WESTERN POWER DISTRIBUTION
Serving the Midlands, South West and Wales

Date of Submission

NIA Project Registration and PEA Document

Notes on Completion: Please refer to the **NIA Governance Document** to assist in the completion of this form. Please use the default font (Calibri font size 10) in your submission. Please ensure all content is contained within the boundaries of the text areas. The full-completed submission should not exceed 7 pages in total.

Project Registration

Project Title

Spatially Enabled Asset Management (SEAM)

Project Reference

Funding Licensee(s)

Western Power Distribution (South West)

Project Start Date

16/11/2020

Project Duration

Years	Months
0	11

Nominated Project Contact(s)

Jenny Woodruff

Project Budget

£426,298

Contact Email Address

wpdinnovation@westernpower.co.uk

Lead Sector

Electricity Distribution	☒	**Gas Transmission**	☐
Electricity Transmission	☐	**Gas Distribution**	☐

Other Sectors

Electricity Distribution	☐	**Gas Transmission**	☐
Electricity Transmission	☐	**Gas Distribution**	☐

Research Area

Network improvements and system operability	☐
Transition to low carbon future	☐
New technologies and commercial evolution	☒
Customer and stakeholder focus	☐
Safety, health and environment	☐

Problem(s)

Geographic Information Systems for utilities have been created by the digitisation of paper records. These have then undergone transformations as data has been moved between one system and another and manipulations such as the corrections to the Master Map background. It is expected that there will be inaccuracies in the current records and these can persist for many years partly due to the lack of visibility of buried assets and partly due to the length of time between sites requiring site visits that would be expected to update the GIS.

Inaccuracies in the GIS system have the potential to impact:

- Accuracy of network modelling
- Accuracy of regulatory reporting
- Field safety
- Network operational efficiency
- Network upgrade/maintenance efficiency
- Accuracy of New Connections information
- Accuracy of information provided to third parties

At the same time the process to correct identified GIS problems, for example a micro-disconnect issue may be highly manual and therefore costly and time consuming. Therefore, the use of an AI engine to police and correct this essential GIS data is a more effective and efficient method of risk reduction.

Similarly, when performing power systems analysis on the networks, data gaps (such as missing cables and asset types) can be a key issue and can greatly increase the analysis time due to time spent fitting data and cleansing the set prior to carrying out the analysis. With the data having potentially many users within and outside of WPD, every time data gaps are filled this not only duplicates effort but is likely to result in different assumptions being made and potentially different conclusions being drawn. A process that can reduce data gaps consistently and accurately would help resolve these issues.

Method(s)

The SEAM project aims to investigate how Machine Learning (ML) can be employed to carry out data cleansing and data gap closure. The ML model will be trained using an existing dataset and then the ability of the model to successfully identify and correct data issues will be evaluated with a separate dataset that has had errors introduced. This will be followed by application to an unaltered WPD dataset with the results being compared

to the errors identified by WPD's Integrated Network Model

Scope

The project will include LV and 11kV networks which represent the bulk of GIS data. 33kV networks are included to provide a comparison between the issues identified by the ML algorithm and those flagged by the Integrated Network Model. To enable the use of INM data, the South West region will be used though the approach could be applied to any area.
The project will investigate model inputs, outputs and different machine learning algorithms with the final model incorporated into a user interface. The design and algorithms used will be documented to enable knowledge transfer.

Objective(s)

In line with the overall objective of creating and testing a machine learning algorithm to identify and propose fixes for GIS data issues, the project objective are to;

- Generate of potential hypotheses to test and use cases for the tool to be applied to
- Understand the data available to support the machine learning proof of concept
- Outline of the model design including selection of machine learning algorithms.
- Create a final cleaned and prepared dataset that will be used to train and develop the model.
- Provide an interim report that sets out early findings from the modelling and direction for the remainder of the project.
- Develop the final version of the PoC model and front end.
- Carry out statistical evaluation of the model and accuracy through comparison of the model outputs with baseline and training datasets.
- Carry out data cleaning and loading of selected network area, including schematics if available in the format of a connectivity and impedance electrical model of EHV, HV and LV networks.
- Provide a summary of key findings, assessment of outcomes against success criteria, recommendations and learnings to be shared.

Success Criteria

The project will be judged successful if the following criteria are met.

- A standalone AI Model has been developed tested and applied to a dataset in the agreed regional area.
- The model performance has been evaluated and the application to the wider GIS data landscape assessed.
- The approach to roll into business-as-usual has been assessed with recommendations.
- Key learnings have been identified and shared with other DNOs.

Technology Readiness Level at Start	Technology Readiness Level at Completion
3	6

Project Partners and External Funding

Capgemini will act as the main contractor providing expertise on machine learning and GIS systems
WSP, engineering consultants, will bring expertise on distribution networks and required DSO capabilities.
No external funding is provided.

Potential for New Learning

The learning generated will relate to the core project aims. i.e.

1) To what degree can machine learning be applied to GIS data to find and fix the more complex errors?
2) Can the ML approach provide useful indices for confidence in the presence of an error or in the proposed solution?

It is expected that additional learning will be generated along that way e.g.

- What types of error are most prevalent in GIS data?
- What types of error can be corrected using only the GIS data itself?
- What types of error require additional data from other systems to provide the context for the ML algorithms?

Scale of Project

The scale of the area required to build and test the model is expected to be equivalent to a BSP group but not to be as large as a single DNO area. The project will focus on 11kV (or 6.6kV) and LV networks with 33kV networks also included. 132kV networks are unlikely to have the volume of data required to support ML techniques.

Geographical Area

This will be determined early in the project once the data has been assessed. The requirement for INM data availability to compare to the 33kV network suggests that the South West region may be the most practical option. The project could be applied to any of the four WPD DNO areas and the types of GIS issue are expected to be similar in all regions though the rules for correction may be different. Data for all regions has now been migrated into Electric Office from the predecessor GIS system, Bentley Microstation

Revenue Allowed for in the RIIO Settlement

Not applicable

Indicative Total NIA Project Expenditure

The project budget is £426,298 with WPDs contribution of £ 42,630 reducing the required funding to £383,668

Project Eligibility Assessment

Specific Requirements 1

1a. A NIA Project must have the potential to have a Direct Impact on a Network Licensee's network or the operations of the System Operator and involve the Research, Development, or Demonstration of at least one of the following (please tick which applies):

A specific piece of new (i.e. unproven in GB, or where a Method has been trialled outside GB the Network Licensee must justify repeating it as part of a Project) equipment (including control and communications systems and software)	☒

A specific novel arrangement or application of existing licensee equipment (including control and/or communications systems and/or software)	☐
A specific novel operational practice directly related to the operation of the Network Licensees System	☐
A specific novel commercial arrangement	☐

Specific Requirements 2

2a. Has the Potential to Develop Learning That Can be Applied by all Relevant Network Licensees ☒

Please answer one of the following:

i) Please explain how the learning that will be generated could be used by relevant Network Licenses.

All DNOs are likely to face similar challenges with the accuracy of data within their GIS and other systems. If an AI approach is found to be successful, then the approach can be replicated even though the format of the data and the prevalence of data issues will be different for each DNO region.

ii) Please describe what specific challenge identified in the Network Licensee's innovation strategy that is being addressed by the Project.

Is the default IPR position being applied?

Yes	☒
No	☐

If no, please answer i, ii, iii before continuing:

i) Demonstrate how the learning from the Project can be successfully disseminated to Network Licensees and other interested parties

ii) Describe how any potential constraints or costs caused, or resulting from, the imposed IPR arrangements

iii) Justify why the proposed IPR arrangements provide value for money for customers

2b. Has the Potential to Deliver Net Financial Benefits to Customers ☒

Please provide an estimate of the saving if the Problem is solved.

Approximately £1m over all the 4 WPD license areas.

Please provide a calculation of the expected financial benefits of a Development or Demonstration Project (not required for Research Projects). (Base Cost – Method Cost, Against Agreed Baseline).

Base Cost. This assumes manual identification and fixing of those GIS errors that are hardest to fix and cannot easily be automatically resolved using existing tools. The cost to resolve "harder to fix" issues is assumed to be £2m based on an estimate of 2 network issues per linear km of cable/line representation and 1% of substations i.e. approx. 88,000 issues to fix and 20% of issues falling into this "harder to fix" category. Assuming GIS operatives with a salary plus oncost of £37,200 per annum having 230 productive days a year and correcting 7 "difficult-to-fix" errors a day suggests a resolution cost for the more problematic errors of approx. £23 an error and therefore £2m to fix the GIS errors.
The Method cost assumes a 50% reduction in time spent addressing these issues and is therefore £1m.
Base Cost less Method cost = approx. £1m per annum.
This excludes the one-off cost for establishing and configuring the machine learning which is estimated to be around £200k per installation subsequent to this project.

Please provide an estimate of how replicable the Method is across GB in terms of the number of sites, the sort of site the Method could be applied to, or the percentage of the Network Licensees system where it could be rolled-out.

This method could be applied by all the other DNOs.

Please provide an outline of the costs of rolling out the Method across GB.

Roll out cost across GB estimated at £200k per DNO group i.e. £1m for the five other DNO groups excluding WPD. While the establishment of a subsequent versions of this tool will be cheaper than the initial creation due to the learning gained, each installation will need to be provided with the relevant data and the machine learning algorithms rebuilt and refined to reflect that data.

2c. Does Not Lead to Unnecessary Duplication ☒

Please demonstrate below that no unnecessary duplication will occur as a result of the Project.

We are aware that other work has taken place in this area, notably NIA_SPEN0016NCEWS (Network Constraint Early Warning System) by Scottish Power Energy Networks which included the use of machine learning approaches to determine likely cable types where this data is missing.
To avoid unnecessary duplication, SEAM will include a workshop to compare approaches, along with representatives of other DNOs actively working in similar areas, after SEAM's initial data evaluation and use case development phase has taken place. If there is sufficient commonality between the datasets available and their quality, the modelling approach used by NCEWS to propose cable types where data is absent will be replicated. Similarly the previous findings for machine learning models that were judged to be less successful will be used to influence the data exploration and model development.

If applicable, justify why you are undertaking a Project similar to those being carried out by any other Network Licensees.

The degree of potential overlap between SEAM and NCEWS appears to be a small proportion of the scope of SEAM, i.e. the population of missing cable type data. The projects have different emphasis in that SEAM is looking to improve the overall quality of data to enable the use by third parties and support wider asset management activity within WPD, such as estimation of health indices and ensuring the best information is available to operational users to understand feeding arrangements. NCEWS was focussed on providing the connectivity model and associated impedance data required for powerflow analysis whereas the creation of an LV connectivity model is not an objective for SEAM.

Additional Governance Requirements

Please identify that the project is innovative (i.e. not business as usual) and has an unproven business case where the risk warrants a limited Research, Development or Demonstration Project to demonstrate its effectiveness ☒

i) Please identify why the project is innovative and has not been tried before

The project is innovative in applying a ML algorithm to identify and correct data inaccuracies. This is a process that would normally require manual intervention and is not yet a widely applied approach to data cleansing.

ii) Please identify why the Network Licensee will not fund such a Project as part of its business as usual activities

This innovation project is a novel approach requiring expertise in machine learning techniques which is not yet part of WPD's business as usual processes or skillsets.

iii) Please identify why the Project can only be undertaken with the support of the NIA, including reference to the specific risks (e.g. commercial, technical, operational or regulatory) associated with the Project

There is a technical risk associated with the machine learning model production. This is thought to be the first time this approach has been applied to GIS data and positive results cannot be assumed.

Additional Registration Questions

These are required for summary section of registration; some areas can be copied from sections above.

Technologies (select all that apply)

☐ Active Network Management	☐ Environmental	☐ Network Monitoring
☐ Asset Management	☐ Fault Current	☐ Overhead Lines
☐ Carbon emission Reduction Technologies	☐ Fault Level	☐ Photovoltaics
☐ Commercial	☐ Fault Management	☐ Protection
☐ Condition Monitoring	☐ Harmonics	☐ Resilience
☐ Community Schemes	☐ Health & Safety	☐ Stakeholder Engagement
☒ Comms & IT	☐ Heat Pumps	☐ Substation Monitoring
☐ Conductors	☐ High Voltage Technology	☐ Substations
☐ Control Systems	☐ HVDC	☐ System security
☐ Cyber Security	☐ Low Carbon Generation	☐ Transformers
☐ Demand Response	☐ LV & 11Kv Networks	☐ Voltage Control
☐ Demand Side Management	☐ Maintenance & Inspection	☐ Gas Distribution
☐ Distributed Generation	☐ Measurement	☐ Gas Transmission
☐ Electric Vehicles	☐ Meshed Networks	☐ Electricity Distribution
☐ Energy Storage	☐ Networks Automation	☐ Electricity Transmission

Project Short Name

SEAM

Project Introduction

Over time, various factors have adversely affected the quality of DNO's Geospatial Information System (GIS) data. If datasets are shared with patterns of error, then different users will fill these error gaps in different

ways leading to inconsistent results from analysis and how the data is exploited by applications. A Machine Learning (ML) tool is proposed to carry out data cleansing and data gap closure of WPD's network GIS and relevant network data. The ML tool will be trained and populated with existing network asset data and run to identify data gaps. The model will initially be developed based on a set of business rules that will be updated based on the outcomes of the ML algorithms. Appropriate will be testing carried out to validate the results and assess the accuracy of the model.

Project Benefits

The project will focus on errors that cannot currently be identified and fixed automatically. By highlighting potential errors and suggesting suitable fixes this will improve the usability of the GIS datasets by third parties, particularly where issues relate to missing data. Similarly, this is expected to reduce the time taken to manually identify and fix these GIS issues.

PEA Version	**1**		
	Name and Title	**Signature**	**Date**
Prepared by	Jenny Woodruff		23/09/2020
Approved by	Yiango Mavrocostanti		