NIA Project Registration and PEA Document

Date of Submission

Oct 2020

Project Reference Number

NIA_WPD_054

Project Registration

Project Title

Spatially Enabled Asset Management (SEAM)

Project Reference Number

NIA_WPD_054

Project Licensee(s)

National Grid Electricity Distribution

Project Start

November 2020

Project Duration

1 year and 0 months

Nominated Project Contact(s)

Jenny Woodruff

Project Budget

£426,298.00

Summary

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.

Third Party Collaborators

Capgemini

WSP UK Limited

Problem Being Solved

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.

Consumer Vulnerability Impact Assessment (RIIO-2 Projects Only)

n/a

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.

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.

Technology Readiness at Start

TRL3 Proof of Concept

Technology Readiness at End

TRL6 Large Scale

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 the RIIO Settlement

N/A

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 Part 1

There are slightly differing requirements for RIIO-1 and RIIO-2 NIA projects. This is noted in each case, with the requirement numbers listed for both where they differ (shown as RIIO-2 / RIIO-1).

Requirement 1

Facilitate the energy system transition and/or benefit consumers in vulnerable situations (Please complete sections 3.1.1 and 3.1.2 for RIIO-2 projects only)

Please answer **at least one** of the following:

How the Project has the potential to facilitate the energy system transition:

n/a

How the Project has potential to benefit consumer in vulnerable situations:

n/a

Requirement 2 / 2b

Has the potential to deliver net benefits to consumers

Project must have the potential to deliver a Solution that delivers a net benefit to consumers of the Gas Transporter and/or Electricity Transmission or Electricity Distribution licensee, as the context requires. This could include delivering a Solution at a lower cost than the most efficient Method currently in use on the GB Gas Transportation System, the Gas Transporter's and/or Electricity Transmission or Electricity Distribution licensee's network, or wider benefits, such as social or environmental.

Please provide an estimate of the saving if the Problem is solved (RIIO-1 projects only)

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

Please provide a calculation of the expected benefits the Solution

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

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.

Requirement 3 / 1

Involve Research, Development or Demonstration

A RIIO-1 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 system 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

RIIO-2 Projects

☐ A specific piece of new equipment (including monitoring, control and communications systems and software)

☐ A specific piece of new technology (including analysis and modelling systems or software), in relation to which the Method is unproven

☐ A new methodology (including the identification of specific new procedures or techniques used to identify, select, process, and analyse information)

☐ A specific novel arrangement or application of existing gas transportation, electricity transmission or electricity distribution equipment, technology or methodology

☐ A specific novel operational practice directly related to the operation of the GB Gas Transportation System, electricity transmission or electricity distribution

☐ A specific novel commercial arrangement

Specific Requirements 4 / 2a

Please explain how the learning that will be generated could be used by the relevant Network Licensees

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.

Or, please describe what specific challenge identified in the Network Licensee's innovation strategy that is being addressed by the project (RIIO-1 only)

n/a

☑ Has the Potential to Develop Learning That Can be Applied by all Relevant Network Licensees

Is the default IPR position being applied?

☑ Yes

Project Eligibility Assessment Part 2

Not lead to unnecessary duplication

A Project must not lead to unnecessary duplication of any other Project, including but not limited to IFI, LCNF, NIA, NIC or SIF projects already registered, being carried out or completed.

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.

n/a

Additional Governance And Document Upload

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.

Relevant Foreground IPR

n/a

Data Access Details

n/a

Please identify why the Network Licensees will not fund the project as apart of it's business and 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.

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.

This project has been approved by a senior member of staff

☑ Yes