Lessons Learnt Report 3

PROJECT SHIELD

Project SHIELD

Synchronising Heterogenous Information to Evaluate Limits for DNSPs

APRIL 2022

Contents

Acknowledgements

The project ‘Synchronising Heterogenous Information to Evaluate Limitations of DNSPs’ (Project SHIELD) is a collaboration led by Luceo Energy (Redback Operations trading as Luceo Energy) and Energex, Ergon Energy, GridQube, the University of Queensland and Essential Energy.

The Australian Government supports Project SHIELD, through the Australian Renewable Energy Agency (ARENA) with $2.6 million in funding over the 2-year project period.

The total project value of Project SHIELD is $5.7 million, consisting of cash and in-kind contribution from ARENA and the consortium partners

Disclaimer

This project received funding from ARENA as part of ARENA’s Advancing Renewables Program.

The views expressed within this report are not necessarily the views of the Australian Government, and the Australian Government does not accept responsibility for any associated information or advice.

Executive Summary

The Lessons Learnt Report No 3 is part of Milestone 3 and concludes Stage 1 of Project SHIELD (Milestones 1 -3). The Lessons Learnt that are presented in this report summarise, at a high level the key learnings made by the project team during Stage 1.

In summary,

- The project was able to identify and utilise data from a set of diverse sources for network purposes.
- There is a diverse range of behind-the-meter, consumer or third party-owned devices that collect data and that can potentially be utilised for network purposes.
- Excluding smart meter data, the density of measurement device coverage on a feeder from any single provider was generally very low.
- A significant amount of integration is required when working with data sources that are non-standardised and heterogeneous.
- Getting access to such data sources and use them for network purposes is as much a technical challenge as it is a legal and commercial challenge.
- Excluding cost of hardware and installation thereof as well as the cost for storage and analytics, the on-going, monthly cost of data per connection point is somewhere between $0.75 and $100.00 (for on average 10-minute resolution, day behind or historical data).
- Using different data sources, data platforms and connected applications is still very novel. Make sure your project team has both expertise and time to resolve unforeseen challenges.
- Access to real-time data from heterogeneous sources is very limited. The project chose to work with historical data to test the processes and assumptions made which are not dependent on access to real-time data.
- The data that was accessible from existing data sources in quality and quantity is not sufficient to allow for a robust conclusion regarding the cost-benefit of the data in estimating the network state and subsequent capacity for additional PV hosting.
- Project SHIELD put forward to ARENA a recommendation to continue the project through Stage 2.

Context and Background

By the end of 2021 more than 3 million Australian households had installed a small-scale, behind-the-meter PV system at their properties. The cumulative capacity of these systems worth around 16GW, now represent the largest generator in the Australian energy system. Australia is the world leader in rooftop solar capacity per person.

Industry analysts forecast that between 2022 and 2030 an additional 2.5GW per annum will be added in this sector.

While the rooftop PV boom is positive for Australia's carbon emission reduction targets, employment in the sector and consumers, the fast and continued growth of rooftop solar PV generation creates a number of challenges for Australia's distribution network operators.

The SHIELD project has picked one of these challenges, rooftop PV hosting capacity on LV network feeders, and looks at how a data-driven, analytical approach can unlock additional PV hosting capacity on LV network feeders without investment by networks in augmentation. True to this investment conscious philosophy, Project SHIELD, in the first stage of the project, that concludes with Milestone 3, is looking to use existing data sources on LV network feeders to run the respective analytics.

The anticipated outcome for Project SHIELD is to develop a robust and reliable model that will allow Distributed Network Service Providers like Energex, Ergon and Essential Energy, to assess the cost-benefit of various data acquisition strategies to support a transition from general, static PV hosting capacity limits to a unique, feeder-individual PV hosting capacity limit. Project SHIELD will aim to reasonably quantify the benefits to network operators and consumers that result from a data-driven approach. The resulting documentation should allow Australia's distribution network service providers to argue their proposed investment in digital assets towards the Australian Energy Regulator as part of their regulatory investment submissions.

Opening up additional PV hosting capacity on network feeders without significant investment for network augmentation will yield tangible benefits to consumers, communities and support Australia's minimum required feeder instrumentation to make such a robust and reliable determination in order to minimise cost to consumers and allowing for the safe maximum penetration of consumer owned rooftop PV with export capabilities. minimum required feeder instrumentation to make such a robust and reliable determination in order to minimise cost to consumers and allowing for the safe maximum penetration of consumer owned rooftop PV with export capabilities. net zero carbon emissions programs.

Since the data collected in Stage 1 of Project SHIELD was insufficient to determine a definitive outcome, the project team, in Stage 2, will move to source and install additional network devices that collect data in real-time at the node level in these feeders. Following installation, the project will test at what degree of data density and fidelity a robust and reliable result can be derived through the analytics process. The aim of the project team is to determine the minimum required feeder instrumentation to make such a robust and reliable determination in order to minimise cost to consumers and allowing for the safe maximum penetration of consumer owned rooftop PV with export capabilities.

Overview of Key Lessons Learnt

During Stage 1 of Project SHIELD, encompassing Milestones 1-3, the following key lessons were identified and are worth sharing:

Lesson #1

There are currently limited "alternate" data sources out there that are suitable for estimating hosting capacity

Expectation

There could be enough data collected from various sources on the LV network to estimate hosting capacity

When embarking on this project, the project team identified many potential sources of energy and power data and theorised that there could be enough "fit-for-purpose" data available from various customer owned or non-network devices to calculate the real PV hosting capacity of a feeder and compare this to the current, static limits applied by industry. The objective is to determine how the data density, including location of devices and frequency of measurement, influences the accuracy of PV hosting capacity calculations. This knowledge will support networks moving to dynamic PV connection options for customers.Sharing Data Between Project Partners

Experience

The project team made an effort to identify alternative data source options. Alternative data source options are data sources that are not network owned sources.

The following, alternative data source options were identified:

- Smart Meters
- NBN Infrastructure
- Connected DER (e.g. Solar PV inverters, home battery system)
- EV car chargers
- Residential data analytics service providers

- Home Energy Management Systems (HEMS)
- Other (eg. Connected Smart Plugs)

The project found the following:

- Home battery systems and EV chargers did not have any data assets in our project trial areas
- NBN data as well as data from service providers such as Solar Analytics seem to be the accessible data source on network feeders, but overall, their penetration levels were very low.
- Whilst still sparse, Smart meters are the most common alternative data source found on network LV feeders

◊ Smart meter data pricing varies by provider and data quality and quantity. Excluding cost of storage and analytics, smart meter data for a residential connection point at 10-minute granularity and day behind provision costs between $0.75 and $1.00 per month.

◊ Getting access to smart meter data requires a commercial arrangement with the respective metering service providers.

◊ In contestable markets there are multiple metering service providers with different types of smart meters with different meter configurations and data collection and upload intervals.

Network-owned data assets in the form of transformer monitoring and LV network monitoring equipment were also available.

- Monitors are not available on all transformers with penetrations up to 30% across the trial feeders.
- Most transformers installed on the network trial areas were not actively monitoring
- New transformers can come equipped with remote monitoring functionality.
- Data from network-owned LV monitoring devices in the feeders (Ergon and Energex only) was available at 1 minute resolution for $1.75 per month (excluding hardware and installation costs).

In summary, the project was at best able to get access to 24% of connection point data on one of the trial feeders using existing data sources but on other trial feeders it was as low as 5% and the type and number of devices found in specific network LV feeders varies significantly.

Lessons Learned

The type and number of devices found in specific network LV feeders varied significantly.

Although there are significant possible options available for collecting network power data, in reality, they are currently installed at very low density across the LV network.

What We Would Do Differently

To improve this result, the project would

1. Allow more time to interact with vendors to secure data

2. Take more time to assess which devices were more readily accessible in what areas in order to maximise the outcomes.

Lesson #2

Getting access to alternative and network data sources is a technical, legal and commercial challenge

Expectation

Accessing data for this data was expected to be complex but not unobtainable

The project expected there would be some challenges associated with collecting the data however given the nature of the project and privacy provisions designed to protect the data, there was an expectation that accessing the data should be relatively procedural. Further, due to the nature of the data sets required (Voltage, Reactive Power) it was assumed there would be less resistance to providing this data.

Experience

The project team found that data concerning the consumption of electricity in small time intervals, continuously supplied and in connection with accurate geographic information was considered highly sensitive and difficult to access. This data was not required for project purposes but when talking to potential providers of data, they often initially assumed this was our need.

Explaining the data needs of this project to non-data experts (business managers) proved challenging when talking format and detail. Often it took us time to reach the subject matter experts.

Currently there is no established and repeatable framework to access data from these alternative sources for use by research projects like Project SHIELD or Distributed Network Service Providers.

- The true value of network data, especially real-time data is still emerging as the industry trials data sources and complementary platforms and analytics solutions.
- With the exception of Victoria, network service providers are not routinely given access to smart meter data.
- Smart meters are deployed foremost for billing purposes, not network analytics.
- Metering service providers can on-sell the off-market smart meter data to network service providers through commercial agreements. The benefits of access to such data on a continuous and granular basis is determined by the underlying business case.

A number of potential data sharing partners were not participating in the project for various technical, commercial or legal reasons:

- Given that Project SHIELD is led by Luceo Energy (Redback Technologies trading as Luceo Energy), companies in the solar PV inverter and home battery market did exhibit IP or confidentiality concerns when it came to data access and sharing.
- As there is no established market for consumer electricity data yet, the value of such data streams is unknown. The pricing of data offered by vendors varied greatly (from free to unviable pricing) Some companies overestimated the value of data from their devices and requested pricing that does not fit the underlying business case.
- Not all companies that collect consumer data through their solutions actually have in place pro-forma agreements that would allow the company to share or sell such consumer data to commercial operators or research projects.

The level of privacy and confidentiality applied to customer-level data made it challenging to acquire the data necessary for the project.

- There are a multitude of opinions, practices and challenges associated with protecting and sharing energy data. Some of these issues are based on legal advice, others are based on interpretation of publicly available information while others are based on fear and risk aversion. Some of the barriers are real while others could arguably be described as perceived. What is clear, however, is that there is no single viewpoint about what is permissible when it comes to sharing data.

The details pertaining to data privacy and customer data rights continue to evolve and constantly change. This has led to uncertainty and caution from companies as to what is and is not permitted.

The pathway to receiving data was long and complex and different for each vendor and it became difficult to keep track of each vendors issues, needs and requests as well as the intricacies of each data set.

Lessons Learned

Consumer data privacy concerns and the applicable laws and regulations need to be carefully considered by any project or enterprise that works in this context.

When working with consumer electricity data, targeting subject matter experts early is highly recommended.

Data is still considered sensitive by some vendors although they are sometimes unable to clearly verbalise the reason for this.

The Energy Industry does not hold a common understanding of the regulations and risks associated with sharing each of the available data sets.

Creating a protocol or guide on this matter could greatly assist industry stakeholders in their endeavours to send or receive customer data. In addition, it would offer future projects a clear expectation of what data they are likely to be able to access and how to do this, as well as ease any legal expenses.

Discussions with the Energy Security Board about their data strategy have identified that they have already recognised the lack of an industry standard around data sharing as a barrier to sharing key power data. They are currently developing a strategy to provide greater clarity and reducing regulatory barriers to allow safe sharing of existing data.

What We Would Do Differently

The project would have appointed a dedicated data broker for the project to manage all negotiations and steward all the data sets through to integration in the project.

Lesson #3

A significant amount of integration is required when working with data sources that are non-standardised and heterogeneous.

Expectation

Integrating data from different providers was expected to be difficult.

Experience

Integrating the data was more difficult than the project team expected.

- Data streams from network owned data assets and smart meters are well defined and specified and simple to integrate.
- Access can be obtained via cloud API or other reasonable form of data ingress through adaptors.
- Most data streams were presented in different time resolution and as a result, the project team required significant integration work to deliver a workable solution where each of these sets could be aligned.
- Data points from various data sources with various data fidelity had to be aligned to present such data in a unified version and make it usable for further processing and visualisation.
- Data adapters needed to be developed and tested to allow the various data formats to be accepted by the platform.
- Luceo needed to adapt the analytics layer to work with the different data sets to attain a set of predetermined outcomes.
- Once integrated, the data needed to be presented to a third-party application, in this case the GridQube State Estimation algorithm, for further processing.
- The low quantity of data sources available in the trial feeders would normally made such integration very cost prohibitive under normal commercial operations.
- Not all companies with data assets have the required infrastructure in place to share such data externally securely and effectively through a cloud API or comparable means. Despite this, using such data for platform-based analytics and related functions such as state-estimation algorithms can be feasible.

Lessons Learned

Allow sufficient time for complex data integrations. The work done on data integrations needs to be documented carefully for future use and refinement.

What We Would Do Differently

Using different data sources, the Luceo Energy data platform and connected applications is still very novel. Make sure your project team has both expertise and time to resolve unforeseen challenges

When combining data from heterogeneous sources it is critical to have clarity over what data is collected including:

- Measurement device and location on the electrical network
- Measurement parameters (e.g. voltage, current, active power)
- Measurement accuracy
- Measurement interval (e.g. records every 5-min, 10-minutes)
- Measurement type (e.g. instantaneous or average)
- Frequency of delivery (e.g. real-time, daily, weekly)
- Whether the time zone of the data and the timestamp represents the start or end of the interval

Lesson #4

Real-time data remains a challenge for now

Expectation

The project would use real time data for its analysis.

Experience

The project hoped to gain access to real time flows of data from the data providers.

Unfortunately, some of the data was not collected in real time and some vendors did not possess the know-how to deliver real time data to the project

- Very few devices and solutions in the project feeders are currently specified and equipped to deliver access to real-time data.
- No real-time capable smart meters were available in the project feeders.
- Real time data streams at scale present a huge challenge to operators with regards to the connected platform and cloud infrastructure as the number of messages scale fast. Many operators are not investing into such capability without a business case that supports such investment.
- The market cost for smart meter data increases the finer the fidelity and the closer to real time delivery the data is serviced. In addition, only few smart meter companies and solution providers are able to handle real time data at scale in their current infrastructure.
- DNSPs have Data Governance frameworks designed to adhere to IPART license conditions that introduce additional labour and process to transact any information without a low dissemination classification. As consequence, the effort required to assess risk exposure due to the receiving parties' cyber security control measures needs to be justified against the value of the outcome of any service provided.

The exception was data from LV monitoring devices owned by Ergon and Energex.

- Ergon and Energex had network owned LV monitoring devices installed at some connection points in the project feeders that provide data in real-time (1 minute resolution). Since the project did not use live data (see above), historical data streams from these devices were used.

While real time data provides the best possible outcomes in a real-life operational environment, in the end, Project SHIELD decided to test the project's assumptions, processes and tools and value of various data sources using historical data.

Lessons Learned

The energy data space is evolving fast. While the technology exists to collect, transfer and manage real time data that can also be used for network purposes, not all companies and solutions are working with real time data at this moment.

What We Would Do Differently

We should have considered the costs and difficulties associated with the use of real-time data and identified from the beginning that historical data would suffice. Alternatively, we could have invited a VIC DNSP into the consortium to broaden our data base.

Lesson #5

Data quality is a problem

Expectation

The project team expected to encounter a range of data quality issues

Experience

As part of the work plan for project SHIELD both the University of Queensland as well as GridQube have run simulations and algorithms using the enhanced data streams provided through the Luceo Energy platform from the data sources mentioned above.

- During these simulations and calculations, it has become evident that data from the various sources has various levels of "usefulness" depending on the combined accuracy of the data
- As expected, data that comes in high fidelity with low measurement tolerance is the most useful data for simulation purposes that are not conducted in real time.
- A combination of connection point data at the household node in sufficient density paired with transformer data seems to provide the best possible outcomes as it allows to run the simulation in both directions and establish the value of data at various locations along the feeder.
- Roughly 60% of the simulations generated had results that were deemed to be unreliable. This was due to a variety of issues, including inaccurate voltage measurements, incomplete transformer monitor measurements, unsynchronized timestamps to name a few.

Lessons Learned

Data quality from third party vendors remains an issue.

What We Would Do Differently

Invest in time at the beginning of the project to better understand the simulation requirements and then benchmark various data sources against those requirements and focus the data acquisition on those with the best fit for purpose.

Lesson #6

There is a lack of unification in terminology

Expectation

The project sought to collect data from vendors and manufacturers of a range of devices and assumed that we would be able to simply, quickly and clearly explain our needs and they could determine their willingness to participate.

Experience

The project found that explaining our needs was not simple, quick or clear. The data was to be used in an electricity network environment and many of the vendors and manufacturers proved to be unfamiliar with that environment or to have different or varied understandings of that environment. Consequently, terms like low voltage, network model, distribution transformers or even Smart DER occasionally created different expectations with the vendor. Often, this disconnect was not identified until the data was actually received.

Even simple terms such as an 'LV feeder' needed clarification as LV feeders are called 'circuits' by United Energy in Victoria and what they called 'feeders' is referred to as '11kv feeders or MV feeders' by the project team.

Lessons Learned

It became evident that consolidating our understanding of terms, jargon and nomenclature as the industry digitises would be beneficial.

What We Would Do Differently

The energy industry is highly complex and, hence, establishing a single lexicon across the field would facilitate clearer and more efficient communication. As the project has progressed, it has become obvious the team needs to be careful about any assumptions regarding definitions and that clear conversations need to be held.

Lesson #7

Building a simulator to help guide networks in their placement of monitoring devices within the LV network proved to be challenging but offered some early discoveries

Expectation

The project wanted to build a simulator that would help network owners to identify the most effective locations to place network LV monitoring to maximise accuracy of results.

The plan was to start with the largest set of data points we could collect and then systemically remove data points and consider the impact of that removal on previously calculated accuracy of the hosting capacity estimation. Work could then be undertaken to understand

- What impact the number of data points had on accuracy
- What impact the context of each removed data source had on the accuracy (improved/reduced.

Experience

Our Upper Coomera trial feeder contained 53 data monitors that were used to feed our state estimation algorithm.

Very quickly, it became obvious that the number of combinations of data sources that could be considered with just 53 data sources was astronomical:

- Mathematically it is 253-1 = 9,000 trillion possible combinations

Clearly it was not possible to create a test case for each of these combinations and then run the combination through the algorithm and analyse the trends in the available time.

UQ's scientists and Electrical Engineers spent some time considering the right approach and settled on taking a smaller sample (1,000 combinations) and seeking to generalise based upon results.

This approach yielded some interesting results which will be explored further in stage 2.

Lessons Learned

Early results using this approach has shown

- In the absence of high-density connection point monitoring data, data collection from Transformer monitors are the most valuable data when utilising state estimation
- Commercial sites, because of their wide variability, are also important
- The number, location and context of residential data points is still being explored.

Conclusion and Summary

This report is the third lessons learned from Project SHIELD which coincides with the end of Stage 1 of the project (utilising third party data to understand LV hosting capacity in low AMI networks),

As the project moves into Stage 2, it will augment the current data sets with additional data points installed across the network and many of the analytics conducted to date will be replicated with the aim of understanding how to improve outcomes.

We expect that Stage 2 will yield deeper insights into some of the lessons to date and provide a contrast between independently sourced heterogeneous data and specifically designed network monitoring devices.

One thing is clear though, the Energy supply chain is starting to transition to a more data driven industry however much is still needed to be put in place to optimise the opportunities that could flow from increased availability of data and increased understanding of its uses.