SP ENERGY NETWORKS

Flexible Networks for a Low Carbon Future

Methodology & Learning Report

Work package 1.2: Data Analytics Trial

July 2015

Contents

1 Executive Summary

This report describes a data analytics trial undertaken in conjunction with IBM. The trial used data collected as part of our Flexible Networks project, together with data extracted from other business systems.

We undertook the trial because, in common with other DNO's, there is an exponential growth in data being collected across the business which is often held in separate systems. As a result there is a need for tools to enable us to pull data from different source systems and efficiently analyse it in a way that provides visualisations and reports that provide business benefit.

Analytics has the potential to improve our cost efficiency, improve our asset management techniques, and facilitate new or improved services to customers.

The original scope of work agreed with IBM is outlined, including the data that SPEN would need to provide and the type of visualisations and reports that the package would provide. For the trial, data relating to the 11kV / 400V secondary networks in St Andrews and Ruabon project trial areas was considered.

The implementation process is described, covering the methods of obtaining the data required and the difficulties that were encountered in obtaining some of the data in the necessary format.

Examples of the output from the analytics package using actual data are provided. These include visualisations of substation location, connectivity and utilisation overlaid on Google Maps. Examples of reports available include a 'Hot Substations' report and a 'Bad Volts and Power Factor' report.

It is concluded that the trial has been a valuable first step towards implementing data analytics and there has been valuable learning on the potential for business benefit and also the difficulties that can arise in obtaining data in the correct format from disparate systems.

2 Background

Commenting on the forthcoming report by the Institution of Engineering and Technology (IET) and the Royal Academy of Engineering (RAE) *"Connecting data: driving productivity and innovation", the IET state "...while there are best practice examples where big data and data analytics have been successfully applied, the area is still largely immature and there remains great potential for innovation and value generation in future years, to the benefit of business and society as a whole."*

All DNO's have been challenged to maximise the use of "Smart" technology to maximise benefits for customers. All projects and business as usual activities that seek to utilise smart technology need a sound data transport, management and analytical basis that allows data to be easily shared between different systems. By allowing all areas of our business to access all data, in a straightforward manner we can both improve our efficiency and enable all future smart projects.

There is an exponential growth in the rate of DNO data intake as illustrated in the graph below. This has been evidenced within the Flexible Networks project through the volume of data generated as a result of enhanced monitoring of secondary substations within the trial areas. However technology such as smart meters will multiply the available data volumes many times.

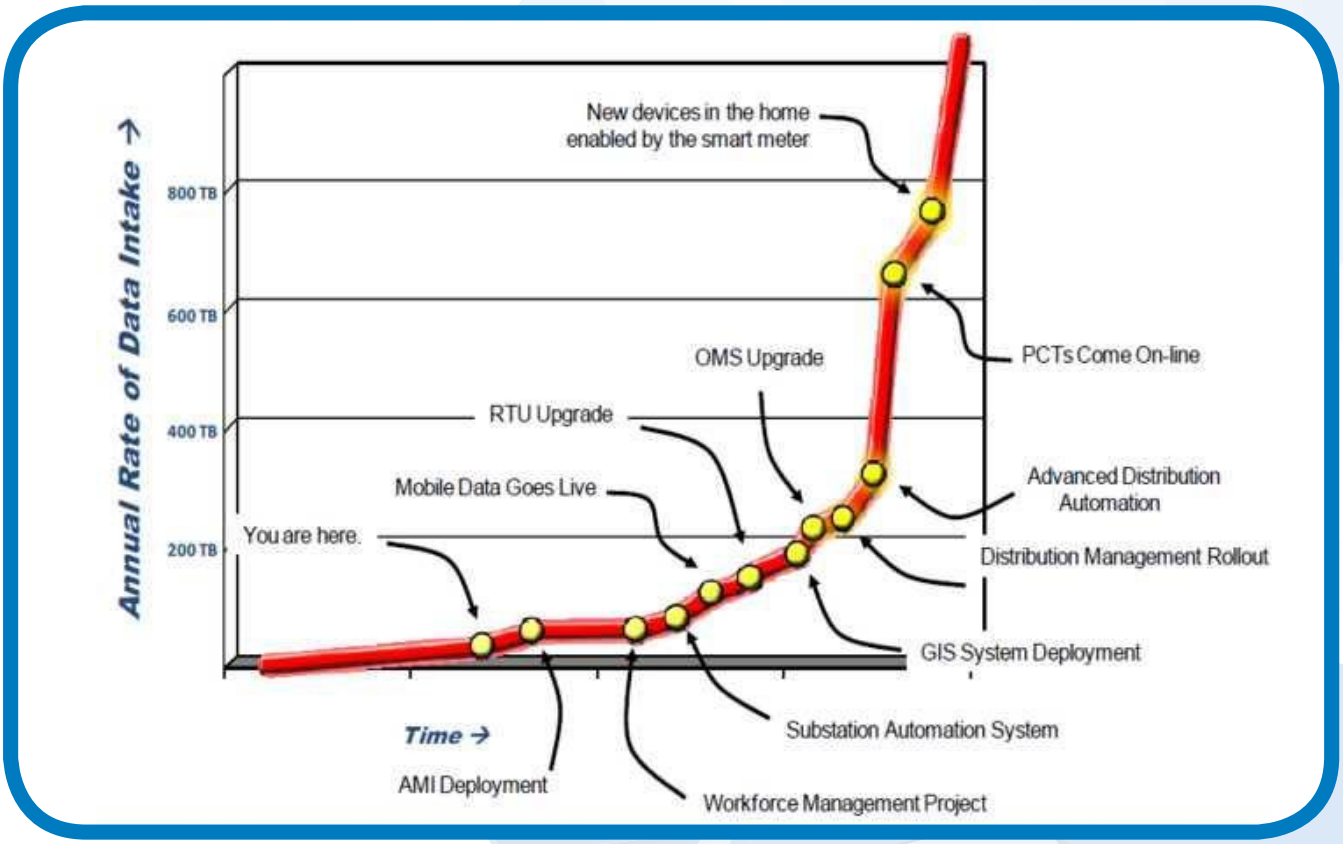


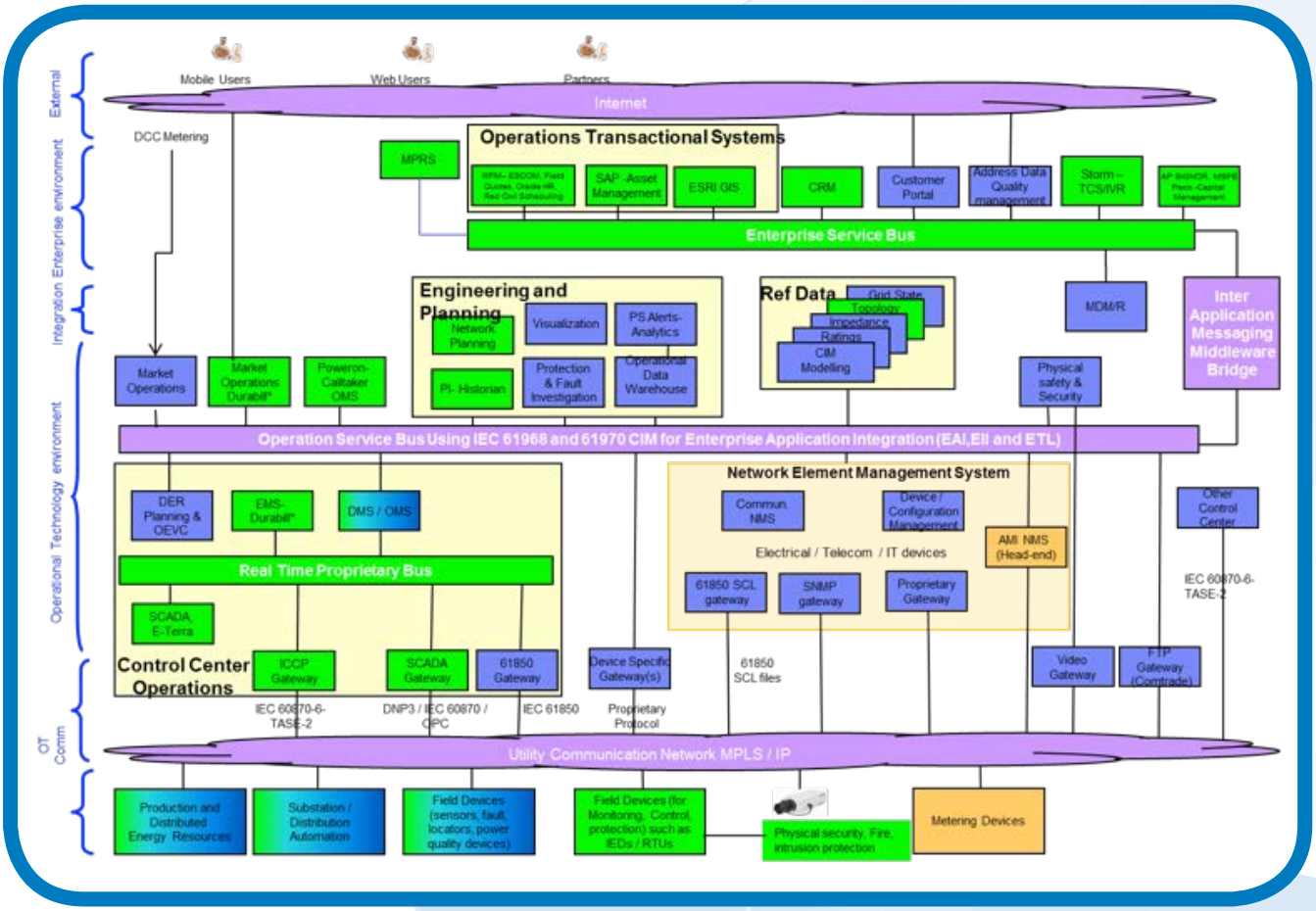
Figure 2.1 Representative data intake for various smart grid milestones. Source: EPRI

2 Background [continued]

The challenge for utilities will be driving insight and actionable intelligence from the vast amounts of data.

Grid Analytics is emerging as a technique to extract value from the data that a DNO holds. We therefore wanted to undertake a trial to help us begin to understand what sort of mechanisms and work practices a DNO should put in place to enable efficient use of new smart data by all areas of the business.

SPEN have already developed a target operational technology model, illustrated in the diagram below, for the transfer of data between systems and the analysis of data. This model was a starting point for development of the trial. The trial explores how we can use data collected from secondary substations as part of the project along with data held in existing systems to create visualisations that create value for the business and customers.



We worked with IBM as a collaborator to trial an implementation of their Distribution Grid Analytics (DGA) package.

3 Details of the work carried out

3.1 Developing the Scope of Work (SOW)

Through discussions with IBM we developed a proposal for IBM to install and support a twelve month pilot to enable users in SPEN to monitor various measurements from selected distribution substations in our distribution grid. The technical solution was to be based on IBM's Distribution Grid Analytics Express (DGA Express) asset and IBM would provide it as a service hosted in an IBM Data Centre (IBM's Smart Cloud Enterprise (SCE) offering) for a period of twelve months.

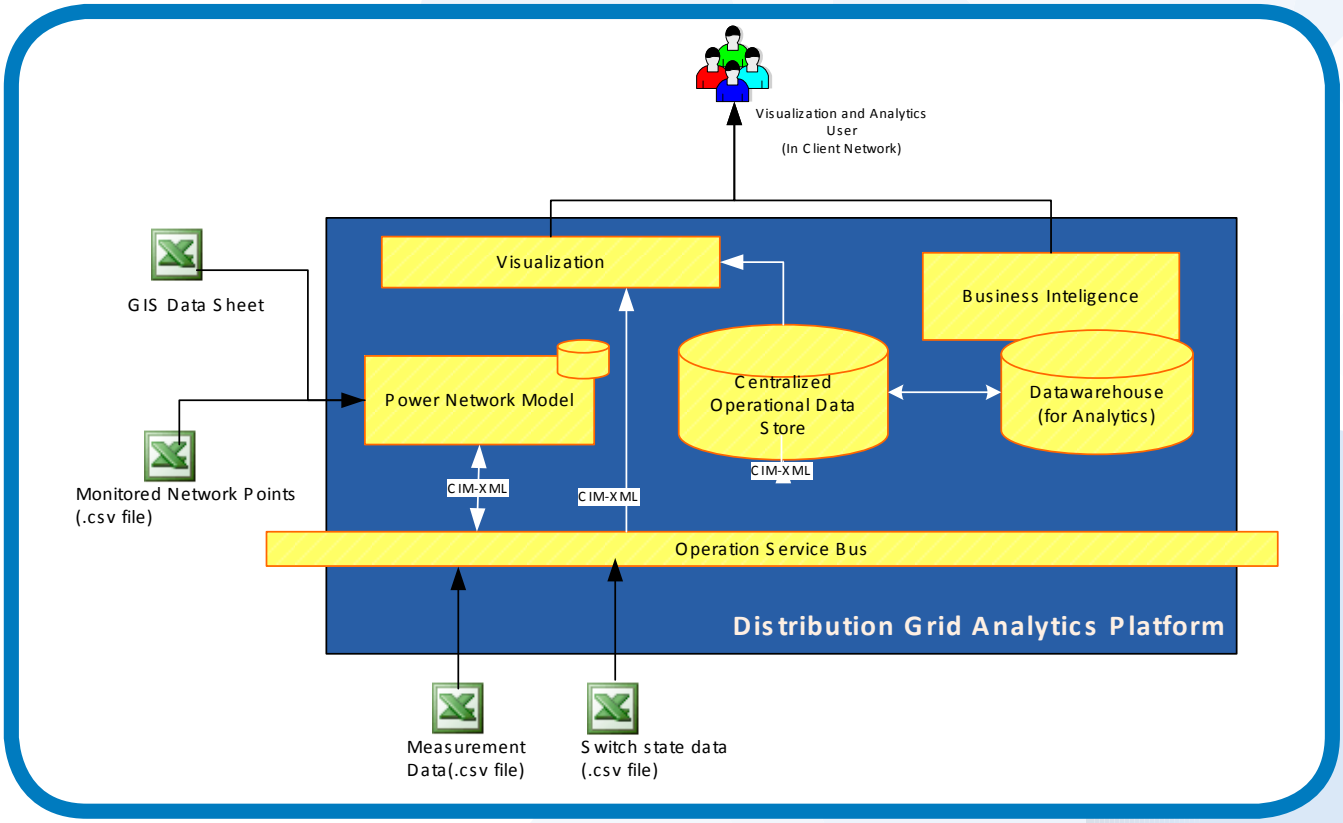
During this time SPEN users would be able to access a web page interface to view;

- grid topology overlaid on Google maps
- grid measurements gathered from existing historical measurement data sources
- analytical reports generated on the data

The services were to be delivered in a progressive fashion over the period of the SOW:

- Initial work to agree the details of the specification.
- An implementation phase.
- A run phase of up to twelve months.

An architectural overview diagram for the pilot solution is provided below.



It was agreed that SPEN users would access IBM hosted applications (Visualisation and Analytics) via a web browser.

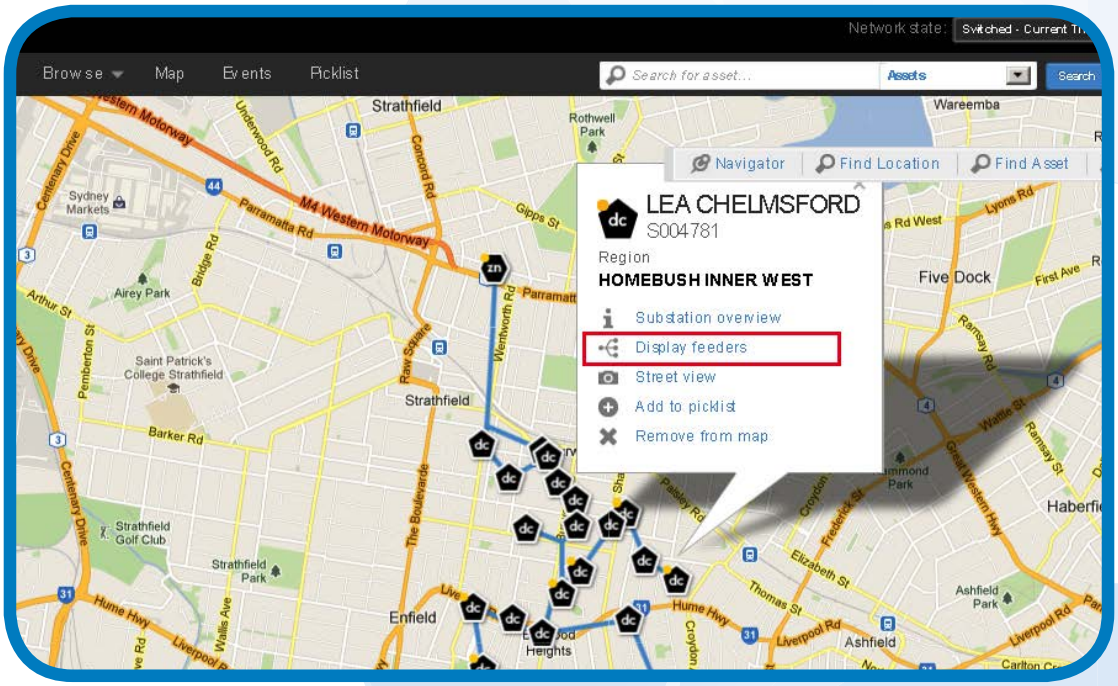
3.1 Developing the Scope of Work (SOW) [continued]

It was envisaged that DGA Express would be provided with four different types of data as described below:

1. Geographical Information System (GIS) Data: To be sourced through a spreadsheet template that IBM specified. SPEN to populate this spreadsheet template and IBM to validate the correctness before loading it into DGA Express.
2. Monitored network points data: To be sourced through a .csv file that IBM specified. SPEN to create this file to let DGA Express know which network points are collecting data.
3. Measurement Data: To be sourced through a .csv file specified by IBM. SPEN to create this file to push measurement data for all monitored network points into DGA Express.
4. Switch State Operational Data: To be sourced through a .csv file specified by IBM. SPEN to create this file to let DGA know about the switching operations in MV network.

It was agreed that the following services would be provided by the trial implementation: -

Visualisation: A front end to view network topology and electrical measurements.



DGA Express Analytics Module

User access to the Cognos reporting module where analytics reports can be run against the historical data (if available). Examples are as follows.

1. **Hot substation Report:** Out of the substations monitored DGA Express will identify the ones which have been over utilized for a selected time frame. It will inform how many times the substation was overloaded and for how long.
2. **Feeder Unbalance Report:** For the substations monitored, DGA Express will prepare a report if there is an unbalance between phases for the selected time frame.
3. **Bad Volts & Poor Power Factor report:** For the substations monitored, DGA Express will prepare a report if there are voltage or power factor excursions outside the configured parameters for the selected time frame.

3.2 Implementation

The first stage in the implementation process was to hold a 'Data Workshop' with participants from IBM and SPEN. The purpose of the workshop was for SPEN to understand the data requirements and format of data required to feed the application. We could then go on to consider the best source for the data and how we could extract it.

The main groups participating in the workshop were:-

IBM Project Management (UK)

IBM Product Specialist (Australia)

IBM Programmers (China)

SPEN Flexible Networks Project Management and Engineering

Nortech (iHost Database)

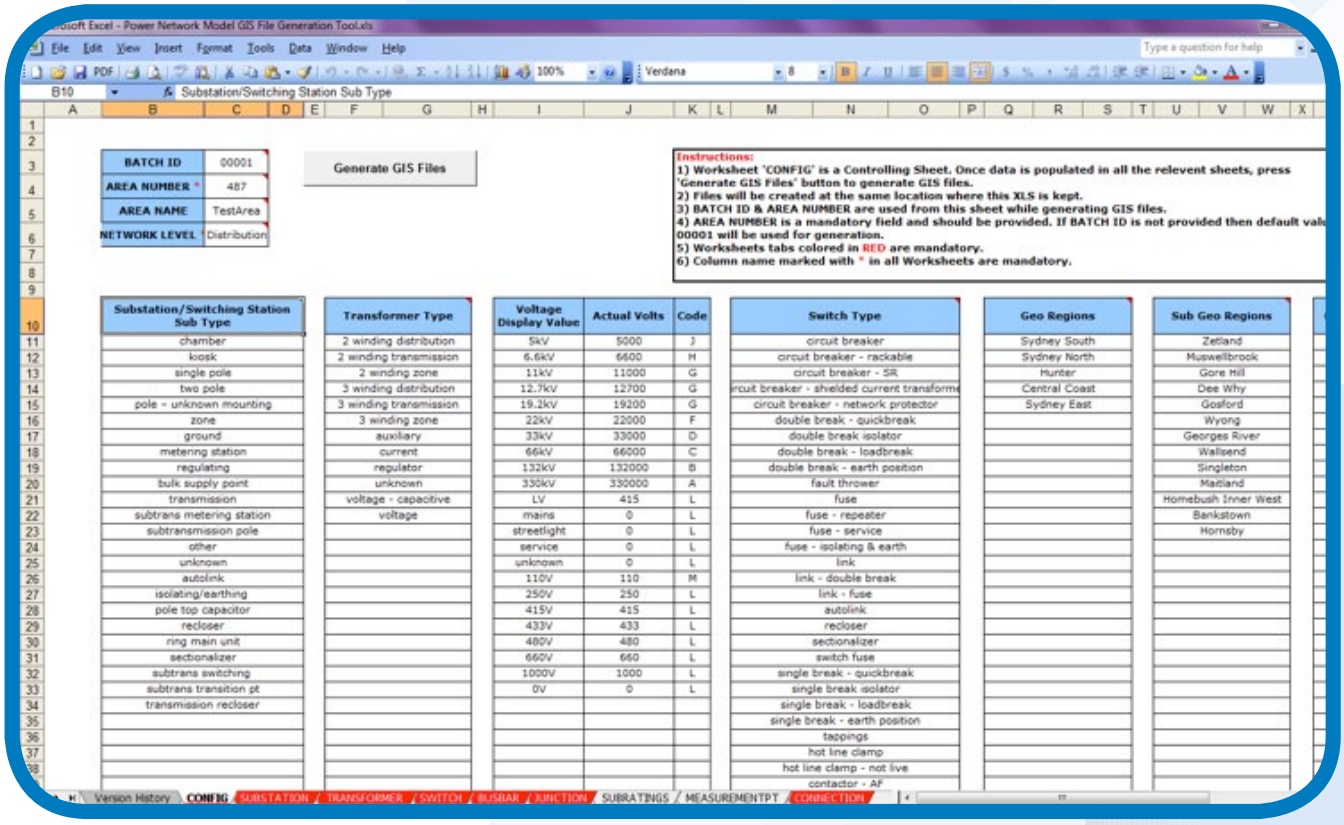
SPEN GIS Analysts

SPEN SCADA Specialists

Considering each of the 4 required spreadsheets in turn, the conclusions of the workshop were as follows: -

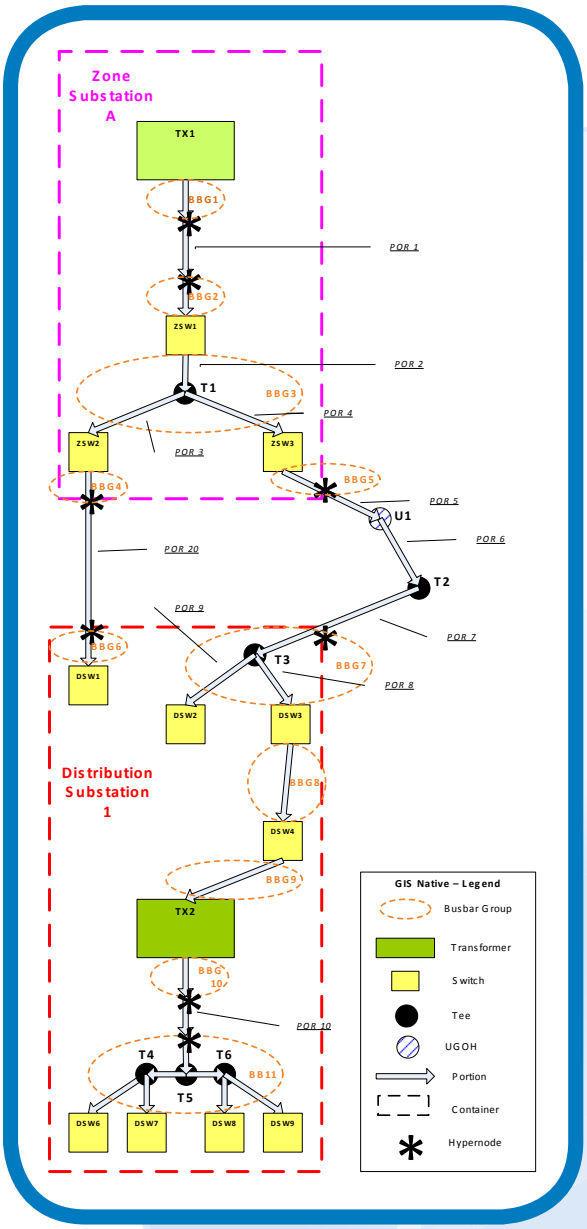
Geographical Information System (GIS) Data

The IBM spreadsheet consists of multiple tabs, an initial 'Config' sheet followed by individual sheets for different network components.



3.2 Implementation [continued]

The diagram below is provided as part of the IBM notes to explain the requirements of the spreadsheet.



The issue that became clear during the workshop was that much of the asset information required for the spreadsheet is available from our ESRI GIS, however connectivity information between network components is required for the spreadsheet and there is no practicable means to obtain this as an output from ESRI. For the SPD area, a report generated from our GE PowerOn Fusion SCADA system helps with connectivity to some extent, however this report is not available for the SPM area. It was therefore concluded that much of the spreadsheet would need to be populated manually by copying and pasting details from the ESRI screens into the IBM GIS spreadsheet.

3.2 Implementation [continued]

Monitored network points data

This spreadsheet effectively provides the link between assets defined in the GIS spreadsheet, and the measurement data, which in the case of the Flexible Networks project is held in the iHost server.

The key is to have a common reference that connects the data and the asset. We found that the best option was to use Energy Networks ID (ENID) numbers. These are held against each asset within the SAP asset register and ESRI GIS. Appropriate ENID numbers were allocated at secondary substation level within the iHost measurement database. It was found that ENID numbers for individual LV feeders are held within ESRI, however these had not been used to identify measurement points within iHost when the system was set up. It was not possible to retrospectively determine the ENID number associated with each measurement point. This is a learning point that will need to be considered in any future installation of monitoring equipment.

Measurement Data

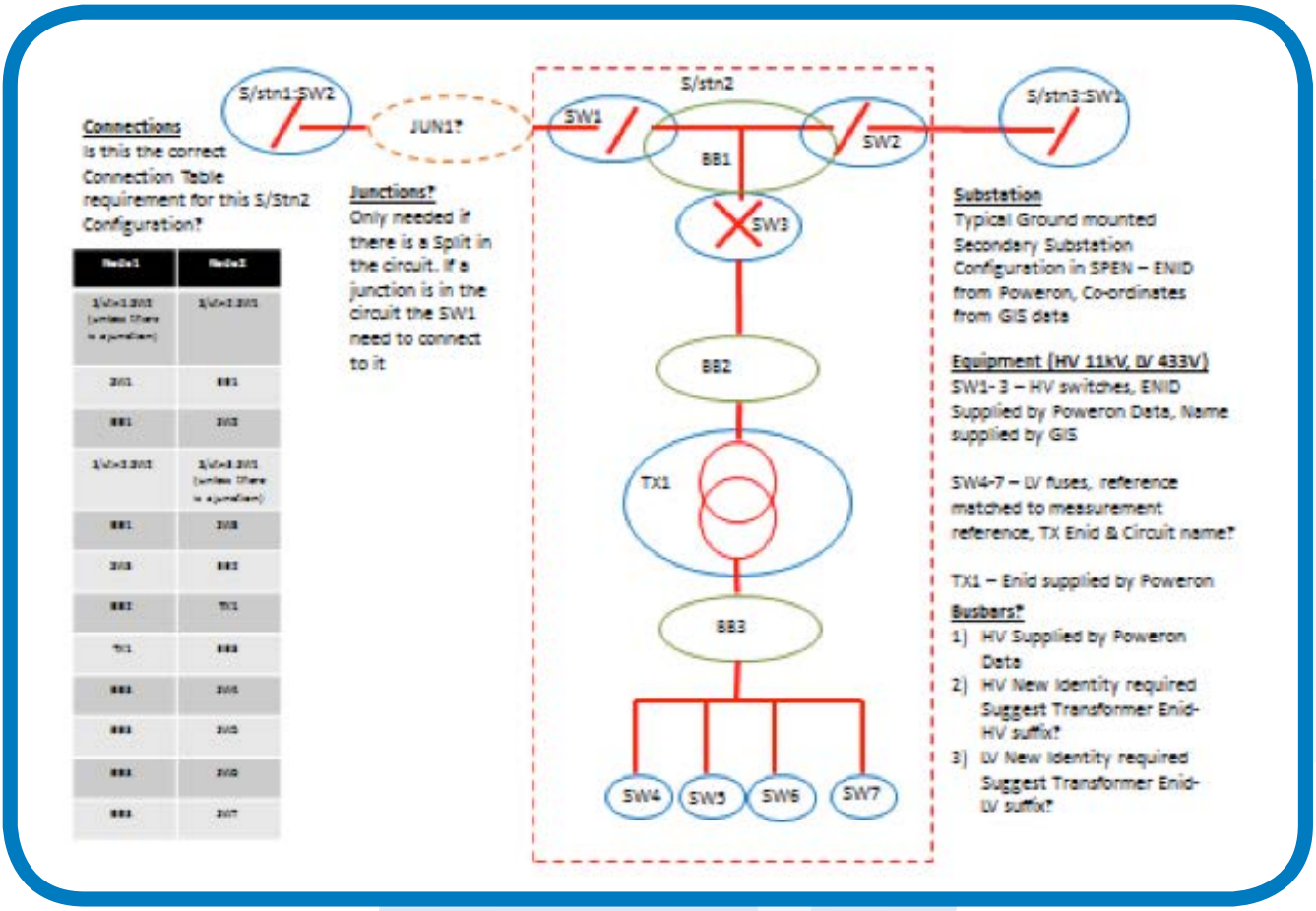
As previously stated the measurement data relevant to the trial is held in a Nortech iHost database. IBM and Nortech were able to agree the format of a csv import file to provide the relevant data required in a suitable format, referenced with the substation ENID number.

Switch State Operational Data

It was determined that a potential source for some of this data might be the PowerOn Fusion SCADA system. However it was found that there was no readily available means to obtain historical switch state data as this is not archived in our data historian. Information on manual switching carried out on the network could potentially be obtained from fault records but this would be highly labour intensive to obtain. It was therefore decided that switch state information could not be imported as part of the trial.

The GIS spreadsheet has proved to be the most demanding to populate. One area of difficulty has been for those members of staff populating the spreadsheet, typically Year in Industry (YII) students to grasp the relationship between the network topology as displayed in ESRI and the model required by the IBM spreadsheet. For example it is often necessary to introduce 'virtual' busbars to satisfy the IBM model where these are not present in the real network (or at least not displayed on ESRI). We developed our own models to illustrate the relationship as shown in the slide overleaf.

3.2 Implementation [continued]

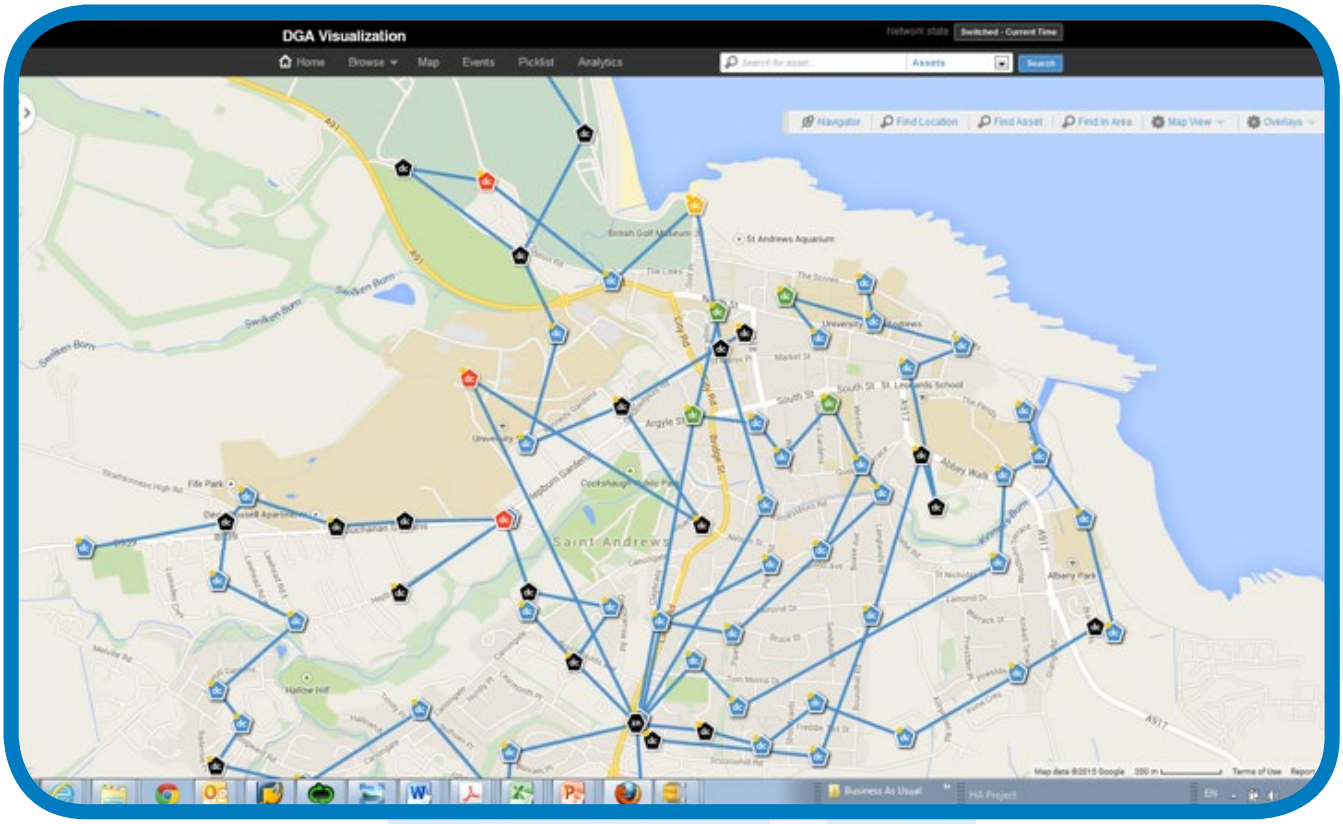


Another learning point that was highlighted during the trial is that the substation asset data held on our ESRI GIS system differs slightly between the SP Manweb and SP Distribution areas. This is because these areas historically had different GIS systems which were migrated onto ESRI. A number of iterations were necessary to arrive at spreadsheets which passed the IBM validation process and were able to be uploaded to DGA.

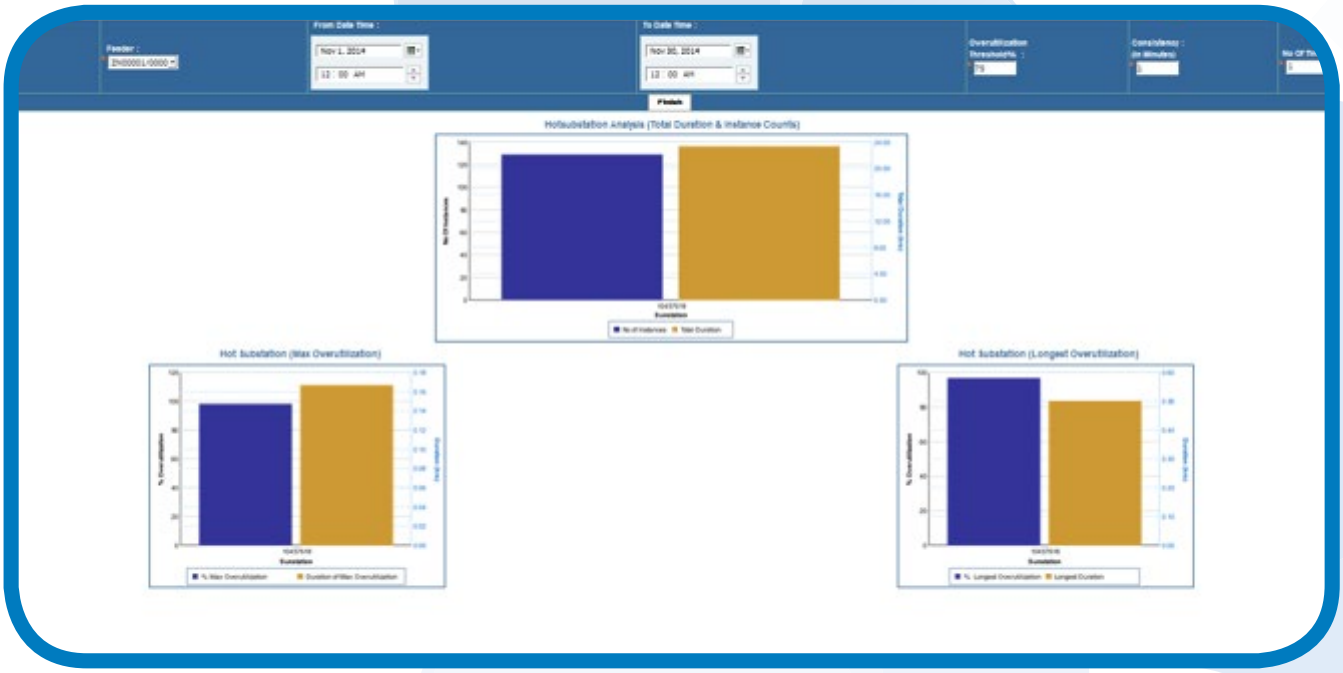
4 The outcomes of the work

4.1 St Andrews trial area

An example of a visualisation of the St Andrews network is shown below showing secondary substations with a substation utilisation overlay which attributes colours to substations in accordance with their level of utilisation.



An example of a 'Hot Substations' report is shown below for one of the 11kV feeders from St Andrews primary substation, highlighting secondary substations that have exceeded 75% utilisation factor.

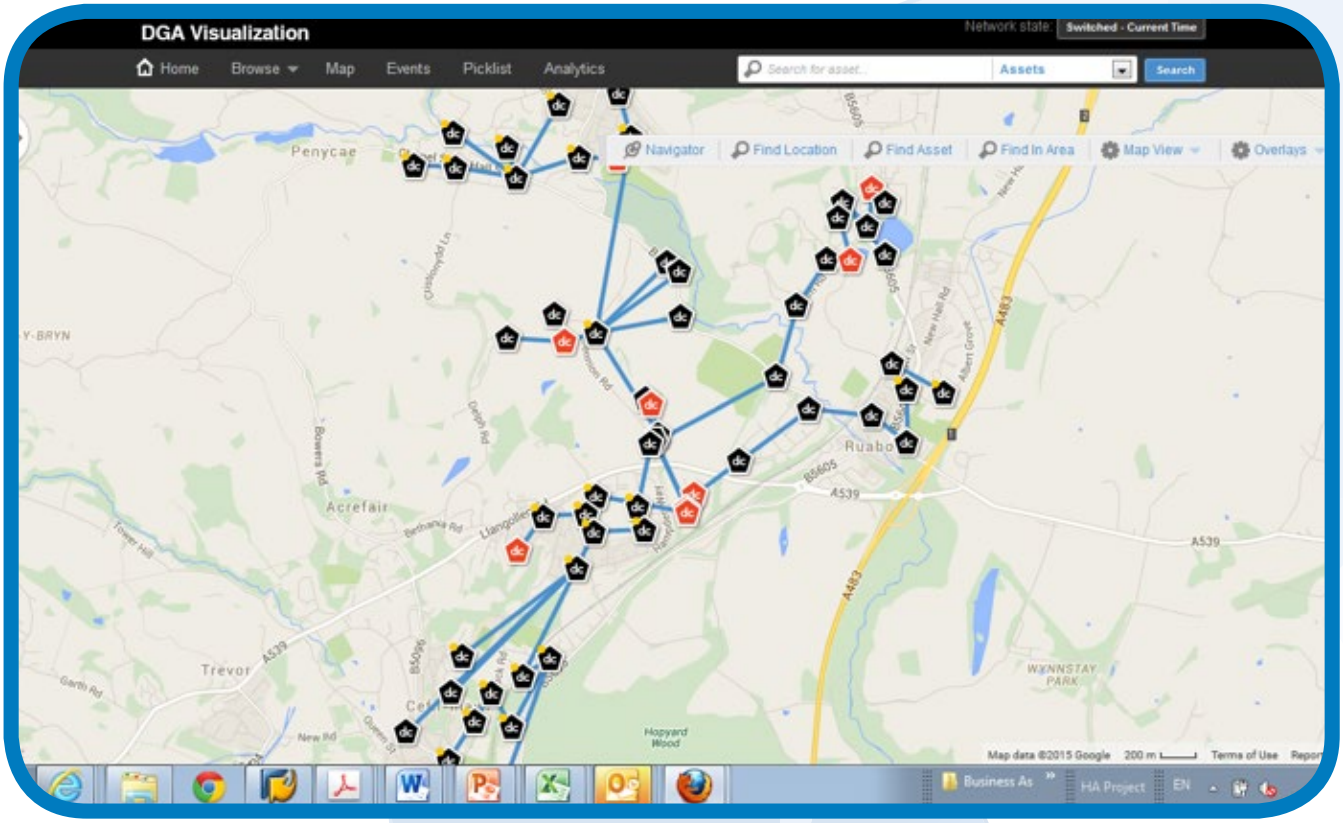


4.2 Whitchurch Trial Area

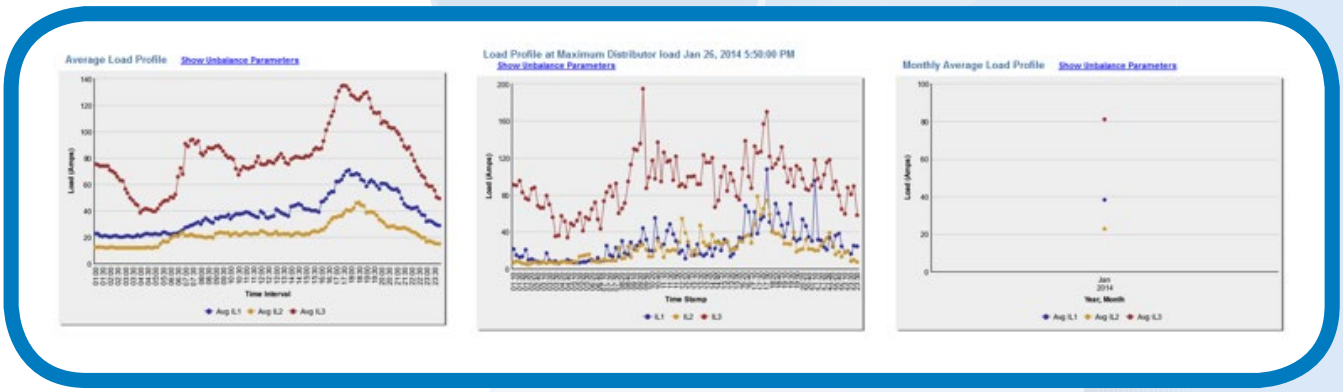
Did not form part of the data analytics trial.

4.3 Ruabon Trial Area

An example of a visualisation of the Ruabon network is shown below showing secondary substations with a substation utilisation overlay which attributes colours to substations in accordance with their level of utilisation.



An example of an 'unbalance Load Profile' report for an LV distributor is shown below. This includes 3 graphs each looking at different aspects of the same data:– Average Load Profile; Load Profile at Maximum Distributor Load; and Monthly Average Load Profile.



4 The outcomes of the work [continued]

The distributor current profile for the corresponding date range is also included below.

5 Internal Dissemination

A number of SPEN personnel from different sections have been involved in the implementation of the data analytics trial. The DGA system has been presented at workshops in SP Manweb and SP Distribution areas to representatives of the business including distribution system designers, outage planning managers and others. Reactions have generally been positive, however in the case of some individuals have been mixed which perhaps shows that the potential benefits of data analytics need further demonstration: -

"For me, ability to view monitoring data on the secondary network is one of the best things to come out of the Flexible Networks project" – Outage Planning Manager

"It's really useful to be able to see the utilisation of secondary substations on a visual overview" – Network Connections Designer

"This tries to solve a problem that doesn't exist" – Senior Design Manager

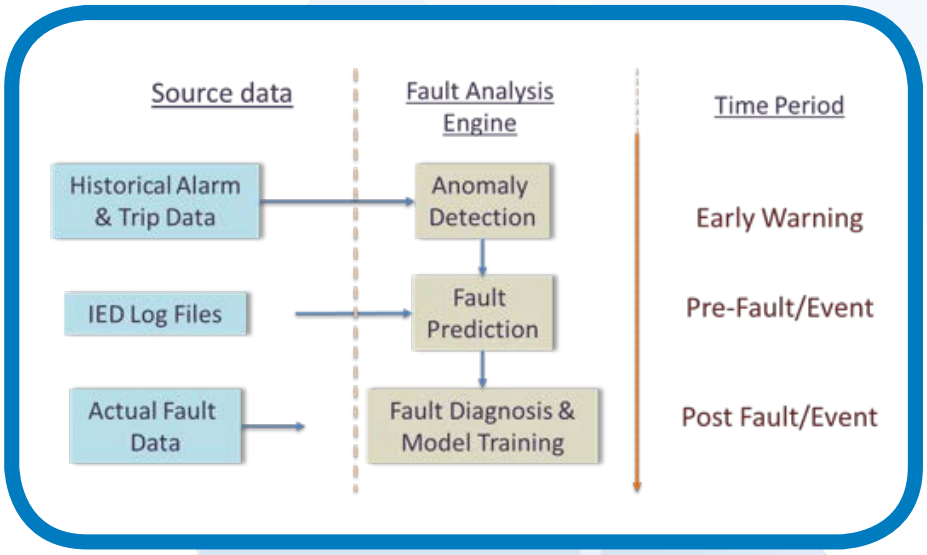
6 Business as usual

This was our first trial of data analytics. We believe that this technique is at least 2 years away from business as usual adoption. Our proposed next step is to undertake a NIA project focussing on a number of 'Use Cases'

The project will address the following problem statements:

- Enable an efficient & repeatable means for the collection and management of high volumes of smart data from multiple sources
- Determine the best mechanisms to uniquely identify data, including network connectivity e.g. CIM evaluation
- How data can be shared internally and externally in a secure and flexible manner so that all areas of the business have full access to all required information e.g. enable DSO role for the DNO
- What types of analysis are beneficial for the business and therefore what types of analytical tools do we require

As an example, one of the use cases we want to trial is the analysis of data from different systems for fault prediction and diagnosis.



7 Potential further work

Following implementation of the DGA trial a one-day workshop was held with IBM to review the trial and to discuss potential further work in this area. The following opportunities were identified for further work: -

- State estimation – IBM have a package available that estimates conditions at points on the network that are not monitored. This is an area that may be of interest to SPEN because there is only limited secondary network monitoring to be installed under our ED1 business plan so the ability to compensate for this may be of interest.
- 11kV Network Data – SPEN have power quality monitors installed at primary substations with the ability to provide large amounts of data. In addition we have large numbers of NOJA pole mounted auto-reclosers on the 11kV network that have data logging as an inbuilt feature. There is potential to examine how this data from these 2 sources might be retrieved and analysed.
- Asset risk management – The potential to analyse data from a number of different systems (including the existing SAP asset management system) for risk management of assets could be considered.
- The analysis of smart meter data is another potential application of data analytics.

8 Conclusion

We have successfully trialled the IBM DGA data analytics package. The trial has demonstrated that it is possible to extract data that resides in different systems and use this to create new visualisations and reports that benefit the business.

Through the trial we have encountered some difficulties in extracting data from existing systems and this has added to our learning. Under a large scale implementation of data analytics it would be impractical to extract data in the highly manual way it was undertaken for the trial. The trial has also highlighted that there are inconsistencies in the way that assets and data are identified between different systems which adds to the complexity of data retrieval.

Overall the trial has been a useful and informative first step for SPEN in the area of data analytics.