

Notes on Completion: Please refer to the appropriate NIA Governance Document to assist in the completion of this form. The full completed submission should not exceed 6 pages in total.

Network Licensees must publish the required Project Progress information on the Smarter Networks Portal by 31st July 2014 and each year thereafter. The Network Licensee(s) must publish Project Progress information for each NIA Project that has developed new learning in the preceding relevant year.

NIA Project Close Down Report Document

Date of Submission

Jun 2022

Project Reference

NIA_WPD_048

Project Progress

Project Title

Presumed Open Data (POD)

Project Reference

NIA_WPD_048

Funding Licensee(s)

WPD - Western Power Distribution (East Midlands) Plc

Project Start Date

January 2020

Project Duration

1 year and 6 months

Nominated Project Contact(s)

Sam Rossi Ashton

Scope

The scope of this project is to deliver on recommendations, made by the Energy Data Task Force in it's a Strategy for a Modern Digitalised Energy System Report.

Recommendation 2 - Maximising the value of data

Government and Ofgem should direct the sector to adopt the principle that Energy System Data should be Presumed Open, using their range of existing legislative and regulatory measures as appropriate, supported by requirements that data is 'Discoverable, Searchable, Understandable', with common 'Structures, Interfaces and Standards' and is 'Secure and Resilient'.

Recommendation 3 - Visibility of data

A Data Catalogue should be established to provide visibility through standardised metadata of Energy System Datasets across Government, the regulator and industry. Government and Ofgem should mandate industry participation though regulatory and policy frameworks.

Objectives

The project has two objectives:

1. **Maximise the visibility of data.** The data hub will make data discoverable, searchable, and provide visibility through standardised metadata.

2. **Maximise the value of data.** The data hub will make understandable by employing common structures and interfaces.

Success Criteria

The project will be deemed successful if an implementation plans for a functional energy data hub is published and an initial set of network data is used by customers to facilitate their own objectives.

Performance Compared to the Original Project Aims, Objectives and Success Criteria

1. Data Discovery & Classification – Document at least 3 of the main datasets from each of the functional areas identified.

· Completed: During the course of the Work Package over 60 datasets were identified and documented in a data catalogue. During work package 2 the catalogue has been expanded and developed and now includes over 100 separate entries

2. Use Case Development – Produce a number of compelling use cases from a range of stakeholders across the energy industry that can be used to demonstrate the value of the data sources identified in work package 1.

· Completed: Approximately 60 external stakeholders attended two POD workshops. These workshops culminate in the generation of 140 use cases which were then rationalised to 109. These use cases have since been used to prioritise WPD's data work programme

3. Data Openness Assessment – produce an easy to understand and easy to use open data triage tool that can be used to advise on the sensitivities of data and the necessary modifications to the data

· Completed: The Data Sharing Assessment Tool developed during the POD project and the associated user guide is now being used by WPD as BaU.

4. Data Hub Development – deploy a data hub which can host all the data necessary for the data science challenge, which includes a login system and a mechanism for

· Completed: The project successfully developed a trial data hub which demonstrated the functionality required to facilitate the release of previously closed datasets. This hub was used for the data science challenge.

5. Data Science Challenge – successfully run a data science challenge which has a comparable number of participants to its peers

· Completed: Over 360 individuals registered for the event with 55 teams consisting of 142 individuals entering at least one submission for the challenge. Of these 37 finished the full challenge. The challenge also received a wide range of participants, from consultants, energy sector data scientists and university researchers, from at least 72 different institutions and 15 different countries.

6. Data Documentation - Draft a data playbook to guide LNOs through the process of identifying, assessing, publishing or sharing and getting feedback on data.

· Completed: A data playbook has been successfully published and WPD plans to use this to develop data treatment policy in the near future.

Required Modifications to the Planned Approach During the Course of the Project

Proposed change: Extend closure date of WP2 from 05th June 2020 to 30th June 2020 to facilitate the orchestration of virtual workshops instead of face-to-face ones.

Reason for change: COVID restrictions prevent face to face workshops

Effect of NOT making change: The project will likely not have enough time for platform testing, invitations, workshops and then write-up before existing WP2 closure date.

CR2 – December 2020

Proposed change: An extension to WP5 (Data Science Challenge) and the addition of a work package to create a documentation suitable to transition learning to policy

Reason for change: Additional Work Package: The driving motivation behind the project was an Energy Data Task Force recommendation to make data Discoverable, Searchable and Understandable. We are keen to implement this as soon as practicable and as such chose to develop documentation for the process outlined below based on the Data Best Practice Guidance, with supporting resources where required, which can be used for policy development.

Effect of NOT making change: Use of project learning and delivery of Energy Data Task Force recommendation would not be optimal.

CR3 – April 2021

Proposed change: An extension to project end date to June 2021.

Reason for change: A late invoice meant that the project had to remain open until it was paid.

Effect of NOT making change: Project finances close and so unable to pay all invoices.

Lessons Learnt for Future Projects

Work Package 1 - Data Discovery & Classification

The primary takeaway from this work package is that WPD has a wealth of valuable information across the organisation, the value of which will be easy to maximise if the barriers to entry are lowered. Some of the decisions that have been taken to make DNOs more secure have a negative impact on their ability to collaborate with external teams and adopt remote working.

Work Package 2 - Use Case Development

Within Work Package 2 the POD team has engaged with a large number of stakeholders across the business and industry. We found that both WPD and external stakeholders were keen to participate in our workshops and were able to offer some very valuable insight. In fact, given the distributed nature of WPD staff and external stakeholders we believe that providing virtual events may have increased the diversity and number of participants.

The workshops have shown that WPD data has significant value to stakeholders outside of the normal groups. From sharing data more effectively within WPD to making more information available to external parties there is significant value to be extracted.

Work Package 3 - Data Openness Assessment

Within Work Package 3, the POD team has engaged with stakeholders across the business to review and test the developed tool with a number of different datasets.

Within this work package the project team have tested the Data Sharing Assessment Tool with real datasets to understand if it meets the need of the end users. This has been critical to identify additional needs, find where the tool is not working as planned or as the user expected and check that the tool delivers valid recommendations.

The concept of Presumed Open is relatively new for the energy sector and does not always come naturally for an organisation that operates critical national infrastructure and is naturally risk averse. In our experience, individuals who have assisted with the triage of datasets have all of the knowledge and skills needed to identify and sensitively mitigate issues however the risk averse nature of the energy sector creates doubt and worry that they have 'missed something'.

For Data Sharing Assessment Tool and the Presumed Open Data project to gain traction and truly become Business As Usual it will be essential to ramp up engagement with stakeholders across the organisations. Providing feedback on assessments completed and additional training where required.

Work Package 5 – Data Science Challenge

The learnings from the data science challenge have suggested some key points for development of future challenges:

- Allowing participants to use external data would make a much more diverse solution and perhaps identify useful data sets for the challenge. This must be carefully considered, since it allows performance to be driven by data accessibility and could lead to uneven playing field for teams with lower resources or accessibility. One option is to split the challenge into two tracks, one with and one without restrictions.
- Time for participants to work on a task should be considered. Participants often take part in the challenge in their own time and hence longer gaps between submissions can be advantageous but also require longer commitments. One compromise could be to reduce the number of tasks but make the test set much larger to allow more robust assessment.
- It is beneficial to choose a realistic problem which can be easily framed as a well-defined competition problem. However, this is not always trivial. More realistic problems are easier to understand but are not always easy to score, and well-defined competitions are not always realistic which can cause confusion. A balance must be sought when developing future challenges.
- Make sure the data is as clean as possible. Although this challenge used high quality data even the small number of erroneous values provided some difficulty and added pre-processing time for some participants.
- Consider skill scores rather than absolute metrics to reduce biases and variations in the tasks.
- A practice challenge is essential. It helps to work out bugs in the submission process and helps clarify the challenge requirements to the participants.
- Regular engagement with the teams is well-received. The LinkedIn Forum and emailing list helped facilitate information sharing and for teams to discuss with each other and clarify minor points from the challenge. It also ensured that information was fairly shared with all participants.

Note: The following sections are only required for those projects which have been completed since 1st April 2013, or since the previous Project Progress information was reported.

The Outcomes of the Project

The outcomes from this project include:

- Learning Reports generated at the end of each of the six work packages.
- Our Data Process Team is now using the Data Playbook developed during Work Package 6 to formalise the policy and process required to fully adopt the Data Sharing Assessment Tool developed during Work Package 3. This will enable third parties to request previously unshared WPD datasets
- The Energy Networks Association (ENA), with the help of WPD, used the Data Playbook developed by the project in Work

Package 6 to create the ENA Data Triage Playbook, this in-turn was used to develop the Energy Data Request Tool which enables a single point for data requests and standardising networks' approach to servicing these requests.

· Additionally, due to the success of the Work Package 5 Data Science Challenge, we are looking to setup additional challenges to engage the data science community with energy problems.

Data Access

The POD project was largely concerned with the study and treatment of existing datasets rather than creating new ones. There was however analysis done on the use cases generated in Work Package 2 to group them into themes. This analysis is available upon request at the link below: www.westernpower.co.uk/Innovation/Contact-us-and-more/Project-Data.aspx

Foreground IPR

The IP generated throughout this project includes:

- The WPD POD Data Catalogue
- The Use Case Assessment Analysis
- The Data Sharing Assessment Tool
- The POD Open Data Hub
- The Data Playbook

Planned Implementation

We are scaling up our capability in digitalisation immediately as a result of this project. Following on from the project there are the following intentions.

Data science challenges

Following the success of the first data science challenge and to incentivise a regular data release schedule, the intention is to run a new data science challenge every quarter with new data published and documented. Over time this will allow the teams involve to upskill and learn by doing practical projects, lead the rest of the energy sector in terms of data publication, and allow WPD to provide new data use cases to guide the digitalisation strategy.

Digitalisation strategy

WPD intends to continue and embed the learnings from this project by continuing to grow the team that are working on digitalisation. By working with external parties that want to access and use data, publishing data on a regular cycle and responding to user feedback, WPD will develop a much closer relationship with the interested parties in the sector and take a leading role in enabling innovation in energy. The data release playbook will be used as a basis for future iterations and the processes will be improved and updated over time as more data is published and further data science challenges are conducted.

Other Comments

N/A

Standards Documents

N/A