for the configuration of metadata and the streaming of data from the sensor to the platform.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - 30%
We have developed a new C37 data ingestor that enables sensors to connect to us, rather than us connecting to them. We are working with Powerside to ensure that the micro-PMU is configured in a way that allows us to easily onboard new sensors and minimize the effort required of sensor hosts.

12/31/2020 - 50%
Onboarding sensors involves two steps: (1) capturing metadata about the sensor, and (2) establishing a safe, secure, reliable, and controlled connection for streaming measurements to the platform. For (1), we need a mechanism to collect and revise metadata -- e.g., to correct mistakes, capture new details and

discoveries, or update information that changes. This task aims to establish a process for sensor hosts to perform these tasks autonomously through an online interface (such as google forms).

Currently, both tasks are performed manually and can only be done by users with administrative access to the platform (i.e., by project team members). Due to limitations with the uPMU firmware, step (1) requires sensor hosts to manually configure firewall settings on their local wireless network to enable us to connect to each sensor (see M2.4.1 updates). As we send out our first round of sensors, we will circulate a google form to collect metadata and detail connection requirements. Lessons learned during these initial deployments will allow us to formalize and automate the onboarding process for future deployments.

03/31/2021 - 60%
We have prototyped a sensor onboarding system using a google form to collect the information we need to establish connections for streaming sensors, and gather associated metadata. We have vetted the form with our control group, and will solicit additional feedback as we deploy additional sensors to project team members.

Minimum requirements to establish a connection for new sensors include the following:

- IP address
- Port
- C37 ID
- Contact information (e.g., for alerting purposes)

06/30/2021 - 70%
This performance period we completed designs for an online portal users may use to turn on various types of ingresses. These include: C37 and GEP protocols, as well as a tool for ingesting data from any URL on the web (see Task 2.1.2 update). This self-service tool collects connection details needed to receive data from streaming sensors (e.g., via C37 or GEP), and to insert metadata about added sensors. Users will also be able to turn on ingresses to pull in real-time data from public data sources. Currently, this offering includes geomagnetic field data published by USGS. Ultimately, this offering could include other data sources of interest to users as well.

We are currently engaging with key platform users responsible for ingress management to determine functionality requirements for our larger customer base beyond NI4AI.

T1.5 - Platform deployment to cloud
Due Q4, Year 1
(Launch) The platform will be deployed on the selected cloud provider. All deployment steps will be thoroughly documented.

12/31/2019 - Not applicable
03/31/2020 - 50%

We are just over half way complete on deploying the platform. The core database systems are present and the systems for ingesting data are present and tested. We still need to deploy additional systems for improving the onboarding of sensors as well as several changes to tailor the system further to the needs of the NI4AI project. We also have several features that will lower the operational cost of the system and make it easier to scale. These are moving along in the design and development stages and will be deployed later in 2020.

06/30/2020 - 80%
We are in the process of doing the final launch of the platform to our selected Cloud Provider (AWS). Our release has been built and our migration tools have been tested. The deployment will be done by July.

09/30/2020 - 100%
We have launched the platform in AWS with data streaming from two OpenPMU sensors installed at a project member's home. We demonstrated the platform during the last quarterly review, and shared a video of the demonstration. We have since recorded a longer video demonstrating the full functionality of the platform. The video is posted on our blog.

Since that time we have added three additional sensors previously deployed as part of the Texas Synchrophasor Network. We socialized the availability of the platform via numerous outreach channels including PowerGlobe, LinkedIn, and sending both broad and personalized emails to relevant members of the community. Outreach efforts are detailed in our milestone report.

M1.5 - Go-No/Go: Platform deployed in selected cloud provider
Due Q4, Year 1
Platform deployed in selected cloud provider and operational, ingesting data from at least 1 external sensor continuously. Demonstration of platform to ARPA-E.

12/31/2019 - Not applicable
03/31/2020 - 50%
Please see above.

06/30/2020 - 80%
Please see above.

09/30/2020 - 100%
See milestone report.

T1.6 - Implementation of additional data ingestors
Due Q6, Year 2
(Launch) The team will build data ingestors needed for the platform to ingest data from the acquired sensors. The team will continue to build ingestors as needed as new sensors are connected to the platform.

12/31/2019 - 1%

The team has been discussing with industry and academia which types of sensors are most likely to be selected for the project. This dialog is helping us determine which data ingestors may be needed in the future.

03/31/2020 - 5%
Team continues discussion with industry and academia to determine types of sensors to deploy and examine the various protocols being used for data transmission. To date, we have not hit anything extraordinary or proprietary that would block us from developing an ingestor.

06/30/2020 - 50%
We have almost finished development of an IEC 61850 protocol ingestor for the platform.

09/30/2020 - 90%
Data ingesters for C37 and GEP are both implemented. C37 is used by the Texas Synchrophasor Network, OpenPMU, and micro-PMU. We developed a new C37 splitter that allows us to accept connections from sensors instead of us connecting to them. This will streamline the process of adding new sensors.

We have a working prototype for IEC 61850 (commonly used for point on wave data), though it has not been deployed as it has not been needed to date.

Partnership with PowerSide may warrant development of an additional data ingestor to accommodate PQube3 data.

12/31/2020 - 100%
See milestone report.

03/31/2021 - 100%
We have had no reason to date to deploy our IEC 61850 ingestor to date. Now that we have acquired the uPMU, we plan to develop an additional ingestor for the PQube (as the same sensor can support both modes).

We also have ingestors for FNET and CPOW sensors. These have been tested, validated, and have been deployed on other instances of the platform. They have not been deployed on NI4AI simply because we have not yet had the need to connect a sensor that requires either ingress.

uPMU deployments under Task 2.3.1 have revealed that certain changes to our C37 ingestor will be needed for the project. These include:

- Ability to customize stream names (as default names are uninformative and confusing)
- Ability to limit which fields from a C37 data frames are ingested and stored. This need arises from the fact that sensors installed in a wall outlet can only measure single phase voltage. However, the device streams “data” (which is really just measurement noise) on all three voltage and current channels. Our ingestor will need to be revisited to allow us to select which values in each data frame to archive, and which to discard.

M1.6 - Sensor data ingestors developed, tested, and deployed on the platform

Due Q5, Year 2

Development, testing, and deployment on the live operational platform of at least two additional data ingestors (i.e., the capability of handling two different types of sensor data[AA(1)]). Demonstration of additional data ingestors to ARPA-E.

12/31/2019 - Not applicable

03/31/2020 - Not applicable

06/30/2020 - Not applicable

09/30/2020 - 90%

Data ingesters for C37 and GEP are both deployed, and we have a working prototype for IEC 61850 but, unfortunately, no sensors that use the protocol. All five sensors currently streaming to the platform use C37. We will continue to develop and deploy additional ingestors as we gain access to new datasets.

12/31/2020 - 100%

See milestone report.

03/31/2021 - 100%

We have developed a new ingestor to support C37 data sent via UDP. We also have sensors to ingest data from Eyedro, FNET, and CPOW sensors. We will deploy these on NI4AI if there is a need to ingest data from these other sensors. Currently our plan is to deploy only uPMUs.

T1.7 - Ingest-stage data quality assessments

Due Q6, Year 2

(Launch) Data quality assessment is key for making real world sensor data useful to downstream consumers and end users. The team will implement an array of different data quality assessment tests that can run on real time as sensor data arrives at the platform.

12/31/2019 - Not applicable

03/31/2020 - Not applicable

06/30/2020 - 30%

We have been developing mechanisms for assessing the general quality of time series data.

09/30/2020 - 50%

We have developed a suite of functions for assessing data quality. These include:

1. Looking for missing data
2. Checking for duplicate time stamps
3. Detecting deviations from the expected sampling interval
4. Sample variability (i.e., to detect gaps or repeated values)
5. Observations beyond sensible bounds
6. Intervals where the signal to noise ratio is abnormally high
7. Non-responsive sensors

Currently, these are implemented as functions in a jupyter notebook. We are evaluating which of these checks are appropriate to do on ingest.

Outreach efforts around the Sept 6 brownout event (discussed in our Milestone 1.5 report) have revealed that there are open questions in the literature regarding PMU data quality during power quality events. We have found that practitioners look to confirm that frequency (and ROCOF) values are within reasonable bounds. The issue stems from the fact that curve-fitting methods used to compute phase angle and frequency break down when waveforms deviate from sinusoidal (e.g., during faults). We have yet to learn of a systematic approach detecting such events from PMU measurements.

We have formulated the problem as an exploratory data analysis exercise for our blog encouraging users to question the quality of phasor measurements by comparing local and wide-area measurements of frequency and ROCOF.

This data quality issue is one of the factors that has garnered interest in point on wave and other power quality measurements. We are continuing to engage with the research community to determine if and how it would be most appropriate to systematically address the issue in a data hosting platform. We are also in discussions with Powerside about modifying the micro-PMU to provide point-on-wave snapshots, or about coupling certain deployments with the PQube3 to monitor power quality in addition to phasors.

12/31/2020 - 80%

Data quality checks we have implemented include:

- Repeated measurements
- Duplicate time stamps
- Density analysis (i.e., measurements per unit time)
- Non-zero values
- Deviation from nominal voltage
- Signal to noise ratio
- Timestamp jitter
- Obviously bad values (e.g., negative magnitudes)

Each check has been tested and run on live data in the platform. The output is an indicator showing if and when a stream violates each check. Two mechanisms for reducing storage cost associated with data quality indicators are: (1) to store data as a sparse data stream indicating only if and where there is a data quality issue, and (2) to aggregate values at higher levels in the tree database -- e.g., across measurements recorded in each 1-second, 1-minute, or 1-hour time intervals rather than for individual measurements.

We are currently working to implement data quality checks in Go so that they can be run using DISTIL. DISTIL is a back end tool which runs calculations on stream data, and archives the results as a separate stream. We are adapting DISTIL to run calculations *on ingest* while individual measurements are already being held in memory to compute other summary statistics. This could reduce computing overhead associated with running data quality checks.

3/31/2021 - 90%
We have built a fully operational data quality assessment system in Python, and have prototyped fully functional visualization capabilities for our front end (under Tasks 1.9.1 and 1.9.2).

We are now moving data quality checks into DISTIL, which was recently deployed in production (see Task 1.11). DISTIL is a foundational software tool that underlies all calculations done at scale. Calculated run as batch process can also be done with far fewer compute resources if calculations are implemented in DISTIL. Currently the tool works on our back end and requires a developer to write functions. We are working towards making a user-friendly feature where customers (or ni4ai users) can specify calculations to be performed in DISTIL (see Task 1.11 for more detail).

We decided to prioritize work on DISTIL after determining that improvements there would allow us to work more efficiently on Milestone 1.7, advance Milestone 1.11, and support work being performed on other projects as well. Now that DISTIL has been released (internally), we are working to move data quality checks from Python into DISTIL. To date we have implemented two of the three checks, specified under Milestone 1.7 -- including zero data and missing (i.e., null) values.

06/30/2021 - 100%
See milestone report.

M1.7 - Data quality assessments developed, tested, and deployed on the platform.
Due Q6, Year 2
Development and testing of a suite of data quality assessment algorithms including at least zero data, null data, and repeating data tests and then deployment on the live operational platform. Demonstration of data quality assessment tools to ARPA-E.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - 30%
We have been developing mechanisms for assessing the general quality of time series data.

09/30/2020 - 50%
See Task 1.7 above.

12/31/2020 - 80%
See task description above.

03/31/2020 - 90%
See Task 1.7 above.

06/30/2021 - 100%
See milestone report.

T1.8 - Platform language bindings

Due Q8, Year 2

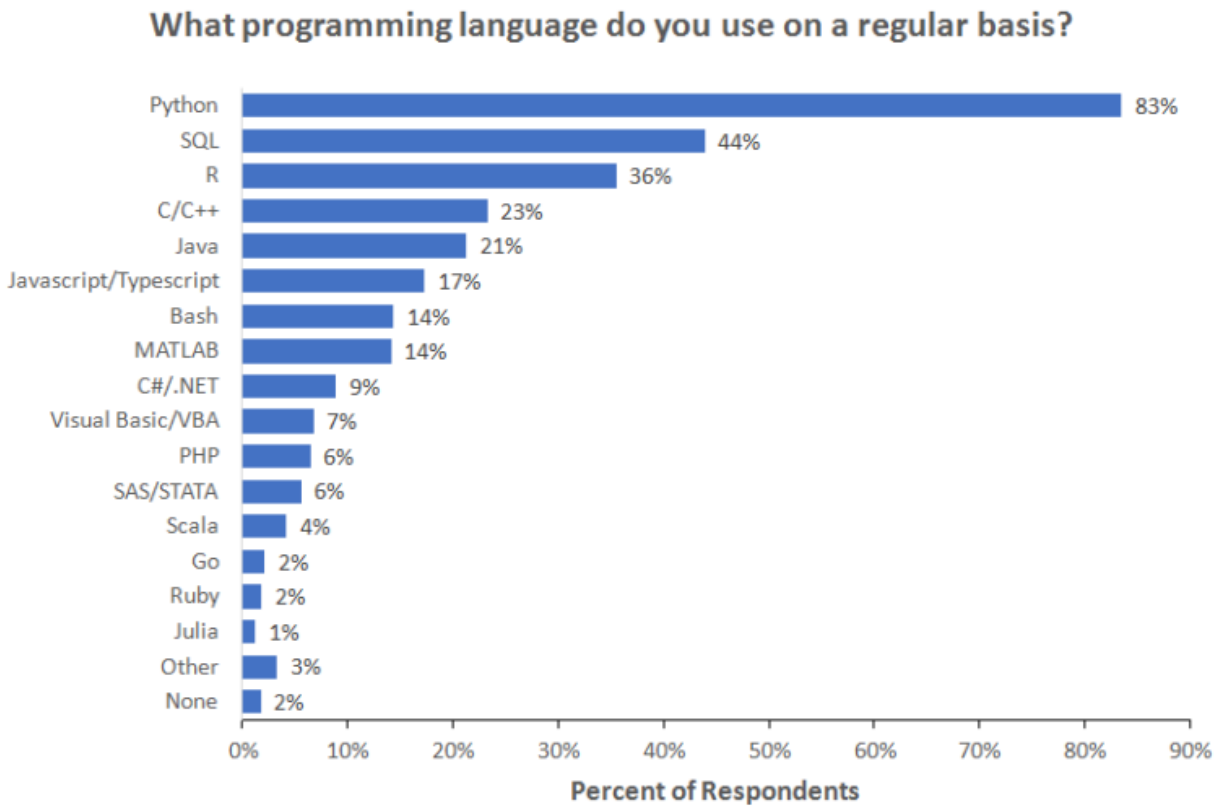
(Post-launch) The team will survey the user base to learn which programming languages must be supported. The team will then develop language bindings for these programming languages.

12/31/2019 - 50%

The team has developed bindings for Go, Python, and Julia discussed below in the associated milestone 1.8. We selected Julia because there is an active and passionate community that is intensely interested in scalable and high performance computing. Also, we know that at least one national lab has started to do grid model development in Julia and has open sourced their work.

A survey has been sent to data scientists within the utility space and within academia. As of 9/30/2019, responses are still coming in. This survey will enable us to determine which languages to concentrate development efforts on in order to maximize benefit to the community.

Top data science languages is also an interest to many others who have also surveyed the broader industry (and not just power engineers). The data below is from Kaggle's very extensive 2018 survey (https://www.kaggle.com/kaggle/kaggle-survey-2018)



Note: Data are from the 2018 Kaggle Machine Learning and Data Science Survey. You can learn more about the study here: http://www.kaggle.com/kaggle/kaggle-survey-2018. A total of 18827 respondents answered the question.

Finally, we have established connections to the OpenECA (Extensible Control and Analytics) team funded by DoE that looked at real time processing of grid data. They also surveyed the power industry to learn which languages must be supported and found the following top three:

1. .NET (C# / F# / VB)
2. C++
3. Java

And then

1. Matlab
2. Python
3. Javascript
4. MS Excel

Please note that their project had different goals and was focused on real time processing.

We had a very positive initial interaction with the Mathworks team but encountered internal opposition at Mathworks to developing a stronger collaboration.

The team has also started to investigate language binding construction for C#.

03/31/2020 - 50%
06/30/2020 - 60%
Work continues to polish the existing bindings.

09/30/2020 - 90%
We have released bindings for Python, Go and Julia. Links to Python and Go documentation are below.

Python documentation: https://btrdb.readthedocs.io/en/latest/
Go documentation: https://pkg.go.dev/gopkg.in/btrdb.v3

Julia is evolving quickly and the language is still relatively unstable. This means that documentation would likely need to be updated. It also means that most Julia users are proficient in other languages, and that Python documentation should be sufficient for most Julia users (at least for now). We will continue to monitor API usage and will revisit the need to document the Julia bindings as appropriate. The upside with Julia is, given the insane performance the language offers, it is well aligned with the benefits offered by the platform.

We have also had some interest in developing additional bindings for R.

12/31/2020 - 100%
See milestone report.

M1.8 - Additional platform language bindings made available.
Due Q7, Year 2
Completion and documentation of at least two platform language bindings for programming languages selected based on community feedback. Documentation submitted to ARPA-E for PD approval.

12/31/2019 - 40%
Initial work on language bindings has begun. We have a prototype of bindings for Julia and production ready bindings for Python and Go. After the survey results have been collated and analyzed, new language bindings will be identified and developed. Based on the incomplete survey dataset, it appears that the next candidate is likely to be Matlab or C#.

Python bindings: https://btrdb.readthedocs.io/en/latest
Go bindings: https://godoc.org/github.com/BTrDB/btrdb
Julia bindings: https://pingthingsio.github.io/BTrDB.jl/latest

03/31/2020 - 40%
06/30/2020 - 60%
09/30/2020 - 90%
See Task 1.8 above.

12/31/2020 - 100%
See milestone report.

03/31/2021 - 100%
This performance period we are wrapping up work on a DOE project funded under FOA 1861 (internally referred to as "AGAVE"). The project enabled us to form a data science team, and gave them a chance to get substantial experience interfacing with the platform. After the project comes to a close, we plan to internalize lessons learned from work on the project to improve our user experience for other data scientists working with the platform.

One of the key areas for improvement includes building out additional capabilities for the Python bindings.

T1.9 - Integration of data quality assessment into platform

Due Q8, Year 2

(Post-launch) Once the sensor data quality assessment is occurring during stream ingest, the information must be made available to the user. The team will make the data quality information (DQI) available in a variety of ways delineated in the associated sub tasks.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - 20%
See Task 1.7 above.

12/31/2020 - 35%
See details on progress below.

03/31/2021 - 50%
See details on progress below.

06/30/2021 - 80%
See details on progress below.

T1.9.1 - DQI integration into core time series visualization functionality

Due Q8, Year 2

Data quality information must be clearly, quickly, and easily communicated to users browsing time series data and therefore must be integrated into the multi-resolution time series visualization engine.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 40%
As discussed under Task 1.7 above, data quality checks return a metric called "stream quality index" which tracks the quality of a given stream over time. Once data quality checks are run, the stream quality index will be stored as an additional stream in the database, which users can select and visualize as they would any other stream. We are

also exploring methods for integrating stream quality information into the standard visualization for each stream, for example in a manner similar to how count data are shown along the top of the display.

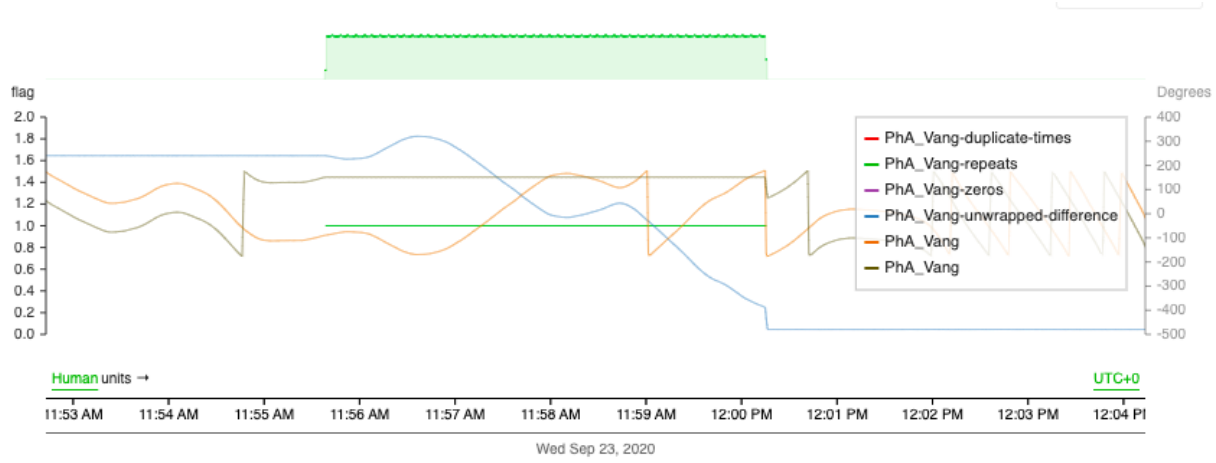
03/31/2020 - 50%
We are currently working to synthesize information about best practices for data quality integration. Lessons learned from that effort will inform next steps for DQI integration. In parallel, we are working to develop DISTIL functions to improve the computational efficiency of generating data quality streams.

06/30/2021 - 100%
Data quality checks have been implemented on ingest under Milestone 1.7.

The results are stored as a time series indicating when possible data quality issues occurred. No values are stored unless a data quality check fails, thus minimizing storage costs associated with data quality streams. Further cost optimizations may be achieved by reducing temporal resolution of the streams if needed.

Data quality streams integrate into our core visualization capabilities by providing a visual queue when issues are present. An example is detailed below.

The figure below shows phase angle time series at two nodes (PhA_Vang), as well as the phase angle difference between the two nodes (PhA_Vang-unwrapped-difference). Phase angle unwrapping is a necessary pre-processing step in phase angle differencing. Various data quality assessments are selected in the visualization interface, including: duplicate-times, repeats, and zeros. These are plotted alongside the raw and derived phase angle streams. The green line and histogram at the top of the plot indicate the presence of repeat values during the time interval shown.



T1.9.2 - DQI integration into sensor fleet management functionality

Due Q8, Year 2

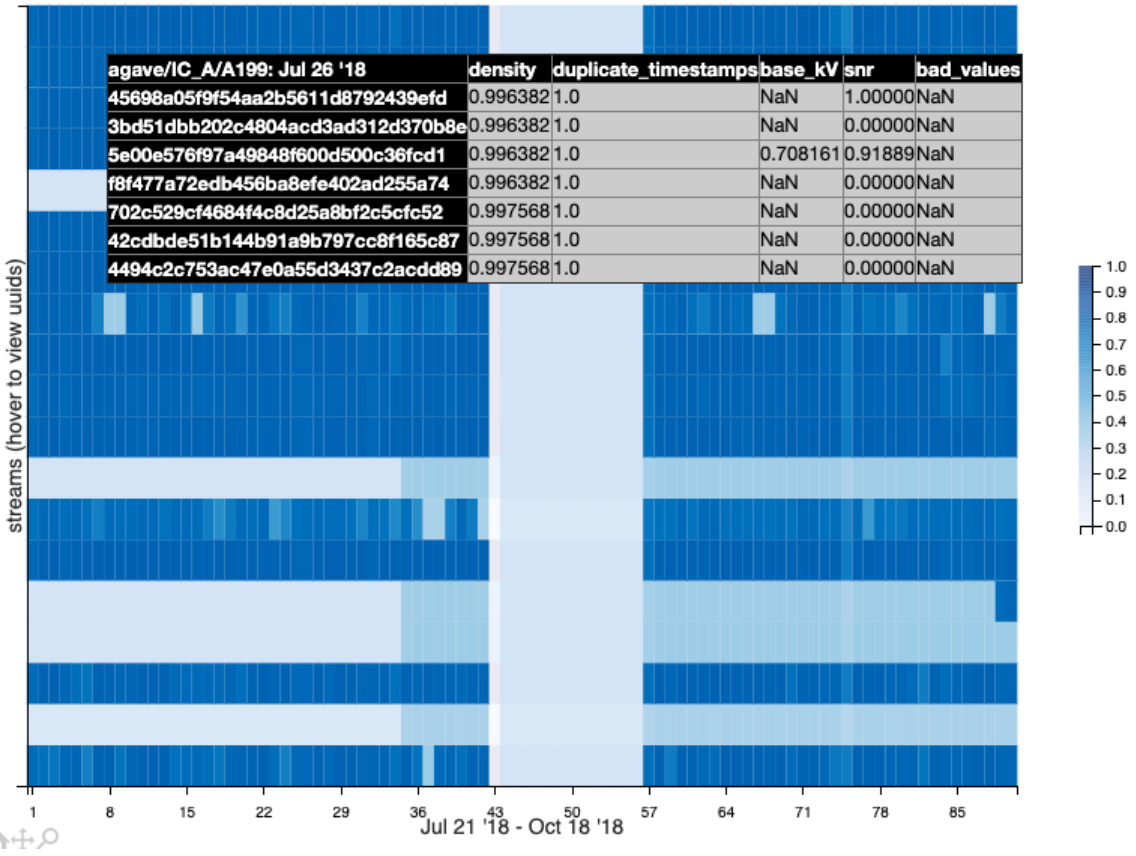
The management of a large number of sensors can consume significant resources. The team will build a web-based application that allows for a global view of the state of sensors that are part of the platform.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable

09/30/2020 - Not applicable

12/31/2020 - 30%

After exploring different visualization options, we found a heatmap to be an intuitive way to show stream quality across a fleet of sensors. We are adding a heatmap to our library of visualization tools to show how data quality (or measurement values) change over time across selected streams. Below is a prototype of the heatmap, indicating the overall SQI value for each stream (y-axis) with respect to time (x-axis).



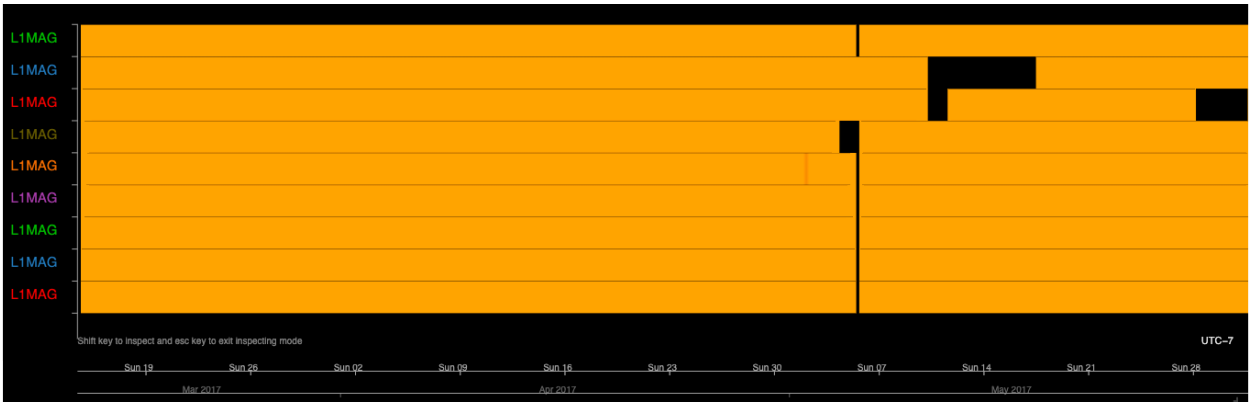
03/31/2021 - 90%

The heatmap is in the final stages of development and will likely be deployed in the upcoming weeks. Feature shows time-varying quantities across user-selected streams -- e.g., streams coming from a fleet of sensors. The heatmap supports visualization of stream quality data (as pictured above) and of values reported in the data.

06/30/2021 - 100%

We have released a beta version of the heatmap which can be used to visualize stream values, counts, or data quality issues. The screenshot below shows a heatmap of data availability for nine sensors in the Golden data set. The figure clearly shows where sensor data is present, and where there are gaps in certain streams. We have provided various heatmaps and visualization schemes which may be suitable for different types of data.

We are currently working to further improve the overall aesthetics and design of the heatmap, and to optimize performance for visualizations of larger numbers of streams that span large sensor fleets.



T1.9.3 - DQI used for alerting

Due Q8, Year 2

A change in the status of the sensor often requires immediate action. The team will develop a web-based application that allows various users to receive notifications based on a defined change in the data quality information associated with a sensor.

12/31/2019 - Not applicable

03/31/2020 - Not applicable

06/30/2020 - Not applicable

09/30/2020 - Not applicable

12/31/2020 - 20%

Our engineering team has developed alerts which notify them when the status of a sensor changes. This allows us to identify and diagnose issues with incoming data, and to reach out to customers if intervention is needed.

This quarter, we experienced connection issues with one of the sensors in the Texas Synchrophasor Network. We received an alert, and contacted the sensor owner to make them aware. They informed us that the issue was due to IT work being done in their building. A simple solution for alerting sensor hosts is simply to draft a canned email which we can forward to them. We plan to explore other options for alerting once we have a better understanding of user needs.

03/31/2021 - 25%

In our last quarterly report (above) we described the current alerting capabilities of the platform, which are used to inform internal personnel about data quality or connection issues. This performance period we have had a number of alerting incidents that have allowed us to gather requirements for exposing these alerts to users. These include:

1. Sending a daily summary of data quality issues that occurred to allow timely investigation into potential issues.
2. Real-time alert when sensor communication has ceased.
3. Limiting the max number of alerts sent (for issues that are intermittent but recurring)
4. Alerting (e.g., via daily summary) on a broad cross section of data quality checks. Issues observed in sensors deployed to date include: zero values, repeat values, out-of-range values, missing values (i.e., dropouts), high signal-to-noise ratio. Each of these has warranted some kind of intervention related to either sensor deployment or data management.

06/30/2021 - 40%

This performance period we have surveyed open source tools for managing messaging queues. Tools evaluated include rabbitmq, kafka, and redis. These services enable the creation of messages about arbitrary events -- such as new user creation, new ingresses, changes in the state of a stream, or arbitrary event detection schemes running on the data. These events are archived and passed to a messaging queue. Messages in a queue may be synthesized in various ways to trigger various types of actions, including email or text message alerts. Alerts may be sent immediately, or as a regular (e.g., daily or weekly) digest based on user-defined preferences.

We are working to better understand what is required of data quality alerts based on the nature of issues detected and stated preferences around notification frequency and speed.

T1.10 - Optimize platform

Due Q8, Year 2

(Post-launch) Examine ways to lower the cost of operating the platform in the selected cloud provider.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 30%

See task descriptions below.

03/31/2021 - 65%

We have made considerable progress on platform optimizations during this performance period. The benefits of these optimizations go well beyond the project itself. These reduce our costs, allowing us to produce more competitive bids on new contracts and differentiating us from our competitors. They also increase our profit margins.

We have prioritized work on Task 1.10 over other tasks (e.g., Task 1.7) because platform optimizations constituted a major opportunity for the company with cross cutting benefits across projects, customer accounts, and proposals.

06/30/2021 - 100%

See milestone report.

T1.10.1 - Cloud compute cost optimization.

Due Q8, Year 2

(Post-launch) The team will benchmark platform with time series benchmark suite, identify bottlenecks for performance optimization, implement performance optimizations for the platform, and upgrade the platform and retest.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 20%

Computing costs stem from data ingests, exports, and computing resources granted to users via jupyter hub instances, and API access. We plan to benchmark the relative contribution of each by tracking where CPU cycles are sent on the Dominion cluster, and determine which optimizations will achieve the greatest cost reductions.

We are also working on various measures to avoid incurring high computing costs due to normal interactions by individual users. These include limiting the size/frequency of API queries, data exports, and constraining computing resources available to users for free. As we integrate jupyterhub into NI4AI and begin to accumulate more users, plan to reduce computing costs by dynamically allocating servers if and when they are needed, and releasing them when they are idle.

03/31/2021 - 50%
The main measure we have taken this quarter to reduce computing costs is to advance work on DISTIL. DISTIL is a back end tool for performing calculations on data in a way that utilizes the structure of the database to maximize computational efficiency (and speed). It performs significantly better than calculations done in Python. Calculations or transformations done using DISTIL are written in Go.

DISTIL reduces computational costs by consolidating the computational burden of common calculations into a single batch process. Thus rather than having replicating database queries and data transformations each time a user signs on, the user can simply generate a stream reporting the results of those transformations. This also allows users to more easily arrive at summary statistics for derived streams. This capability is critical to performing certain tasks -- such as phase angle monitoring and model validation tasks (e.g., topology change detection, impedance calculation, power flow monitoring, parameter estimation, etc.).

We may also choose to use this task to optimize our process for spinning up new instances of the platform. Here the “optimization” would be principally to reduce labor costs.

06/30/2021 - 100%
See milestone report.

T1.10.2- Cloud storage cost optimization
Due Q8, Year 2
(Post-launch) Look at ways of integrating slower, lower-cost storage into the platform in a tiered caching system that looks to provide a particular performance/cost ratio.

12/31/2019 - Not Applicable
03/31/2020 - 10%
Given that cloud storage costs are such a fundamental driver of expenses for this project, we have begun examining ways that we can mitigate this project risk, a development which will allow us to potentially collect and host additional project data.

06/30/2020 - 10%

09/30/2020 - 30%
We have begun to optimize cloud storage costs. We will provide a detailed report as the work progresses.

12/31/2020 - 40%
We are developing the capability to use tiered storage to reduce storage costs for data that users do not need to interact with quickly or frequently. We have also continued to make advances in compression.

Tiered storage uses a sliding time window to move older data from fast but expensive “hot” storage mediums (like RAM, or solid state drives) to slower-to-access but lower cost “cold” storage mediums such as S3-style buckets. This allows us to grant rapid access to recent measurements, while still permitting users to access the full historical record as needed. This capability could potentially adjust what data is kept in hot vs cold storage dynamically based on user preferences, or machine learning based event triggers.

03/31/2021 - 80%
One of our big achievements (as a company) this performance period was to deploy tiered storage. Though we have not yet deployed the feature on NI4AI, we have begun rolling it out to customers. This feature will significantly reduce data storage costs across all instances of the platform.

An important question that needs to be answered before deploying tiered storage is: which data to place in “hot” vs. “cold” storage. Hot storage costs more per GB, but exceeding a certain number of queries on data in cold storage may lead to additional costs. We have developed a tool for monitoring queries to determine which data users request and how often. We will investigate these usage patterns to inform which data are stored in which tier. As new data are added, usage patterns will need to be monitored to optimize which data are stored where.

06/31/2021 - 100%
See milestone report.

M1.10 - Platform optimization documented
Due Q8, Year 2
Evaluation and complete documentation of platform performance as a function of cost using different tiers of cloud storage and compute optimizations. This evaluation will include the use of the benchmark suite developed for M1.1. Submission of documentation to ARPA-E for PD approval. If cost effective, the platform will be upgraded with the enhancement(s).

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 30%
See task descriptions above.

03/31/2021 - 65%
See task descriptions above.

06/30/2021 - 100%
See milestone report.

T1.11 - Integration of leading machine learning and deep learning libraries
Due Q8, Year 2
(Post-launch) The team will survey both the industry and the current platform users to understand which libraries require integration into the platform to speed up use case development. Currently, Apache Spark and TensorFlow are the two leading candidates.

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 20%
See task descriptions below.

03/31/2021 - 45%
See task descriptions below. Note that Task 1.11.4 poses a substantially larger work effort than tasks 1.11.1-3.

06/30/2021 - 50%
See task descriptions below.

T1.11.1 - Community survey of ML and DL needs
Due Q8, Year 2
The project team will conduct a survey of members of the utility industry that are using or considering the use of ML and AI to understand which libraries are most used.

12/31/2019 - Not Applicable
03/31/2020 - 20%
We have had many informal discussions with individuals in the utility sector concerning which analytics libraries are of interest. To date, we have learned that there is not considerable expertise among the power engineering community and, thus, they tend to follow larger trends in the ML/AI community. In other words, it would seem that they are looking more for guidance and recommendations than anything else.

06/30/2020 - 20%
09/30/2020 - 30%
Given the need for guidance in this area, we are exploring the possibility of using the platform as a mechanism for driving change in the AI/ML capabilities in the industry. We recently completed a series of

AI/ML workshops targeted at industry practitioners, and are engaging with those who participated to better understand (and potentially influence) the market for new AI/ML libraries.

12/31/2020 - 40%
As discussed in previous quarterly reports, we have found limited pull from the industry to integrate AI/ML libraries. Library selection has largely been informed by internal expertise of PingThings data scientists who have prior experience using AI/ML libraries in their own workflows. In the electric power sector, low job turnover and less focus on rigorous technical interviews means that practitioners tend to be less familiar with new and state-of-the-art data science tools. We are working to develop training materials showing how to integrate these tools with platform workflows, and to connect users with external resources to learn more about the capabilities of these various tools.

03/31/2021 - 80%
Work on the DOE FOA 1861 project provided valuable in-house expertise using the platform for AI/ML workflows. Monitoring progress of other teams working on the same project has also allowed us to understand both user needs and competitor offerings. Participating teams include prominent researchers (e.g., University of CA Riverside) and direct competitors (e.g., GE and Siemens). A panel session put on as part of the Spring NASPI WG meeting featured discussion from about half of the project participants about work performed under the project, including lessons learned that will ultimately inform platform improvements.

06/30/2021 - 100%
We have continued to talk with customers and potential customers to better understand their needs, and evaluate which AI/ML libraries and frameworks are of most interest. In terms of algorithmic needs, utilities are most strongly interested in event detection methodologies. In terms of software frameworks, we continue to find Ray, Spark, and TensorFlow at the top of the list but must also include Amazon Web Service's Elastic Map Reduce (EMR). We have consistently found that the most AI/ML workflows in the industry are founded upon Python-based libraries, first and foremost scikitlearn.

T1.11.2 - Survey of best of breed and current state of the art in ML and DL
Due Q8, Year 2
The project team will perform a literature and online review and analysis of existing state of the art ML and DL techniques and libraries that have applicability to the national infrastructure project.

12/31/2019 - Not Applicable
03/31/2020 - 50%

The team has kept a very close watch on the broader trends and dynamics around the development, refinement, and success of various machine learning and deep learning libraries both broadly and in the power engineering community. Note that we are restricting this search to only include open source libraries given the need for adoption and integration into our platform. We have seen that libraries cluster around different categories of functionality or capability including:

1. General large scale analytics and machine learning
2. Deep learning (both training and inference)
3. Distributed computing and application development

Category 1 - Apache Spark appears to be the clear winner given the incredible commercial success that Databricks, the company spun out of UC Berkeley's RISELab has had to date including a $400 million raise at a $6.2 billion valuation last October. Please note that Apache Spark was briefly considered for the platform's time series-specific processing needs but was not selected due to the slow performance.

Category 2 - The consistently top 3 libraries for DL are:

1. TensorFlow (Google)
2. Pytorch (Facebook)
3. Keras (research)

Google's TensorFlow appears to have maintained its lead in the marketplace and in community mindshare and is now one possible backend for Keras. There are numerous meta articles available online that directly address and answer this question by assembling online information about library preferences. One such article is here:
https://www.kdnuggets.com/2019/05/which-deep-learning-framework-growing-fastest.html

Category 3 - While we may be biased a bit due to the UC Berkeley RISELab connection, we have been watching Ray, which is now Ray, rapidly grow, receive significant funding (https://techcrunch.com/2019/12/17/anyscale-ray-project-distributed-computing-a16z/), and accelerate its penetration of the general marketplace. Ray is "a fast and simple framework for building and running distributed applications. Ray is packaged with RLlib, a scalable reinforcement learning library, and Tune, a scalable hyperparameter tuning library."

We will continue to track the market, we have selected the following libraries for integration:

1. Ray
2. Google TensorFlow
3. Apache Spark

06/30/2020 - 50%
09/30/2020 - 50%
12/31/2020 - 60%
The project is looking to leverage code developed by PingThings data scientists using Ray to parallelize workflows. The challenge here is that integrations between the two tools are not necessarily generalizable across different types of analysis tasks. This poses a learning barrier as users look towards adapting the code to new analytical user cases. As we expose AI/ML tools to users, we will need to direct them to external tutorials where they can learn about the inner workings of these tools.

03/31/2021 - 80%
Our understanding of community needs has not changed significantly since the last performance period. Rather, information shared by other FOA 1861 participants about how they plan to support more advanced AI/ML workflows has validated our beliefs about which data science libraries to support.

We are continuing to monitor data science libraries and will explore suitability for the platform as new technologies are released.

06/30/2021 - 100%
Given our participation in DoE FOA 1861 - Big Data Analysis of Synchrophasor Big Data - we have been able to see what the other seven teams have used in terms of analytical approaches and software frameworks over the last 18 months. This has reaffirmed our original assessment of the range of AI/ML techniques of interest in utility-focused organizations and the associated requirements.

Beyond Ray, scikitlearn (and other Python ML libraries), Google's TensorFlow, and Spark, Amazon's Elastic Map Reduce is not a software framework but a cloud-based big-data as a service that could be a compelling tool to support with more direct or native integration into the PingThings Platform.

T1.11.3 - Library selection
Due Q8, Year 2
Machine learning and deep learning libraries will be selected for integration into the national infrastructure based on a number of criteria including but not limited to: (1) requests from the platform community, (2) technical suitability for integration into the platform, and (3) overall success in the marketplace (beyond the utility industry).

12/31/2019 - Not Applicable
03/31/2020 - 50%
Please see Task above.

06/30/2020 - 50%
09/30/2020 - 50%
12/31/2020 - 60%
See task 1.11.2 above.

03/31/2021 - 80%
See task 1.11.2 above.

06/30/2021 - 100%
After careful consideration including customer and potential customer surveys and significant market research, we have ranked the ML/AI libraries in the following order:

(1) Ray is a high-performance distributed execution framework targeted at large-scale machine learning and reinforcement learning applications. It achieves scalability and fault tolerance by

abstracting the control state of the system in a global control store and keeping all other components stateless. It uses a shared-memory distributed object store to efficiently handle large data through shared memory, and it uses a bottom-up hierarchical scheduling architecture to achieve low-latency and high-throughput scheduling. It uses a lightweight API based on dynamic task graphs and actors to express a wide range of applications in a flexible manner. We put Ray on the top of the list because of its flexibility, broad capabilities, and reasonable learning curve.

(2) SciKitLearn (and other Python-based ML Libraries) - Python is the language for data science and SciKitLearn is quite possibly the most widely used and comprehensive ML library on the planet.
(3) Google TensorFlow remains the choice for deep learning
(4) Apache Spark

Our stated project milestone was to integrate two of the above libraries, making them available via the platform.

T1.11.4 - Library integration into platform
Due Q8, Year 2
The selected libraries will be integrated into the national infrastructure.

12/31/2019 - Not Applicable
03/31/2020 - 25%
Software engineering work has begun on the integration of the libraries described above.

06/30/2020 - 25%
09/30/2020 - 25%
12/31/2020 - 30%
We are beginning to develop code integrating AI/ML libraries into the platform. Further work is needed to determine how to expose these features to users in a manner that ensures proper and efficient use. Inefficient or uninformed use could lead to unnecessarily high computing costs. To mitigate this risk, we plan to restrict the resources we provide to users for free (e.g., API queries and computing resources).

03/31/2021 - 50%
We have documented a solution for using Ray to parallelize calculations here. The integration relates to defining functions (in Python) that query the database and perform calculations, and then passing these functions to Ray. Ray then stores these functions in a remote database where they can be retrieved by remote workers which then perform the specified queries/calculations in parallel.

This implementation needs to be revisited for three reasons:

1. Recent versions of Ray (i.e., 1.0+) do not support serializers.
2. Running queries remotely means passing user credentials such as an API key to a remote server. For user groups where these remote resources are shared, this may expose credentials to other users.

We are currently working to define the scope of changes to the btrdb python library that will be necessary to support Ray integration.

06/30/2021 - 90%
We have nearly completed integration of Ray into the platform and will document the final implementation in our upcoming milestone report.

Efforts this quarter have focused on streamlining big data workflows by improving performance (via memory allocation and parallelization) of the Python bindings. The Python bindings are absolutely critical integration enablers for both Ray and for sklearn. These improvements make it possible for users to perform faster queries on longer time histories across a larger number of sensors. They also streamline the process for transforming these data into arrays which can then more easily be passed to numpy to perform matrix operations, or to sklearn, statsmodels, tslearn, patsy, and other libraries for building AI/ML pipelines.

A key component of platform integration has been to create training materials for new and existing users to familiarize themselves with AI/ML tools for improving their workflows. We are continuing to develop training materials and provide robust support for users wishing to further develop their skills.

M1.11 - ML/AI Open Source Library Integration
Due Q8, Year 2
Two selected machine learning and/or deep learning libraries have been implemented into the platform and are available for immediate use.

12/31/2019 - Not Applicable
03/31/2020 - 20%
06/30/2020 - 20%
09/30/2020 - 20%
12/31/2020 - 30%
See task updates above.

03/31/2021 - 45%
See task updates above.

06/30/2021 - 95%
See task updates above.

T1.12 - Additional platform enhancements requirements gathering
Due Q12, Year 3
(Post-launch) The team acknowledges that it cannot predict additional requirements that will emerge after the launch of the platform. Thus, we will continuously collect feedback from the community through organic means

and directly solicit feedback periodically. A list of potential enhancements will be maintained. This will continue through the life of the project.

12/31/2019 - Not Applicable
03/31/2020 - 20%
We have been soliciting comments and feedback about our platform from our commercial customers, potential commercial customers, university partners, and potential university partners. These discussion include:

- Programming Languages and Libraries
    - Which programming languages do you use?
    - Which ML and AI libraries do you use or are most interested in?
- General Analytic Concerns
    - What are your largest concerns when working with large data sets?
    - What are capabilities that you wish you had when working with large amounts of sensor data?
    - What associated metadata is required for sensor analytics?
    - Is grid topology required for sensor data analytics?
    - Do you have any specific problems that you believe can be solved by grid sensor data?
- Sensor Fleet Management
    - How do you manage your existing fleet of sensors?
    - What capabilities do you wish you had to manage your sensor fleet?
    - What would that interface look like?
- Overall Platform Concerns
    - How important is cost?
    - How much data or what length of time of data do you currently retain?
    - How important is performance?
    - What are your existing platforms?
    - What are your existing data processes?

06/30/2020 - 20%
09/30/2020 - 20%
12/31/2020 - 50%
Since deploying the platform, we have been able to collect more detailed information about user needs and enhancements. We have had several interested users reach out to request a demo of the platform, including users at Schneider Electric and ABB. These discussions have informed how we understand the needs and priorities of our user base.

These interactions underscored the need to develop robust materials which new users can use to familiarize themselves with the platform and workflows (see Task 3.5). We have also identified two enhancements that will make the platform easier and more intuitive for new users to navigate. The first provides a portal showcasing various features of the platform and project (see Task 3.. The second enhancement integrates jupyterhub into the online web interface, eliminating the need for users to spin

up jupyter servers on their local machines. Though most python users are familiar with jupyter notebooks, these workflows are new to users who are transitioning from Matlab. While there is a general willingness to make this transition, learning this new workflow may be a frustration. Jupyterhub integration will ease the learning curve for Matlab users.

Jupyter hub instances will, however, need to be limited in size and functionality to prevent abuse. This will involve customizing our jupyterhub integration -- e.g., by restricting internet access.

03/31/2021 - 60%
See task/milestone updates below.

06/30/2021 - 80%
This performance period we launched several pilots under Task 2.3.3 which have provided new insights into platform requirements to enable a wider array of applications. Details about these pilots are included in our milestone report for Milestone 2.3.2. These pilots included evaluations of the platform for enabling real-time control applications, and exploration of insights gained from an analytical use case initially developed by Dominion. This underscored an opportunity for exposing standard techniques for signal processing, anomaly detection, and change point detection to help improve and scale customer workflows across a wide range of possible analytical use cases.

Gathering information from pilot participants about the criteria they are using to evaluate the platform has provided new insights into possible platform enhancements that potential customers would wish to see developed.

1. Improve performance for multi-stream queries (e.g., time-alignment and parallelization)
2. Enable re-streaming of data to simulation testbeds to enable low-latency simulation
3. Provide analytical toolkits (e.g., python notebooks or other visualizations) to detect change points, anomalies, or incongruencies between data streams. A survey of analytical methods underlying use cases of primary importance to the industry suggest that these three general purpose time series analysis capabilities could be foundational across a wide range of use cases. Efficient event detection is also critical to deploying these analytics at scale.
4. Issue real-time alerts based on user-defined threshold violations
5. Monitor API usage and employ rate-limiting to flag inappropriate or excessive use (note: this enhancement was identified internally as a necessary prerequisite for exposing Jupyterhub broadly to NI4AI users. It is as yet unclear whether this capability will also benefit PingThings customers)
6. Improve real-time visualization capabilities by more frequently updating data in the plotter
7. Provide access to distillate streams, jupyter notebooks, or helper functions to perform common calculations/transformations -- e.g., to calculate phase angle differences, power spectral density, phase imbalance, and network impedance.
8. Improve metadata and network model data stores to draw more explicit linkages between data streams (based on measurement quantity, network connectivity, siting, etc.) to improve capacity for performing multi-stream calculations at scale.

9. One pilot participant wished to benchmark the performance / speed of the platform for ingesting a large historical data set. To support that demonstration, we developed a new data ingestor which increased ingestion speed by 10x. This capability will help to eliminate bottlenecks in ingesting very large data sets, allowing us to move more quickly on future data ingestions.
10. Our C37 TCP data ingestor is among the most heavily stress-tested in operation, followed by GEP. As we have begun to deploy our C37 UDP ingestor more broadly, we have encountered new issues not previously known to exist. Though UDP is known to be less reliable for streaming data transmissions than TCP, results to date indicate that the ingstor is less robust to multi-sensor fleets expected. There is a need to further develop our C37-UDP ingestor to maximize reliability of data streaming.

M1.12 - Additional platform enhancement gathering complete.
Due Q10, Year 3
Completion of survey of current users for desired platform enhancements complete with findings documented, submitted to ARPA-E for PD approval, and publicly available.

12/31/2019 - Not Applicable
03/31/2020 - 20%
See above Task 1.12

06/30/2020 - 20%
09/30/2020 - 20%
12/31/2020 - 50%
We have identified two enhancements, and are continuing to refine the scope of work involved in deploying these features.

03/31/2021 - 60%
Additional enhancements we wish to explore include application development, and platform enhancements what will streamline user workflows for developing applications of their own. We will continue to explore the opportunities and requirements as we begin piloting the platform with new organizations.

06/30/2021 - 80%
See task update above.

T1.13 - Implementation of required platform enhancements
Due Q12, Year 3
(Post-launch) Enhancements from 1.1.11 above will be implemented as time and resources allow.

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not applicable

09/30/2020 - Not applicable
12/31/2020 - 10%
See milestone updates below.

03/31/2021 - 20%
See milestone updates below.

06/30/2021 - 30%
See milestone update below.

M1.13.1 - Additional platform enhancement 1 completed
Due Q10, Year 3
Development, testing, and launch of additional platform enhancement 1 given requirements collected in T1.12. Demonstration of additional enhancements to ARPA-E.

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 30%
The first enhancement will expose a jupyterhub integration making it more intuitive for users unfamiliar with python workflows to more easily gain API access. Work to date has focused on defining needs and requirements to prevent mis-use of jupyterhub instances, and to avoid incurring undue computing costs once this feature is exposed to users. For example, we plan to limit the size of computing clusters that are accessible to users for free. As the project evolves, we may also need to explore the possibility of an approval process, which would allow us to verify the identity of users before granting them jupyterhub access.

03/31/2021 - 40%
We have deployed jupyterhub on the NI4AI server to allow us to do demos that showcase a fully featured version of the platform. We have not yet exposed this offering to users, as additional work would be needed to prevent users from (potentially inadvertently) imposing an excessive burden on project funds.

During our next performance review, we wish to lay out a plan for “re-architecting” the project in a way that would eliminate much of the work that has stalled progress on this task to date. Specifically, we wish to explore the possibility of deploying instances of the platform with project partners (e.g., ABB, or a utility) to help us solicit confidential data that would be valuable for advancing use case development.

Shifting the focus of the project to deploy the platform to a more controlled user group would significantly change our work plan for exposing jupyterhub allocations to users. We have paused work on this task while the goals and outcomes of the project are revisited.

06/30/2021 - 50%
In previous performance periods, we had scoped this task to include development of API tracking tools that would enable us to provide EC2 instances to users via jupyterhub, without risking undue expenses due to potential misuse.

Requirements gathering has shown us that this capability, though necessary if we are to provide open access to jupyterhub instances for NI4AI users, is not a capability that will necessarily benefit our customers. We are continuing to make jupyterhub instances available to a select number of trusted users engaged in platform pilots and their collaborators. Due to our focus this quarter on working more closely with these users (rather than on widespread user recruitment), we have determined that API rate limiting is of lesser importance than other platform enhancements which could meet both project and customer needs. In the upcoming months, we plan to prioritize the development of these features rather than further developing API capabilities to enable more widespread jupyterhub access.

Requirements identified to date are detailed above.

M1.13.2 - Additional platform enhancement 2 completed
Due Q11, Year 3
Development, testing, and launch of additional platform enhancement 2 given requirements collected in T1.12. Demonstration of additional enhancements to ARPA-E.

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 5%
We are still working to define the scope of this enhancement. Options under consideration include migrating to Apache Arrow 3.0, or continuing to develop and improve the python language bindings.

03/31/2021 - 10%
We are holding off on specifying additional project enhancements as we look towards defining what the aims and scope of the project will be moving forward. This milestone could potentially be focused on developing industry specific applications to run on the platform. Acquiring the data to support robust application development would, however, require a shift in focus from generating open access data to hosting confidential data provided by project partners. This task could include further developing use cases prototyped in-house at Dominion to demonstrate their value to other organizations. It could also focus on developing altogether new use cases to increase our competitive offerings for a new and more diverse customer base.

Options include:
- Transmission (BPA, ATC, ComEd, or others)
- Distribution (ComEd, SDG&E)
- Solar (Origis)

- Storage (Stem)
- Equipment manufacturers (ABB)
- Sensor hardware providers (SEL, Itron)

Here, we list potential partnerships we could explore from among the organizations that have previously expressed interest in the platform or project.

06/30/2021 - 20%

We are continuing to engage with key stakeholders engaged in platform pilots to determine what feature enhancements could give us the greatest advantage in driving sales. A list of possible enhancements is included under Task 1.12. Early prototypes of some of these features have been developed, including event detection, real-time alerting, real-time visualization, and others. We are working to evaluate the scope and requirements for translating these prototypes into user-friendly tools that will bring new customers to the platform.

M1.13.3 - Additional platform enhancement 3 completed

Due Q12, Year 3

Development, testing, and launch of additional platform enhancement 3 given requirements collected in T1.12. Demonstration of additional enhancements to ARPA-E.

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - Not applicable

We are still defining the scope of this enhancement.

03/31/2021 - 10%

We are still working to scope out this enhancement. One option we are considering would be to develop a second application to run on the platform; this could target a different user base, for example distribution or the solar industry. Another option would be to provide greater support for users to build applications of their own.

06/30/2021 - 20%

As discussed under M1.13.2, we are exploring several of the requirements listed under M1.12 before further defining the scope of the current milestone.

Task 2

Note that all text in *blue Italics* is from the last full SOPO Attachment that was exchanged between PingThings and ARPA-E. Also, the tasks appear ahead of the Milestones for which they lead into.

Task 2 - Hyperscale grid sensor data collection

Deliverables include data set generation which covers three types of time series data: (1) in silico - data generated from simulation (also simulated or virtual data); (2) in vitro - data generated by a sensor within a lab or other controlled environment; and (3) in vivo - data generated by a sensor deployed to the grid. Deliverables also include a data escrow service which covers (1) common legal documents (non-disclosure agreement, data sharing agreement); (2) platform augmentation (encryption, obfuscation, user interface); and (3) validation and testing reports.

T2.1 - Previously generated sensor data set collection

Due Q10, Year 3

Previously generated sensor data sets will be identified, collected where possible, and made available through the platform.

12/31/2019 - 5%

T2.1 has been slightly altered as part of Milestone 0.1.

03/31/2020 - 15%

We have made progress identifying and collecting historical sensor data sets detailed below.

06/30/2020 - 25%

We have made progress identifying and collecting historical sensor data sets detailed below.

09/30/2020 - 40%

We have made progress identifying and collecting historical sensor data sets detailed below.

12/31/2020 - 50%

We have ingested several previously generated data sets, and are continuously identifying new data sets. See milestone report for M2.1.2 for more details about the data that are available.

03/31/2021 - 60%

We are continuing to ingest previously generated data sets as they become available to us. See task updates below.

06/30/2021 - 70%

We are continuing to ingest previously generated data sets as they become available to us. See task updates below. We are ingesting previously generated data for pilot projects launched under task 2.3.

T2.1.1 - Identify previously generated sensor data sets

Due Q10, Year 3

The team will utilize literature and industry and academic relationships to identify previously generated sensor data sets and determine their suitability for the project and varying levels of public release. The purpose of this search is to not only identify the actual sources of data but to engage the community members that simulated and/or collected the data and might be interested in analyzing it in the platform. The team will generate publicly available content, most likely via the project blog, based on search results.

12/31/2019 - 30% - In Progress

Initial research to identify potential simulation efforts has begun. Project announcements have resulted in researchers mentioning previously collected sensor data sets that they either know about or they themselves collected. Thus, we have identified several potential previously generated data sets for the project including those below:

- RPURef – 6uPMUs – 1.5 years of data
- MITRE – 90 TB of continuous POW collected over a 4 year period by approximately half a dozen sensors
- EPRI Power Quality

03/31/2020 - 60%

Liaisoned with EPRI regarding the PQ Dashboard/NI4AI. – Fostered discussion between PingThings and EPRI regarding incorporating features from the EPRI PQ Dashboard into the NI4AI platform. Conversely, discussed using the NI4AI as a data source for the EPRI PQ Dashboard.

Liaisoned with Whisker Labs regarding Ting/NI4AI. – Fostered discussion between PingThings and Whisker Labs regarding incorporating data from their TING sensor into the NI4AI platform.

Discussed the possibility of hosting previously generated uPMU data with Powerside, formerly PSL.

06/30/2020 - 80%

We have been broadly assessing public datasets and prior efforts to collect and share data. We conducted over two dozen interviews to better understand what data are shared, under what terms, and with whom. We are using the insights gained from these interviews to both explore potential partnerships, and to inform broader outreach efforts.

Key partnerships include:

- Electrical utilities (IOUs, transmission operators, munis/co-ops, CCAs)
- Regulatory agencies that govern data sharing
- Researchers working with or collecting sensor data
- Funding agencies that oversee sensor research programs

We have learned that it is not uncommon for utilities to enter into data-sharing relationships. AMI data are more readily shared than other data streams, though these data also come with restrictions due to customer privacy. SCADA data are protected by NERC's CIP standards and are not readily shared except with vendors. Though it is not uncommon to share PMU data with trusted research partners, there is a general lack of understanding as to whether widespread PMU data sharing poses a security risk. A number of interviewees mentioned DOE's Big Data initiative as an opportunity to begin answering these questions. The number of utilities that have a track record for entering into data-sharing partnerships is relatively small. These efforts are usually spearheaded by individuals within the utility who are willing to push through legal barriers to establishing data-sharing relationships.

Terms and conditions under which utilities are willing to share information vary from one utility to another. Terms under which people have expressed greater willingness to share information include:

- Anonymously, with network data and sensor location withheld
- Historical data (i.e., from last year)
- Event snapshots, rather than continuous measurement data
- Dedicated sensors deployed for purposes of data sharing (i.e., NI4AI sensors)

In addition to operational grid data, we are exploring datasets collected primarily for research purposes. Several examples are listed below. We have contacted dataset owners to get a sense for how actively each dataset is being used, and what educational/research value they could have.

- Fast solar inverter measurements
  https://ieee-dataport.org/open-access/one-year-submillisecond-fast-solar-database
- Fault detection in underground lines
  https://www.kaggle.com/c/vsb-power-line-fault-detection/data
- Voltage sags https://ieee-dataport.org/documents/real-life-power-quality-sags
- Non intrusive load monitoring (e.g., http://portoalegre.andrew.cmu.edu:88/BLUED/, http://redd.csail.mit.edu/)

09/30/2020 - 90%

We have continued discussions with large dataset owners including utilities, EPRI, and FNET. PingThings has discussed the possibility of hosting FNET data in PredictiveGrid. It is highly unlikely that the data will be made available through NI4AI given the current data access model employed by the project responsible for FNET. We have also been in discussion with the program manager for DOE FOA 1861 and with utilities who contributed data around the possibility to make this data available to a broader audience.

Despite the historic lack of data sharing in the industry, discussions with utilities have been met with enthusiasm, particularly among specific "champions". We have spoken with champions at three large utilities, who have agreed to share either real-world or experimental data. Despite the enthusiasm, concerns around CIP restrictions, customer privacy, and the risk of liabilities create general uncertainty about what data can safely be shared and with whom.

This uncertainty poses a risk in seeing data sharing agreements through, but may also create an opportunity to become a voice of authority and thought leadership about data sharing. We are beginning to engage with NERC and with JSIS to learn more about current data sharing practices. We are also hosting a panel at the Fall NASPI working group meeting Nov 3 on the theme of “data sharing and collaboration”. The panel will bring in several perspectives that will motivate the need for open access data.

In parallel with targeted outreach to potential data providers, we are doing broader outreach to industry via Energy Central and other news outlets to spread thought leadership about external data sharing (motivating NI4AI) and about ease of access for internal use (motivating PredictiveGrid).

An article published in Energy Central had received over 550 views as of 10/30/2020 and was listed on 10/30/2020 as a featured article on digitalization. We anticipate that more persistent engagement with the site will lead to additional visibility. The site will be interviewing Theo Laughner (a project industry advisor) about the project in November. For industry news outlets, we are framing the project and the platform as a mechanism for early-career practitioners to differentiate themselves and get ahead of the industry with regards to “digitalization” and data usage.

Last quarter we explored the possibility of diversifying the data we host -- for example to include load monitoring, weather, and other data. We have not excluded this possibility, but have not chosen to pursue them this quarter for the following reasons:

- Engaging with the non-intrusive load monitoring community told us that research interest in the space is diminishing.
- Though there is a large audience for weather data, there are also already existing platforms that are custom-built for weather data. Given the substantial amount of work involved in processing raw data to provide a rich user experience, we have focused our efforts elsewhere for now.

We have also identified a repository of power quality event data (http://map.pqube.com/).

12/31/2020 - 100%

We’ve developed three new data ingestors this quarter. In anticipation of a presentation we gave to the IEEE Power Quality Data Analytics working group, ingested data sets they had made available via google drive. During the talk, we gave a demo of the platform and made them aware that we had ingested their data.

We will continue to develop new ingestors as new data sets come to our attention.

03/31/2021 - 100%

During a previous reporting period, we identified a utility partner interested in using the project as a proof of concept to demonstrate the capabilities of the platform. This performance period they shared a longer time history of data, including 7-days of continuous monitoring data for four generating stations. Dates were falsified and the data scrubbed of any information that could be used to determine the location of the plants or the identity of the utility.

The data provider was interested in exploring a use case that Dominion had developed, but did not have the staffing or expertise to do the analytics in-house. Meanwhile, Dominion was interested in deploying their analytics at other utilities to demonstrate business value and advance the technology readiness of the use cases they have explored.

In other cases we have found that providing open access to the data is a barrier to getting data contributions. During our next quarterly meeting, we wish to discuss the possibility of deploying several instances of the platform for various utilities (and other partner organizations) as a "proof-of-concept" to host confidential data for them, while also allowing us to use their data and expertise for application development. This change in focus would involve ingesting confidential data, and brokering data sharing agreements with partners.

06/30/2021 - 100%

This performance period we acquired data from a CPOW sensor which recorded 24-hours of data during the Bonneville Power Association's annual Chief Joseph brake resistor test, which has been known to cause wide-area oscillations throughout WECC. Data was provided by a DOE project called GridSweep which is developing instrumentation to monitor high-frequency dynamics related to grid disturbances.

T2.1.2 - Develop necessary data ingestors

Due Q10, Year 3

As previously generated sensor data sets are identified, the team will have clarity to determine which types of data ingestors must be created. This is established based on the file-formats of simulator output and file-formats of existing time series. The team may deem it necessary to build a single ingestor (perhaps CSV) and then translate all of the data sources into CSV before ingest. This decision will be based on factors such as the volume or significance of the format the data currently exists, whether the ingestor may have value to the community in the future, or the cost-benefit analysis of developing a custom ingestor for a particular data type relative to a standard like CSV.

09/30/2019 - Not Applicable
12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - 50%

We have developed data ingestors for the following protocols:

- CSV with dynamic schema to cope with differences in how sets of data are mapped on to CSV
- Parquet
- Powerside uPMU binary file

We have found that many datasets have non-standard representations, so have addressed this by creating a generic framework for inserting large amounts of data into the database and dealing with the heterogeneity in encoding formats on an as-needed basis. By leveraging this framework, we can add support for a previously unseen format with as little as 10 lines of code.

Datasets ingested:

- Sunshine (6 uPMUs)
- Patient EKG data
- USGS seismic data

09/30/2020 - 80%

We developed ingestors for two datasets acquired this quarter, bringing the platform up to a total of six previously generated datasets. New datasets were provided by a transmission utility and by the Texas Synchrophasor Network.

The “events/” data was shared by an industry practitioner to demonstrate the capabilities of the platform internally to build a case for licensing the platform commercially. This conversation is ongoing.

The second dataset was shared to help us better understand the geographic scope (and accuracy) of data from a frequency disturbance that one of our own sensors recorded.

Collection	Sensors	Duration	Ingest time	Description
events/	23 PMUs	3x 1-min event snapshots		This dataset was shared by a transmission co-operative. Data are scrubbed of sensitive information by withholding the location of the sensors, masking time stamps, and removing current measurements and line flows from the raw data provided.
brownout/	3 PMUs	1 hour		This dataset came from the Texas Synchrophasor Network’s data archives. We witnessed a frequency deviation in one of the ni4ai sensors, and reached out to verify whether their sensors had also witnessed the event. The dataset forms the foundation of the following blog post (link).

We have submitted a milestone report tracking the ingestors developed to date including these and three others developed in previous performance periods. We will continue to develop additional ingestors on a case-by-case basis as we acquire new datasets under Task 2.1.1.

12/31/2020 - 90%

We have developed ingestors for several data sets currently live in the platform (see milestone report for M2.1.2). We are in the process of developing a new ingestor for a time synchronized point on wave data from a MITRE project called “Sentinel” which collected wide area monitoring data, primarily on the Eastern Interconnect, between 2012 and 2016.

03/31/2021 - 100%
This performance period we developed ingestors for two new datasets:

- "monitoring" data includes one week of anonymized data including voltage, current, and power output from four generators.
- "texas" includes voltage and frequency measurements for the month of February from 10 TING sensors located in Texas. We anticipate that data from a larger network of sensors will also be provided, though exporting the data poses a significant lift for the data providers. These measurements are not time synchronized.

While we are continuing to search for new data to be made available, and are working to streamline this process with a Python library for specifying data structures and ingesting data from CSV files or S3.

We are also vetting a new ingest process whereby data providers are given write access to a dedicated collection in the database where they can ingest and QAQC the data themselves. This will be more efficient, as the providers themselves are already familiar with the data formats. It will also eliminate bottlenecks related to transferring the data via email or FTP. This will allow for more streamlined deployment of the platform as a proof-of-concept with industry partners (as discussed under Task 2.1.1).

This python ingestor is available via the following github repository.
https://github.com/PingThingsIO/pgimport

M2.1.1 - Asynchronous data ingestors developed for selected previously generated sensor data sets.
Due Q4, Year 1
The bulk, asynchronous data ingestors have been developed, tested, and deployed on the platform. Test bulk data ingest has been benchmarked and documented. Documentation submitted to ARPA-E for PD approval.

09/30/2019 - Not Applicable
12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - 50%
Work continues to develop the ingestors necessary to import previously generated data sets.

09/30/2020 - 100%
We have submitted a milestone report tracking the ingestors developed to date (see Task 2.1.1 above), and will continue to develop additional ingestors on a case-by-case basis as new datasets are acquired.

03/31/2021 - 100%
This quarter we developed a Python library to streamline the process of ingesting data from CSV. For very large datasets, we would still recommend using the Go ingestor to improve computational speed and efficiency.

We have also developed ingestors for two new datasets (`texas` and `monitoring`) and are working to test the Python ingestor. We are now ingesting a library of real and synthetic wide-area oscillations. Some of these were released to support a data competition being run by PNNL.

06/30/2021 - 100%
See task updates above.

T2.1.3 - Ingest previously generated sensor data sets
Due Q5, Year 2
For each of the identified data sets, the team will use the previously created data ingestors to ingest the data into the platform. The team will document ingestion performance.

09/30/2019 - Not Applicable
12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - 25%
We have ingested some previously generated data including the sunshine data set of uPMU data.

09/30/2020 - 100%
To date we have ingested data for five previously generated datasets as documented in the milestone report for M2.1.1. In response to a recommendation issued in our quarterly PD memo, we have published a blog post summarizing each of the datasets hosted to make it easier for users to find the collections and the data they are looking for.

Although we have completed this particular task, we remain open to ingesting additional previously generated data sets.

03/31/2021 - 100%
See M2.1.1 above.

06/30/2021 - 100%
All data sets identified under Task 2.1.1 and 2.1.2 have been ingested into the platform.

M2.1.2 - Previously generated sensor data set ingested and available
Due Q5, Year 2
The first previously generated sensor data set is ingested into the platform and made broadly available to the community.

09/30/2019 - Not Applicable
12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - 25%

As described in task 2.1.2, we have ingesters for CSV and Parquet files, and a framework for adding new formats rapidly**.** There are currently three datasets available in the platform.

09/30/2020 - 100%
We have ingested six previously generated datasets, as described in Tasks 2.1.1 and 2.1.2. These are all currently available in the platform, and are summarized on the blog. The blog post will be updated as new datasets are ingested.

We will continue to look for previously generated datasets and ingest them as appropriate.

03/31/2021 - 100%
Both the `texas` and `monitoring` data are ingested and available. We expect to receive additional data potentially from both providers.

06/20/2021 - 100%
See task updates above.

T2.1.4 - Socialize availability of previously generated sensor data sets
Due Q12, Year 3
This represents a significant project milestone that should be shared with the community. This data may also be used as the basis of a data science competition or hackathon if possible.

09/30/2019 - Not Applicable
12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - 10%
We have identified several outlets for socializing the platform and have a record of individuals who have attended NI4AI events or expressed interest in receiving updates about the project. We will send a broad announcement to these public channels and to the NI4AI community when the platform launches.

09/30/2020 - 50%
We have launched a LinkedIn page, circulated two PowerGlobe announcements about the project, and ran an email campaign which went out to over 200 individuals at 100 universities.

These announcements focused primarily on advertising our CIGRE workshop, though they included a broad description of the project as well. During the CIGRE event we listed datasets available in the platform. See screenshot from our slide deck below. Note that the EKG and Geomag data are not listed here, as the focus of the talk was on grid measurement data.

Datasets currently available to you

	Collections	Sensors	Duration	Features of interest
Real-time	ni4ai/ texas_pmus/	5	April-now	Real-time streaming data
Point on Wave	epfl/	3	6 events	Battery charging & discharging
Wide Area Events	transmission_events/	23	3 events	Switching events Oscillation
Archive	sunshine/	6	18 months	Distribution feeder with a PV array

46

12/31/2020 - 60%

We have continued to announce data sets available in the platform through talks, emails, and in conversations with stakeholders. An updated screenshot of the data sets slide is pictured below. Note that the way we categorize and describe data sets may change somewhat depending on the audience.

The Data

Type	Collections	Sites	Duration	Features
Streaming	ni4ai/ texas_pmus/	6 (so far)	April-now	Real-time Continuous
PQdata	underground/ epri/	1 300 events	24-hours 2 seconds	POW data Faults & failures
Events	events/ epfl/	20+ 3	3 events 6 events	Wide-area Battery charge
uPMU Pilot	sunshine/ golden/	6 10	18 month 3 months	Distribution grid PV array

PingThings https://blog.ni4ai.org/tags/ni4ai-data 14

03/31/2021 - 70%

We have continued to socialize the availability of new datasets via conference presentations and other outreach efforts. In the upcoming performance period we hope to issue more widespread announcements as the Sentinel data set comes online, and as additional sensors are deployed.

We have improved website materials to make it easier for users (and for the general public) to find information about datasets hosted. We had previously maintained a blog post which listed data collections along with a brief overview of the data set and links to additional documentation where relevant. We have now released a "datasets" page on the ni4ai website (https://ni4ai.org/datasets) to make this content more visible from our home page, and to include relevant use cases and code. The design and contents of the site are inspired by https://www.kaggle.com/datasets.

06/30/2021 - 80%
We have identified an opportunity to leverage data already ingested into the platform to gain fast "wins" as we conduct platform pilots under Task 2.3. Historic data allows us to showcase the capabilities of the platform for working with novel instrumentation (e.g., CPOW sensor) and with historical data archives that are much longer than it would be feasible to ingest under a short demonstration project.

T2.2 - In vitro sensor data collection
This task covers laboratory created sensor data generated during controlled testing of project sensors.

09/30/2019 - Not Applicable
12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - See below
12/31/2020 - 80%
See task updates below.

03/31/2021 - 90%
See task updates below.

06/30/2021 - 100%
See task updates below.

T2.2.1 - Select sensors for laboratory generation of in vitro time-series
Due Q2, Year 1
The team will determine which sensors can be setup in the laboratory to create in vitro streaming time-series data. Understand requirements for in vitro data generation and how this impacts which community members should be engaged for the best results.

See T2.3.1 for more details
The sensors that will compose the in vitro data generation will be those sensors that we are considering for field deployment for the in vivo data generation tasks. The project team has no desire to acquire

sensors that we only test and not deploy. Thus, this task strongly overlaps with T2.3.1 for which initial progress has been made. It might make sense to merge these tasks together.

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - 100%

T2.2.2 - Incentivize community member for in vitro data set generation

Due Q8, Year 2

The team will engage with and incentivize community members from T2.2.1 to crowdsource the generation of in vitro time series. While the team will focus on data generated from project provided sensors, we will enable community members to create any in-vitro data sets from their own sensor hardware as well.

09/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/30/2020 - 25%
We have, through our numerous presentations, reached out and discussed the possibility of securing individuals and/or organizations that would be interested in testing/evaluating sensors. No official partnerships have been created and we continue to conduct the in vitro testing ourselves.

06/30/2020 - 25%
09/30/2020 - 30%
We have connected two in vitro sensors and are in the process of onboarding an OpenPMU installed in a laboratory in Switzerland.

12/31/2020 - 100%
See milestone report for M2.2.

03/31/2021 - 100%
This performance period we have begun to explore the possibility of using the project to deploy dedicated instances of the platform where project partners can ingest and interface with data internally. Data would be shared confidentially, and (optionally) with other collaborators such as Sascha von Meier or Kevin Jones.

06/30/2021 - 100%
This performance period we have focused on launching “proof of concept” deployments of the platform to enable possible customers to explore the capabilities of the platform in supporting their own internal workflows. The incentive here is to enable possible customers to evaluate the speed / performance of the platform and explore capabilities in developing use cases of interest to them. The ability to deploy analytics developed by existing users of the platform has also proved to be an appealing mechanism for recruiting community member participation.

T2.2.3 - Develop necessary streaming ingestors for capturing laboratory data

Due Q4, Year 1

Development and testing of streaming ingestors for all sensors is more easily done first in a controlled environment. Additionally, this task is necessary to stream in vitro data from laboratory sensors into the platform.

09/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not applicable
06/30/2020 - 50%
This is in progress for the needed sensors.

09/30/2020 - 50%
We have deployed C37 and GEP ingestors, and a prototype to ingest point on wave data via IEC 61850. We have found more interest (and greater need) for real-world data than laboratory generated data, and have focused on Task 2.3 over Task 2.2. We will revisit Task 2.2 as high-value laboratory generated data are identified.

12/31/2020 - 100%
See milestone reports for M2.2 and M1.6.

03/31/2021 - 100%
We have identified a couple of needs for further developing our C37 ingestor. These include:

- *Re-naming streams.* Stream names are currently inherited from C37 data frames sent by the sensor, and default nomenclature is inconsistent (and sometimes confusing).
- *Ability to sub-select streams to be archived.* The microPMU is capable of recording on up to 20 channels. However, installing sensors in wall outlets means that only single-phase voltage phasors report meaningful measurements. The additional channels simply report noise and should not be ingested. Our ingestor will need to be adapted to store data from certain channels, while discarding others.

T2.2.4 - Ingest in vitro time series data

Due Q8, Year 2

The team will ingest in vitro time series data created by the community by establishing a streaming connection directly with the platform. The team will document the performance of the ingestors and demonstrate successful ingestion through platform tools for exploration and administration.

09/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - 20%
We are currently ingesting streaming data from two sensors which we deployed (see Milestone 1.5), and from three sensors deployed as part of other research initiatives.

12/31/2020 - 100%
See milestone update for M2.2.

03/31/2021 - 100%
We are in the process of testing new firmware for the microPMU which will allow us to establish streaming connections via UDP rather than TCP. This change allows us to establish two-way communication with the device, and eliminates the need for sensor hosts to configure port forwarding on local routers. We have demonstrated that the new firmware can support UDP connections and have successfully installed the firmware on two of the sensors deployed to date. We have developed a new C37 ingestor to accept data via UDP.

M2.2 - First in vitro data is streaming to the platform.
Due Q6, Year 2
The first in vitro data set has been streamed into the platform.

09/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - 50%
We have deployed two OpenPMU sensors which are currently streaming data to the platform.

12/31/2020 - 100%
See milestone report.

03/31/2021 - 100%
This performance period we deployed two microPMUs which are streaming data into the platform.

T2.2.5 - Socialize in vitro data acquisition
Due Q5, Year 2
09/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - 50%
We have made announcements via LinkedIn and PowerGlobe about streaming data. We also made an announcement at our Oct 21 CIGRE workshop. We will do so again at both SmartGridComm and at the NASPI Fall working group meeting.

We have been using these streaming data as a mechanism for garnering interest in the project as ours is the only source of open access real-time phasor measurement data. We have received some interest from NERC. We also expect that there could be interest among companies that sell data, such as

Genscape. We are investigating data requirements for these different audiences, and plan on using these requirements to inform our deployment strategy.

12/31/2020 - 60%
We have continued to tell the community about streaming sensor data available in the platform, including at talks and in conversations. As we continue to onboard new sensors deployed under Task 2.4, we will issue broader announcements (e.g., via PowerGlobe and other channels).

03/31/2021 - 70%
We are waiting to issue a broad announcement once we have achieved broader geographic coverage and can be sure that sensors will not be taken offline for firmware testing. We have continued to make individuals aware that the work is underway, and have had several stakeholders volunteer to host sensors in their homes.

06/30/2021 - 80%
We have continued to socialize the availability of streaming data to various audiences. The most promising partnership identified to date is NERC. Their primary interest is in our wide-area sensor deployment, as it would provide easy access to streaming data which cannot easily be obtained from existing PMU deployments -- either due to limited geographic coverage or limitations of conventional mechanisms for sharing data among transmission entities. We will continue to engage with NERC about data availability as additional sensors are deployed.

T2.3 - In vivo sensor deployment and data capture
Due Q12, Year 3
There will be three major components to this task. These include: (1) procurement of sensor hardware; (2) deployment (locations, schedules, shipping, & remote setup); and (3) maximizing operational uptime to collect data. This data will be made available via the platform.

09/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not applicable
06/30/2020 - 10%
Progress towards identifying sensors and host sites is detailed below.

09/30/2020 - 20%
Progress towards identifying sensors and host sites is detailed below.

12/31/2020 - 40%
See task updates below.

03/31/2021 - 50%
We have continued to explore hardware options (Task 2.3.1), have acquired microPMU hardware (Task 2.3.2), and continue discussions with potential host sites (Task 2.3.3).

While we have identified a couple of host sites, general feedback is that most private companies are unwilling to provide open access to their data. Open access data has been an important pillar of the project to date. Moving forward, we intend to lift this project requirement in the interest of bringing more companies to contribute data to the platform. This will allow us (and our collaborators) to capitalize on their data and/or expertise to advance use case development. It will also allow them to test the capabilities of the platform for internal use.

06/30/2021 - 60%
See task updates below.

T2.3.1 - Identification of sensor hardware
Due Q2, Year 1
The team will select sensor hardware to generate the national coverage (wide-area) data set and the boutique location (site-specific) data sets. This should happen in parallel with T2.3.2. Considerations will be made for vendor diversity, hardware diversity, sensor population size, sensor cost and availability of special pricing. The team envisions adapting this mix and population throughout the project. For example, sensors deployed in Year 1 may be different than sensors deployed in Year 2 or 3 if there is strategic benefit

12/31/2019 - 30%
We have acquired and begun testing of three different sensors, each at a very different price point that would allow varying sized sensor fleets to be deployed. These sensors are:

1. Eyedro (https://eyedro.com/home-electricity-monitors/)
2. uPMU (https://powerside.com/products/micropmu)
3. OpenPMU - a sensor developed at university that offers potentially very good capabilities at a much lower price point than the uPMU

The search for additional sensor vendors, both commercial and academic, is still ongoing and is partially coupled to our project announcements at various conferences and workshops.

03/31/2020 - 50%
We have continued discussions in depth with both Powerside and the small team behind the OpenPMU. With the OpenPMU, we have aggressively pursued moving this instrument from one suitable for a research lab to one that can be deployed for this project. This effort has been invested because the OpenPMU is still significantly cheaper than the Powerside uPMU. However, significant obstacles remain due to UL certification requirements and other potential legal obstacles (all of which we are now learning about).

We have also identified additional sensors including:

- https://gomeshnet.com/
- Synaptec - a fibre-based time synchronized PMU and Point on Wave measurement system

06/30/2020 - 100%
Given obstacles with UL certification and a pause in certifying new sensors due to the pandemic, we have decided to move forward with the uPMU as opposed to other grid sensors. New partnerships may lead us to revisit new sensor technologies as the project evolves or as new sensor hardwares come to our attention.

09/30/2020 - 100%
We are in the process of re-budgeting the project to reflect our partnership with Powerside. We are also closely following research on point on wave data. We suspect that disseminating point on wave data will create a broader audience for the data, and are looking to acquire previously generated point on wave datasets.

12/31/2020 - 100%
As we continue to discuss the project with different stakeholder groups, we are continuing to learn about new hardware providers that are entering the market. We are continuing to explore new opportunities to form strategic partnerships with hardware vendors.

03/31/2021 - 100%
We have continued to explore hardware options. While microPMU is an excellent choice for both distribution and transmission grid applications, it is over-built for installation in wall outlets, as we currently stream on only two of twenty channels. We have continued to explore other hardware options, and connected with two sensor providers: Whisker Labs and Iotawatt.

Whisker Labs' TING sensors plugs into a wall outlet and records very fast (MHz) voltage measurements. From these they calculate frequency. Sensors achieve decent time-synchronization (presumably using NTP), though the company has not explored applications that would require more precise GPS time sync. Whisker Labs recently signed a contract that will require a substantial scale-up in production to fulfill. They've made it clear that they do not have the capacity to fulfill other orders at this time.

IoTaWatt maintains a low cost open-source sensor designed for home energy monitoring. We reached out to them to explore what changes (e.g., custom hardware and firmware) would be needed to adapt the sensor for our purposes. Changes would include increasing the sample rate, and improving time synchronization either via NTP, or with the addition of a GPS module. Though technically possible, these capabilities would not serve the company's primary customer base and they can offer limited support for customizing firmware. Given that the sensor is open source, it would be feasible for us to make these changes ourselves. However, to develop, test and maintain firmware would require additional staff with relevant expertise.

We have determined that the microPMU continues to be the most promising hardware to support the current needs of the project. Should we discover new long-term revenue streams for independent grid monitoring sensors (installed in wall outlets), the IoTaWatt sensor is a viable option to be considered.

06/30/2021 - 100%
We have continued to move forward with the microPMU. During our M3.7.2 workshop, we identified a potential partner at Schneider Electric. They have since reached out to donate an ION 9000 power meter to the project. They wish to use the deployment to showcase analytics that can be performed.

The sensor will stream 1Hz voltage measurements and will send COMTRADE files to an FTP site capturing point on wave data during power quality events -- such as faults, sags, swells, and harmonics. It will be deployed in conjunction with a microPMU to report phasor values from the same deployment site.

T2.3.2 - Identification of sensor locations and collaborators
Due Q10, Year 3
The team will develop a strategic roadmap identifying the best approach for sensor deployment. The roadmap will include an initial rollout to environments controlled by team members to help perfect distribution and operation. It will also include electrical, physical, geographical, and interconnection diversity for the national coverage and boutique sets of sensors. Where necessary, partnership agreements will be established with sensor hosts. The team will document this process.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - 10%
We have started this process and recruited Laurel Dunn out of UC Berkeley to help lead this effort.

06/30/2020 - 40%
We are developing a roadmap for the project to begin targeted outreach to recruit sensor hosts. . The table below details different types of partners we are exploring. We have conducted about three dozen interviews to understand the value proposition to each type of institution, and the concerns each may have around hosting data publicly. The table below lists different types of hosts we are considering and details about each deployment. The sections that follow summarize what we've learned from talking to each.

Conversations with different types of institutions suggests that broad recruitment efforts will be most successful if we can identify one (or two) partners willing to serve as early adopters, and develop analytical tools that allow them to quickly realize the benefits of participation. We have been focusing on understanding the analytical use cases that will most benefit specific partners to ensure that we can set up a pipeline for making these early deployments successful.

While recruiting institutional partners (e.g., college campuses, utilities) to host sensors on grid assets (i.e., at substations, etc.) provides a pathway for collecting high-quality data and for operationalizing tools, these partnerships will also take time to pan out. Recruiting individuals to host sensors in wall sockets presents a much faster pathway for getting sensors in the field, and eliminates many of the institutional barriers to sharing data.

Sensor host	Sensor siting	Deployment type
College campuses	Substation/distribution grid	Wide area/Boutique
Co-operative utilities	Substation/distribution grid	Wide area/Boutique
Municipal utilities	Substation/distribution grid	Wide area/Boutique
Investor owned utilities	Substation/distribution grid	Boutique
Research partners / faculty	Wall socket	Wide area
Grid customers impacted by high wildfire risk	Wall socket	Boutique
Renewables installations	Inverter	Boutique

College campuses

College campuses operate their own mini-grids with interesting/distinctive features such as on-site generation, renewables installations, and microgrids. The Association for the Advancement of Sustainability in Higher Education (AASHE) collects records of these sites and offers a ranking of sustainability programs -- including innovation and integration of campus energy systems with educational curriculum. We are mining these reports to inform outreach to college campuses. AASHE may also provide a platform for socializing deployments to reach additional campus hosts.

Colleges with evolving energy needs or ageing generation assets often contract with engineering firms to modernize their systems. Examples include Clearway Energy Group, Arup, Siemens, and POWER Engineers. Engineering proposals often include “living laboratory” components whereby sensors, dashboards, or other innovative technologies are integrated with engineering design projects to develop educational curriculum. Engineering firms create a natural partnership, as they may also be able to leverage the data to reduce capital costs.

Co-operative and Municipal Utilities

The advantage of working with small utilities is that they are more agile and can see projects through with buy-in from relatively few individuals. We plan to target mid-size utilities that have enough employees to shoulder the work of installing sensors. Decision-making processes at co-ops and munis are often very different from IOUs, and community priorities, local grid challenges, long-term contracts, or institutional values carry considerable weight. As a result, there are certain utilities that have become a testbed for innovative technologies. Examples include SMUD, Kit Carson, Holy Cross Energy, and Green Mountain Power.

Investor Owned Utilities

Conversations with investor owned utilities suggest that there is very little willingness to host sensors that would involve public data-sharing. Legal and regulatory restrictions are often cited as barriers. Though unwilling to contribute data, IOUs have expressed a willingness to support educational content

that could improve the analytical capabilities of early career professionals. We are exploring opportunities to leverage this interest to garner support for the platform.

Research Partners/Faculty
Researchers studying power systems provide a natural partnership for hosting sensors. Recruiting individuals to host sensors in their offices at national labs and college campuses would provide excellent wide-area coverage. Engaging with host departments and institutions could also create opportunities to recruit partners for contributing educational content and host hackathons.

Grid Customers
We are extremely interested in a boutique dataset focused on wildfire risk, and have been exploring the possibility of partnering with local communities (e.g., neighborhood groups, or local businesses) in Northern California to host sensors. Given that most local businesses are closed due to the pandemic, we are focusing on winemakers, as they remain operational. We are also exploring public data released by PG&E and SCE including distribution system maps and ignitions data to identify customers on feeders with long stretches of overhead lines where wildfire risk is high.

Renewables Installations
We have had limited success getting in touch with renewables developers and have prioritized other leads. Deployments on college campuses with renewables installations provide a natural opportunity to develop analytical tools that could demonstrate the value of participation to a wider audience.

09/30/2020 - 45%
Our initial deployment will focus on research partners that are geographically distributed across the country to facilitate the creation of the “broad” data set. This strategy eliminates potential risks associated with legal documentation, data ownership, installation costs, and challenges in establishing partnerships with utilities or other institutions given inability to travel. Focusing on deploying sensors in wall outlets allows us to leverage existing relationships without the need to garner institutional support (and associated permissions) at a time when institutions are heavily taxed and when the risk of further shut downs is high.

Sensors installed in wall outlets do not collect current or three-phase measurements, but our research suggests that there is considerable value in collecting voltage and frequency data. Providing sensors to individuals who are enthusiastic about the cause -- or who have research applications already in mind for the data -- also serves as a “community-building” mechanism, and creates a natural audience for the data.

We have drafted outreach materials for recruiting sensor hosts, and are in the process of finalizing a sensor hosting agreement to establish terms of ownership for data and sensor hardware.

12/31/2020 - 70%
Our first wave of sensors will be sent to project team members to be installed in a wall outlet at their homes. The deployment will aim for broad geographic coverage, while also capturing some interesting

local grid dynamics. One project team member lives on a distribution circuit with a hydroelectric power station, while another lives in PG&E's high fire threat district (HFTD).

The next wave of sensors will go to members of the community interested in siting sensors to support their own research or commercial ventures.

We are also looking for sites (e.g., campuses or testbeds) that are generating data already. In particular, we hope to find sites willing to share information about the topology of their networks, as this data would be incredibly valuable to the research community. A key challenge in obtaining this data is that transmission network topology is carefully guarded, while distribution network topologies are often unknown. This limits the number of locations that are willing and able to provide topological data.

03/31/2021 - 80%

Since the start of the project we've had a number of conversations with organizations interested explore what insights could come from the microPMU. We have also had sensor owners (at utilities) interested in the capabilities of the platform. The most promising partners identified to date have been:

- Individual contributors
- Washington State University (academia)
- University of Vermont (academia)
- University of Alaska (academia)
- Heila Technologies (microgrid operator)
- Camus Energy (vendor)
- Packetized Energy (vendor)
- Schneider Electric (vendor)
- American Transmission Company (transmission utility)
- Bonneville Power Association (transmission utility)
- Dominion Energy (transmission utility)
- Origis Energy (solar developer/operator)

We also have leads we have not yet pursued in storage, electric vehicles, and industrial electrification. A key challenge in recruiting partners (particularly among private companies) is that they are largely unwilling to open access their data.

Deploying sensors in wall outlets is faster, easier, and eliminates institutional barriers to data sharing. However, measurements recorded on secondary distribution networks provide limited visibility into dynamics present on medium- and high-voltage grids. This makes the data less useful for engineering applications.

Wall outlet data does hold high marketing value for the company, as it provides a living proof of concept which anyone can explore. The data may also be readily used by researchers and students, creating a pathway for more widespread adoption students enter the workforce. There may also be commercialization opportunities which we have not yet explored. Stakeholder interviews suggest that the data could be used in regulation (e.g., NERC) and electricity markets (e.g., Genscape).

06/30/2021 - 90%
Discussions with ARPA-E during our last quarterly review centered around focusing the project to grow our user base within organizations that could become commercial partners in the long term. Here, we revisit deployment strategies discussed in previous performance periods with this end-goal in mind.

College campuses / Research testbeds
We have launched a pilot with an NSF funded project out of UCSD called DER Connect. They are installing instrumentation that will enable researchers to use their campus as a DER testbed. The project is geared at real-time control, and the project team is evaluating use of the platform for control applications. While we are excited to support research and demonstration projects such as this, they do not have the same budget or longevity as enterprise contracts. These types of partnerships are not our primary focus.

Municipal and Co-operative utilities
Though smaller utilities may benefit from access to real-time data visualizations, they typically do not have the in-house analytical expertise needed to develop or deploy experimental applications. This means that they tend to seek platforms that offer out-of-the-box functionality rather than flexibility to adapt the software to meet their own individual needs. Given the variety of needs identified to date, we have prioritized the development of a platform that can perform the latter.

We are in discussions with a utility called Kit Carson Electric Cooperative which is heavily engaged in research and development of next-generation distribution grid operations tools. There, the utility has existing partnerships in place to support them in operationalizing the data. We are currently negotiating terms of deploying sensors with them under Task 2.3.2.

Investor Owned Utilities
IOUs have been the primary focus of commercial sales efforts to date and we have been advancing discussions with them about piloting the platform under NI4AI. This partnership model addresses a key pain point that we have identified in marketing efforts to date: potential customers do not grasp the full potential of the platform. Given that the performance we offer is radically different from legacy systems, familiarizing potential costumers with the performance capabilities (and scalability) of a cloud-based platform is foundational to advancing sales efforts.

Grid Customers
We are continuing to deploy sensors in a residential setting to generate a rich and open access wide-area monitoring data set. Among deployment sites selected to date, several of these are expected to exhibit interesting dynamics due to local generation. This deployment strategy eliminates institutional barriers against sharing data. The key disadvantage here, however, is that there is less engagement / buy in from utilities regarding applications developed using the data.

Independent System Operators
Previous performance periods focused largely on identifying partners to deploy sensors that would generate new data. ISOs were not a good strategic fit, as the infrastructure they monitor is not theirs to

instrument. We have had ISOs (including CAISO and SPP) express interest in data exploration, and in decreasing level-of-effort involved in collaborating with external partners.

T2.3.3 - Acquire sensor hardware
Due Q4, Year 1
Equipped with knowledge of the desired sensor hardware from T.2.3.1, the team will procure desired hardware.

9/30/2019 - 30%
We have acquired and begun testing three different sensors, each at a very different price point. The first is the PSL/POWERSIDE uPMU (developed under an ARPA-E Open Innovation 2012 project), which the team has extensive experience with, but is relatively costly. The second is the OpenPMU, which is an academic project that shows promise but poses some challenges with regards to certification (UL) and production scale. The third is the EyeDRO energy monitor which is cheap and mass produced but may not produce sufficiently high quality data for the analytics we wish to do.

12/31/2019 - No update
03/31/2020 - No update
06/30/2020 - 60%
As discussed above, we have decided to use the uPMU for early deployments. We have also started negotiations with PowerSide.

09/30/2020 - 70%
Powerside has agreed to ship sensor hardware directly to sensor hosts, preconfigure sensors to our specifications, and provide technical support during installation (as needed).

12/31/2020 - 70%
We did not acquire any new sensor hardware during the current performance period.

03/31/2021 - 80%
We acquired seven new microPMU sensors during this performance period.

06/30/2021 - 90%
We have worked out the needed administrative processes with PowerSide to acquire all of the uPMUs budgeted for this project. While manufacturing slowdowns due to the current semiconductor chip shortage have impacted our ability to acquire sensors, we are working diligently with PowerSide to acquire the rest of the sensors needed for this project. In 2018 when the project was originally authored we envisioned acquiring all of the sensors in Year 1. Due to the pandemic and associated supply chain / personnel shortages, we slowed the roll out of sensors. In parallel, a strategic change towards leveraging data supplied by existing sensor fleets has meant that procuring sensors is not a barrier to collecting high-value data sets that we had initially intended for sensor deployments to capture.

This performance period we acquired an ION9000 sensor from Schneider Electric. This sensor was shipped to a home with solar PV. A new data ingestor will also need to be developed in order to receive data streaming via either Modbus or IEC 61850.

T2.3.4 - Test and validate one of each type of sensor hardware

Due Q6, Year 2

Before deployment to the field for in vivo data capture (T2.3.5 through T2.3.7) as well as before in vitro data generation (T2.2.1 through T2.2.4), the team will perform verification testing of the hardware to ensure proper functionality and understanding of the sensor configuration and features. This task is closely related to the development of streaming ingestors from T1.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 100%
We have extensively validated the uPMU built by Powerside.

12/31/2020 - 100%
We have identified a bug in the micro-PMU firmware which makes it logistically challenging to onboard new sensors (see M2.3.1 update below). We are working with Powerside to address the issue.

06/30/2021 - 100%
We have tested a new hardware addition that uses an adapter to stream data from the sensor via wifi (rather than ethernet) to a local internet router. This relaxes certain siting constraints which include: ethernet connectivity, GPS visibility, and proximity to a power outlet. These constraints made it difficult for some sensor hosts to find a suitable location.

T2.3.5 - Deploy sensor hardware to control group (project partners)

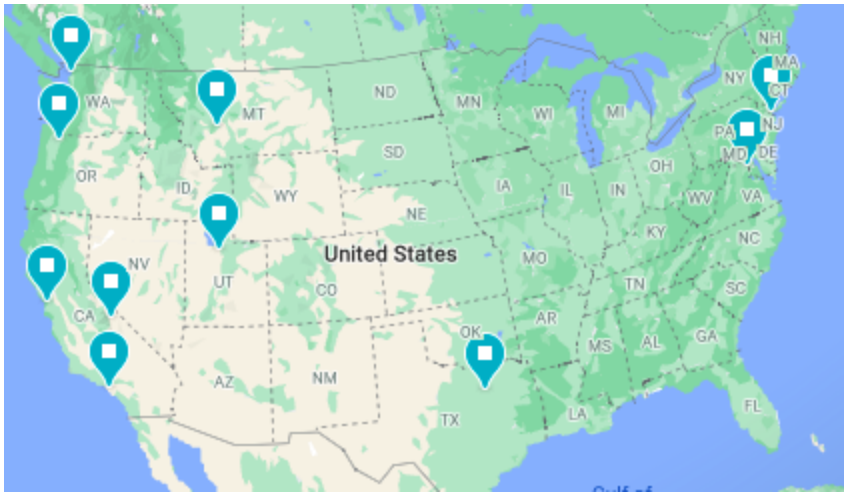
Due Q12, Year 3

Sensor hardware, particularly for the national coverage data set, will be first deployed to a pilot group of sensor hosts. This pilot group will be primarily composed of project team members. The purpose of this step is to enable the team to perfect their end-to-end processes of testing, packaging, shipment, installation, configuration, and connection in a controlled environment before shipping to collaborators.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 25%
The initial group of sensor hosts have been identified, contacted, and have confirmed interest in hosting sensor hardware and streaming data.

12/31/2020 - 40%

We have re-distributed sensor hardware acquired to date to achieve broader geographic coverage, and to get more practice installing sensors in a residential setting. We will be sending micro-PMUs to PingThings employees to refine the installation process in a way that will make it easier to scale. We have selected sites that will also allow us to begin generating wide-area monitoring data. The map below shows several sites selected to date; the first wave of sensors will go to a subset of these sites.



03/31/2021 - 60%

We have acquired seven microPMUs. These are distributed as follows:

- One is located in Oakland, and has been streaming data to the platform for over a month. We are now using the sensor to test a new release of the microPMU firmware which will enable us to connect via UDP (rather than TCP). This will eliminate the need for sensor hosts to enable port forwarding on their local area networks.
- Two of the sensors were shipped to key project personnel also involved in testing the new firmware and confirming that the new capabilities will meet project needs. This testing is underway.
- The remaining four microPMUs are currently in Oakland and will be shipped to the appropriate project personnel once a stable version of the firmware can be installed.

06/30/2021 - 100%

We have installed seven sensors to date. These have been distributed to obtain wide-area coverage. Five are sited with a control group of key project personnel. The other two were sent to family/friends to vet deployment procedures with less technical partners. Sensor deployment took at most 45 min from start to finish (from “unboxing” sensor components to streaming data) when streaming data via UDP.

T2.3.6 - Deploy sensor hardware to national coverage (wide-area) locations

Due Q12, Year 3

Sensor hardware is deployed to the identified locations to generate the National Coverage (wide-area) data set.

9/30/2019 - Not Applicable

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 10%
We have selected the initial wave of sensor hosts.

We have also connected three PMUs that are part of the Texas Synchrophasor Network, and are exploring opportunities to host data for new and existing micro-PMU owners.

12/31/2020 - 20%
We have re-distributed sensor hardware acquired to date to provide broader geographic coverage. We are currently waiting on Powerside to make a commitment to fixing a firmware issue (see Task 2.3.3) before we send additional hardware to additional project team members (see Task 2.3.5).

03/31/2020 - 30%
We have now deployed three of the initial seven sensors, and will ship the remaining four as soon as a stable firmware release can be installed. Host sites were selected with wide-area monitoring in mind.

06/30/2021 - 40%
We have now deployed seven sensors and selected sites for another 13 sensors to obtain wide-area coverage of the grid. TT&O efforts to date have connected us to a broad community of enthusiasts that we can easily tap for additional deployments. We have prioritized deployments with project personnel or other key stakeholders and have held off on deployments with more strategic partners (e.g., utility practitioners) until we deploy and test our web interface for connecting sensors. This will enable us to put our best foot forward in marketing capabilities of the platform to potential customers.

T2.3.7 - Deploy sensor hardware to boutique (site-specific) locations
Due Q12, Year 3
Sensor hardware is deployed to the identified locations to establish a collection of boutique locations of sensor data on the grid.

The sensor host will re-package the sensor hardware for shipment. The team will facilitate the shipment (logistics and cost) of the sensor hardware to the subsequent facility. The sensor host will provide support in this process.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 10%
We have begun outreach to a number of utilities and possible host sites. We are in discussion with Powerside, potential hosts, and with the research community to identify deployments that maximize the value of the data we collect across various stakeholder groups.

We have also begun outreach to micro-PMU owners to discuss the possibility of hosting data from existing sensors. Given that many of the sensors are installed for R&D purposes, these installments could provide interesting and valuable data for boutique datasets related to distributed generation, microgrids, and perhaps other applications.

12/31/2020 - 20%
We have reached out to several people in our network -- including Heila Technologies and Camus Energy -- about siting sensors in locations where the data would be beneficial to them. We have also had fruitful conversations with UC San Diego and the University of Alaska about leveraging their partnerships and/or campus infrastructure. We have also been in discussion with groups that have already installed micro-PMUs about sharing their data with the project. We are currently working to define data sharing terms and data handling requirements (see Task 2.4).

03/31/2020 - 20%
We are waiting to complete firmware testing before we deploy sensors with project partners. While we have identified a couple of host sites willing to provide open access to their data, this has been a limiting factor for other partners.

Several other potential partners have expressed willingness to enter into data sharing agreements to explore the value of the sensors and platform. This change in focus could give us access to new host sites, and could lead to the creation of data sets that will help us to develop and demonstrate new high-value use cases through closer collaborations with sensor hosts.

06/30/2021 - 50%
As detailed in our milestone report for M2.3.2, we have focused boutique data acquisition on industry and research groups that already have sensors deployed. This strategy is advantageous in that it creates an opportunity to ingest a longer time history of data than would be possible to generate within the scope of the project, and enables us to collect more data than would be feasible to generate with sensors deployed under the project.

Leveraging existing sensor installations also creates a natural mechanism for stronger partnerships with organizations and key personnel in the industry who are pioneering sensor deployment and analytics. We have leveraged the project to establish connections with these organizations in hopes that these connections will help to drive sales.

M2.3.1 - Go-No/Go: Successful sensor deployment to control group
Due Q6, Year 2
The first wave of sensors to create the wide area or national data set have been acquired, tested, and deployed with team members or trusted collaborators. Data is streaming to the platform. Demonstration of streaming data to ARPA-E.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable

03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 50%
We have developed the necessary ingestors, identified sensor hosts, and are ready to ship sensors once negotiations with Powerside regarding the cost of ordering sensors in bulk are complete.

12/31/2020 - 60%
We have re-distributed sensor hardware acquired to date to achieve more widespread geographic coverage. We now have a sensor in Los Angeles, and another in Oakland. As we configure the sensor in Oakland, we are also working to streamline and simplify a sensor onboarding process which we can refine as the next round of sensors is deployed.

We are currently in discussions with Powerside about fixing a firmware bug in the micro-PMU that interferes with bidirectional communication with the sensor. The device does not by default send information necessary to decode the data it transmits and this information must be explicitly requested periodically. To do so, however, requires two-way communication with the device, a capability which has not been fully developed.

To date, micro-PMU installations have avoided this issue by having sensor owners manually configure firewall settings to enable port forwarding to the device, a process requiring sensor hosts to customize their network settings. For small deployments, this hack has been adequate but, at scale, especially when taking into consideration that sensors may be moved to new locations, is not viable and would likely generate significant additional and unbudgeted technical support requests. The best solution is to simply fix the bug, which our engineers estimate is confined to a few lines of code.

Powerside is aware of the issue and we are currently waiting for them to make an internal decision about how to move forward. Powerside has made it clear that their commitment to supporting the device is limited.

03/31/2021 - 100%
See milestone report.

M2.3.2 - Successful sensor deployment for national coverage data set
Due Q8, year 2
The first 20 sensors have been acquired, shipped, and deployed to generate the wide-area data set. Data is actively streaming to the platform. Demonstration of streaming data to ARPA-E.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 10%

We are close to having the first wave of the initial 20 sensors acquired and deployed. The rest will quickly follow.

12/31/2020 - 20%
See task updates above.

03/31/2021 - 30%
See Task 2.3.2 update above.

06/30/2021 - 50%
See Task 2.3.2 update above.

The completion of this milestone has been delayed due to the continued semiconductor chip shortage that is slowing the acquisition of the sensors needed for the wide area dataset.

M2.3.3 - Successful sensor deployment for boutique data set
Due Q7, Year 2
At least half the sensors for the first boutique data set have been acquired and deployed for the creation of the first boutique (site-specific) data set. Data is streaming into the platform . Demonstration of streaming data to ARPA-E.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - 20%
We have identified a number of sites that could serve as sources of boutique data for the project -- including microgrids, college campuses, and distribution co-ops. Targeted outreach efforts have shown us that many researchers and start-ups are interested in collecting micro-PMU data, and are interested in receiving sensors from us. We have had promising discussions to date with UC San Diego and the University of Alaska about connecting to existing PMU installations, both of which are sited on microgrids. We are also in discussions with two start ups -- Camus Energy and Heila Technologies -- about siting sensors on grids that they operate.

03/31/2021 - 30%
See Task 2.3.2 update above.

06/30/2021 - 100%
See milestone report.

<u>T2.4 - Data escrow platform services</u>
Due Q12, Year 3

This task has goals of (1) establishing an industry standard Non-Disclosure Agreement (NDA) and Data Sharing Agreements (DSA) to facilitate data exchange; and (2) development of a service on top of the platform to ingest, clean, optionally obfuscate, and provide to the collaborators identified by the original owners of the data.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - Not Applicable
See task updates below.

03/31/2021 - 20%
As originally scoped, the primary goal of the project was to generate data that could readily be shared via the platform. Moving forward, we wish to build deeper collaborative relationships centered around use case development. This task will include deploying separate instances of the platform with individual companies interested in a proof-of-concept. It would allow them to explore use cases in house, and to test and deploy use cases developed by collaborators (e.g., Dominion, UC Berkeley, or other groups using the platform). This task will focus less on broad data sharing, and more on focused collaborations that will improve the capabilities of the platform both to support use case development and to build out-of-the-box data analysis tools.

06/30/2021 - 40%
Even in 2021, utilities are still exporting data from legacy historians to large and inefficient CSV files, copying these files to hard drives, and physically shipping the drives to collaborators. This is archaic for a large number of reasons. Exporting data and shipping drives is incredibly resource intensive. Further, and more importantly, the utility is physically shipping its data to a third party and, in the process, losing all control over that data. This last fact mandates robust and restrictive legal paperwork between the two parties, which can take years to complete.

The original vision for Data Escrow was to use the PingThings platform as a more efficient mechanism for utilities and other grid asset owners to share data with interested parties such as researchers. Utilities could upload the data to the NI4AI platform where it would be securely held for the approved third party collaborator, an idea very similar to putting money into escrow to facilitate a large financial transaction.

This idea has evolved given our experiences with our commercial customers. There is still significant effort to get data into the PingThings platform, on par with exporting the data to disk. However, once the data is in the platform, it can easily be shared with multiple organizations. Further, export restrictions can be enforced that mandate the data remain in the cloud where all work will be done in the platform. First, data escrow is no longer a representative term as the data never changes hands. Instead, a much better label to use is data collaboration. Further, since the data is not being transferred to a third party, we believe that much less legal paperwork will be required. Thus, the legal paperwork has not been as much of a focus for the project as originally projected.

Finally, if we are to go through the effort of having a utility push data into our platform, we should use this as an opportunity to sell the utility on the platform. This is the perfect chance to conduct a pilot or proof of concept with the utility so that they can also use the data while it is in the platform. This led us to change the focus on enabling collaborations (a positive side effect that we have seen with our commercial customers) to launching a series of platform pilots.

Pilots to date have been geared at enabling organizations to interact with their data internally or at providing open access to anonymized data, and thus have not required NDA/DSA documents as expected. The only legal documentation needed to date is a sensor hosting agreement establishing terms of data ownership for sensors purchased under the NI4AI project, which we have included as an attachment to this report.

T2.4.1 - Collect sample NDA and DSA documents

Due Q4, Year 1

The team will engage with electric utilities, universities, research laboratories, to gather previously used or standard NDA and DSA and to evaluate each stakeholder as a potential early adopter. The team will create marketing material to support this initiative.

9/30/2019 - 10%

Document collection has begun.

12/31/2019 - 10% - no change

03/31/2020 - 10% - no change

06/30/2020 - 70%

We have a number of NDA and DSAs from past PingThings projects and have identified prior efforts to collect and consolidate legal agreements.

09/30/2020 - 100%

While collecting NDAs, we became aware of an initiative by GridBright to publish standardized data sharing documents for other projects (like ours) to use.

In our last quarterly review we discussed the possibility of building on the work GridBright has already done to synthesize common elements from existing documents. The project has already drafted a standard agreement, which is currently undergoing review. It was agreed that our project would use the GridBright documents once they are finalized, rather than drafting our own.

06/30/2021 - 100%

GridBright has made their NDA available for us to use should the need arise.

T2.4.2 - Legal analysis of sample NDA and DSA documents

Due Q6, Year 2

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - Not Applicable

03/31/2021 - 10%
We wish to re-focus Task 2.4.2 (and Milestone 2.4.1) to support data sharing agreements between PingThings and project partners. The company has experience entering into confidential data sharing agreements, and will review them on a case-by-case basis. Where appropriate, we may consider using the GridBright NDA and DSA documents which were recently released.

06/30/2021 - 100%
See milestone report for M2.4.2.

M2.4.1 - Legal document collection and analysis
Due Q4, Year 1
Collection of at least 5 separate non-disclosure agreements and data sharing agreements from appropriate organizations. Legal analysis of the 5 separate documents to identify commonalities and essential components.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - 20%
See above.

09/30/2020 - 50%
GridBright has shared with us a draft of their NDA which they submitted for legal review. As discussed in our quarterly PD memo and ab, we will leverage these documents once they are finalized rather than pursuing Milestone 2.4.1 in parallel with the GridBright initiative.

12/31/2020 - 50%
We still intend to use the GridBright NDA for the data escrow aspect of the project. However, a new potential opportunity has arisen that needs to be addressed.

The original intent of the project was to generate “open” data accessible to third parties without the need of common legal encumbrances such as a DSA or NDA. However, in our search for sensor hosts, we have found a class of host, typically in industry, that while open to serving as a sensor host, is not amenable to complete openness. Given that these hosts could enable the creation of very valuable data sets, we are exploring the possibility of embracing data handling requirements that would allow us to move forward with these partnerships.

We are continuing to gather information about data handling requirements as these partnerships emerge. Candidate hosts have voiced concerns about providing open access to the following:

- the precise location of the sensors,
- the identity of the sensor host,
- real-time streaming measurements.

We are working to define data sharing terms that will enable us to collect data from prime deployment sites without compromising on collecting information that would be needed to support specific use cases. This could involve using approximate locations or obfuscating the locations of certain sensors, or adding a time delay (e.g., 24 hours) before data become publicly available.

03/31/2020 - 50%
See Task 2.4.2 update above.

06/30/2021 - 100%
As stated during our 9/30/2020 task update above, we became aware of a parallel effort by GridBright to collect and analyze legal documents to facilitate data sharing. By the time we became aware of the effort, a common NDA had already been drafted and was under legal review. We have included the final NDA as part of our milestone report for M2.4.2.

T2.4.3 - Develop common NDA and DSA documents
Due Q7, Year 2
The team will draft a document which considers all three parties (producer, consumer, and escrow service) involved in the data transfer including NDA, DSA, and Terms of Service. Feedback will be solicited from stakeholders identified in T2.4.1. The resultant documents will be made publicly available.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - 20%
We have drafted a data sharing agreement for the first round of sensors which will be installed in people's homes. We are working with candidate host sites who may have more stringent data handling requirements -- such as college campuses and distribution grids -- to understand potential concerns and requirements. We are working to define a policy that will be amenable to some, without compromising on the utility of the data we collect.

03/31/2021 - 20%
Hosting confidential data in the platform poses several requirements that the NI4AI platform (as currently deployed) does not meet. Data supplied by different organizations would need to be hosted in distinct instances of the platform to provide confidence that only users within the organization (and

perhaps trusted collaborators) will be able to gain access to confidential data. This is the same approach we use to deploy the platform with paying customers.

06/30/2021 - 100%
See milestone report.

M2.4.2 - "Common" NDA and DSA developed and released
Due Q6, Year 2

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - Not Applicable
03/31/2021 - 10%
See task updates above.

06/30/2021 - 100%
See milestone report.

T2.4.4 - Determine data handling functional requirements
Due Q6, Year 2
The team will engage with previously identified stakeholders to determine requirements for the data escrow service including encryption, obfuscation, and user interface.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - 40%
As discussed above, we have been working with potential partners to understand data handling requirements necessary to make data available without an NDA. We have also identified several precautionary measures that we can take to build confidence with risk averse data providers. These measures include removing sensor locations, adding a delay so that real-time data are not available, modifying the time stamps (e.g., to offset them by a year), or providing access to voltage but not current streams so that power flow cannot be calculated. Several of these measures significantly limit the value of the data.

Outside of transmission, data handling requirements are set by institutional policies rather than by regulations. Even federal regulations are quite open to interpretation, rather than being prescriptive. This means that policies vary from one institution to another, and that many institutions do not have

well-defined data handling requirements. We are in discussion with a number of possible host sites to better understand what concerns they have, and to define a blanket policy that will be amenable to some. We remain open to the possibility of negotiating terms with individual organizations as prime deployment sites or compelling partnership opportunities emerge.

To date, host sites have expressed concerns about revealing the precise location of sensors, and of providing access to real-time data. Data handling procedures may mask the precise location of the sensors, or may hide their locations altogether. The latter is to be avoided if possible, as it means the data can no longer support wide-area monitoring applications. We are exploring options for maintaining enough information to support wide area monitoring without compromising the identity of the data provider. These discussions are ongoing.

03/31/2021 - 45%
If we deploy separate instances of the platform for each partner organization, a blanket policy is no longer needed as access and permissions can be easily customized for each deployment. We expect that the existing capabilities of the platform will exceed expectations. As we bring different users onto the platform, we will continue to gather information about their interest in sharing data more broadly, and requirements for doing so.

Deploying new instances of the platform repeatedly under the project would allow us to develop a simple and streamlined approach for onboarding new customers. At scale, we anticipate that this process (sans data ingest and sensor onboarding) could take as little as 1 hour of developer time.

06/30/2021 - 60%
Data handling functional requirements will continue to be identified and defined as we launch new pilot projects. Given that the project has recently refocused to support platform pilots across various organizations, the original timeline no longer makes sense.

Pilot discussions to date have focused on using the open access NI4AI platform to ingest anonymized data, and on using dedicated instances of the platform to host proprietary data. The former provides access to data which other projects/partners may use, while the latter enables us to better showcase the capabilities of our product.

We have found that many possible partners are looking for leadership regarding best practices for cloud secure data management. This has led us to revisit our own processes for user account and password management both as a company and for our customers. We now require two-factor authentication for all PingThings personnel, and are revisiting processes for user account creation, password resets, and account management for admin users.

T2.4.5 - Implement required platform functionality
Due Q8, Year 2
The team will implement the functional requirements identified in T2.4.4.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - 30%
We have the functionality to grant universal access across all users, or to restrict how certain users can view or interact with data in the platform. Currently, we can restrict user access to certain streams or collections. We use this functionality internally in developing new data ingestors. Data are first ingested to a collection on the NI4AI cluster that is visible only to team members. Once the ingestor is vetted and the ingest complete, the data are made available to users by moving it to a publicly accessible collection.

Data handling requirements established to date include masking locations and adding a time delay so the data are not available in real-time. Masking locations simply involves excluding lat/lon information from metadata fields (i.e., "annotations"). A time delay is not currently supported, however. We are investigating two possible options. One is to ingest data into a private stream that is accessible only to the sensor host, and copy data to a public stream in batches (i.e., once every 24 hours). The drawback of this approach is that it would increase storage costs. We are also exploring an alternative approach which would set permissions on the basis of stream *version*, thus restricting user access to newer versions of the data. This functionality is not currently supported, and would need to be developed.

03/31/2021 - 40%
We will work on broadening our support for restricting collaborator access as we learn more about the requirements of different partner organizations.

This quarter we deployed a portal through which administrators can add users, create user groups, and modify permissions granted to different groups. Permissions can be granted (or revoked) for each collection (or group of streams) in the database. A collection could include an entire data set, a single measurement site, or calculations derived from the data. Currently, users can be granted permissions to read, visualize, issue API queries, insert, delete, or "obliterate" (i.e., permanently delete) data. Additional data permissions will be implemented as we gather further information about customer needs.

06/30/2021 - 50%
See task update above. We fully anticipate that we will continue to encounter new requirements for user account management as both the number and variety of pilot participants (e.g., IOUs, co-ops, ISOs, etc.) continues to grow.

T2.4.6 - Validate data escrow functionality
Due Q8, Year 2
The team will validate the end-to-end data escrow service through two approaches. First, the team will guide a sample data set through the process while validating the functional requirements. Secondly, the team will work with an early adopter to perform the same validation with "real" data.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - 20%
As discussed under Task 2.4.5 above, the project team is currently using this functionality to create a staging environment to test and deploy new data ingestors. Additional work is needed to add a time delay before streaming data become available to users.

03/31/2021 - 30%
See task 2.4.5 update above.

06/30/2021 - 50%
See task 2.4.5 update above.

M2.4.3 - Deployed and tested data escrow functionality.
Due Q8, Year 2
Data escrow requirements implemented as a service running on top of the platform with a successful exchange of data between two separate parties.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - 20%
As discussed above, we have already deployed and tested some of this functionality for internal use.

03/31/2021 - 30%
See Task 2.4.5 update above.

06/30/2021 - 50%
See task 2.4.5 update above.

T2.4.7 - Market data escrow service
Due Q12, Year 3
9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable

12/31/2020 - 10%
Our ability to grant differential access to certain users is one of the core features that makes the platform an effective collaborative tool, and we have been making the community aware of this feature. As we build relationships with host sites, we are continuing to learn more about data handling requirements, and about their interest / willingness to share additional data (e.g., about network topology) with the general public, or with trusted collaborators. Stakeholder outreach efforts suggest that there is a large community of researchers whose work could benefit substantially from gaining access to this kind of data.

03/31/2021 - 20%
We highlighted the collaborative features of the platform during our Spring NASPI WG vendor talk. We believe this message will become much more compelling if we can demonstrate this directly with project partners. To say that logistics around data transfer and anonymization have been a pain point in the past would be an understatement. Despite these challenges, most utilities remain skeptical about the security of hosting data in the cloud.

Bringing more utilities onto the platform via the project will allow them to experience these benefits first hand. Where existing IT policies permit cloud-hosting of sensitive data, we are suggesting that data providers could ingest anonymized data as a proof of concept. This could also allow for broader data sharing with research partners via the platform; effortless data sharing is one of the obvious benefits of cloud hosting data, and could potentially bring new advocates to the table.

06/30/2021 - 30%
The Spring NASPI working group meeting included a series of talks highlighting logistical challenges around data preparation and transfer for FOA 1861 projects. In response to these talks, we highlighted “data escrow” capabilities for data-sharing, anonymization, and collaboration during our NASPI talk.

Additional marketing efforts are primarily focused on onboarding new potential users and demonstrating key functionality of the platform to them. These demonstrations have showcased user access controls, alerting, and other key functionality that are central to leading demonstrations of platform capabilities that will be compelling from a business perspective.

Note that all text in *blue Italics* is from the last full SOPO Attachment that was exchanged between PingThings and ARPA-E. Also, the tasks appear ahead of the Milestones for which they lead into.

Task 3 - Community building

Task 3 yields the following set of deliverables: (1) a community of engaged individuals and organizations; (2) project-related content; (3) a project launch and associated materials; (4) a project-focused blog; (5) a project-focused website; (6) platform-related content including tutorials, videos, documentation, distil programming guide, literature reviews, bibliography, and aggregated relevant content from other fields; (7) conference and related-event attendance and presence; (8) hosted side events and meetings at established conferences; and (9) data science competitions and hackathons related to the data captured by the project.

The Community is defined to be a group of engaged, active researchers, some from within the power engineering field and others from external fields working to solve existing problems and develop new use cases for analytics, ML, and AI applied to grid data - is the ultimate deliverable for this task. The community will, over the course of the three year-project, help develop the content listed above.

The Project Team will focus on building a community around the platform. The community will be brought together through publication of data analyses, publication of papers, conference presentations aimed at academic and industry audiences, system documentation to facilitate use by researchers, sponsorship of research contests, and three meetings to showcase success stories and continue to grow the community.

M3.0 - Go-No/Go: Outreach documentation provided to ARPA-E
Due Q8, Year 2
Documentation provided to ARPA-E for PD approval detailing T2M efforts including identification of applications for platform, partnerships (existing and planned), as well as general community engagement and feedback, including feedback given in Task 1 and 2.

06/30/2021 - 90%
T2M efforts are detailed in our forthcoming milestone report.

T3.1 - Announce project launch
Due Q2, Year 1
The project launch provides an excellent opportunity to begin community building immediately.

9/30/2019 - 40%
Even though we have not created all of the content listed below in subsequent sub tasks, we have aggressively announced and described the project via:

- Papers
- Conference booths
- Abstracts
- Presentations
- Direct emails

This information is described in greater detail in the T2M section of the quarterly report.

12/31/2019 - 100%

The Milestone 3.2 completion report details our efforts in greater detail, especially focusing on the effectiveness of conference presentations versus booths. We found that booths were far more effective (although more expensive on a cash but not time basis) in generating more and varied in-person connections.

T3.1.1 - Create content for the project announcement

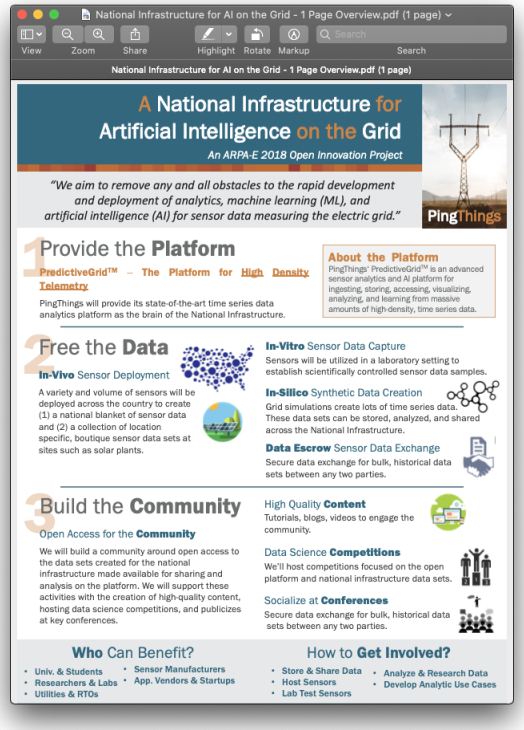
Due Q2, Year 1

The team will create press releases and additional marketing material to announce the project and its goals with clear direction on the next steps for individuals and organizations that would like to get involved at this time or later in the project.

9/30/2019 - 40%

We have created:

- Text for emails
- A 60MB, 20 minute presentation (and several other shorter variations)
- A 1-page flyer overviewing the project (shown below)



12/31/2019 - 100%

In addition, we have created:

- Many more emails
- Many variations of the presentation for use with different audiences

T3.1.2 - Make broad project announcement

Due Q2, Year 1

The announcement content will be disseminated using traditional ARPA-E channels, via social media (including LinkedIn, Twitter, and YouTube), and announcement via the website and blog.

9/30/2019 - Not Applicable

We have not focused on a social media-based announcement of the project yet.

12/31/2019 - 100%

Our broad project announcements have focused on the use of the project website and blog and direct email communications. We have made announcements via LinkedIn but not Twitter. We are looking for additional, effective communications channels moving forward. We believe that YouTube could be an effective avenue in the future once we have generated video content as stipulated by later tasks.

T3.2 - Community recruitment

Due Q12, Year 3

The next phase of community building is to try to jumpstart the community with individuals from across fields and organizations whose beliefs are well aligned with the underlying project ideologies. Recruitment will come in two primary forms: (1) targeted outreach through personal relationships, word of mouth, and direct connections and (2) organic outreach that entices new members to explore, learn, and participate.

9/30/2019 - 10%

While we have a list of individuals in mind for the first wave, only informal recruitment has commenced.

12/31/2019 - 25%

Direct recruitment of specific individuals, institutions, utilities, and vendors has begun, initiating a number of conversations with:

- Dominion Energy
- EPRI
- Schweitzer Engineering Labs
- PowerSide (formerly Power Standards Labs)

03/31/2020 - 30%

Recruitment has continued at a somewhat slower pace due to the pandemic. PingThings is bringing on multiple UCBerkeley interns to accelerate outreach.

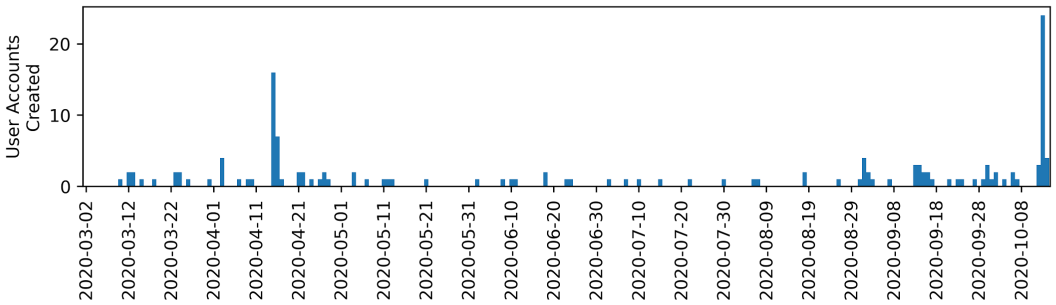
06/30/2020 - 40%

Recruitment has continued, focusing on targeted outreach to specific individuals. Conversations to date have focused on building partnerships within DOE and at research institutions. Interns have helped with developing a strategy for doing targeted outreach to different types of partners and with drafting outreach materials.

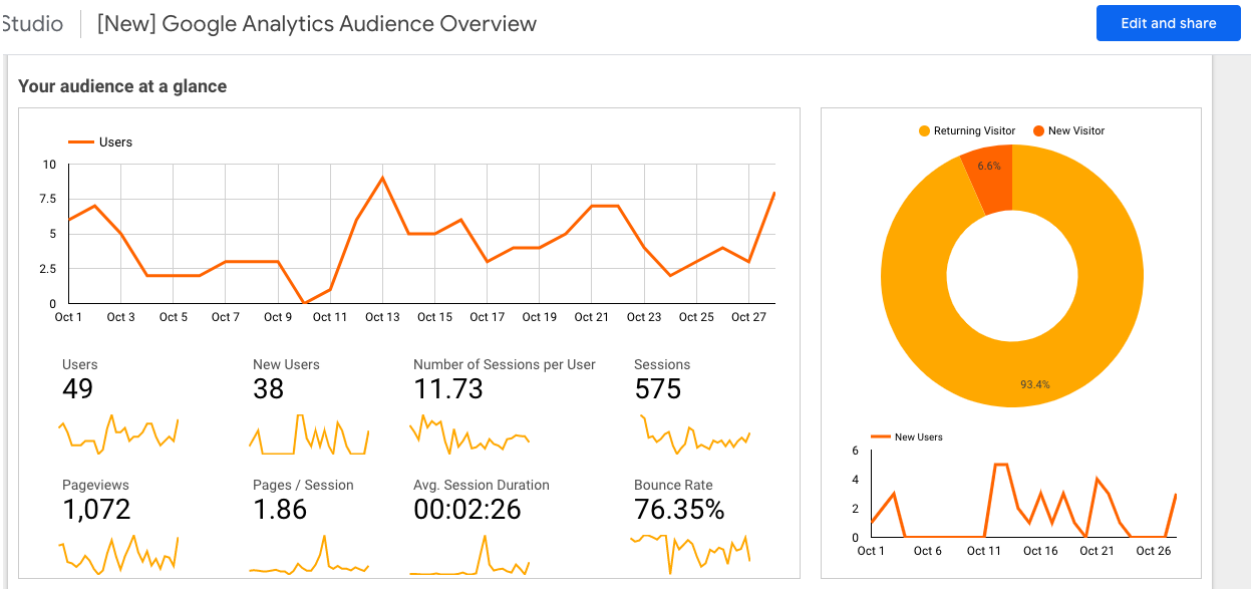
09/30/2020 - 50%

We have been experimenting with different outreach strategies including targeted outreach, email campaigns, announcements circulated to listservs, and outreach to individuals who attended past events. We are tracking several metrics of success including: (1) new user accounts, (2) web traffic, and (3) workshop attendance.

New user accounts



Web Traffic



Workshop Attendance

- Spring NASPI working group: 200+ attended
- September NASPI webinar: 100+ attended
- CIGRE Academy Webinar (PMU Fundamentals): 50+ attendees
- CIGRE Academy Webinar (AI on the Grid): 80+ attendees
- IEEE SmartGridComm: TBD
- Fall NASPI panel: TBD

We are also developing scripts to monitor API usage that will allow us to track which users are active, and what data they interface with.

12/31/2020 - 60%

We have continued to announce the project via conferences and targeted emails. These outreach efforts have four primary objectives:

1. Make people aware of what we are doing.
2. Teach people how to use the platform.

3. Showcase datasets we host.
4. Share information about how people are using the platform.

With regards to [1], stakeholder feedback has been overwhelmingly positive. More challenging is to get people to engage more deeply with the project and the platform.

With regards to [2], we have in the past hosted on-site workshops with PingThings customers to familiarize them with the platform and workflows. We have repurposed that content and integrated new content into our workshops (Task 3.7) and github tutorials (Task 3.5). We are also working on platform enhancements that will make it much easier for new users to onboard themselves (Task 1.12).

With regards to [3], though we are continuing to identify new stakeholder groups, we are now at a point where we understand many of the stakeholder needs and have ingested data sets that are interesting and valuable to the community. As the project proceeds, we are shifting the emphasis of outreach efforts to provide more detailed information about individual data sets we host, and about how they can be used.

With regards to [4], most of our outreach to date has focused on work done by the project team. As more users spend time with the data and become familiar with the platform, we hope to solicit much more content (blog posts, github contributions, panel presentations, etc.) from users. We plan to use data competitions as a mechanism for learning about work users are doing, and soliciting content we can showcase.

03/31/2021 - 65%
We have hosted talks and a workshop during this performance period, though we saw limited attendance. We have received a number of introductions from both close collaborators and members of the broader community as they hear of people who are looking for data. We plan to issue another broad announcement about the project once we have made a bit more progress on our wide-area deployment.

06/30/2021 - 80%
We hosted talks this quarter at the NASPI and JSIS working group meetings, a NERC staff meeting, and various ad hoc meetings with various companies about the platform and project. We also delivered a workshop to the IEEE SGSMA (Smart Grid Synchronized Measurement and Analytics) conference, and participated in two panels at IEEE Power and Energy Society General Meeting.

Talks highlighted capabilities of the platform, testimonials from users, presentations on analytical methods used to synthesize data in the platform, and broader discussions about platform capabilities and project offerings. This quarter, outreach efforts have been focused on securing pilots enabling practitioners in industry (including utilities, vendors, regulators, ISOs, etc.) to explore their data in the platform by streaming data from existing sensors, installing new sensors, or bulk data uploads.

Outreach efforts have continued to announce the availability of open access data sets, and to facilitate connections/collaborations between new and existing users of the platform.

T3.2.1 - Develop marketing content and tools to be used for targeted recruitment

Due Q3, Year 1

This would include professionally designed slides, images, one-pagers, and diagrams that can be distributed in material or used for the website and for the blogs.

9/30/2019 - 40%

While the adjective “professionally” could be argued, we have put forth substantial effort to develop an overview presentation describing the project and made a one-page project overview, complete with graphical elements. We expect work on this to continue with more content getting developed. The content has been described in more detail in other tasks/milestones, including screenshots.

12/31/2019 - 50%

We have iterated on and created multiple versions of the presentation content used to discuss and announce the project and its goals at conferences.

03/31/2020 - 55%

We have enhanced the website with a conferences page, collecting all relevant conferences and posting them online. Unfortunately, the pandemic happened. We have also developed a poster overviewing the platform.

06/30/2020 - 60%

We have drafted announcements for circulation and slide decks geared at introducing the project to different types of audiences (i.e., analysts, utilities, students). We have drafted emails targeted at recruiting different types of partners (i.e., funding agencies, utilities, researchers), and for soliciting coverage for the project on public media outlets.

09/30/2020 - 100%

Marketing content including a slide deck, announcements, and flier were developed and have been circulated through both targeted and widespread outreach efforts.

We are continuing to develop new outreach materials to share information about project developments, events, and “ideas” providing thought leadership relevant to the success of the project and platform. These efforts are currently targeted at normalizing the idea of building collaborations, and motivating the need for open access datasets to support. We are also issuing broader recommendations about data hosting and archiving to facilitate every-day data access.

12/31/2020 - 100%

We have continued to develop outreach materials, including a slide deck and canned emails which we are sending to potential partners and sensor hosts. These materials will continue to evolve as the project advances and the desired outcomes from outreach initiatives change.

03/31/2021 - 100%

We are developing additional marketing material to support our booth at ARPA-E summit in May.

06/30/2021 - 100%
We have continued to develop new marketing material targeted at new audiences we address. We have found messaging to be more successful when we place the project in the broader context of complementary initiatives with which potential partners are engaged.

T3.2.2 - Select first cohort of potential research partners from across academia, industry, and utilities.
Due Q2, Year 1
The team needs to jump start the creation of this community by identifying key individuals within utilities, academia, consulting firms, and government that may find the project of particular interest. This list of individuals and organizations can be informed through knowledge from the literature which identifies who is working on the most relevant work that could benefit from the value proposition of the platform.

9/30/2019 - 20%
While we have a list of individuals in mind for the first wave, only informal recruitment has commenced via emails and in person conversations. We would like to have a bit more up and running before officially inviting individuals to collaborate with us.

12/31/2019 - 50%
We have identified key individuals for outreach but believe that further digging will result in additional folks to contact.

03/31/2020 - 60%
Outreach continues at a somewhat slower pace due to the pandemic. We are transitioning this task to be led by UCB.

06/30/2020 - 100%
We are maintaining a presence in online conferences, working group sessions, and webinars and have been following up with key individuals identified through these channels. We are also pursuing partnerships geared at elevating the profile of the project in industry and academia by tapping in to trade organizations, DOE research initiatives, and universities with prominent power engineering programs.

We have compiled a list of over 50 strategic partners working in this area in academia, industry and government. A list of this first cohort of partners in available here: https://docs.google.com/document/d/1jtEnKNBlQ5pBYX0iypXyiBKOqPUdLuN4ehy64PPiC7Q/edit?usp=sharing

12/31/2020 - 100%
Workshops, informational talks, and participation in industry meetings is continuing to bring new partners to light.

03/31/2021 - 100%
As we look toward shifting towards confidential data sharing with data providers, we will have more opportunity to work directly with stakeholders to understand their needs and solicit collaborations with specific research groups on the basis of their particular areas of expertise.

06/30/2021 - 100%
Platform pilots launched under the project to date are detailed in our milestone report for M2.3.2. Further detail on how these partnerships advance the strategic positioning of the company and of the project will be included with our forthcoming milestone report for M3.0.

T3.2.3 - Targeted partner outreach to the first cohort
Due Q3, Year 1
The team will directly engage with identified organizations and individuals. In addition to working towards successful recruitment into the community, throughout this process the team will learn about what matters to these members of the community, what the best ways to communicate the project objectives and value proposition are, and who makes up the next cohort for outreach.

9/30/2019 - Not Applicable
12/31/2019 - 50%
Emails have been sent to all individuals identified in T3.2.2

03/31/2020 - 60%
Outreach continues at a somewhat slower pace due to the pandemic. We are transitioning this task to be led by UCB.

06/30/2020 - 90%
Outreach to individuals identified in T3.2.2 continues.

09/30/2020 - 100%
We have continued to contact key individuals identified in T3.2.2, as well as individuals whom they have connected us to.

We saw an uptick in user accounts after the NASPI working group meeting in April, 2020 and had a handful of individuals reach out -- including at LANL, PNNL, and at the University of Nevada.

We proceeded to contact individuals identified through NASPI, team members' professional networks, and other channels. These conversations were aimed at building awareness of the project, and at determining what messaging drives the most interest in the project. Outreach included potential sensor hosts, data contributors, vendors, researchers, and educators developing curriculum at the intersection of big data and power systems.

More recently, outreach has focused on more widespread announcements to find new audiences that we have not previously connected with. Our hypothesis is that this tactic will help us to recruit early

adopters across a range of institutions and across different subsets of the industry that we might not otherwise meet.

12/31/2020 - 100%
03/31/2021 - 100%

T3.2.4 Continue targeted partner outreach

Due Q12, Year 3

The team will continue to directly engage with identified organizations and individuals as they are discovered through outreach activities. Part of this task will be to notify the community of new content that has been created.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable

06/30/2020 - 10%

We are now formally tracking individuals that have expressed interest in the platform through webinar attendance and personal interactions with the project team, and have circulated an announcement soliciting subscriptions to our newsletter. The majority of subscribers to date are junior faculty members who requested updates about new datasets and upcoming webinars. Some subscribers also asked for more information about opportunities to contribute content or to host sensors.

We have prepared a newsletter template we can use to alert project partners to new blog posts, events, or opportunities for contributing to the platform. We have identified the need to structure the website so that different types of potential partners (e.g., sensor hosts, data providers, researchers) can easily learn about those aspects of the project that are most relevant to them.

09/30/2020 - 20%

Targeted outreach is ongoing. As discussed in Task 3.2.3 above, we have begun to explore opportunities for more widespread outreach (rather than to targeted individuals) geared at encouraging specific individuals to reach out and contribute their ideas, vision, data, or skills to advancing the project.

12/31/2020 - 40%

This quarter we hosted two workshops, a panel session and both formal and informal talks. We attended a number of industry conferences and webinars, reaching an audience of several hundred attendees. Given the radical nature of what we are trying to do (i.e., open sourcing grid data), these broad outreach efforts are fundamental to normalizing the idea, and to articulating how contributing data, hosting sensors, or otherwise engaging with the project can bring value to the industry.

At each talk, we have had at least one or two successful partnerships emerge. These have allowed us to tap into new networks, identify new data sets, and engage with broader data-sharing initiatives which were already underway. Tapping into these existing initiatives puts us in a more strategic position as a bridge between industry and research institutions. To date, we have identified and engaged with

industry data sharing initiatives spearheaded by the IEEE Power Quality Data Analytics Subcommittee, and by WECC's Joint Synchronized Information Subcommittee (JSIS).

03/31/2021 - 50%
Our most successful outreach efforts this quarter have focused on soliciting data from industry partners interested to test analysis methods that Dominion has developed in-house. Moving forward, we wish to work with Dominion to build applications leveraging these methods to deliver similar insights to other customers via the platform. Targeted recruitment of project partners will allow us to validate the tools and further develop them to streamline their use and broader deployment.

We have continued to engage with the JSIS working group, and have found strong support for open data sharing initiatives. While we are working to support that work, it remains unclear whether stakeholders will be able to circumvent the legal hurdles that the effort was originally orchestrated to avoid.

06/30/2021 - 60%
This quarter we continued outreach efforts leveraging user developed analytics (e.g., from Dominion and UC Berkeley) to bring more companies to the project. These efforts allow us to showcase how the platform has enabled users from different domains to use a wide range of analytical methods to solve problems that were of interest to them.

As we have fostered deeper relationships with key thought leaders in industry, we have found that analytics most widely discussed in the industry do not necessarily solve problems that these individuals find most compelling. Rather, the most forward thinking individuals are less intrigued by existing tools than they are by the possibility of leveraging analytics to solve new problems. Targeted outreach efforts this quarter have thus focused not only on highlighting specific analytics demonstrated to date, but also on providing software and onboarding materials needed to enable new users to more quickly explore analytics that are most compelling to them.

M3.2 - Industry/academia project launch announcement
Project announcement across industry, academia, and utilities. This includes the dissemination of the initial press releases via ARPA-E and direct contact to at least three dozen select individuals and organizations of interest. This also includes outreach to potential project partners.

9/30/2019 - 50%
We have announced the project at numerous conferences as described elsewhere in this report and reached out to at least a dozen different individuals concerning participation in this project. We are learning what aspects of the project most resonate with individuals.

12/31/2019 - 100% - Completed
Please see Milestone Completion report

03/31/2021 - 100% - Completed

We are working on updating our project announcement to reflect current goals and achievements to date.

T3.3 - Create, update, and market the project blog

Due Q10, Year 3

As much of the standard reporting documentation as possible required by ARPA-E will be pushed to a project and technical blog with the aim of posting new content at least once every two weeks (26 times per year). Further, this blog will be the place that we detail our efforts and the discussions and findings of the project.

9/30/2019 - 20%

The NI4AI blog has been launched at the following URL - https://blog.ni4ai.org/post/ - with content posted.

12/31/2019 - 30%

The blog has been advertised on the LinkedIn page for PingThings as well as broadcast via email.

03/31/2020 - 35%

We continue to update the blog on our desired schedule and advertise it via naturally occurring opportunities.

06/30/2020 - 40%

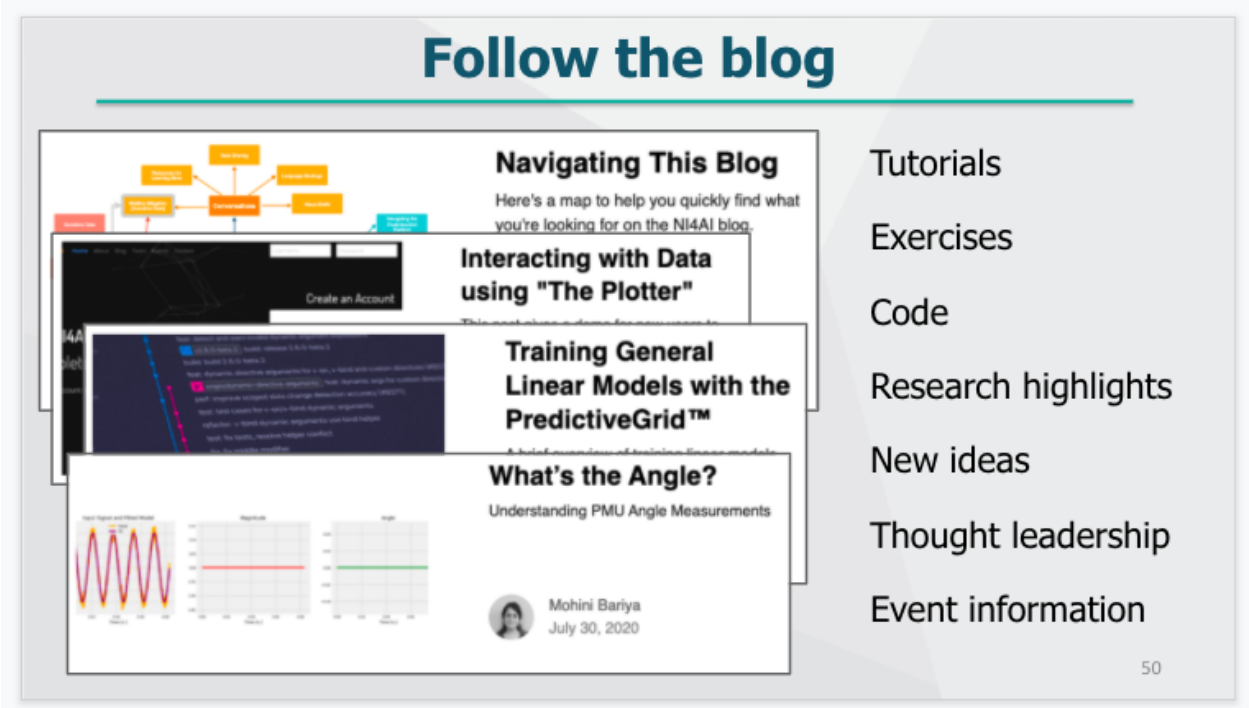
We continue to update the blog and advertise it via naturally occurring opportunities.

We have established a newsletter template and are compiling a mailing list to support more structured outreach, including a newsletter we plan to circulate at regular intervals. The newsletter will direct the community to new blog posts, datasets, and other content. The newsletter will also allow us to advertise events and publicize work that project partners are doing.

In addition to blog posts we are looking at marketing the project launch by getting media attention from more formal outlets. We have been in communication with editorial staff at UtilityDive and T&D World.

09/30/2020 - See updates below

We have continued to update the blog. We developed a suite of exercises to support a recent workshop, and present work that is also hosted on the blog. We directed attendees to these resources during the event. The screenshot below features a slide about the blog from our Oct. 21 CIGRE workshop.



12/31/2020 - 50%

We have continued to update the “datasets” post to provide an up-to-date record of the datasets we host, as well as the “blog map” to reflect recent additions. This quarter we have written several new posts listed below.

03/31/2021 - 70%

See task updates below.

06/30/2021 - 80%

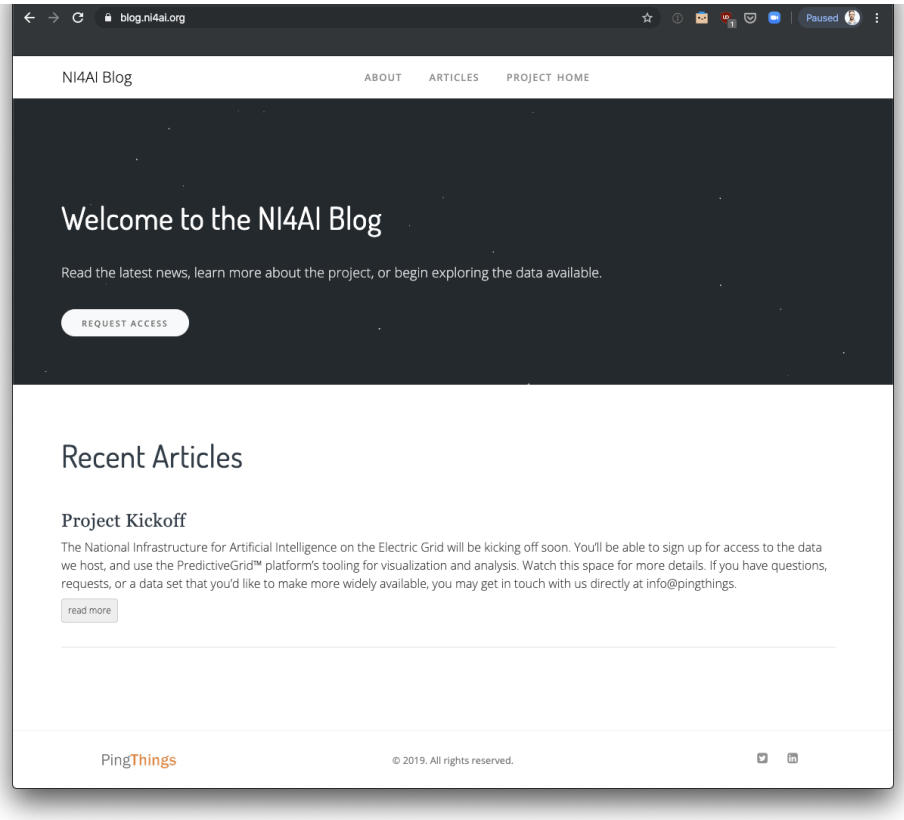
See task updates below.

M3.3.1 Project blog launched

Successful design, creation, and structuring of the project blog with at least one initial post.

9/30/2019 - 100%

The NI4AI blog has been launched at the following URL - https://blog.ni4ai.org/post/ - with content posted. We may revisit the overall graphic design of the blog at a later date.



M3.3.2 Year 1 blog content creation completed

Due Q4, Year 1

Successfully created and posted 20 blogs over the course of the year, with roughly even spacing throughout the calendar.

9/30/2019 - 5%

An initial post has been written and pushed to the blog.

12/31/2019 - 20%

Four total blog posts have been written and posted online.

3/31/2020 - 55%

11 total blog posts have been written and posted online.

06/30/2020 - 75%

Four new blog posts were posted online during this reporting period including:

- Exploring EKG Data in the PredictiveGrid
- Voltage Sag Safari
- NI4AI Webinar
- A Brief Walkthrough of the Sunshine Data

We have a backlog of blog posts to publish from our interns this summer that will get posted in early August.

09/30/2020 - 100%
See milestone report.

M3.3.3 - Year 2 blog content creation completed
Due Q8, Year 2
Successfully created and posted 20 blogs over the course of the year, with roughly even spacing throughout the calendar.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 45%
We currently have 29 posts, the first 20 of which are summarized in the M3.3.2 milestone report. We have started to recruit guest authors to contribute to the blog, and have two posts currently underway.

12/31/2020 - 70%
We are halfway through year 2 and have generated 34 blog posts to date. Additions this quarter include a discussion on phasor calculation (see "What's in a Phasor"), an analysis of the frequency deviation that occurred during the Blue Cut Fire incident (see "Blue Cut Fire"), and a tutorial on how to compute decompose phasors into positive, negative, and zero sequence vectors (see "Symmetrical Components").

We continue to use the posts "NI4AI Data Collections" and "Blog Map" to help new users navigate the data and content that is available to them. These posts are periodically updated to reflect the latest additions to the blog.

03/31/2021 - 80%
We have written four new blog posts (total: 38). These are in response to questions that have come up during stakeholder discussions. They cover:

- A practical guide to PMU placement
- Spectral analysis (parts 1 and 2)
- Phase angle differencing

06/30/2021 - 95%
We have written five new blog posts this quarter (total: 43). These will be posted in the coming weeks. They include:

- Spectral analysis of point-on-wave data
- Guide to data quality assessment tools

- SGSMA workshop materials
- Application based guide to microPMU siting
- When to use windows vs. aligned windows queries

M3.3.4 - Year 3 blog content creation completed
Due Q12, Year 3
Successfully created and posted 20 blogs over the course of the year, with roughly even spacing throughout the calendar.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - Not Applicable
03/31/2021 - Not Applicable
06/30/2021 - Not Applicable

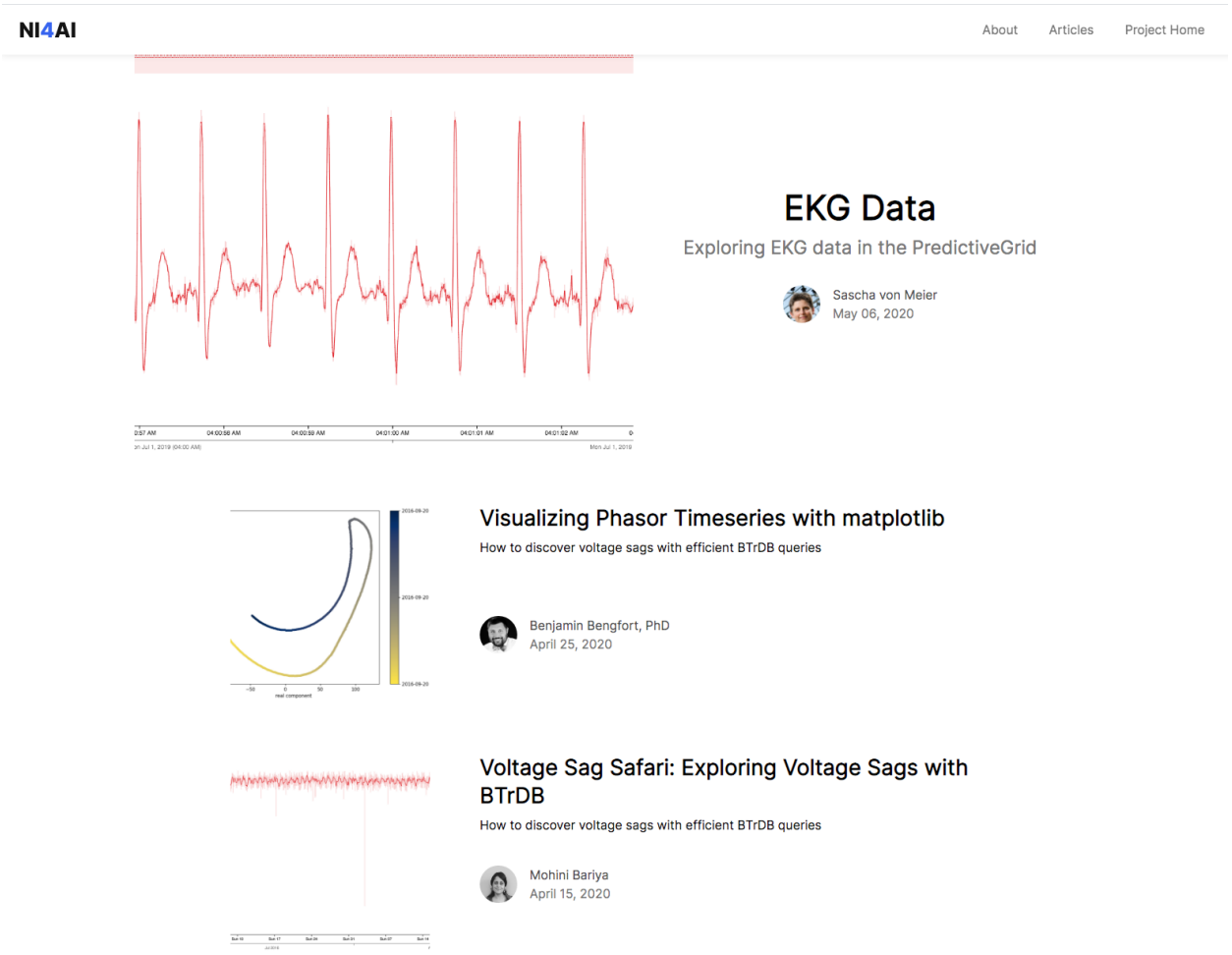
T3.4 - Create and update project website
Due Q12, Year 3
The project website will serve as a central repository of information and orchestration for the project. Further, it will provide easy navigation and discovery of the content that will be created in this project.

9/30/2019 - Not Applicable

12/31/2019 - 25%
The website has been built and launched. Please see the Milestone 3.4 Completion report for more information. We will continue to update the website throughout the project.

03/31/2020 - 30%
We added a conference page to the website that listed details and links to over a hundred different industry events.

06/30/2020 - 50%
We completely redesigned the blog and updated the underlying technology stack used in its creation. A screenshot of the front page of the blog with its new design language is shown below.

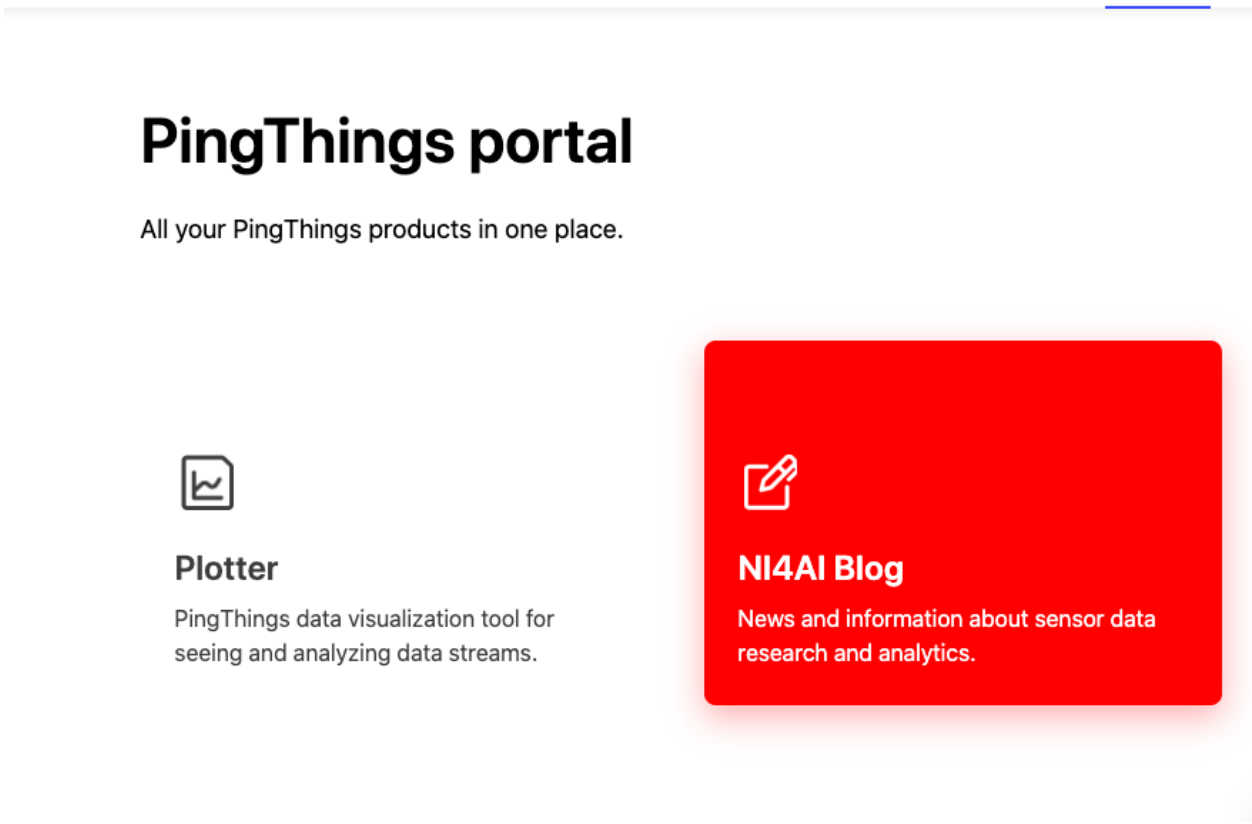


09/30/2020 - 60%

We have written two posts for the blog which are indexed so as to always appear first (other posts appear chronologically). These provide a high-level overview of content on the blog, and a detailed description of datasets we host and how to find them in the platform.

12/31/2020 -70%

We have deployed a beta version of a "portal" (see screenshot below) for directing users to different project offerings. As this feature is developed, it will include links to launch jupyterhub, and to access training materials on our github repository. We may also use the portal to advertise data competitions and workshops we host. We are exploring various aspects of the project that would be beneficial to highlight here.



03/31/2021 - 80%
Where datasets were previously listed on the blog, we have now deployed a “datasets” tab on the website to make this information much easier to find. See https://ni4ai.org/datasets (modeled after https://www.kaggle.com/datasets).

We have also increased the capabilities of the portal, which now includes links to the datasets page, API documentation, workshop videos, events page, blog, github tutorials, demo jupyter notebooks, and a user’s “profile” page where they can retrieve their API key.

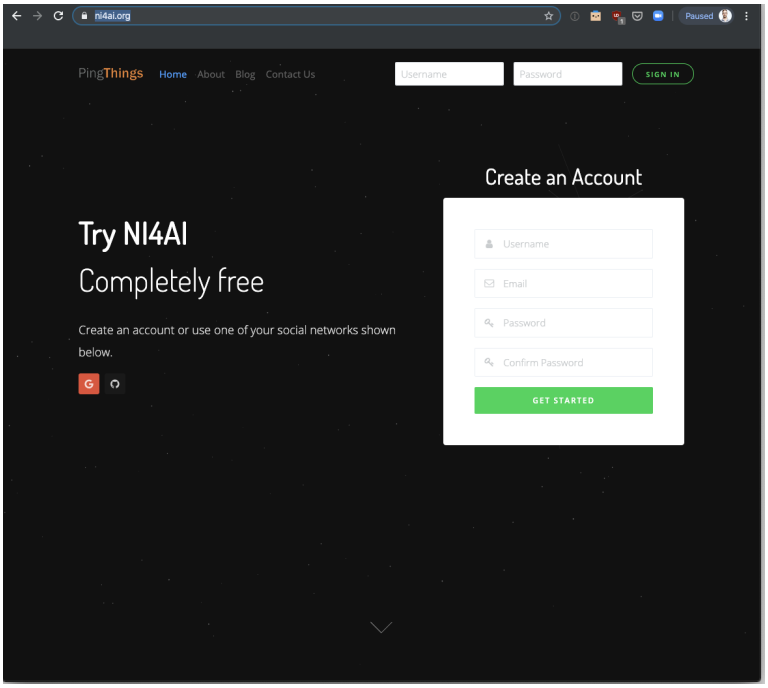
06/30/2021 - 90%
This performance period we have continued work on the Portal feature, and have developed a front-end interface for adding ingresses for new sensors (Task 1.4).

M3.4 - Project website launched
Due Q2, Year 1
Project website designed, built, and launched, complete with a project overview section and a team page.

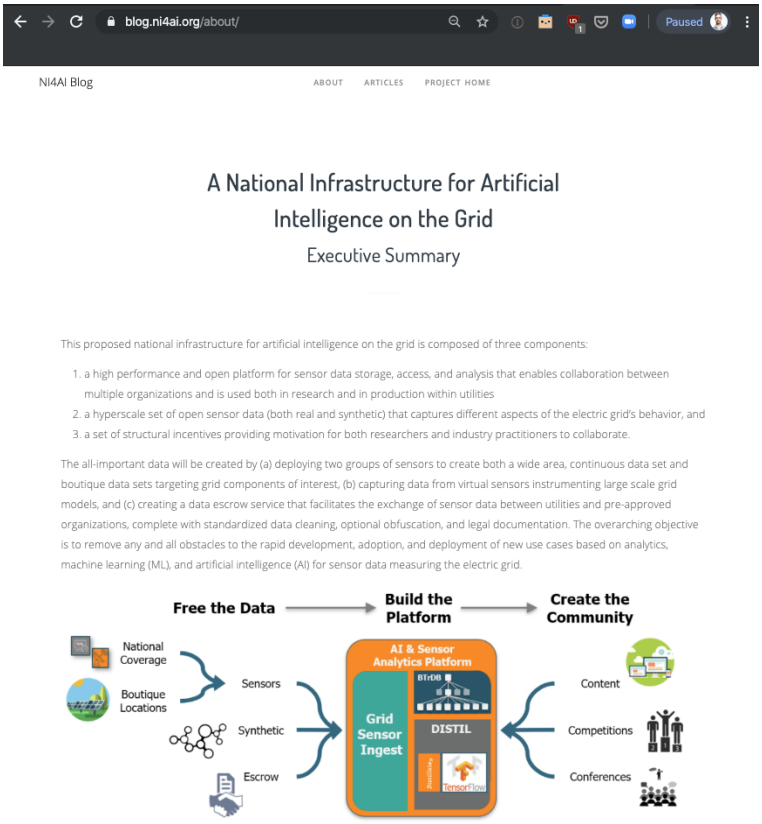
9/30/2019 - 100%
We have built a website that serves two main purposes; it provides:

1. Content, including the blog, to the community
2. Access to the demo platform and, eventually, the production platform

Additional content will be added and new functionality may be added over time.



Website landing page



Project overview page

T3.5 - Create platform documentation and training material

Due Q12, Year 3

Adequate documentation and training material is an essential ingredient to building a community around a particular technology platform. This content will include a diverse array of material.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - 20%

We have created documentation for bindings and published code on the blog to demonstrate their use.

In addition to blog posts, we have designed a 4-hour workshop training a broad audience on how to use the data and the platform. We submitted a workshop proposal to several conferences (see screenshot below), and delivered the first of these on Oct 21 & Oct 28. Attendees earned continuing education credits for attending the course.

More detail is provided in the workshop description below, and in the milestone report which will be submitted in the upcoming quarter (as the workshop was after the end of the performance period on 09/30/2020).

Workshop materials were recorded, and will be posted on the blog.

12/31/2020 - 70%

We have launched a github repository and have a content plan that will streamline the process of onboarding new users, and highlight analytical capabilities of the platform. We have generated several jupyter notebooks illustrating how to access the API and use the database bindings (see "Tutorials"), and how to integrate these into machine learning and power engineering workflows (see "Analytics"). Since our quarterly review, we revisited our content plan to address an identified need to provide a more comprehensive beginner's guide suitable for users who may be new to Python. Notebooks highlighted in blue are on the roadmap but are still under development. The repository is accessible here: https://github.com/PingThingsIO/ni4ai-notebooks

Tutorials

1. Finding collections & streams
2. Working with streams
3. Working with StreamSets
4. Basic plots
5. Working with StatPoints
6. Memory Efficient Queries
7. Exploring Sunshine data

Analytics

8. Voltage sag detection
9. Voltage change detection
10. Symmetrical components
11. *Small signal analysis*
12. *Time series analysis*
13. *Frequency event detection*
14. *Oscillation detection*
15. *Load modeling*

03/31/2021 - 100%

This performance period we hosted another project workshop, using content developed from previous workshops. We also had a proposal accepted to host a workshop at SGSMA. For this we will present new materials detailing analytical methods and data analysis workflows suitable for a more academic audience.

We have also released a new jupyter notebook about small signal analysis.

06/30/2021 - 100%

We have continued to develop new content based on identified needs and interests of teams that are evaluating the platform. This quarter, we released a new jupyter notebook providing guidance on benchmarking the platform. The notebook highlights various contributors to latency, limitations of various benchmarking strategies, and opportunities to improve performance.

T3.5.1 - Create API documentation

Due Q6, Year 2

The team will strive to produce API documentation that uses - https://stripe.com/docs - as an aspirational model.

9/30/2019 - 5%

This task is coupled to T1.8. Please see details there.

12/31/2019 - 5%

03/31/2020 - 5%

06/30/2020 - 5%

09/30/2020 - 90%

Documentation has been developed for Python and Go bindings, as detailed in Task 1.

12/31/2020 - 100%

See milestone report for Task 1.8.

T3.5.2 - Aggregate the best external educational content

Due Q6, Year 2

The power and utility engineering community would be greatly enhanced with additional awareness to the advances in other fields. As part of this project, we will aggregate and facilitate the diffusion of good content from the machine learning and AI research areas.

9/30/2019 - Not Applicable

12/31/2019 - Not Applicable

03/31/2020 - Not Applicable

06/30/2020 - Not Applicable

09/30/2020 - 90%

We have started a blog post titled “expertise for experts” which synthesizes reports, webinars, and other materials relevant to teaching users to work with PMU data. We have found limited content about PMU data targeted at new data analysts, and recently hosted a workshop designed to create some of this content. The workshop proposal was well received by the organizing committee, attendees, and received attention/attendance from prominent researchers and educators in the PMU space. The workshop was just recently completed (Oct 28), and a recording of the event which will be posted on our blog.

We have also been communicating with educators (including Sascha von Meier, among others) about repurposing video content developed during Fall 2020 for remote teaching purposes. We are beginning to form a list of educators and outreach materials to encourage others to use the platform as an educational tool in curriculum or to support student projects/research during the Spring 2021 semester.

12/31/2020 - 90%

Our investigation of user needs regarding AI/ML library integration demonstrates that work on this task will be key to exposing these features to users. Providing links to high-quality tutorials for users to learn

how to leverage these tools effectively will allow us to focus on demonstrating their integration into industry-relevant workflows.

03/31/2021 - 100%
We will add more content as the project continues, starting with content that we have found to be helpful in our own internal data science work flows.

T3.5.3 - Create 3 Jupyter Notebooks

Due Q6, Year 2

Jupyter Notebooks are the lingua franca of the analytics and data science communities and serve as the perfect medium to capture explanations, images, and executable code in one easy to interpret form. The team will create Jupyter Notebooks to demonstrate: (1) the basic use of the platform for reading and writing data; (2) exploring data; (3) performing fundamental power engineering computations; (4) and training machine learning algorithms

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - 10%
Several of our blog posts include Python code that readers can copy to replicate analytics that were done. Our NASPI Webinar included a live-coding demonstration in a Jupyter notebook corresponding to the blog post "Voltage Sag Safari". We are collecting jupyter notebooks to make available once we have a suitable platform for code-sharing.

09/30/2020 - 50%
We have created two jupyter notebooks for voltage sag detection and for analyzing frequency disturbances. We also have notebooks under development for spectral analysis, topology estimation, and data quality assessment. We plan to create a github repository for hosting these workshops.

In addition to developing our own notebooks, we have posted (on our blog) a series of exercises designed to familiarize new users with the data and with the platform. These offer exploratory data analysis exercises that they can use to build intuition around the data and common data analytics that are done, without having to imagine questions or use cases for themselves.

12/31/2020 - 100%
We have created a github repository which currently includes 10 jupyter notebooks. Of these, three feature analytical use cases for the data.

Github: https://github.com/PingThingsIO/ni4ai-notebooks

03/31/2021 - 100%
This performance period we developed a new jupyter notebook on spectral analysis.

While the notebooks developed to date offer a useful demonstration of how to use python libraries to work with data in the platform, discussions with potential partners have told us that their organizations are not necessarily in a position to be developing the analytics themselves. This finding suggests that it may be worthwhile to allocate more internal staff towards developing analytics or towards re-purposing analytics that other collaborators (e.g., Dominion or Berkeley) have developed in order to deliver insights to data providers. This will allow us to further develop and pilot new analytics which could ultimately become applications that run on top of the core platform.

T3.5.4 - Create video content

Due Q6, Year 2

Video content is an excellent way of bringing in new members to the community and the team plans on making video content to explain the platform and the project and its goals.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - 66%

The 90 minute long NASPI workshop focused on the NI4AI project held in the spring of 2020 was recorded. This event featured a number of different speakers talking about different aspects of the NI4AI project and related work, including a live-coding demonstration walking through a jupyter notebook posted on our blog.

We also created a 90 minute talk on opportunities for applying AI to the power industry, and the relevance of big data and the platform that we built to handle this problem. We delivered the presentation to an audience at a company called POWER Engineers, and are refining more condensed versions of the talk targeted towards different types of audiences that can be delivered by different members of the project team.

09/30/2020 - 80%

We developed a 4-hour workshop including seven talks between 20 and 45 minutes. These talks are described in detail below. The session was recorded to allow us to repurpose the materials for other virtual tutorials, and to host the content on our blog.

12/31/2020 - 100%

We have generated about three hours of video content, which are summarized in a blog post titled “Workshop on PMU data analysis”. We will add to or update these videos as we improve this content for future workshops. As we come to learn more about the interests and needs of the user community, we also plan to create additional posts bringing in new material, external resources, or synthesizing existing material in new ways to help users more easily find material they are looking for.

Blog: https://blog.ni4ai.org/post/2020-10-31-workshops/

Youtube: https://www.youtube.com/playlist?list=PLqSksWmTnkQ64OW0zreAvMIYefARkfrdT

03/31/2021 - 100%
This performance period we hosted a workshop for ISGT which requested pre-recorded talks. We re-used several of the talks recorded during our Fall CIGRE workshop.

We are currently working to organize another workshop at the end of May which will include all new content.

06/30/2021 - 100%
This performance period we recorded 2.5 hours of new video content during our SGSMA workshop. This included new analytics on spectral analysis, time series clustering, and case studies using small signal analysis to detect and diagnose control issues for inverter-based resources and other nonlinear loads.

M3.5 - Platform content created
Due Q6, Year 2
The team has created at least 3 videos totaling 90 minutes in length, a dozen different Jupyter Notebooks for the platform, and comprehensive API documentation.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - 50%
The 90 minute long NASPI workshop focused on the NI4AI project held in the spring of 2020 was recorded. This event featured a number of different speakers talking about different aspects of the NI4AI project and related work, including a live-coding demonstration walking through a jupyter notebook posted on our blog.

We have also created a 90+ minute talk on AI and its applicability to the power industry and the relevance of big data and the platform that we built to handle this problem. We are currently refining this talk.

09/30/2020 - 70%
See updates for Tasks 3.5.1 through 3.5.4 above.

12/31/2020 - 95%
Videos and API documentation are complete. We are working to polish off the jupyter notebooks currently posted on our github, and to develop the two additional notebooks that remain.

03/31/2021 - 100%
See milestone report.

T3.6 - Establish project presence at relevant conferences
Due Q10, Year 3

Subtasks 3.6 and 3.7 are part of a larger goal of maximal in-person exposure for the project and project personnel interaction and engagement to build the community and understand existing and future data and analytics needs.

12/31/2019 - 25%

We have announced and discussed this project at the following events:

- Spring 2019 NASPI Working Group Meeting
- International Union of Geodesy and Geophysics in Montreal, Canada
- Fall 2019 NASPI Working Group Meeting in Richmond, VA
- IEEE SGSMA 2019 in College Station, TX
- IEEE PES General Meeting in Atlanta, GA
- Defense Tech Connect event in Washington, DC
- ARPA-E Innovation Summit 2019 in Denver, CO
- CIGRE Grid of the Future 2019 in Atlanta, GA

03/31/2020 - 30%

We have announced and discussed this project at the following events:

- Clemson Power Systems Conference, Clemson, SC
- ARPA-E Open Grid Data Kick Off, New Orleans, LA
- NSF Workshop: Forging Connections between Machine Learning, Data Science, & Power Systems, Alexandria, VA

06/30/2020 - 60%

We have discussed the project at the following events:

- EPRI - Electrification - Virtual
- Spring 2020 North American Synchrophasor Initiative Working Group Meeting - Virtual
- NREL Industry Growth Forum

09/30/2020 - 80%

We have continued to attend conferences virtually, and have announced the project at the following events:

- AEG Resilience Cities (Chicago)
- IEEE PES General Meeting
- CELI EmPOWER20

12/31/2020 - 100%

Conferences attended this quarter include:

- NASPI Fall working group meeting (presenter & sponsor)
- NERC Synchronized Measurement Subcommittee (attendee)
- WECC Join Synchronized Information Subcommittee meeting (attendee)
- IEEE Power Quality Data Analytics Working Group Meeting (presenter)

03/31/2021 - 100%

Conferences attended this quarter include:

- NASPI Spring working group meeting
- IEEE ISGT

06/30/2021 - 100%

Conferences attended this quarter include:

- JSIS Spring working group meeting (panelist)
- IEEE SGSMA (tutorial)
- ARPA-E Summit (exhibitor)
- PACWorld (presenter) [PAC: Protection Automation and Control]

T3.6.1 - Select conferences to maximize growth in community and overall cost

Due Q12, Year 3

9/30/2019 - 20%

We have attended a number of conferences described above and are trying to determine the impact and output of the individuals met at each. Part of the function of attending conferences is to continue to engage with key individuals and growing relationships. Another function of these conferences is the opportunity to meet new people who may be beneficial. Certain events are more likely to have previously “unencountered” individuals and groups than others.

12/31/2019 - 30%

We have attended additional conferences through the fall of 2019 and are keeping a detailed list of conferences coming up in 2020, particularly the first 6 months of the year, to identify additional conferences to potentially attend and host a booth.

03/31/2020 - 35%

Back to the proverbially drawing board given the pandemic, which, on the bright side, has forced us to think outside of the box, aggressively targeting virtual conferences that can allow us to attend multiple “events” on the same day.

06/30/2020 - 60%

In the new post pandemic world, we have tried a number of different, virtual approaches including direct-to-a-company presentation, presentations to individuals in the community, presentations to academic groups, and presentations to broader audiences via conferences.

We have (or plan to) establish an NI4AI presence by hosting talks, tutorials or panel sessions at conferences listed on the project website. Key conferences include: IEEE SmartGridConn, IEEE PS General Meeting, CIGRE Grid of the Future, IEEE ISGT, and NASPI. We will continue to build this list as new opportunities come to light.

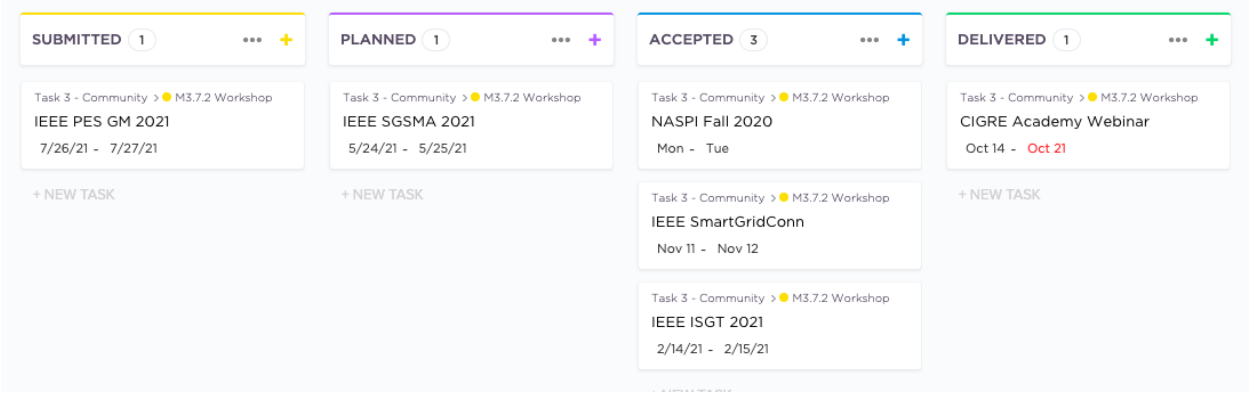
09/30/2020 - 60%

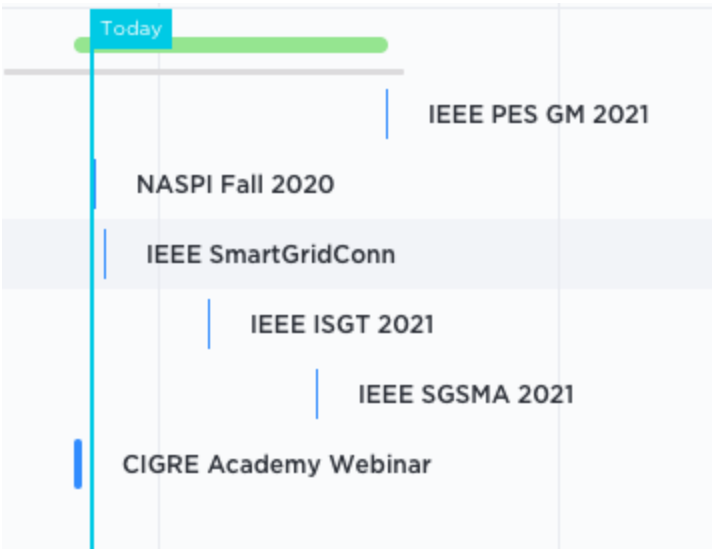
Despite the pandemic, we have maintained an active presence at relevant conferences. We submitted workshop proposals to several conferences, and have engaged with conference organizers. In spite of CIGRE Grid of the Future's cancellation, we were able to deliver the same content in a modified format as part of a fall CIGRE Academy session which granted continuing education credit for attendees.

Planned, scheduled, and past workshops:

- CIGRE Grid of the Future 2020 (workshop delivered)
- IEEE PES GM 2021 (submitted)
- IEEE PES ISGT 2021 (accepted)
- IEEE SGSMA 2021 (planned submission)
- NASPI Fall WG Meeting (proposal accepted)

Current status and schedule of workshops in project management platform. After this week's workshop we are well ahead of schedule for workshop attendance, and expect to hit our workshop quota for the project well ahead of schedule. Given that material delivered during virtual workshops persists via youtube, we are considering focusing Year 3 on coding challenges and hackathons instead of workshops.





12/31/2020 - 70%
In the last quarter, we hosted two workshops and gave a project update at the Fall NASPI working group meeting. Given travel restrictions, cost is not a limiting factor in attending conferences or hosting events. However, the effectiveness and utility of participating in remote events is variable. Community engagement depends on pre-covid cohesiveness of the group, the configuration of the conferencing platform, and current events.

Hosting events through CIGRE and IEEE has allowed us to capitalize on the advertising capabilities of each organization. We have found, however, that attendance tends to be limited to pre-existing members. This poses a barrier to participation for new entrants to the space. As the project advances, we may want to consider hosting our own events to make it easier for a broader audience to register and join.

03/31/2021 - 80%
We have seen successful partnerships emerge since our participation in the JSIS working group meeting last Fall, and will continue to engage with that group as well as their national counterpart: the NERC Synchronized Measurement Subcommittee. We have had fruitful conversations with an international ISO after one of their data scientists attended a workshop we hosted. We also plan to interface more extensively with ARPA-E projects and partners during Summit.

06/30/2021 - 100%
We have continued to maintain a virtual presence at conferences identified to date, though we have found virtual attendance to be less fruitful for identifying partners and building relationships.

T3.6.2 - Determine most cost effective, available method of impact
Due Q11, Year 3
There are many ways to engage with a conference, each requiring different levels of resource expenditures including attending, giving a talk, participating on a panel, or sponsoring a booth.

9/30/2019 - Not Applicable

12/31/2019 - 25%
We have made good progress in understanding the effectiveness of conference presentations versus booths. While booths cost more money upfront, they appear to be much more effective in initiating conversations with potential collaborators from industry, academia, and utilities.

03/31/2020 - 35%
We are now beginning to test the efficacy of virtual meetings and events.

06/30/2020 - 60%
In the new post pandemic world, we have tried a number of different, virtual approaches including direct-to-a-company presentation, presentations to individuals in the community, presentations to

academic groups, and presentations to broader audiences via conferences. The current metric of effectiveness is the number of emails and other outreach that the event generates.

Something to note as we transition to virtual -- there seems to be a balance of impact between attending an in-person conference and a virtual event. The in-person conference affords not only the opportunity to present directly to an audience but also to have very random, unexpected conversations before and after the talk. Of course, this opportunity comes at a high cost of travel and time consumed. Virtual presentations do not create the opportunity for random connections to form through chance hallway encounters or by sitting next to someone at another talk. Another disadvantage of the virtual format is that there is no way for attendees to approach the speaker afterwards. We will need to find new ways of capturing that type of immediate interaction.

Virtual presentations also present several new opportunities. [1] They greatly minimize the time and resources consumed, allowing for greater participation among NI4AI project partners. The same benefits are true for attendees -- and strategically promoting events may allow us to attract audience members that would not otherwise attend. [2] Virtual presentations often result in new content generated and potentially captured in a re-usable format. [3] They afford the possibility to track, research, and potentially follow up with those in attendance, and chat windows allow a broader cross-section of audience members to pose questions. [4] We are finding that sending targeted follow-up emails may allow us to engage more strategically with audience members than could readily be done at live conferences.

Areas for improvement include:

- Need to create the virtual equivalent of hanging around after a talk to speak directly with the presenter (engagement at this point is top of mind)
- Systematic way of tracking contacts (CRM)
- Better lighting for video presentations
- Better audio for video presentations
- Official screen capture software for when the organization hosting the event does not capture the talk

09/30/2020 - 70%

We have explored three main outreach channels: PowerGlobe, targeted emails, and broadly disseminated email campaigns. We saw a spike in web traffic and user account creation after two announcements sent to the PowerGlobe listserv, and plan to continue announcing the project via that channel as there are new developments to report. We have also found it effective to tap into existing industry networks including CIGRE and NASPI, and have seen engagement after events hosted with each organization. Broad email campaigns have been less effective than outreach to individuals who we know, through targeted outreach to individuals is time consuming and may preempt us from identifying partners that we would not otherwise be aware of.

We will continue to engage with platform users at UC Berkeley, Dominion, and other collaborators to give talks about their work.

12/31/2020 - 80%
In the past we have sponsored a booth at NASPI, which we did again for the working group meeting in November. This gave us a 30 minute speaking slot during which time we gave an overview of the company, and an update about the project. The talk was well received and led to several follow-up discussions with members of the community. One attendee was even inspired to launch an initiative to anonymize and share event data. We are following the effort closely, and plan to ingest the data when it becomes available.

03/31/2021 - 90%
We again sponsored the NASPI virtual working group meeting, and will be hosting a booth at ARPA-E summit. This performance period we also sponsored a booth at a UC Berkeley energy conference, which proved to be a useful networking opportunity -- both in terms of hiring and project partnerships.

06/30/2021 - 100%
This performance period we participated in virtual conferences in several capacities. Here, we detail the advantages and disadvantages of each.

Panelist: Participating as a panelist involved relatively less preparation, enabling us to develop new content targeted at the particular audience. We find that attendance in virtual panel sessions is relatively high, as rapid fire presentations enable broad exposure to work/ideas without a substantial time commitment to being on zoom.

Tutorial organizer: We have found tutorial participation to be highly variable, depending on the conference schedule (i.e., does the tutorial lead or follow the main conference session), advertising materials (i.e., whether the tutorial is advertised as part of the main agenda) and fee structure (i.e., whether tutorial registration requires an additional fee).

Sponsorship: Sponsorship brings brand visibility and often includes an alloted speaking time. We have continued to sponsor NASPI working group meetings, and will consider sponsorship as a mechanism for engaging with new communities when in-person events re-commence.

Exhibitor: At in-person events, the exhibition hall is a space where conference attendees go to seek out exhibitors that interest them, or simply to pass time in between other events. This creates a natural opportunity for building relationships. Attendance in virtual exhibition booths depends highly on the virtual conference platform that is used, and on the degree to which exhibition time is integrated into the conference agenda. We look forward to exhibiting at conferences once in-person events recommence.

<u>T3.6.3 - Attend and impact the community</u>
Due Q12, Year 3

9/30/2019 - Not Applicable

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - 20%
We hosted a NASPI webinar to announce the project to strategic industry partners. The panel was hosted by Sascha von Meier and included an introductory overview to the project (by Sean Murphy), two talks by research partners about use cases they have developed (by Luigi Vanfretti and Mohini Bariya), and a talk about running a utility data analytics program (by Kevin Jones). The session was well received and will serve as a model we can replicate at future conferences.

Given that the audience for this particular talk (NASPI) already has high analytical maturity, this conference was largely informational. Iterating on this model, we have proposed a more "training" oriented talk geared at recruiting users and familiarizing them with how to develop their own analytics using the platform.

09/30/2020 - 30%
We have continued to attend conferences and host workshops, as discussed above. We have also begun a broader outreach campaign providing thought leadership via industry news outlets. These efforts are geared at building up support in the industry for aspects of the project which are more innovative and new.

We are also finding some success framing the project as an educational initiative geared at promoting curriculum and course projects that will help to differentiate students and early career professionals as they enter the job market or begin their careers.

12/31/2020 - 50%
We have continued to attend conferences and events to build awareness about the project. We have also found it useful to include a short overview describing how open access data are used, and the merits of making such data available.

We have made several connections through these conferences which have helped us to find new data sets, and to influence thought leaders in the industry. The content of our talks, and our broad outreach strategy may need to evolve as our focus moves from data acquisition to engaging with data analysts.

03/31/2021 - 60%
Outreach efforts have begun to focus less on broad announcements and more on targeted information gathering. We are working to better understand how we can position the project to support analytical use case development that will broker a safe and reliable renewables transition. This includes interfacing with a much broader cross section of the energy sector. This change in tactic suggests that broad outreach is less critical at this stage in the project. Rather, we are focused on understanding the needs of specific organizations so that we can demonstrate the value of the data and platform. Once the value is demonstrated, we will again focus on socializing these success stories at conferences.

06/30/2021 - 70%

Outreach efforts this quarter have continued to focus on establishing deeper relationships with key industry thought leaders. Discussions have focused on a much broader range of partnership models -- including platform pilots with historical, streaming, synthetic, anonymized, or open access data. These different models allow us to cater the details of data acquisition efforts to the capabilities and requirements of a much wider range of organizations wishing to explore the capabilities of the platform.

M3.6.1 - Conference attended
Due Q2, Year 1
9/30/2019 - 100%
We have announced and discussed this project at the following events:

- Spring 2019 NASPI Working Group Meeting
- International Union of Geodesy and Geophysics in Montreal, Canada
- IEEE SGSMA 2019 in College Station, TX
- IEEE PES General Meeting in Atlanta, GA
- ARPA-E Innovation Summit 2019 in Denver, CO

M3.6.2 - Conference attended
Due Q4, Year 1
12/31/2019 - 100%
We have announced and discussed this project at the following events:

- CIGRE Grid of the Future 2019 in Atlanta, GA
- Fall 2019 NASPI Working Group Meeting in Richmond, VA
- Defense Tech Connect event in Washington, DC

M3.6.3 - Conference attended
Due Q6, Year 2

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - 50%

- ARPA-E Grid Kick Off Meeting, New Orleans, LA
- Clemson Power Systems Conference, Clemson, SC
- NSF Workshop: Forging Connections between Machine Learning, Data Science, & Power Systems, Alexandria, VA

06/30/2020 - 100%
We have discussed the project at the following events:

- EPRI - Electrification - Virtual
- Spring 2020 North American Synchrophasor Initiative Working Group Meeting - Virtual

- NREL Industry Growth Forum

Please note that we are technically 6-months ahead of where we need to be, by design. We started announcing and discussing and presenting about the project at various workshops and conferences after the award had been announced but before the project officially kicked off. Relevant conferences are clustered in the spring and in the fall and we started presenting in the spring of 2019.

Given the pandemic and the likelihood of conferences remaining virtual for the rest of 2020 and into 2021, I question whether "conferences attended" is the best milestone name or if "conferences" is the best metric or proxy for quantifying the project efforts at broadcasting the message. A different, more direct metric, may be better. Examples include the number of individuals messaged or the number of presentations given.

M3.6.4 - Conference attended
Due Q8, Year 2

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - 20%
This quarter we attended the following industry meetings and conferences:

- NASPI Fall working group meeting
- IEEE SmargGridComm
- NERC Synchronized Measurement Subcommittee meeting
- WECC Joint Synchronized Information Subcommittee

03/31/2021 - 100%
This quarter we attended the following industry meetings and conferences:

- IEEE ISGT
- NASPI Spring working group meeting
- IEEE PES Joint technical subcommittee meeting
- GA Tech Gault and Disturbance Analysis Conference

06/30/2021 - 100%
This quarter we attended the following industry meetings and conferences:

- IEEE SGSMA
- JSIS Spring working group meeting
- PACWorld
- ARPA-E Summit

M3.6.5 - Conference attended
Due Q10, Year 3

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - Not Applicable
03/31/2021 - Not Applicable
06/30/2021 - Not Applicable

M3.6.6 - Conference attended
Due Q12, Year 3

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - Not Applicable
06/30/2020 - Not Applicable
09/30/2020 - Not Applicable
12/31/2020 - Not Applicable
03/31/2021 - Not Applicable
06/30/2021 - Not Applicable

T3.7 - Plan and host workshops associated with relevant conferences
Due Q12, Year 3
The utility industry is driven by personal networks and face-to-face meetings. Hosting in-person events, focused on the National Infrastructure project, will be one of the best ways to build and reinforce the community. Given the resources and focus of this project, the team does not seek to start an entirely new conference but to leverage existing large events that attract different segments of the community that is to be built. The NASPI working group meeting, the CIGRE Grid of the Future, and the IEEE Smart Grid Synchronized Measurements and Analytics are three potential target conferences. Additionally, these project-focused workshops could be held at utilities or at ISO's or other organizations.

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable

03/31/2020 - 10%
A workshop opportunity has arisen due to the change from in-person to virtual that we are working to exploit.

06/30/2020 - 33%

We delivered a virtual workshop at the Spring NASPI 2020.

We are adapting a similar workshop model for delivery at future conferences. We had a tutorial accepted at SmartGridConn 2020, and are submitting similar proposals to CIGRE Grid of the Future 2020, IEEE PES General Meeting 2021. We have also been asked to deliver a webinar to IEEE Young Professionals, and are exploring similar webinar series targeting industry professionals. We hope to take advantage of the virtual presentation format to do more frequent conferences and to reach a more diverse audience.

09/30/2020 - 66%
See task updates below.

12/31/2020 - 80%
See task updates below.

03/31/2021 - 90%
See task updates below.

06/30/2021 - 100%
See task updates below.

T3.7.1 - Evaluate available conferences
Due Q12, Year 3

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable

03/31/2020 - 25%
Given the move to virtual events, we are re-evaluating our previously held beliefs about which conferences will be best suited for a workshop.

06/30/2020 - 33%
We have put together an agenda for a 3-hour tutorial that we will deliver at SmartGridConn 2020. We have connected with conference organizers around hosting similar sessions at CIGRE Grid of the Future 2020 and IEEE PES General Meeting 2021.

In light of the pandemic, virtual presentations could give us the opportunity to engage with targeted groups such as local chapters of professional development/networking organizations. Reaching these audiences could present an opportunity to connect with early-career professionals and engage with the companies they work for. We are adapting existing presentations to deliver focused sessions targeting different types of professional groups. We have already reached out to IEEE Young Professionals in Energy, and plan to organize webinars with CELI, CIGRE NGN, and other similar organizations. We plan to cast a broad net targeting professional and volunteer organizations focused on energy and AI.

09/30/2020 - 60%
The current performance period has focused on developing content for conferences that were already planned. We have not submitted proposals to new conferences. We are finding that remote conference attendance is not as effective at building new connections as in-person conferences would be, though we have continued to attend conferences virtually as it creates a natural opportunity to stay engaged with broad professional networks. We have also used these events as an opportunity for building new connections though barriers to establishing new relationships are considerably higher than they would previously have been.

12/31/2020 - 70%
While we are still soliciting ideas and strategic support from the industry, we have also come to better understand existing needs and to identify new stakeholder groups. We started the project with targeted outreach to organizations and individuals where we knew we would find strong support. As the project progresses, we are expanding the scope of conferences we attend in order to reach a broader audience. We have already begun this process, and are attending new conferences and industry working group meetings recommended to us by project partners who engage with different cross sections of the industry.

03/31/2021 - 70%
We have found virtual conferences to be less fruitful for partnership building. While we have maintained a presence at conferences we participated in in the past, we have not been actively seeking new virtual conferences to attend.

06/30/2021 - 80%
We have continued to maintain a presence at virtual conferences. We are using these as a platform for disseminating information about analytics that have been demonstrated in the platform, educating the community on big data workflows, and providing thought leadership regarding the capabilities of cloud technologies.

T3.7.2 - Reach out and connect with conference organizers for approval
Due Q12, Year 3

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable

03/31/2020 - 33%
We have reached out to the NASPI leadership to host a workshop at the Spring 2020 NASPI working group meeting.

06/30/2020 - 33%
We have reached out to the organizers of SmartGridConn and GoTF to host tutorial sessions and received positive reviews from both organizations. We have a tutorial session confirmed for SmartGridConn which will take place virtually in November 2020.

09/30/2020 - 50%
Our planned workshop schedule, and current status of submissions is detailed above under task 3.6. To date, our proposed workshop agenda has been well received by both conference organizers and attendees. We will continue to iterate on the material as deemed appropriate.

12/31/2020 - 60%
After connecting with several members of the IEEE power quality data analytics subcommittee, we reached out to Tom Cooke (EPRI) about securing a time to speak about the project at the IEEE Joint Technical Subcommittee meeting. We gave a 30-minute talk at the subcommittee meeting in January about the company, platform and project.

03/31/2021 - 70%
Our SGSMA workshop proposal was accepted.

06/30/2021 - 80%
We have reached out to the organizers of CIGRE Grid of the Future, and plan to submit a workshop proposal for their conference this Fall. It was recently announced that the conference would be a hybrid event. We hope to attend in-person.

T3.7.3 - Develop agenda for event
Due Q12, Year 3

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - 33%
We have developed an agenda for our first proposed workshop.

06/30/2020 - 33%
Based on our experience hosting the NASPI workshop, we have developed a workshop/tutorial agenda that can be repeated or expanded upon for longer sessions. Shorter sessions will focus on outlining the vision of the project and highlighting use cases that NI4AI users have developed. We may choose to highlight use cases based primarily on academic merit or operational value to utilities, depending on the target audience. Longer sessions will include hands-on tutorials where attendees will learn how to access the API and may begin developing analytical use cases for themselves.

We have proposed an agenda for a 3-hour workshop we will host at SmartGridConn 2020. The tutorial will aim to achieve the following goals:

1. Educate attendees on the contents of PMU data
2. Get attendees to access the API on their personal computers through interactive exercises
3. Describe use cases developed to date, recruiting NI4AI users to present their work
4. Motivate attendees to begin using the platform in their own work

The agenda is broken into 20-30 minute sessions. These sessions can be removed or substituted depending on the expertise of the audience. The specific use cases we choose to highlight could make the session suitable for a broad audience, for a research audience, or for an audience of industry practitioners.

09/30/2020 - 66%

We developed content based on the agenda described last performance period with some modifications. We realized that interactive exercises designed to use the API during the event were impractical due to logistic complications, and the possibility that attendees would be unable to access from behind utility firewalls. Instead, we brought on platform users to step through jupyter notebook workflows and to provide tips and tricks for efficient analytics such as parallelization and leveraging the structure of BTrDB to develop computationally efficient analytics.

The workshop material was well received by both workshop organizers and attendees. We plan to re-use the content and will consider revisiting the agenda or developing new content as needed. The workshop agenda for the CIGRE Academy workshop is pictured below.

12/31/2020 - 90%

See milestone report for M3.7.2. We will continue to revise the workshop agenda as our understanding of user needs and interests continues to evolve.

03/31/2021 - 100%

The agenda for our SGSMA workshop will feature talks by Dominion about their small signal analysis toolkit and by UC Berkeley about distribution PMU use cases. We will also give a brief overview of the project and of resources that are available to participants.

06/30/2021 - 100%
Our CIGRE tutorial last year offered an introduction to PMU data analytics. This year, we will include additional content on developing scaleabing data analysis workflows. This will likely draw from prior workshops about event detection and spectral analysis, highlighting differentiating capabilities of the platform for handling large-scale queries and for parallelizing workflows.

In addition to public-facing workshops, we have begun to refine our user materials to facilitate the process of onboarding new users/organizations engaged in pilots. These materials include a checklist of workflows that users may go through -- such as new user/group creation, addition of metadata, etc. -- in order to get the most out of the platform. This includes demonstrations of features such as user access controls, geospatial data visualization, and event triggers.

T3.7.4 - Develop content required for the event
Due Q4, Year 3

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - 15%
Content development for the first event is under development by all speakers.

06/30/2020 - 33%
We are developing a model for hosting webinars that builds on the same format as the NASPI event hosted in April. Content from that event can be re-used in future events. Berkeley has also developed an educational overview of synchrophasor measurements, and we are developing short presentations to accompany use cases highlighted on the blog. In addition to these, tutorials with adequate time (2-4 hours) will include a live coding demonstration and interactive exercises that will get attendees to begin engaging with the NI4AI blog and API during the session.

09/30/2020 - 66%
Content required for the event was developed as described in Task 3.7.3. Slides and video recordings of the event are hosted on the CIGRE website.

https://cigre-usnc.org/cigre-academy-webinars/

12/31/2020 - 90%
See milestone report for M3.7.2. We may continue to revise workshop content as our understanding of the needs of the industry continues to evolve.

03/31/2021 - 90%

Content for our SGSMA workshop is still under development.

06/30/2021 - 100%
See milestone report for M3.7.3 for a summary of new content presented this performance period. As discussed under task updates above, we are continuing to develop and refine content to enable new and potential customers to get the most out of the platform and out of resources that are available to them in the cloud.

T3.7.5 - Marketing and outreach for event
Due Q4, Year 3

9/30/2019 - Not Applicable
12/31/2019 - Not Applicable
03/31/2020 - 15%
Marketing for the NASPI workshop is in progress via email and phone calls.

06/30/2020 - 33%
We are compiling an email list based on subscriptions, personal communications, and attendance at past NI4AI events. We have a newsletter template we can use to alert our user community to upcoming events. We plan to announce events more broadly by leveraging PowerGlobe and newsletters for other professional organizations (e.g., IEEE YPE).

09/30/2020 - 66%
We marketed the event via various channels including PowerGlobe, email campaigns, personal communications, and dissemination of workshop announcements via other existing channels (e.g., the NASPI, CIGRE, and PSERC email lists).

12/31/2020 - 70%
Most of the marketing for the event occurred in the previous performance period, during the months leading up to the workshop. We will continue to use a similar marketing strategy to promote upcoming workshops and other events.

03/31/2021 - 80%
We will begin socializing via PowerGlobe and other channels the registration deadline approaches.

06/30/2021 - 100%
We have continued to socialize upcoming events on our website and blog, as well as other outreach channels.

T3.7.6 - Film presentation made for content generation purposes
Due Q12, Year 3

9/30/2019 - Not Applicable

12/31/2019 - Not Applicable
03/31/2020 - Not Applicable.

06/30/2020 - 33%
A recording of our NASPI webinar is hosted on the NASPI website and on the NI4AI blog.
https://blog.ni4ai.org/post/2020-04-07-ni4ai-webinar/

09/30/2020 - 66%
The CIGRE workshop was recorded and will be hosted on the CIGRE website. The workshop is also being made available through IEEE SmartGridComm. Once the IEEE conference comes to a close, we plan to post the materials on our blog.
https://cigre-usnc.org/cigre-academy-webinars/

12/31/2020 - 90%
Workshop materials were recorded. We have posted them on our youtube channel and blog.

Blog: https://blog.ni4ai.org/post/2020-10-31-workshops/
Youtube: https://www.youtube.com/playlist?list=PLqSksWmTnkQ64OW0zreAvMIYefARkfrdT

03/31/2021 - 90%
Though we did host a workshop this performance period, we used material filmed during a previous workshop. We will generate new material during the upcoming performance period.

06/30/2021 - 100%
This performance period we recorded 2.5 hours of new content during our SGSMA tutorial.

M3.7.1 - Project-focused workshop hosted
Due Q4, Year 1

12/31/2019 - Not applicable
03/30/2020 - Not applicable
06/30/2020 - 100%
See milestone report.

M3.7.2 - Project-focused workshop hosted
Due Q8, Year 2

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 100%
See milestone report.

M3.7.3 - Project-focused workshop hosted
Due Q12, Year 3

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - Not applicable
03/31/2021 - 30%
Our last workshop deliverable will be met with the SGSMA workshop in May. From there we may continue to host workshops either re-using content or developing new content as necessary.

06/30/2021 - 100%
See milestone report.

T3.8 - Plan and host data science competitions/data hackathons for the national infrastructure
The original Netflix Prize and Kaggle have demonstrated that providing open data sets and open questions to a community of researchers with a specific goal and performance metrics is a great way to build temporary communities. Thus, data science competitions provide a nearly perfect mechanism to blend the many aspects of this project. Initial thinking is to have competitions around the following three topics: (1) wide area competition; (2) point of wave competition; and (3) solar farm competition However, the topics above are only suggestions and will be changed depending on available sensors, data sets, and industry and university interests. Further, it is unlikely that there will be enough new data for a competition in year 1 so the team's focus is year 2 and year 3.

As a key goal is the development of new use cases for data, there is interest in structuring a data science competition that is more open ended and directed toward the discovery of applications. It is also possible that the team will seek partnerships with utilities and universities to host local events.
Due Q12, Year 3

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 10%
See task update below.

03/31/2021 - 10%
Data science competitions were originally designed to get students to begin using the platform and developing use cases which would ultimately translate to a vision they could bring to future employers. This was a mechanism for training young engineers to become agents of change for the industry.

We delayed our first data competition for several reasons:

1. Participation in multi-day virtual events is nearly impossible given current levels of “Zoom fatigue”
2. Remote hackathons don’t offer the same fun/engaging experience
3. We had previously submitted an application to host a workshop which happened in February; given that this coincided with a concerted effort to meet M2.3.1 (a go/no-go milestone) on time, we decided to focus attention on other tasks
4. An imminent change in strategy and focus would allow us to better leverage data competitions going forward

We wish to focus data competitions to support in-house development of analytical tools to address the needs of partner organizations, while also lifting the requirement that all data provided be made openly accessible to users. This will allow us to allow various project partners to “pilot” the platform for internal use, while also supplying the company with data to support use case development.

Use cases will build on work Dominion has done and on the work of research collaborators (e.g., UC Berkeley, or others), depending on the needs and requirements of partner organizations. Applications will be deployed on partner data (likely using a Jupyter notebook initially). After generating results, we will work with partner organizations to determine how they might act upon them. Here, we will recruit industry experts to issue recommendations and guidance where appropriate.

We believe more targeted use case development (rather than broad, open ended data competitions) will increase the success of sensor deployments in delivering value to partner organizations.

06/30/2021 - 20%
The original intent of hosting data competitions was to foster participation from a much broader community of analysts around developing new analytics to advance the industry. What we have learned to date, is that for many of the questions the industry is interested to explore in their data do not require methodological advancements to solve. Rather, what is needed is the development of algorithmic tools that are transparent, easy to use, and adaptable to a wide range of analytical inquiries. A key barrier to widespread adoption of AI/ML tools is that domain experts in power engineering often lack skills in computer science and data science that would be needed to scale up their workflows. Furthermore, insiders stress the importance of interpretability and transparency. Industry practitioners will not deploy a tool unless they fully understand how it works.

We have delayed this task in hopes that it will be possible to host it in person. The key advantage there is that it will enable us to engage more fully with participants. Data science competitions will focus on pair programming with new and existing users to better understand how they work with their data and what use cases they are most interested in. This model will focus heavily on building relationships with our users, and identifying analytical methods which need to be integrated into platform onboarding materials and jupyter notebooks. This hands-on engagement with new and existing users will further

help us to understand where there are skill gaps that prevent users from doing analytics at scale, and how we can focus platform enhancements (Task 1.13) to make this much easier.

T3.8.1 - Host an online data science competition
Due Q6, Year 2

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 30%
We have explored a couple different models for the first data competition. Given the pandemic, we do not expect a 2-day "hackathon" style event to solicit much participation. Instead, we intend to host an event that will run for a longer period of time.

Industry partners have identified several open research areas which coding challenges could focus on. Leveraging these ideas poses a compelling opportunity to recruit industry participation and support. Several ideas have been posed:

- Advanced phasor calculation techniques that reduce (or report) information loss
- Frequency event detection
- Frequency calculation
- Oscillation detection
- Fault event classification and labeling
- Synthesis of PMU and POW data

In many of these cases, the questions that have been posed do not have clear answers. For example, there is no such thing as a "true" phasor which we can use to evaluate phasor calculation techniques. Thus they do not necessarily lend themselves to a Kaggle-style competition. Instead, we plan to host a hackathon where we propose several project ideas and allow participants to work on open ended questions of their own choosing. We plan to circulate an announcement to faculty who teach relevant coursework in AI/ML, power systems, and signal processing.

The goal of the first data competition will be to vet our process for hosting competitions to ensure that future competitions are a success. Future competitions will focus on soliciting more widespread participation.

03/31/2021 - 30%
See task 3.8 update above.

06/30/2021 - 50%
This first data competition has focused on scheduling dedicated time to work with users to get started with writing code needed to explore various use cases in the data. These have focused more on workflows -- e.g., for exploratory data analysis, code performance, and more fully leveraging capabilities

of the cloud -- than on any specific analysis question or task. We have found, rather, that the interests of individual organizations are widely variable. These interests stem from regulatory pressure, emerging local risks (e.g., inverter dynamics, intermittent generation, wildfires, etc.), internal priorities, or funded projects. In short, we have found that the primary limitation in extracting value from data is not the analysis methods, but establishing data analysis workflows themselves.

Data competitions will focus on gathering requirements and sharing targeted insights into how users may most effectively use the platform to enable new use case discovery.

T3.8.2 - Host an online data science competition
Due Q9, Year 3

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - Not applicable
03/31/2021 - Not applicable
06/30/2021 - Not applicable

T3.8.3 - Host an online data science competition
Due Q12, Year 3

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - Not applicable
03/31/2021 - Not applicable
06/30/2021 - Not applicable

M3.8.1 - Data science competition or hackathon hosted
Due Q6, Year 2

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - 30%
See task update above.

03/31/2021 - 30%
See task 3.8 update above.

06/30/2021 - 50%
See task 3.8 update above.

M3.8.2 - Data science competition or hackathon hosted
Due Q1, Year 3

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - Not applicable
03/31/2021 - Not applicable
06/30/2021 - Not applicable

M3.8.3 - Data science competition or hackathon hosted
Due Q4, Year 3

12/31/2019 - Not applicable
03/31/2020 - Not applicable
06/30/2020 - Not applicable
09/30/2020 - Not applicable
12/31/2020 - Not applicable
03/31/2021 - Not applicable
06/30/2021 - Not applicable

Milestone 1.7 - Data quality assessments developed, tested, and deployed on the platform

Development and testing of a suite of data quality assessment algorithms including at least zero data, null data, and repeating data tests and then deployment on the live operational platform. Demonstration of data quality assessment tools to ARPA-E.

Introduction

Data quality issues create a bottleneck in any data analysis pipeline. Issues are inevitable in streaming data captured by sensors monitoring the behavior of real world assets where communications dropouts may introduce gaps in time series. They may also arise when measurement conditions,such as temperature, voltage, current, deviate from design conditions. Sensors may report meaningless values when a signal is too weak to measure, when the sensor itself is misconfigured, or when assumptions underlying measurements prove to be false.

This milestone focused on implementing a suite of functions that generate new streams to indicate if and when a particular data quality issue arises.

Implementation

We began work on this milestone by developing a library of seven data quality checks implemented in Python. The table below lists checks developed, with bold text indicating checks that were selected for implementation on ingest.

Num	Check	Definition
1	**Repeat measurements**	The same value is reported repeatedly
2	**Gap detection**	Sensor experienced gaps in data reporting
3	**Duplicate timestamps**	Stream reports multiple values with the same time stamp
4	Deviations from expected point density	Data streaming deviates from expected sample rate
5	**Zero values**	Values are reported to be equal to zero

6	Deviations from nominal voltage	Reported values are above/below expected bounds
7	Signal-to-noise ratio	Signal to noise ratio is higher than expected
8	Timestamp jitter	Timestamps deviate from expected time resolution/frequency
9	Out-of-range values	Values reported are clearly outside of acceptable range

The four checks selected for fulfillment of this milestone were chosen based on issues observed in the data to date and based on their broad relevance across domains beyond the power sector. Checks that would require user-defined inputs (e.g., specifying the expected sample frequency) or user interventions (e.g., by changing the nominal voltage reported in metadata) were excluded because the functional form of calculations performed on ingest cannot easily be modified by users at this point.

Once functions were developed and tested using Python, they were then implemented using DISTIL. DISTIL is a data processing tool that enables arbitrary Go code to be run on streams in the platform. Data quality checks output a binary flag indicating where a particular data quality check has failed and write the results back to the database. Given that we assume most streams will be of reasonable data quality, this means that the data quality streams will be sparse. The platform has a very efficient storage engine for sparse streams so that the storage requirements for the platform will likely not dramatically increase. Data quality streams are linked to their parent streams both by collection name and via metadata.

Gap detection was implemented differently than other checks, as it did not lend itself to implementation in DISTIL. DISTIL uses data versioning to identify data insertions and other changes, and triggers calculation (or updating) of distilled streams when changes are detected. When a gap occurs, it does not lead to a change in the database, and thus does not trigger DISTIL to perform calculations on chunks of time where data are missing. Gaps, however, can easily be detected from aggregates already stored in the data (namely using the `count`).

Platform integration

Data visualization interface

Under this initial deployment, data quality streams are exposed to users via our data visualization interface in the same manner as other streams. Data quality streams are linked to their parent stream both via metadata and via the collection name, both of which can be used as filtering criteria for selecting streams. Ultimately, we envision that the presentation of data quality streams could differ from the presentation of other streams. For example, one option considered

is to create a “toggle” which would enable a user to pull in data quality visualizations for streams they have selected, without the need to select individual data quality streams.

In the coming months, we plan to experiment with different ways of interacting with data quality streams both internally and with customers to better understand how these data are used and how they can be presented in a way that will be intuitive to users.

Python API

We have developed an extension to our Python API that enables users to easily search for data quality streams, and determine whether issues are present. This functionality is currently stored in a library called BTrDB Extras, where experimental features and capabilities are deployed. This allows us to capitalize on the features internally and to provide access to key super users without exposing them broadly until we fully understand what functionality is needed.

The current implementation includes the following functionality:

- Defines new `DQStream` and `DQStreamSet` objects giving added functionality to traditional `Stream` and `StreamSet` objects.
- Includes `DQStream.list_distillates()` function which lists data quality checks and other streams derived from a given parent stream using DISTIL.
- `DQStream.contains_event(“zeros”)` reports a boolean value indicating whether a stream contains a particular type of data quality issue. The function also includes an optional keyword specifying start and end times of interest.
- `DQStream.contains_any_event()` reports a boolean value indicating whether any data quality issues are present.
- `DQStreamSet.describe()` provides a list of data quality checks (and other distillates) which are active on a given stream.

Examples

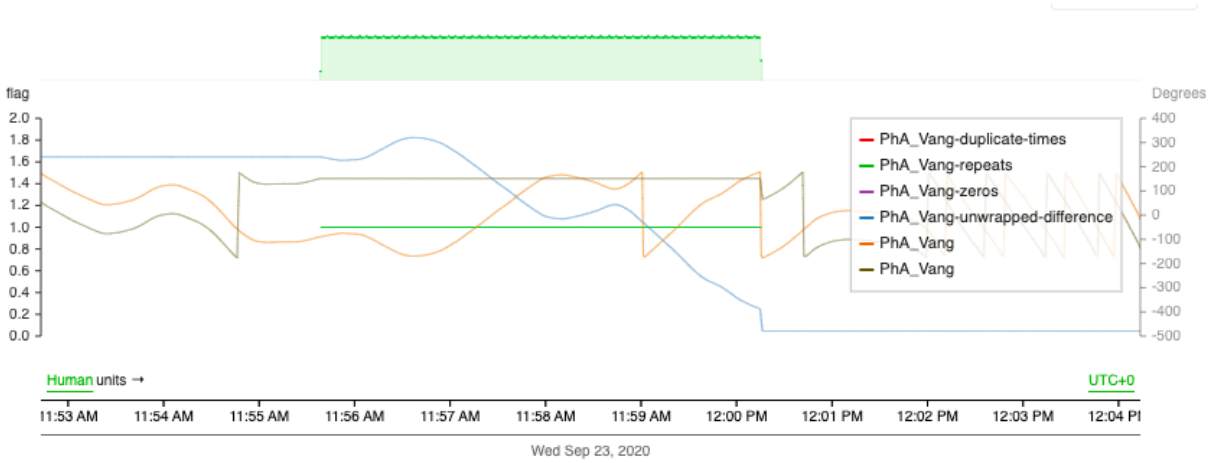
Repeat Values

Figure 1 below shows an example of a data quality issue that has been flagged. The orange and brown lines labeled “VA_ang” show phase angle measurements from two different PMUs. The blue line labeled “VA-unwrapped-difference” shows the phase angle difference between the two streams. Phase angle differencing is a necessary step in processing phase angle data for many applications, and can enable visual detection of topology changes, stability issues, and other events which would be difficult to detect from raw phase angle data.

Phase angle differences typically follow a stationary distribution unless there is a change in the system. After performing the calculation on these streams, we saw a large change (>360 deg) in the phase angle difference. This was not expected, and further investigation revealed that one

of the phase angle streams had reported repeat values for some time. This led the downstream phase angle differencing calculation to report bad results.

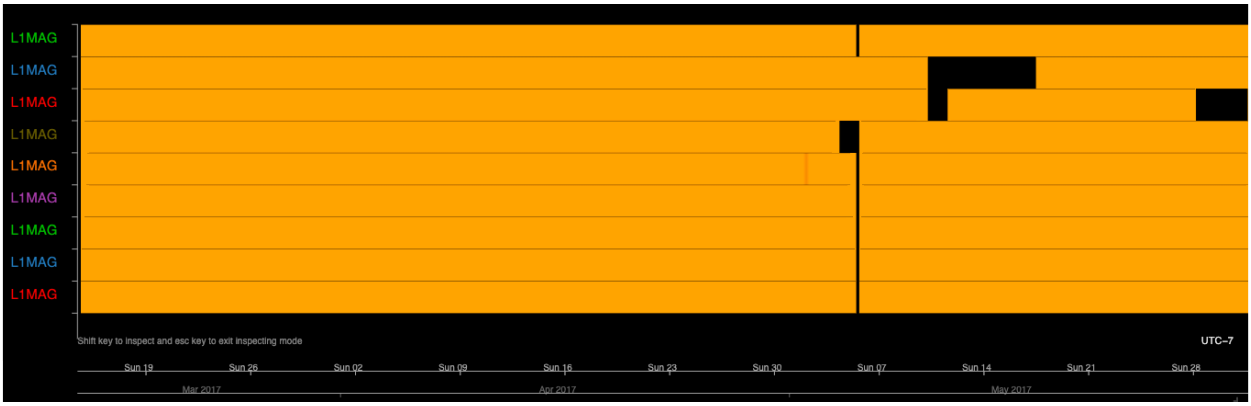
The visualization below shows data quality streams (including duplicate-times, repeats, and zeros) for the problematic stream. The green line and histogram at the top clearly indicates the presence of repeat values. This visual queue significantly streamlines the process for honing in on the presence of issues such as this. This will also enable integration of data quality streams into downstream calculations -- such as phase angle differencing -- to ensure that exceptions are handled appropriately and results are robust to any issues observed.



Gap Detection

The “golden” collection offers a rich dataset of 9 sensors deployed on a single distribution feeder. The time interval was selected from a longer deployment based on high availability of all 9 sensors in question.

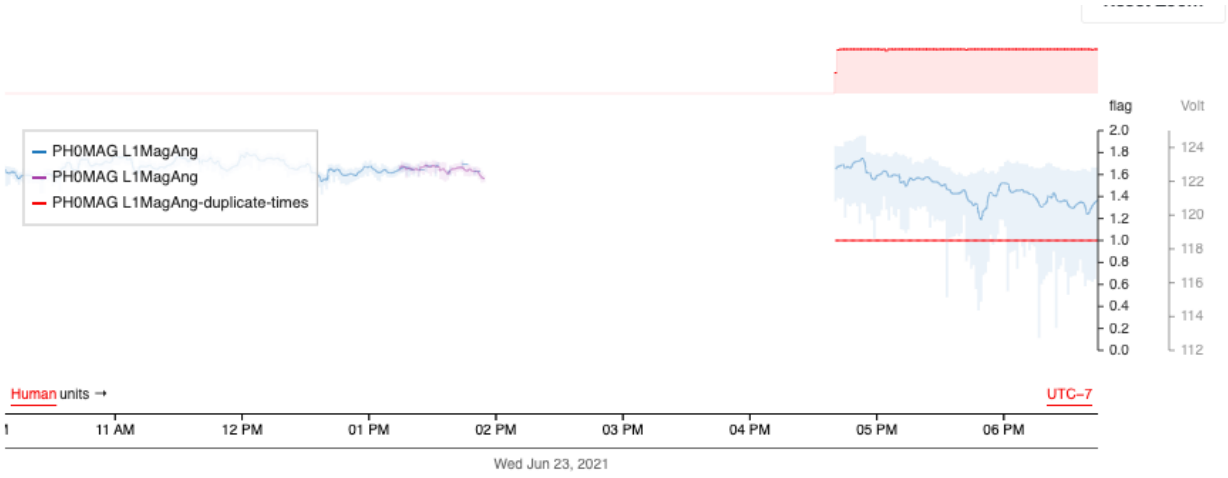
Despite its richness, the data does include gaps. The following screenshot provides a heatmap showing the number of values recorded during discrete time windows from the entire fleet of sensors (see Task 1.9.2). The clear indication where there are gaps in the data could help with sensor selection, or with selecting windows of time where data availability is high.



Duplicate timestamps

Transitioning from TCP to UDP created a new stress test for data ingestion, as we have primarily ingested data via TCP in the past. We initially transitioned two sensors for testing, and on June 23 began to update the firmware for remaining sensors. The time interval in question is shown in the figure below.

Once the firmware was updated, we saw unexpected behavior where both sensors were inserting data to the same stream, resulting in duplicate measurements for each time stamp. We found that the two streams were being transmitted to us without any differentiating information between them. Where our TCP ingestor uses the host IP address and port to poll data which could easily be sourced to a particular sensor, there was a need for additional information to determine where data sent via UDP had originated from. We updated our configuration process to address the issue, setting a unique C37 ID for each sensor so that it can be determined from the C37 data frame where the data originates from.

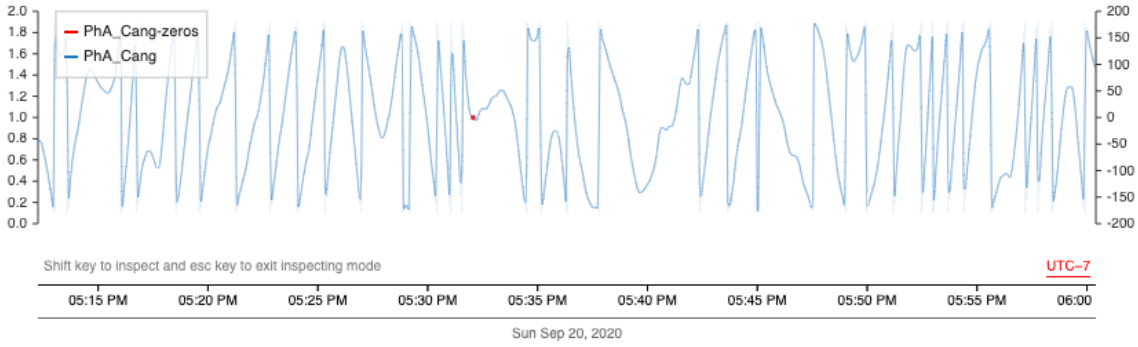


Zero Values

Our zero values data quality flag has largely been absent from data, with the exception of phase angle streams. Here, we periodically see cases where point values are reportedly zero. We do not see these zeros over sustained time intervals. One such instance is pictured below, with the zero value flag shown in red.

The figure clearly shows that zero crossings are not a data quality issue, but are in fact an artifact of the data. While voltage or current measurements would likely only equal zero when something is amiss, or when something noteworthy is happening in the data, phase angle “measurements” cross zero regularly. We might expect to see similar behavior with point on wave data.

This investigation told us that our interface for enabling data quality checks will need to be user customizable, where certain checks may be disabled for certain streams based on expected behaviors given the nature of the data in question.



Concluding remarks

Data quality assessments provide a valuable tool for users to quickly detect and locate data quality issues. This can then support more robust exception handling, data analysis, and debugging as in the examples cited above.

Now that data quality checks have been deployed, we will continue to experiment with techniques for integrating these checks into our workflows to identify simple and intuitive mechanisms for exposing this information to users as part of our core API and visualization functionality.

Downstream work on this milestone includes integration with Task 1.9 which enables sensor fleet visualization and notifications when data quality issues are observed. Fleet visualization will enable analysts to more easily focus their analysis on streams where data quality is high, while notifications will enable sensor hosts to more quickly intervene to address connectivity issues or other issues that could be causing data quality issues.

Milestone 1.10 - Platform Optimization

Evaluation and complete documentation of platform performance as a function of cost using different tiers of cloud storage and compute optimizations. This evaluation will include the use of the benchmark suite developed for M1.1. Submission of documentation to ARPA-E for PD approval. If cost effective, the platform will be upgraded with the enhancement(s).

Introduction

This milestone report describes platform optimizations geared at reducing the cost of data storage and computing. Computing optimizations were foundational to our completion of Milestone 1.7 (Data quality assessment) and provide new opportunities to support data exploration and analytics at scale. Here, we lay out work performed under this milestone.

As an extra bonus, we lay out some additional plans for ongoing work as cost-competitiveness is an absolutely critical feature for the commercial viability of the platform.

Tiered storage (storage cost optimization)

Costs in a cloud-based big data platform tend to be dominated by storage costs. Each gigabyte of data stored incurs a monthly charge. Amazon and other providers charge different fees depending on the performance characteristics of the provided storage along a number of dimensions including access latency, read- and write-speeds, durability, and replication factor. Our overall highly logical plan is to leverage lower cost storage tiers with poorer performance characteristics for less frequently accessed data.

Storage costs have been optimized by leveraging a variety of data storage tiers available in AWS. These tiers provide access to a variety of physical storage resources, enabling platform users to trade off storage cost against data access and query speeds. For data where immediate or frequent access is warranted, for example, data can be stored in SSD drives. For data that users are less likely to access (e.g., historic data), we may capitalize on lower-cost alternatives. Our initial release of tiered storage uses a 90-day horizon for fast storage; data older than 90 days is moved to “cold” storage which may result in slower query speeds but at a reduced cost per GB per month. Initial benchmarking suggests that the difference in query times speeds between different storage tiers is 10% or less, an amount that will likely be imperceptible for the vast majority of users.

Ongoing work on this task will expand flexibility in defining what data are stored in which storage tier. Currently, storage tiers are set based only on timestamp. We plan to expand this capability to enable specified collections or different types of data streams that are less readily used (e.g., STAT flags) to be transitioned either directly or more quickly to low-cost storage.

DISTIL (computing cost optimization)

Computing costs have been optimized by developing and deploying a general-purpose tool called DISTIL, which enables resource efficient calculations on streaming or historical data. DISTIL uses the inherent structure of the database to perform calculations on time-partitioned “chunks” of data. Calculations are “stateless”, meaning results are independent of data from other times. Query speed and efficiency are maximized by enforcing queries aligned (in time) with the inherent time partitioning of the database.

Performing calculations in DISTIL improves computational efficiency, facilitates parallelization, and enables us to more easily leverage “spot instances” (or unused AWS servers) to perform calculations at times when low-cost compute is available.

DISTIL helps to optimize computing cost in a number of ways:

1. Implementation in Go (the native system language of the platform) reduces computational overhead associated with queries issued via Python and other language bindings
2. DISTIL passes data in discrete chunks based on the inherent structure of the database. This enforces time-aware queries that eliminate overhead associated with naively querying the database. Once a distiller is enabled, the calculation is automatically performed over the entire time history of the stream and is updated when changes are detected -- e.g., from streaming data, changes to date, or out-of-order insertions.
3. Stateless calculations eliminate the need to pass information between calculations performed on different time intervals; this improves computational speed and allows distillate calculations to be easily parallelized.
4. Calculations are triggered using data and distillate versioning. This makes it easy to determine if/where in the data a calculation needs to be performed, or updated based on a change in the data or in the distillate code. This opens up the possibility of using AWS spot instances to perform calculations at a fraction of the cost. Spot instances provide intermittent access to low-cost computing resources at times when there is an excess of AWS server availability. Inherent data versioning enables DISTIL processes to be easily interrupted and restarted as needed. PingThings is actively pursuing this optimization strategy and expects it to be ready before the end of this project.

DISTIL was initially prototyped during the microPMU project funded under Open Innovation 2012, and was documented in an academic paper. Work under this milestone focused on building the tool into a production-ready feature, and on enabling it to run in operation. A number of distillates which can be used “out of the box” were implemented. We further developed a distiller template and have improved documentation to more readily enable users to write distillers of their own.

Concluding remarks

Both DISTIL and tiered storage allow us to more thoroughly utilize a variety of cloud offerings to optimize our costs by picking and choosing resources that are most appropriate for a given task. These resources optimize cost by trading off other dimensions of performance -- such as speed or availability -- which may not be necessary. We plan to continue work on these milestones to enable more flexibility where warranted to achieve further cost reductions. As we explore further opportunities for performance gains, we plan to revisit our benchmarking suite developed under Milestone 1.1 to evaluate the benefits and consequences of cost optimizations made.

Milestone 2.3.2 - Boutique sensor deployment

At least half the sensors for the first boutique data set have been acquired and deployed for the creation of the first boutique (site-specific) data set. Data is streaming into the platform. Demonstration of streaming data to ARPA-E.

Introduction

The core idea for this milestone - Boutique Sensor Deployment - and the associated Task 2 tasks and subtasks, was to create finite, narrowly-scoped time series data sets capturing the behavior of a local portion of the grid for the express purpose of creating value from that data for relevant stakeholders such as a utility, an ISO, a regulatory body, or even a third party. In 2018, we envisioned creating these boutique sensor data sets by deploying project-purchased sensors and collecting data. The rate limiting step in this process is obviously the acquisition and deployment of sensors.

It is important to note that the creation of value is not as simple as it seems and is at least a two step process. The first step is the obvious step: using the data to do something often not possible before. The second step, which is almost always overlooked, arises from the fact that value creation is a social construct. For something to be valuable, relevant stakeholders must perceive and believe that it is valuable and, in the case of this project, the relevant stakeholders are often utilities. Thus, it is not enough to do something new, novel, or interesting with data but the relevant stakeholders must perceive the value that is created by the use case.

We conducted a large number of stakeholder interviews to better understand analytical needs, data requirements, and partnership structures that could readily enable the development and demonstration of new use cases explored using data generated under this task. We have learned that applications and use cases derived from sensors deployed by a third party and not directly attached to utility owned equipment will not be as readily accepted by key stakeholders as applications arising from utility data.

We found that the most impactful way to provide data access via the platform would be to execute proof of concept projects with utilities and other relevant organizations, enabling them to leverage their own sensors and their own data to solve problems they care about. This approach creates more value for the industry than could be achieved deploying sensors for research applications, while also helping PingThings achieve commercial success by enabling possible customers to experience the platform. This has the added benefit of focusing project funds not on sensor acquisition, deployment, and maintenance but on new features, application and use case development for our pilot partners. Thus completion of this milestone has included a range of partnership models that go far beyond recruiting sites to deploy sensors that will stream data to us.

This milestone initially focused on identifying key locations on the grid where measurement data could enable development and exploration of high-value use cases. This milestone report details progress towards securing boutique data sets to enable the development and demonstration of use cases, while also fostering partnerships with key individuals and organizations providing thought leadership to the rest of the industry. While work on this milestone was initially geared at logistics around facilitating sensor deployments, the much larger effort to date has been geared at technology transfer and outreach (TT&O) activities.

Boutique data contributors

Sensor hosts that have contributed data (either streaming or historic) under this task are listed below, along with additional context about use cases each data contributor was interested to explore.

- **American Transmission Company (ATC)** - renewables integration, oscillation detection, and generation station monitoring
- **University of California, San Diego** - DER integration testbed including distribution PMU deployment and 1Hz building-level metering.

Additional partnerships currently moving forward (though without data having been exchanged):

- **Schneider Electric** - provided ION9000 power quality meter to stream data to the platform and enable demonstrations of analytical applications for the data; meter will be installed on-site at a single-family home outfitted with rooftop PV.
- **Kit Carson Electric Cooperative** - distribution PMU deployment to enable asset monitoring, renewables integration, and DER integration.

We are in discussion with a number of other companies including but not limited to large-scale utilities. Our milestone report for Milestone 3.0 (to be submitted next performance period) will include further details about planned partnerships, and lessons learned from those that have been ruled out.

Deployment strategies

Outreach efforts focused on exploring five main deployment strategies, each with very different benefits, and with different constraints. The advantages and disadvantages of the strategies considered are detailed below. While we continue to move forward with partnerships targeted at sensor deployment, we have also aggressively pursued partnerships geared at demonstrating value of the platform to organizations that already have their own sensor hardware deployed. Various partnership models explored to date under the project are detailed in the following section.

Discussions with possible host sites revealed a fundamental trade-off between the openness of the data and the commercial potential of the partnerships involved. We found that utilities are unwilling to provide open access to streaming data due to security concerns. Wall outlet installations eliminate these constraints, but the data generated contain measurement noise and other dynamics that may not be present farther up on the distribution or transmission grids.

The most relevant data, we determined, is data that utilities generate themselves. We have also found that many utilities interested in exploring the capabilities of distribution PMUs have already installed sensors themselves, thus eliminating the need for us to deploy sensors in order to generate the data needed to enable application development. In parallel with these discussions, we have continued to explore partnerships geared at sensor deployment in order to bring new utilities to the table to explore and demonstrate the operational value to be gained from the technology.

	Advantages	**Disadvantages**
Wall outlets	- Few restrictions on data usage / commercialization - Easy installation - Hosting sensors is a hook for community building - Creates sensor network for grid monitoring independent from the utility	- No current measurements - Voltage measurements highly influenced by load - Low observability into transmission or distribution grid dynamics - Secondary partnerships needed for TT&O
Distribution grid	- Data sharing is at the discretion of utilities - Distribution asset monitoring - Novelty of use cases - Ability to garner interest of utility practitioners - Opportunity to demonstrate use cases in operation - High research value for DER integration studies	- Policies for data sharing are inconsistent and ill-defined - Data access and sharing is governed by utility (much more limiting) - Limited institutional familiarity with PMU technology / use cases - Less appetite for data given limited precedent for real-time or high-resolution data
Transmission grid	- Long track record for exploring PMU use cases - Possibility for partnerships with key thought leaders - Strong appetite for making data available to enable new application development - Long-standing research	- Data are protected as critical infrastructure data - Efforts to create open data sets still in early stages - Practitioner mindset heavily influenced by limitations of legacy technologies - Institutional policies against data sharing / cloud hosting

	collaborations are in place - Potential partners have excellent data resources and domain expertise, but lack software tools to enable analytics at scale (i.e., good strategic fit)	persist - Data sharing terms likely to include restricted access to data, posing new requirements for project
Research testbed	- Fewer restrictions on data access - Guaranteed visibility into dynamics of interest - High research value - Possibility for researchers to deploy apps with on data from other providers	- Less engagement / buy in from industry stakeholders - Extensibility to real-world grids is unclear
Leverage pre-existing sensor deployments	- Eliminates hardware costs - Provides access to historical data - Enables partnership formation with leading innovators - Mutual alignment with partner companies - Prior investment increases depth of partnership / odds of success - Possibility for driving sales	- Increases data storage costs - Much higher bar for demonstrating “success” by delivering value to data provider - May require new feature development to advance partnerships with possible customers (Task 1.12) - Eliminates possibility of making streaming data broadly available - May require anonymization (which diminishes utility of data for some applications)

Partnership models

Type 1: Knowledge sharing.
The most basic partnership model involves simply sharing information and insights into promising use cases for sensor data. Partnerships formed under this model have enabled the project team to better understand data requirements and analytical methods needed to enable certain use cases. Where there is mutual alignment with project goals, use cases explored under this partnership model will be prototyped and shared with partners, encouraging further participation in data sharing partnerships in order to deploy the analytics on their own data.

Type 2: “Proof of concept” platform pilot.
This partnership model includes deploying an instance of the platform for pilot participants to begin exploring the capabilities of the platform for supporting use cases with their own data.

Participants are supplied with new-user training, onboarding materials, and ongoing support in their use of the platform. These partnerships enable knowledge sharing as well, providing

Type 3: Streaming sensor pilots.
This partnership model includes recruitment of host sites to stream data into the platform. Pilot participants may stream existing sensors, or may install new hardware procured under the project. These pilots are geared at demonstrating the operational value of the data, and at building institutional expertise with data exploration and analytics.

Conclusions

Completion of this milestone included connection of streaming sensor and ingestion of historic data from a number of thought leaders in industry seeking to advance the use of data analytics within their organizations. Moving forward, we will continue to foster these relationships to support requirements gathering (Task 1.12) and technology transfer and outreach targeted at marketing the platform.

Milestone 3.7.3 - Workshop Hosted

Introduction

On May 24, 2021 the project hosted its third workshop titled *Tutorial on PMU and Time Series Data Analysis* as part of the IEEE Smart Grid Synchronized Measurement and Analytics conference. The workshop ran for 2.5 hours and included 45 participants. The abstract can be found on the SGSMA website here: https://www.sgsma2021.org/tutorial-3/.

The workshop featured analysis performed by some of the most experienced users of the PredictiveGrid platform, including Dr. Mohini Bariya, Dr. Chetan Mishra, and Miles Rusch. Panelists gave “research” presentations summarizing analysis methods used to conduct analytics with practical real-world applications. They also highlighted the value of the platform in enabling rapid iteration and efficient data workflows. The talk closed with a discussion of the NI4AI project itself, highlighting data sets and code made available via the project for workshop participants to gain more familiarity with analytical methods which panelists demonstrated using relevant data sets hosted in the platform.

About SGSMA

SGSMA runs every two years and was launched in 2019 by NASPI leadership. The conference was formed out of an identified need to foster greater discussion among academic communities about synchronized measurements. Panel presentations were offered primarily by academics, though attendees included both academics and practitioners. The in-person event had been planned for Croatia, and the conference agenda remained targeted at incorporating international audiences.

Agenda

This tutorial was designed to train researchers and practitioners on data analysis methods and applications for synchrophasor and point on wave data. The course was divided into three sections, covering (1) fundamentals of synchrophasor measurement, (2) data analysis tools and techniques, and (3) practical applications in industry. The course aimed to help attendees put concepts from power engineering and data science into practice to establish efficient workflows using the platform to analyze and visualize time series data at scale.

Fundamentals synchrophasor measurement

Speaker: Alexandra von Meier, PhD
Duration: 15 min
Topic: Physical foundations for phasor measurement

The first part of the course described foundational concepts for phasor measurement units (PMUs) and synchrophasors. We covered basic concepts from physics and power engineering to provide intuition about how PMU measurements relate to physical phenomena on the grid. This served as a quick refresher of relevant concepts for attendees familiar with power engineering and PMU data, and provided a brief introduction for attendees new to the field.

Data analysis and machine learning workflows

Speakers: Mohini Bariya, PhD and Miles Rusch
Duration: 60 min
Topics: Time series clustering and spectral analysis

The second part of the course demonstrated analytics performed on synchronized measurement data from distribution systems. The focus was to describe methods and show analytics that participants could reproduce with NI4AI data, while providing code for attendees to apply and adapt to their specific purposes and needs. We discussed methods for unsupervised event detection and classification with time series clustering of voltage magnitude and frequency. These algorithms prioritize transparency and communication to the human user. Panellists demonstrated the PredictiveGrid platform in action, and highlighted how workflows in the platform helped to streamline and enable their work.

PMU applications in industry

Speakers: Chetan Mishra, PhD and Kevin Jones, PhD
Duration: 60 min
Topics: Small signal analysis for grid modernization and control

The third part of the course covered case studies from a recent initiative at Dominion Energy to uncover dynamics observed under ambient conditions using synchrophasor data. Traditional methods for integrating new resources onto the grid rely on generic models which fail to account for the reality of the dynamics that can occur in operation. Utilities often lack transparent models for characterizing the dynamic behaviors of certain devices — including solar inverters, FACTS devices, DERs, and industrial loads.

This observation motivated an initiative to use continuous monitoring data to improve visibility and inform advancements in grid integration and controller design. Findings reveal that devices are often deployed with poorly set controllers which may become unstable. Panellists described the use of signal processing techniques to alert the utility to possible issues occurring under ambient conditions which would typically be brushed off as “noise”. The tutorial covered the challenges faced when working with actual measurements as opposed to simulation data where often, the existing analysis techniques need to be adapted due to violation of underlying assumptions.

Putting it to practice with open access data

Speaker: Laurel Dunn, PhD
Duration: 15 min
Topics: PredictiveGrid platform and data made available through NI4AI

The tutorial closed with an overview of the NI4AI project and offerings. This portion of the workshop featured NI4AI data sets that were used to support analytics presented by other panelists. The talk also highlighted core capabilities of the PredictiveGrid platform for both data visualization and analytics, and workflows for rapid data exploration and application development. The talk also directed attendees to resources on our blog, youtube channel, and github repository, as well as collaboration opportunities which the project affords.

Panelists

Laurel Dunn (PingThings) recently received her PhD in Civil Engineering from UC Berkeley on risk-aware decision making. She has worked closely with industry and with regulators about putting state-of-the-art methods from statistics and data science into practice. Currently, she is working to advance the use of next generation sensors that can remove guess-work from decision making processes. She is working with a startup called PingThings to lead an initiative called NI4AI. NI4AI is an ARPA-E funded project which provides data, computational tools, and educational content to advance the use of time series data for grid applications.

Alexandra "Sascha" von Meier (University of California, Berkeley) is an Adjunct Professor in the Department of Electrical Engineering and Computer Science, where she teaches a course on Electric Power Systems. She is also Director in CIEE's Electric Grid program area focusing on power distribution systems, Smart Grid issues, and the integration of distributed and intermittent generation. Her current research projects center on the use of high-precision micro-synchrophasor measurements for situational awareness, diagnostics and control applications in distribution grids.

Kevin Jones (Dominion Energy) received his Ph.D. in Electrical Engineering from Virginia Tech as a Harry Lynde Bradley Fellow in 2013. He created and helped commercialize the open source linear state estimator and has led the development and integration of a variety of innovative technology solutions in his role at Dominion Energy including the cloud deployment of PredictiveGrid for synchrophasor analytics in Electric Transmission. He currently serves as Manager of ET Operations Engineering Support leading the Fault Analysis, Sensor Data Communication/Engineering/Analytics, Special Studies, and Web Development teams.

Chetan Mishra (Dominion Energy) is an electric transmission engineer with research interests in nonlinear dynamics and control, synchrophasors and renewable energy. He earned his PhD in Electrical Engineering from Virginia Tech in 2017 and has been with Dominion for over five years.

Mohini Bariya (University of California, Berkeley) was a PhD candidate (now graduate) focusing on the use of novel, high-resolution measurements for improved situational awareness in the electric grid. She has worked extensively with real PMU datasets. She has experience teaching science and engineering concepts to different audiences, including as a graduate student instructor for the Electric Power Systems course at Berkeley.

Miles Rusch (University of California, Berkeley) is a PhD student in Electrical Engineering at UC Berkeley. His research involves using unsupervised learning algorithms to perform data analysis on electric power systems.

Conclusions

The workshop was very well received and created new content targeted at a much more academic audience. Highlighting research efforts underway at UC Berkeley and Dominion created a natural mechanism for workshop attendees to identify potential collaboration opportunities with other organizations who are engaged with the project. Discussion of specific case studies from Dominion further helped to highlight the need in the industry for robust research tools to begin delivering operational value to utilities. Workshop content is featured on our youtube channel.

Milestone 2.4.2 - Common NDA and DSA Released

Introduction

This milestone was initially focused on standardizing legal documents used to facilitate data sharing among companies using the platform, and researchers developing applications for the data. Collaborations such as these often require drafting non-disclosure agreements and data-sharing agreements which can be lengthy and expensive to negotiate.

While collecting legal documents under Task 2.4.1, we became aware of a parallel effort by GridBright funded under ARPA-E award DE-AR0001030. The project had already drafted a common NDA which was undergoing legal review. On April 28, 2021, GridBright released three NDAs including a group, two-way and one-way NDA. The latter is included as an attachment to this report.

Under NI4AI, we have focused less on data-sharing among collaborators than initially expected and instead on creation of open access datasets which can be made available to anyone without the need for legal contracts or agreements. To support that effort, we have released a sensor hosting agreement granting PingThings full ownership over data generated by sensors purchased under the project. Sensor hosts are granted full access to their data in the platform, including access to our data visualization interface, Python API, and other functionality. The agreement also reserves our right to revoke access, for example due to misuse of the platform. The document has undergone legal review, and is included as an attachment to this report.

Disclaimer

This Non-Disclosure Agreement (NDA) was drafted by GridBright, Inc. as part of the Department of Energy (DOE) Advanced Research Project Agency-Energy (ARPA-E) project DE-AR0001030, "Secure Grid Data Exchange (SGDX)", and with input from the Secure Grid Data Exchange Working Group (SGDX WG) of BetterGrids Foundation, Inc. Visit arpa-e.energy.gov for more information about ARPA-E, and GRIDEON.com for more information about the SGDX project.

An NDA is used to protect the disclosure of sensitive grid data to unauthorized parties, thereby reducing the risk of physical and cyberattacks on the grid. Many organizations use their own customer NDA. The motivation for creating the attached NDA is to provide a publicly available NDA that could be used by all organizations free of charge, including smaller organizations that do not have their own custom NDAs and prefer not to spend money on legal fees to create one. The team working on this NDA has made best effort to base it on best practices used across the industry, including NDAs used by several ISOs and transmission reliability organizations. If the attached NDA receives wide adoption in the utility industry, it could lead to a "standard" NDA that would become generally acceptable to all parties, saving the utility industry unnecessary legal fees and time spent in negating thousands of custom NDAs each year.

Neither GridBright nor BetterGrids Foundation claims any intellectual property rights in the NDA: GridBright and BetterGrids Foundation claim no copyright ownership in the NDA, and they explicitly declare their express intention to submit this NDA to the public domain as a document free of and unprotected by copyright. Accordingly, you may freely copy, use, and modify this NDA for all your purposes. IF YOU CHOOSE TO USE IT, YOUR ACT OF USING THE NDA SIGNIFIES THAT YOU ACKNOWLEDGE AND AGREE (A) THAT YOU ARE USING THE NDA SOLELY AT YOUR OWN RISK; (B) THAT YOU ASSUME ALL LIABILITY IN REGARD TO YOUR USE OF THE NDA; (C) THAT NEITHER GRIDBRIGHT NOR BETTERGRIDS FOUNDATION MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND ABOUT THE LEGAL SUFFICIENCY OF THE NDA; AND (D) THAT NEITHER GRIDBRIGHT NOR BETTERGRIDS FOUNDATION SHALL HAVE ANY LIABILITY OF ANY KIND WHATSOEVER EITHER TO YOU OR TO ANY PERSON IN ANY WAY RELATED TO YOUR USE OF THE NDA.

GridBright is a small business focused on Secure Grid Integration. GridBright mission is to help the electric industry keep the lights on in the face of increasing intensity of natural disasters, frequency of cyber-attacks, and penetration of renewable and distributed energy resources. Visit GridBright.com for more information on GridBright.

BetterGrids Foundation is a non-profit public research charity dedicated to the support of the BetterGrids Grid Data Repository. The Repository provides grid researchers easy access to many non-sensitive grid models. GridBright created the Repository with funding from the DOE under the ARPA-E project DE-AR000716. Visit BetterGrids.org for more information about the BetterGrids Repository and the Foundation.

The SGDX WG was created under the Technical Committee of BetterGrids to represent BetterGrids in the DOE ARPA-E SGDX project and to help create uniform and standardized processes and non-disclosure agreements (NDAs) for access to sensitive grid data by legitimate parties including grid researchers. SGDX project team members include GridBright, BetterGrids Foundation, Midcontinent independent system operator (misoenergy.org, and Mid-Carolina Electric Cooperative (www.mcecoop.com).

More than 100 organizations in the utility industry have voluntarily contributed time and input to the research activities referenced above.

ONE WAY NON-DISCLOSURE AGREEMENT

This Non-Disclosure Agreement (***"Agreement"***) is entered by and between the signatories of this agreement on _____ ("***Effective Date***"), individually referred to as party, and collectively as parties.

WHEREAS, Disclosing Party is prepared to disclose confidential information under this Agreement to Receiving Party in connection with Disclosing Party's business with, or possible engagement of, Receiving Party (the "Purpose");

WHEREAS, Receiving Party represents that it desires to receive confidential information pursuant to this Agreement;

NOW THEREFORE, in consideration of the mutual promises, covenants, representations and agreements contained in this Agreement and other good and valuable consideration, the receipt and sufficiency of which are hereby acknowledged, the parties agree to the following:

i) standard terms in Schedule A (***Standard Terms***),
ii) the exceptions to Standard Terms in Schedule B (***Exceptions to Standard Terms***), and
iii) the list of employees, agents, or consultants in Schedule C who are subject to the Agreement (***Employees***).

The representative of each party hereby acknowledges and agrees that he/she is duly authorized to execute this Agreement on behalf of the party and that this Agreement shall bind and be enforceable by and against the Employees. The authorized representative of party further acknowledges and agrees that only those individuals who are listed on the attached Schedule C incorporated herein shall be authorized to receive confidential information directly from Disclosing Party and that he/she will notify Disclosing Party in writing of any modification to Schedule C prior to releasing Confidential Information to those individuals listed on Schedule C.

This Agreement shall be construed and governed by the laws of the state of <State> (State), without giving effect to its conflicts of law principles. The parties hereby submit to the personal jurisdiction of the State and agree that any legal proceeding with respect to or arising under this Agreement shall be brought solely in, the state or federal courts located in the State. If any legal action or proceeding is commenced in connection with any dispute arising under, relating to or otherwise concerning this Agreement, the prevailing party, as determined by the court, shall be entitled to recover its attorneys' and experts' fees and all costs and necessary disbursements actually incurred in connection with such action or proceeding.

Each party acknowledge that it has read the Agreement, had the opportunity to discuss it with counsel, and is executing it with an understanding of its provisions. This Agreement may be executed in two or more counterparts, each of which will be deemed an original and all of which together will constitute one and the same document.

IN WITNESS WHEREOF, the parties hereto have caused this Agreement to be executed by their duly authorized representatives effective as of the Effective Date.

DISCLOSING PARTY: <NAME>

By_____
Name_____
Title _____
Date_____
Address_____
City/State/Zip_____
Fax_____

RECEIVING PARTY <NAME>

By _____
Name_____
Title_____
Date_____
Address _____
City/State/Zip_____
Fax_____

SCHEDULE A

Standard Terms

1. **Confidential Information.** "Confidential Information" as used in this Agreement means all information disclosed to Receiving Party by Disclosing Party in connection with the Purpose. Confidential Information includes, without limitation, (i) any and all business, technical, marketing, financial or other information, whether in electronic, oral or written form; (ii) trade secrets, business plans, techniques, methods, or systems, data, know-how, formulae, compositions, designs, sketches, mock-ups, prototypes, photographs, charts, graphs, forms, documents, drawings, samples, inventions, ideas, research and development, customer and vendor lists (including, without limitation, the identity, characteristics, contact persons, product and service needs thereof), rates, price lists, computer software programs and systems, financial statements, and budgets; (iii) all memoranda, summaries, notes, analyses, compilations, studies or those portions of other documents prepared by Receiving Party to the extent they contain or reflect such information of, or the contents of discussions with the Disclosing Party ("Receiving Party's Material"), including the contents or existence of discussions or negotiations related to the Purpose; (iv) information not generally known or readily ascertainable; (v) information that provides a competitive advantage for Disclosing Party; and (vi) information that is marked "Confidential" or nonpublic information which under the circumstances surrounding disclosure a reasonable person would conclude should be treated as confidential.

Confidential Information shall not include information that (a) is or becomes part of the public domain other than as a result of disclosure by Receiving Party, (b) becomes available to Receiving Party on a non-confidential basis from a source other than Disclosing Party, provided that, to the best of Receiving Party's knowledge, such source is not prohibited from transmitting such information by a contractual, legal, or other obligation, or (c) was in Receiving Party's possession prior to disclosure of the same by Disclosing Party.

2. **Nondisclosure Obligations.** Receiving Party agrees not to disclose, discuss, use, reproduce, duplicate, distribute, copy, reconstruct or in any way communicate, directly or indirectly, the Confidential Information for purposes other than in connection with the Purpose. Receiving Party shall not disclose, discuss, use, reproduce, duplicate, distribute or in any way communicate, directly or indirectly, the Confidential Information to any other party and will use all reasonable efforts to protect the confidentiality of such information. Receiving Party will require that Receiving Party's employees, officers, directors, agents, contractors, representatives, consultants and advisors who need to have access to such Confidential Information in order to assist Receiving Party in connection with the Purpose (1) are aware of the Receiving Party's confidentiality obligation hereunder, and (2) agree to be bound by such confidentiality obligations. Receiving Party shall notify Disclosing Party immediately of any loss, misuse, or misappropriation of any Confidential Information of which Receiving Party becomes aware.

3. **Ownership and Return.** All Confidential Information, including Receiving Party's Material, shall be and remain the property of Disclosing Party, and no right or license is granted to Receiving Party with respect to any Confidential Information. No transfer or creation of ownership rights in any intellectual property comprising Confidential Information is intended or shall be inferred by the disclosure of Confidential Information by Disclosing Party, and any and all intellectual property comprising Confidential Information disclosed and any derivations thereof, shall continue to be the exclusive intellectual property of Disclosing Party. Upon the termination by any Party of the Purpose, or sooner if so requested, Receiving Party agrees to immediately return all Confidential Information, including Receiving Party's Material, to Disclosing Party or to destroy all Confidential Information, including all copies of the same, however, Receiving Party shall not be required to destroy Confidential Information that has become embedded in Receiving Party's planning models. Upon request, the fact of any such destruction shall be certified in writing to Disclosing Party by Receiving Party. Nothing in this Agreement obligates Disclosing Party to disclose any information to Receiving Party or creates any agency or partnership relation between them.

4. **Warrantee.** ALL CONFIDENTIAL INFORMATION FURNISHED UNDER THIS AGREEMENT IS PROVIDED BY DISCLOSING PARTY "AS IS, WITH ALL FAULTS." DISCLOSING PARTY DOES NOT MAKE ANY WARRANTIES, EXPRESS OR IMPLIED, REGARDING THE ACCURACY, COMPLETENESS, PERFORMANCE, MERCHANTABILITY, FITNESS FOR USE, NONINFRINGEMENT OF ANY INTELLECTUAL PROPERTY RIGHTS, OR ANY RIGHT OF PRIVACY, ANY RIGHTS OF THIRD PERSONS OR OTHER ATTRIBUTES OF ITS CONFIDENTIAL INFORMATION.

5. **Compliance and Protection of Confidential Information**. Receiving Party represents and warrants that it has practices and procedures adequate to protect against the unauthorized release of Confidential Information received. Receiving Party must educate its employees, agents, and assigns in the provisions of this Agreement and provide to Disclosing Party upon request any information necessary to determine compliance with the terms of this Agreement.

6. **Indemnification.** Receiving Party agrees to indemnify, hold harmless and defend Disclosing Party, its employees, principals (owners, partners, shareholders or holders of an ownership interest, as the case may be), agents, contractors, representatives, consultants and/or advisors against any and all liability, loss, costs, damages, expenses, claims or actions, joint or several, arising out of or by reason of any breach of this Agreement by Receiving Party and/or Receiving Party's employees, agents, contractors, representatives or consultants, or arising out of or by reason of any act or omission of Receiving Party and/or Receiving Party's employees, agents, contractors, representatives or consultants in the execution, performance, or failure to adequately perform their obligations under this Agreement. For purposes of this Section, to "indemnify" means to defend and pay all expenses (including reasonable attorneys' fees) and satisfy all judgments (including costs and reasonable attorneys' fees) which may be incurred or rendered against Disclosing Party, its employees, principals (owners, partners, shareholders or holders of an ownership interest, as the case may be), agents, contractors, representatives, consultants and/or advisors.

7. **Compelled Disclosure.** If Receiving Party is requested or required by legal or administrative process to disclose any Confidential Information, Receiving Party shall promptly notify Disclosing Party of such request or requirement so that Disclosing Party may seek an appropriate protective order or other relief. In any case, Receiving Party will (a) disclose only that portion of the Confidential Information that its legal counsel advises is required to be disclosed, (b) use its reasonable efforts to ensure that such Confidential Information is treated confidentially, including seeking an appropriate protective order, and (c) notify Disclosing Party as soon as reasonably practicable of the items of Confidential Information so disclosed.

8. **Remedies.** The Parties acknowledge that remedies at law may be inadequate to protect Disclosing Party against any actual or threatened breach of this Agreement by Receiving Party, and, without prejudice to any other rights and remedies otherwise available to Disclosing Party, agree to the immediate granting of preliminary and final injunctive relief (without prior notice and without posting any bond) in favor of Disclosing Party to enjoin and restrain any breach or violation, either actual or anticipatory, of this Agreement.

9. **Limitations.** None of the Parties will be under any legal obligation of any kind whatsoever with respect to the Purpose by virtue of this Agreement, except for the matters specifically agreed to herein. No representation or warranty is made by the Disclosing Party as to the accuracy or completeness of any information provided to the Receiving Party.

10. **Term and Termination.** Receiving Party's obligations under this Agreement shall be effective on the Effective Date and shall be perpetual, notwithstanding any expiration, cancellation or termination of this Agreement. Upon termination of the Agreement, Receiving Party shall either promptly (1) deliver or cause to be delivered to Disclosing Party or (2) certify to the Disclosing Party the destruction of, all Confidential Information, including all copies of the Confidential Information in Receiving Party's possession or control including, without limitation, originals and copies of documents, customer lists, prospect lists, price lists, operations manuals, and all other documents reflecting or referencing the Confidential Information, as well as all other materials furnished to or acquired by Receiving Party to facilitate the Purpose of the Agreement.

11. **Agency.** This Agreement is binding on Receiving Party, its employees, agents, contractors, representatives, consultants, advisors, successors and assigns. In the event of a dispute regarding liability for breach of this Agreement, common law agency principles apply.

12. **Waiver.** No waiver of any of the provisions of this Agreement will be deemed or will constitute a waiver of any other provision, whether or not similar, nor will any waiver constitute a continuing waiver. No waiver will be binding unless executed in writing by an authorized representative of the Party making the waiver. The failure of either Party in any one or more instances to insist upon strict performance of any of the terms and conditions of this Agreement will not be construed as a waiver or relinquishment, to any extent, of the right to assert or rely upon any such terms or conditions on any future occasion.

13. **Modification.** This Agreement may not be amended except in a writing signed by authorized representatives of both Parties.

14. **Severability and Survival.** Should any clause, portion or paragraph of this Agreement be unenforceable or invalid for any reason, such unenforceability or invalidity will not affect the enforceability or validity of the remainder of this Agreement, and any court having jurisdiction is specifically authorized and encouraged by the Parties to hold inviolate all portions of this Agreement that are valid and enforceable without consideration of any invalid or unenforceable portions hereof. The headings of the sections in this Agreement are for the purposes of convenient reference only and are not intended to be part of this Agreement, or to limit or affect the meaning or interpretation of any of the terms hereof.

15. **Assignment and Succession.** This Agreement shall inure to the benefit of and be binding upon the successors and permitted assigns of the Parties hereto. Any successor to or assignee of Disclosing Party shall assume its rights and obligations under this Agreement with or without notice to Receiving Party. Receiving Party may not assign its rights hereunder without the written permission of Disclosing Party.

16. **Attorney's Fees.** If Receiving Party breaches or defaults in the performance of any of the covenants, agreements, representations, or warranties described in this Agreement, then in addition to any and all of the rights and remedies which Disclosing Party may have against Receiving Party, Receiving Party will also be liable to and pay Disclosing Party its court costs and reasonable attorney's fees incurred in enforcing Disclosing Party's covenants, agreements, representations and warranties herein.

17. **Notices.** All notices and other communications hereunder shall be in writing and shall be deemed given if delivered personally or by commercial delivery service, or mailed by registered or certified mail (return receipt requested) or sent via facsimile (with acknowledgment of complete transmission) to the parties.

18. **Entire Agreement.** The Parties agree that this Agreement, including Appendix A incorporated herein and as modified, constitute their entire agreement with respect to the subject matter hereof and that it supersedes any prior agreements or understandings between them, whether written or oral.

SCHEDULE B

Exceptions to Standard Terms

The parties acknowledge and agree that the Provisions below of this Schedule "A" shall form an integral part of the Agreement as if contained therein and are incorporated into the Agreement by this reference. To the extent of any inconsistency or discrepancy between the Agreement and the Provisions below of this Schedule "A", the Provisions below shall govern and take precedence. Unless otherwise noted, capitalized terms contained in this Schedule "A" shall have the same meaning as in the Agreement.

Article/Heading	Provision

SCHEDULE C

Employees

<Receiving Party> Employees, Agents, or Consultants subject to Confidentiality Agreement as of this_____day of _____, 20_____:

Print Name	Title	E-mail Address
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		

(Attach Additional Pages If Necessary)

Sensor Hosting Agreement

for the

National Infrastructure for Artificial Intelligence on the Grid ("NI4AI") Project

This Agreement sets forth the terms of Recipient's agreement to host one or more sensor(s) on behalf of NI4AI and funded by the ARPA-E project DE-AR0001104 (such hosting, the "Program").

As requested by Recipient, one or more sensor(s) will be shipped to Recipient on behalf of NI4AI at the following address:

_____

(Name, mailing address, email and phone number of Recipient)

Recipient agrees to connect the sensor(s) to a power source at a location of their choice, connect the sensor(s) to the Internet, and initiate streaming sensor data to NI4AI upon receipt, without unnecessary delay. The GPS receiver, if included, must be sited near a window or outdoors where it can acquire a suitable GPS lock. If necessary, an extended GPS cable will be sent. PingThings, Inc agrees to provide advice and support for any technical questions via support@pingthings.io.

The sensor(s) will be set up to stream data continuously. Recipient will be responsible for any costs associated with communications, i.e. internet connection, data streaming charges, and the like. Recipient expressly acknowledges that the data stream will be made available via NI4AI and disclaims any ownership rights in and to the data stream.

If one or more of the sensors are a synchrophasor, data from a single-phase voltage measurement should be streamed by Recipient (rms magnitude and angle, and frequency). If multiple phases and/or current measurements are also available, Recipient is encouraged to provide such information. Recipient is responsible for any costs and risks associated with installation and operation of the micro-PMU and the streaming of the data.

Recipient will provide location data for the sensor(s) to PingThings, including GPS coordinates and a qualitative description of the electric circuit. Recipient may suggest suitable language with a more vague, approximate description of the location for public release if need be.

Recipient may register for an NI4AI account that will grant them access to certain streamed data and analytics features via the PredictiveGrid™ platform. Ping Things and NI4AI reserve the right to limit such access if they deem appropriate.

Recipient may move the sensor(s) at their convenience. If disconnected, the sensor(s) will be re-connected at a new location meeting the above criteria, and PingThings notified about the new location, without unnecessary delay.

In case of data stream interruption, Recipient and PingThings agree to cooperate to address any problems in a timely manner.

The expectation is that data streaming will continue for the duration of the project, until Aug 11, 2022. If at any time prior, Recipient wishes to end data streaming, Recipient may return the sensor(s) to PingThings at the address below without delay. A pre-paid shipping label can be provided upon request.

The sensor(s) and all data associated with and generated by the sensor(s) are considered to be property of the NI4AI project and Recipient disclaims any ownership or right of use thereof. The sensor(s) will be returned to PingThings at the address below (or to its designee) within fourteen days of PingThings' request to Recipient, unless otherwise agreed by PingThings on behalf of NI4AI. A pre-paid shipping label can be provided upon request.

Recipient agrees that its participation in the Program is voluntary and without any compensation and agrees to indemnify and hold harmless PingThings, Inc. and NI4AI from and against any and all liability, damages, costs, losses and expenses arising out of or in connection with any personal injury, including death, or any damage to property or business which occurs due to Recipient's participation in the Program, including due to the installation, use, or operation of the sensor(s).

PingThings, Inc.
1220 S Street, Suite 150
Sacramento CA 95811

_____ _____

Recipient (signature, date) for PingThings (signature, date)

Schedule - Updates

WBS	Task Name	Mile-stone	Baseline Start	Baseline End	Actual Start	Actual End	% Comp
M0	Refine tasks and milestones (if applicable)	▼	11/11/2019	11/11/2019	8/12/2019	12/31/2019	100%
Task 1	Data platform deployment and enhancement		8/12/2019	8/11/2022	8/12/2019		75%
T1.0	Deployment of existing platform as project demo		8/12/2019	11/11/2019	8/12/2019	12/1/2019	100%
T1.1	Open source time series benchmark		8/12/2019	11/11/2019	8/12/2019	11/7/2019	100%
M1.1	Time series benchmark suite v1.0 released	▼	11/11/2019	11/11/2019	8/12/2019	11/7/2019	100%
T1.2	Platform cloud provider selection		8/12/2019	2/11/2020	8/12/2019	6/30/2030	100%
M1.2	Cloud provider selected	▼	2/11/2020	2/11/2020	8/12/2019	6/30/2030	100%
T1.3	Enhanced time series compression algorithms		11/12/2019	5/11/2020	10/1/2019	6/30/2020	100%
M1.3	Time series compression algorithm testing completed	▼	5/11/2020	5/11/2020	8/12/2019	6/30/2020	100%
T1.4	Development of an online method to onboard a new sensor		11/12/2019	2/11/2021	7/1/2020		70%
T1.5	Platform deployment to cloud		11/12/2019	8/11/2020	11/30/2019	8/11/2020	100%
M1.5	Go-No/Go: Platform deployed in selected cloud provider	▼	8/11/2020	8/11/2020	11/30/2019	8/11/2020	100%
T1.6	Implementation of additional data ingestors		2/12/2020	2/11/2021	1/1/2020	12/30/2020	100%
M1.6	Sensor data ingestors developed, tested, and deployed on the platform	▼	11/11/2020	11/11/2020	1/1/2020	12/30/2020	100%
T1.7	Ingest-stage data quality assessments		11/12/2019	2/11/2021	12/1/2019	6/30/2021	100%
M1.7	Data quality assessments developed, tested, and deployed on the platform.	▼	2/11/2021	2/11/2021	12/1/2019	6/30/2021	100%
T1.8	Platform language bindings		8/12/2019	8/11/2021	8/12/2019	12/31/2020	100%
M1.8	Additional platform language bindings made available.	▼	5/11/2021	5/11/2021	8/12/2019	12/31/2020	100%
T1.9	Integration of data quality assessment into platform		2/12/2020	8/11/2021	7/1/2020		80%
T1.9.1	DQI integration into core time series visualization functionality		2/12/2020	8/11/2021	4/1/2021	6/30/2021	100%
T1.9.2	DQI integration into sensor fleet management functionality		2/12/2020	8/11/2021	10/1/2020	6/30/2021	100%
T1.9.3	DQI used for alerting		2/12/2020	8/11/2021	4/1/2021		40%
T1.10	Optimize platform		2/12/2020	8/11/2021	10/1/2020	6/30/2021	100%
T1.10.1	Cloud compute cost optimization.		2/12/2020	8/11/2021	10/1/2020	6/30/2021	100%

DE-AR0001104 : PingThings Inc. - Murphy
Q3 of FY 2021: April 1, 2021 - June 30, 2021

WBS	Task Name	Mile-stone	Baseline Start	Baseline End	Actual Start	Actual End	% Comp
T1.10.2	Cloud storage cost optimization		2/12/2020	8/11/2021	10/1/2020	6/30/2021	100%
M1.10	Platform optimization documented	▼	8/11/2021	8/11/2021	4/1/2021	6/30/2021	100%
T1.11	Integration of leading machine learning and deep learning libraries		8/12/2020	8/11/2021	10/1/2020		60%
T1.11.1	Community survey of ML and DL needs		8/12/2020	8/11/2021	1/1/2020	6/30/2021	100%
T1.11.2	Survey of best of breed and current state of the art in ML and DL		8/12/2020	8/11/2021	1/1/2020	6/30/2021	100%
T1.11.3	Library selection		8/12/2020	8/11/2021	1/1/2020	6/30/2021	100%
T1.11.4	Library integration into platform		8/12/2020	8/11/2021	7/1/2020		90%
M1.11	ML/AI Open Source Library Integration	▼	8/11/2021	8/11/2021			95%
T1.12	Additional platform enhancements requirements gathering		8/12/2019	8/11/2022	1/1/2020		80%
M1.12	Additional platform enhancement gathering complete.	▼	2/11/2022	2/11/2022	1/1/2020		80%
T1.13	Implementation of required platform enhancements		8/12/2021	8/11/2022			30%
M1.13.1	Additional platform enhancement 1 completed	▼	2/11/2022	2/11/2022	1/1/2021		50%
M1.13.2	Additional platform enhancement 2 completed	▼	5/11/2022	5/11/2022			20%
M1.13.3	Additional platform enhancement 3 completed	▼	8/11/2022	8/11/2022			20%
Task 2	Hyperscale grid sensor data collection		8/12/2019	8/11/2022	8/12/2019		60%
T2.1	Previously generated sensor data set collection		8/12/2019	8/11/2022	8/12/2019		70%
T2.1.1	Asynchronous data ingestors developed for selected previously generated sensor data sets.		8/12/2019	2/11/2022	8/12/2019	12/31/2020	100%
T2.1.2	Develop necessary data ingestors		2/12/2020	2/11/2022	8/12/2019	12/31/2020	100%
M2.1.1	Asynchronous data ingestors developed for identified simulated data sets.	▼	8/11/2020	8/11/2020	8/12/2019	8/11/2020	100%
T2.1.3	Ingest previously generated sensor data sets		8/12/2020	11/11/2020	8/12/2019	3/31/2021	100%
M2.1.2	Previously generated sensor data set ingested and available	▼	11/11/2020	11/11/2020	8/12/2019	9/30/2020	100%
T2.1.4	Socialize availability of previously generated sensor data sets		5/12/2020	8/11/2022			80%
T2.2	In vitro sensor data collection		8/12/2019	8/11/2022	8/12/2020	6/30/2021	100%
T2.2.1	Select sensors for laboratory generation of in vitro time-series		8/12/2019	2/11/2020	8/12/2019	12/31/2020	100%

DE-AR0001104 : PingThings Inc. - Murphy
Q3 of FY 2021: April 1, 2021 - June 30, 2021

WBS	Task Name	Mile-stone	Baseline Start	Baseline End	Actual Start	Actual End	% Comp
T2.2.2	Incentivize community member for in vitro data set generation		8/12/2019	8/11/2021	8/12/2019	12/31/2020	100%
T2.2.3	Develop necessary streaming ingestors for capturing laboratory data		11/12/2019	8/11/2020	8/12/2019	12/31/2020	100%
T2.2.4	Ingest in vitro time series data		8/12/2020	8/11/2021	7/1/2020	12/31/2020	100%
M2.2	First in vitro data is streaming to the platform.	▼	11/11/2020	11/11/2020	7/1/2020	12/31/2020	100%
T2.2.5	Socialize in vitro data acquisition		8/12/2020	8/11/2022	7/1/2020		80%
T2.3	In vivo sensor deployment and data capture		8/12/2019	8/11/2022	1/1/2020		50%
T2.3.1	Identification of sensor hardware		8/12/2019	2/11/2020	8/12/2019	6/30/2020	100%
T2.3.2	Identification of sensor locations and collaborators		8/12/2019	2/11/2022	1/1/2020		90%
T2.3.3	Acquire sensor hardware		11/12/2019	8/11/2020	8/12/2019		90%
T2.3.4	Test and validate one of each type of sensor hardware		2/12/2020	2/11/2021	7/1/2020	9/30/2020	100%
T2.3.5	Deploy sensor hardware to control group (project partners)		8/12/2020	8/11/2022	7/1/2020	3/31/2021	100%
T2.3.6	Deploy sensor hardware to national coverage (wide area) locations		8/12/2020	8/11/2022	7/1/2020		40%
T2.3.7	Deploy sensor hardware to boutique (site-specific) locations		8/12/2020	8/11/2022	7/1/2020		50%
M2.3.1	Go-No/Go: Successful sensor deployment to control group	▼	2/11/2021	2/11/2021	7/1/2020	3/31/2021	100%
M2.3.2	Successful sensor deployment for national coverage data set	▼	8/11/2021	8/11/2021	7/1/2020		50%
M2.3.3	Successful sensor deployment for boutique data set	▼	5/11/2021	5/11/2021	7/1/2020	6/30/2021	100%
T2.4	Data escrow platform services		8/12/2019	8/11/2022			40%
T2.4.1	Collect sample NDA and DSA documents		8/12/2019	8/11/2020	8/12/2019	9/30/2020	100%
T2.4.2	Legal analysis of sample NDA and DSA documents		2/12/2020	2/11/2021		6/30/2021	100%
M2.4.1	Legal document collection and analysis	▼	8/11/2020	8/11/2020	8/12/2019	6/30/2021	100%
T2.4.3	Develop common NDA and DSA documents		8/12/2020	5/11/2021			100%
M2.4.2	“Common” NDA and DSA developed and released.	▼	2/11/2021	2/11/2021		6/30/2021	100%
T2.4.4	Determine data handling functional requirements		2/12/2020	2/11/2021	10/1/2020		60%
T2.4.5	Implement required platform functionality		8/12/2020	8/11/2021	10/1/2020		50%
T2.4.6	Validate data escrow functionality		11/12/2020	8/11/2021	10/1/2020		50%

DE-AR0001104 : PingThings Inc. - Murphy
Q3 of FY 2021: April 1, 2021 - June 30, 2021

WBS	Task Name	Mile-stone	Baseline Start	Baseline End	Actual Start	Actual End	% Comp
M2.4.3	Deployed and tested data escrow functionality.	▼	8/11/2021	8/11/2021	10/1/2020		50%
T2.4.7	Market data escrow service		8/12/2019	8/11/2022	10/1/2020		30%
Task 3	Community building		8/12/2019	8/11/2022			80%
M3.0	Go-No/Go: Outreach documentation provided to ARPA-E	▼	8/11/2021	8/11/2021			100%
T3.1	Announce project launch		8/12/2019	2/11/2020	8/12/2019	12/31/2019	100%
T3.1.1	Create content for the project announcement		8/12/2019	2/11/2020	8/12/2019	11/5/2019	100%
T3.1.2	Make broad project announcement		8/12/2019	2/11/2020	8/12/2019	11/5/2019	100%
T3.2	Community recruitment		8/12/2019	8/11/2022	8/12/2019		80%
T3.2.1	Develop marketing content and tools to be used for targeted recruitment		8/12/2019	2/11/2020	8/12/2019	9/30/2020	100%
T3.2.2	Select first cohort of potential research partners from across academia, industry, and utilities.		8/12/2019	2/11/2020	8/12/2019	6/20/3020	100%
T3.2.3	Targeted partner outreach to the first cohort		8/12/2019	8/11/2020	8/12/2019	9/30/2020	100%
T3.2.4	Continue targeted partner outreach		8/12/2020	8/11/2022	8/12/2019		60%
M3.2	Industry/academia project launch announcement	▼	11/11/2019	11/11/2019	8/12/2019	11/1/2019	100%
T3.3	Create, update, and market the project blog		8/12/2019	8/11/2022	8/12/2019		80%
M3.3.1	Project blog launched	▼	11/11/2019	11/11/2019	8/12/2019	11/11/2019	100%
M3.3.2	Year 1 blog content creation completed	▼	8/11/2020	8/11/2020	8/12/2019	9/29/2020	100%
M3.3.3	Year 2 blog content creation completed	▼	8/11/2021	8/11/2021	9/29/2020		95%
M3.3.4	Year 3 blog content creation completed	▼	8/11/2022	8/11/2022			0%
T3.4	Create and update project website		8/12/2019	8/11/2022	8/12/2019		90%
M3.4	Project website launched	▼	2/11/2020	2/11/2020	8/12/2019	10/31/2019	100%
T3.5	Create platform documentation and training material		8/12/2019	8/11/2022	8/12/2019		100%
T3.5.1	Create API documentation		8/12/2019	2/11/2021	8/12/2019	12/31/2020	100%
T3.5.2	Aggregate the best external educational content		8/12/2019	2/11/2021	8/12/2019		100%
T3.5.3	Create Jupyter Notebooks		8/12/2019	2/11/2021	8/12/2019	12/31/2020	100%
T3.5.4	Create video content		8/12/2019	2/11/2021	1/1/2020	12/31/2020	100%
M3.5	Platform content created	▼	2/11/2021	2/11/2021	8/12/2019	3/31/2021	100%
T3.6	Establish project presence at relevant conferences		8/12/2019	8/11/2022	8/12/2019		100%

WBS	Task Name	Mile-stone	Baseline Start	Baseline End	Actual Start	Actual End	% Comp
T3.6.1	Select conferences to maximize growth in community and overall cost		8/12/2019	8/11/2022	8/12/2019		100%
T3.6.2	Determine most cost effective, available method of impact		8/12/2019	8/11/2022	8/12/2019		100%
T3.6.3	Attend and impact the community		8/12/2019	8/11/2022	4/1/2020		70%
M3.6.1	Conference attended	▼	2/11/2020	2/11/2020	8/12/2019	10/25/2019	100%
M3.6.2	Conference attended	▼	8/11/2020	8/11/2020	8/12/2019	12/31/2019	100%
M3.6.3	Conference attended	▼	2/11/2021	2/11/2021	1/1/2020	6/30/2020	100%
M3.6.4	Conference attended	▼	8/11/2021	8/11/2021	10/1/2020	3/31/2021	100%
M3.6.5	Conference attended	▼	2/11/2022	2/11/2022			0%
M3.6.6	Conference attended	▼	8/11/2022	8/11/2022			0%
T3.7	Plan and host workshops associated with relevant conferences		8/12/2019	8/11/2022	1/1/2020		100%
T3.7.1	Evaluate available conferences		8/12/2019	8/11/2022	1/1/2020		80%
T3.7.2	Reach out and connect with conference organizers for approval		8/12/2019	8/11/2022	1/1/2020		80%
T3.7.3	Develop agenda for event		8/12/2019	8/11/2022	1/1/2020	3/31/2021	100%
T3.7.4	Develop content required for event		8/12/2019	8/11/2022	1/1/2020		100%
T3.7.5	Marketing and outreach for event		8/12/2019	8/11/2022	1/1/2020		100%
T3.7.6	Film presentation made for content generation purposes		8/12/2019	8/11/2022	4/1/2020		100%
M3.7.1	Project-focused workshop hosted	▼	8/11/2020	8/11/2020	4/1/2020	6/30/2020	100%
M3.7.2	Project-focused workshop hosted	▼	8/11/2021	8/11/2021	10/1/2020	11/15/2020	100%
M3.7.3	Project-focused workshop hosted	▼	8/11/2022	8/11/2022	3/1/2021	5/25/2021	100%
T3.8	Plan and host data science competitions/data hackathons for the national infrastructure		8/12/2019	8/11/2022	10/1/2020		20%
T3.8.1	Host an online data science competition		8/12/2019	2/11/2021	10/1/2020		50%
T3.8.2	Host an online or onsite data hackathon		2/12/2020	11/11/2021			0%
T3.8.3	Host an online data science competition		11/12/2020	8/11/2022			0%
M3.8.1	Data science competition or hackathon hosted.	▼	2/11/2021	2/11/2021	4/1/2021		50%
M3.8.2	Data science competition or hackathon hosted.	▼	11/11/2021	11/11/2021			0%
M3.8.3	Data science competition or hackathon hosted.	▼	8/11/2022	8/11/2022			0%

Additional Performance Updates

A) ISSUES, RISKS, AND MITIGATION

There are two major risks to report.

The first risk is pandemic-related. The initial hope was that the vaccine would bring about an end to the pandemic and a return to normalcy. Unfortunately, the Delta variant is spreading like wildfire across the US and additional variants could catch a foothold. Delta and subsequent variants are capable of escaping vaccine-induced immunity with a certain probability and their spread has and will force some set of restrictions on travel and work that will likely escalate with declining fall temperatures in the northern hemisphere later this year. Not only does this situation pose a direct risk to employees and family members, even the ones that are fully vaccinated, but it also is causing us to re-examine travel plans for in-person events and conferences planned for this fall. This is particularly impactful for the PingThings' long overdue company retreat, important given that the company grew substantially during the pandemic and many employees have never met face to face. We will continue to employ the same risk mitigation techniques as before with a focus on virtual events, workshops, and meetings.

Beyond personal safety and team cohesion, another round of lockdowns makes it harder to get sensors deployed. We have expanded the definition of Task 2 to include boutique data sets created not only from sensors deployed by the project but also those created by private parties, such as utilities or commercial vendors.

The second risk is one provided by growth. As a startup, growth is not only to be expected but is also required for survival. However, PingThings is beginning to grow faster than we can hire, especially given the ultra competitive landscape for technical talent that has been exacerbated by the pandemic. Thus, we are being forced into a position of triaging significant amounts of work. To mitigate this, we are working to ensure that commercial growth tasking is well aligned with our stated NI4AI work and milestones.

Contains protected data: □ Protected Rights Data □ SBIR/STTR Data ■ No Protected Data

B) CHANGES IN APPROACH

In previous performance periods, we were encouraged to re-evaluate how the company could more effectively leverage project milestones to advance strategic partnerships that could help to drive sales. This clarification in focus has led us to prioritize and broaden the scope of certain Task 2 milestones. Most significantly, we have focused on building closer partnerships with key thought leaders in industry, using tasks geared at historic data ingestion to enable them to work with their own data in the platform. These key strategic partners have been less interested in sensor deployment, as they already own and maintain large sensor fleets. While we are still working with individuals and with smaller utilities to deploy sensors, we have prioritized efforts that could lay the groundwork for recruiting much larger companies to begin working with their data in the platform.

Contains protected data: □ Protected Rights Data □ SBIR/STTR Data ■ No Protected Data

C) CHANGES IN PROJECT PERSONNEL

During the summer we hired three interns to support the project. These included two undergraduate computer science interns and a graduate intern from the University of California, Berkeley Haas School of Business to help with market outreach.

Jeff Maxim, our front end lead, has recently left the company.

Technology To Market (T2M)

A. PUBLICATIONS

N/A

B. OTHER PROJECT OUTPUT

N/A

C. FOLLOW-ON FUNDING

N/A

Administrative and Legal Updates

DISCLOSURES

N/A

DE-AR0001104 : PingThings Inc. - Murphy
Q3 of FY 2021: April 1, 2021 - June 30, 2021

Cost Updates - LEAD: PingThings Inc. (DE-AR0001104)

A National Infrastructure for Artificial Intelligence on the Grid

ARPA-E RECORDS ONLY (invoices edited by performer are highlighted)

Invoice #	Date	Total	Federal Share	Performer Share	Shortpays	TT&O	Personnel	Fringe Benefits	Travel	Equipment	Supplies	Contractual	Construction	Other	Indirect
AR0001104-001	9/6/2019	$852,596.82	$95,904.54	$756,692.28	($1,017.74)	$8,798.66	$24,917.75	$5,357.32	$0.00	$0.00	$0.00	$54,995.97	$0.00	$758,798.66	$8,527.11
AR0001104-2	10/20/2019	$128,014.80	$127,464.80	$550.00	($104.50)	$9,563.75	$52,536.31	$11,295.31	$2,326.11	$0.00	$693.29	$41,106.25	$0.00	$9,563.75	$10,493.79
AR0001104-110013	4/5/2020	$112,959.52	$111,859.52	$1,100.00	($9,772.63)	$8,741.84	$44,867.31	$9,646.47	$0.00	$0.00	$0.00	$30,381.25	$0.00	$27,964.48	$100.00
AR0001104-110014	7/29/2020	$167,830.03	$84,548.51	$83,281.52	($47.47)	$1,583.78	$33,601.02	$7,224.22	$0.00	$0.00	$0.00	$33,097.50	$0.00	$11,575.77	$82,331.52
AR0001104-110017R	8/14/2020	$480,148.68	$437,088.47	$43,060.21	$0.00	$7,170.64	$272,584.07	$58,605.58	$0.00	$0.00	$0.00	$90,162.45	$0.00	$16,661.37	$42,135.21
AR0001104-110018R	12/24/2020	$342,148.51	$311,640.86	$30,507.65	$0.00	$0.00	$184,618.43	$39,692.97	$0.00	$0.00	$0.00	$66,265.07	$0.00	$22,514.39	$29,057.65
AR0001104-APRIL2021REV1	6/8/2021	$696,291.86	$646,072.96	$50,218.90	$0.00	$50,046.67	$319,884.18	$68,775.10	$0.00	$0.00	$25,711.90	$87,529.70	$0.00	$146,772.08	$47,618.90
AR0001104-MAY2021	6/16/2021	$134,872.23	$123,250.17	$11,622.06	$0.00	$14,233.12	$73,339.46	$15,767.98	$0.00	$0.00	$0.00	$8,530.00	$0.00	$26,162.73	$11,072.06
AR0001104-JUNE2021 (Pending)	7/31/2021	$188,112.20	$172,205.97	$15,906.23	$0.00	$22,390.67	$93,857.48	$20,179.36	$0.00	$0.00	$0.00	$25,774.00	$0.00	$33,145.13	$15,156.23
Total:		$3,102,974.65	$2,110,035.80	$992,938.85	($10,942.34)	$122,529.13	$1,100,206.01	$236,544.31	$2,326.11	$0.00	$26,405.19	$437,842.19	$0.00	$1,053,158.36	$246,492.47
% Expended in Budget Category:		49.64%	42.20%	79.39%		36.51%	50.04%	50.04%	3.44%	0.00%	7.12%	32.09%	0.00%	75.25%	65.37%
Remaining:		$3,147,748.35	$2,889,964.20	$257,784.15		$213,074.54	$1,098,354.99	$236,146.69	$65,261.89	$0.00	$344,536.81	$926,452.81	$0.00	$346,403.64	$130,591.53

Overall Proposed Cost Share: 20.01 %

Invoiced Cost Share to Date: 32.00 %

Proposed TT&O to Federal Share: 6.71 %

Actual TT&O to Federal Share: 5.81 %

Agree that the information provided is accurate and correct: ■ Yes □ No

Certification of Compliance

I certify that I have the authority to make the following certification and to submit this Research Performance Progress Report on behalf of the Prime Recipient. On behalf of the Prime Recipient, I further certify that the information provided in this Research Performance Progress Report is accurate and complete as of the date shown below. I understand that false statements or misrepresentations may result in civil and/or criminal penalties under 18 U.S.C. § 1001.

Signature: Laurel Dunn

Submitted by: Laurel Dunn (ldunn@berkeley.edu), PingThings

Submitted Date: 8/10/2021 2:14:07 PM