

# Milestone 2

Creating an Integrated Distribution  
System Analytical Model from Planning  
and Geographical Data

By Alan McMorran

20/01/2020

 **Open** Grid Systems



This report details how SSEN 11kV and 33kV network model and GIS data have been converted into CIM format to create interconnected models of the Fort William and Oxfordshire circuits being examined for the MERLIN project.



# Contents

<b>1</b>	<b>Executive Summary</b>	<b>6</b>
<b>2</b>	<b>Introduction</b>	<b>7</b>
<b>3</b>	<b>Target Data</b>	<b>8</b>
<b>4</b>	<b>Extra High Voltage Data</b>	<b>9</b>
4.1	Definition	9
4.2	Scottish Hydro Electric Power Distribution	9
4.2.1	Long Term Development Statement	9
4.2.2	Extracting the Target Area	12
4.2.3	Process Validation	12
4.3	Southern Electric Power Distribution	<b>12</b>
4.3.1	PSS/E Data	12
4.3.2	Extracting the Target Area	13
4.3.3	Process Validation	13
4.4	Geographical Information System Data Integration	<b>14</b>
4.4.1	Conversion to CIM	14
4.4.2	Correlating Planning and Geographical Data	14
4.4.3	Substation and Junction Mapping	14
4.4.4	Line Routing	16
4.4.5	Automatic Geographical Integration Challenges	18
4.4.6	Process Validation	22



<b>5</b>	<b>High Voltage Data</b>	<b>23</b>
5.1	Geographical Information System Data Conversion	23
5.1.1	Network Reference Number Hierarchy	23
5.1.2	Deriving Network Topology	23
5.2	Line Parameter Data	25
5.3	Distributing Load Data	25
5.3.1	Equalised Distribution	25
5.3.2	Distributing by Customer Connections	25
5.4	Limitations of HV Network Creation	26
5.5	Process Validation	26
<b>6</b>	<b>Conclusions</b>	<b>30</b>



## Table of Figures

Figure 1	PINEGR Automatically Generated Schematic	10
Figure 2	PINEGR Schematic Extended with Connected Buses	11
Figure 3	Example Planning Network Topology	18
Figure 4	Example GIS Network Topology	18
Figure 5	Route Identified from A to B	19
Figure 6	Route Identified from B to E	19
Figure 7	Routes Identified from A to E	20
Figure 8	Initial Route Identified from C to D	20
Figure 9	Illustration of Gap Identified in Route from C to C	21
Figure 10	Secondary Substation Topology Routes	24
Figure 11	Secondary Substation Topology from Connectivity Analysis	24

## Table of Tables

Table 1	Example of Buses defined in SHEPD LTDS Data	10
Table 2	SHEPD Substation Mapping Table Extract	14
Table 3	SHEPD Junction Mapping Table Extract	15
Table 4	SEPD Planning to GIS Substation Mapping Table Extract	16

# 1. Executive Summary

This is the second stage of the data discovery and conversion exercise for the MERLIN project. The first stage examined the data sources available within SSEN and identified the best sources for creating the distribution system analysis network, and systems that can provide historical measurement data. These sources were described in the Milestone 0 report “SSEN Network Data Analysis”.

This report describes the second stage of building the distribution system analysis model from the 33kV planning data and the geographical datasets exported from the Geographical Information System. This requires the conversion of the planning and geographical data from their source formats into an IEC standard structure, then the challenges identified and addressed with integrating electrical connectivity and geographical data.

Extending the network model to include the 11kV network is required to support detailed analysis of the impact of distributed generation. This report describes the approach taken to automatically extract the network topology from the geographical data and integrate this with the 33kV network.

The result is a full interconnected network model from the Grid Supply Point down to the secondary substations, automatically extracted and integrated from the source datasets defined in an IEC standard structure.





## 2. Introduction

The economic analysis modelling for MERLIN requires detailed network models covering both the Extra High Voltage (EHV) and High Voltage (HV) for areas of the Scottish Hydro Electric Power Distribution (SHEPD) and Southern Electric Power Distribution (SEPD) service areas. The data being used for MERLIN is to be an open, standard structure, made available in open, standard file formats and Application Programming Interfaces (API).

A previous report, SSEN Network Data Analysis published for Milestone 0 of the MERLIN project, discussed the sources of data available and the target structure, that of the IEC 61968-11/61970-301 Common Information Model (CIM) standard. The sources of network data available were discussed and analysed, with the conclusion that the planning data that forms the Long Term Development Statement (LTDS), published by SSEN, and the Geographical Information System (GIS) data were the most suitable sources to use in building the core distribution system analytical model. This model will underpin the simulation and analysis work for MERLIN.

This document will describe how the data was analysed, enhanced and merged to create the integrated EHV model, combining the electrical connectivity and geographical data into a single, integrated model for analysis. This includes the challenges posed by variable data quality, divergent policies and patterns for naming and identification, and aligning real-world physical equipment with abstract mathematical equivalents.

As utilities around the world deal with more Distributed Energy Resources (DER) and Electric Vehicles (EV) being connected at lower voltages, so too must the planning and analytical models be extended to these voltages. Understanding operation of the network closer to the point of consumption is key to recognising where the grid will need enforcement or more intelligent control. To support this, this documents describes the process taken to extending the EHV data by integrating HV data automatically extracted from the GIS data.



## 3. Target Data

The target data structure for the distribution system analytical model, is the data model defined by IEC 61968-11 and IEC 61970-301, commonly referred to as the Common Information Model (CIM).

These standards are widely used throughout the world for exchanging and storing power system data in an open structure and format. ENTSO-E, the European Network of Transmission System Operators for Electricity, uses CIM as the core of their Common Grid Model Exchange Standard (CGMES), used by 42 TSOs across Europe to exchange electrical network models. It is also used by distribution and transmission network operators world-wide as the foundation of their Enterprise Information Model and for system integration.

The datasets used by, and produced by the analysis undertaken for MERLIN will be in a CIM structure which will be accessed via open standard data formats (IEC 61970-552 CIM RDF XML) and open standard APIs (OData v4.0). The target version of CIM is CIM v16 as this aligns with the CGMES v2.4.15 standard and currently has the widest industry support. The data used for MERLIN will be for a balanced, three-phase system, which aligns closely with the CGMES v2.4.15 standard. The CIM has a standard for distribution, IEC 61968-13 (Common Distribution Power System Model), but the last published version of this standard from 2011 is based on an earlier version of the CIM and the differences are more applicable to unbalanced, low voltage modelling which is typical in North America.

The approach taken is to convert the EHV and HV data into CIM, using the classes from CGMES v2.4.15, but allowing for additional data to be incorporated either from other classes within the CIM data model or through custom extensions where appropriate. This allows for the core network data to be defined by the international standards, while allowing for additional data to be retained and accessible for those systems that support extensions.





## 4. Extra High Voltage Data

### 4.1 Definition

The EHV data covers the 33kV and above network. As was discussed in the SSEN Network Data Analysis paper, the optimum source of this data was identified for both SHEPD and SEPD as being the network planning data maintained internally for the LTDS. For SHEPD this data covers the 33kV network connecting to the transmission system at the Grid Supply Points (GSP). For SEPD this data covers the 33kV and 132kV network with connections to the transmission system at GSPs and then Bulk Supply Points (BSP) defined as the 132/33kV boundary.

The networks for SHEPD and SEPD are maintained separately. For each network there are two sets of LTDS data:

- Detailed internal networks that include commercially sensitive data
- Publicly available network data with commercially sensitive data removed

This commercially sensitive data includes information on generators and large customers where public disclosure could impact the fair operation of the market. The removal of commercial data includes the substitution of detailed network components for simplified components that are electrically equivalent including aggregating multiple loads/generators into a single equivalent so that individual generators and their capabilities are not disclosed.

The target of the data being created for MERLIN is initially to support internal SSEN studies and as such the detailed internal data was requested and used in this exercise.

### 4.2 Scottish Hydro Electric Power Distribution

#### 4.2.1 Long Term Development Statement

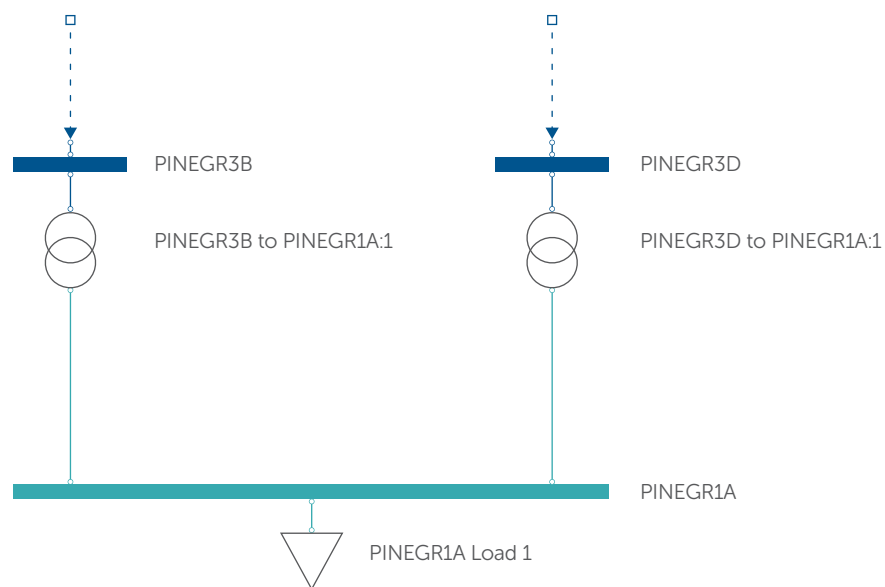
The SHEPD LTDS data is maintained as a spreadsheet in Microsoft Excel format. As was discussed in the SSEN Network Data Analysis paper, this data covers:

- The branch data including electrical impedance and rating data.
- The transformer data including the electrical impedance, vector, tap and rating data.
- Load data split by substation including historical and forecast demand and the capacity information.
- Fault Level information including system impedance, fault currents and circuit breaker ratings.
- Generation data on the distribution system covering connected and accepted (planned) generation, its type, and capacity.
- Interest in connection data covering offered and accepted connections for load and generation on the network.

Custom scripts were written to extract this data and to build an initial network connectivity model.

This initial conversion and direct conversion of data is sufficient for power-flow analysis and connectivity analysis, but as the data is focussed on planning analysis applications it is focussed on buses, load and generation, lacking data points for correlating the components with real-world locations such as substations or cable/line routes.

Figure 1 PINEGR Automatically Generated Schematic



Naming conventions are used for buses, with prefixes indicating the station buses are tied to. Initially these were looked at for grouping into substation, however it was found that buses were linked to substations even when they were connected via lines with defined length and impedance. For example, Table 1 shows a small sample of buses that are defined within the LTDS data.

'PINEGR3E'	33kV
'PINEGR3A'	33kV
'PINEGR3B'	33kV
'PINEGR1A'	11kV
'PINEGR3C'	33kV
'PINEGR3D'	33kV

Table 1 Example of Buses defined in SHEPD LTDS Data

When the connecting branches were analysed it was found that of these buses PINEGR1A PINEGR3B and PINEGR3D were interconnected via zero-impedance/length connections or transformers, with the other three buses connected via branches.

The auto-generated schematic for the PINEGR substation is shown in Figure 1.

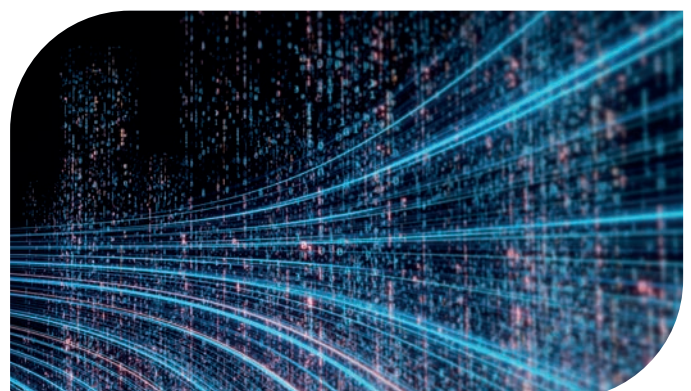


Figure 2 PINEGR Schematic Extended with Connected Buses

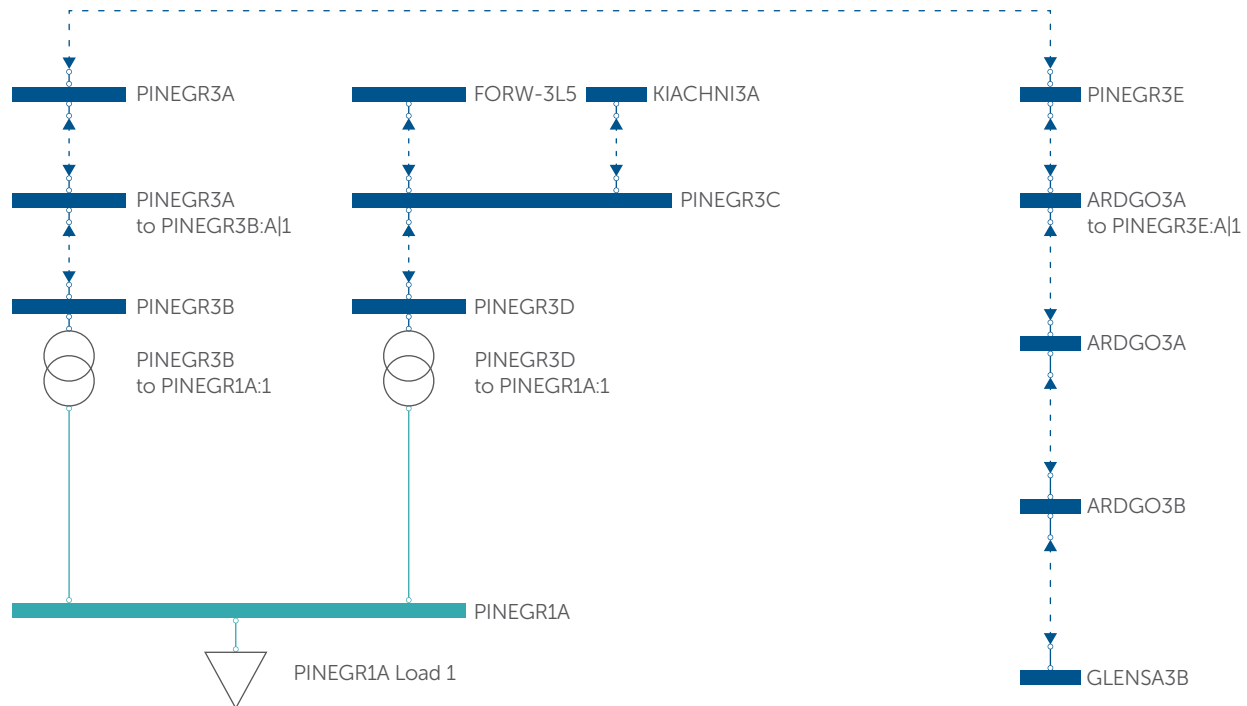


Figure 2 shows the PINEGR schematic with additional buses and branches added showing connectivity beyond the internals of the substation. This diagram shows that while buses with a common name are electrically connected, they are not necessarily within the same geographical location.

The line between PINEGR3A and PINEGR3E for example is defined as being 7.4km long. Other lines however are shorter (200m between PINEGR3C and PINEGR3D) indicating a point potentially outside a substation boundary fence but still geographically close by.

A further step of analysis was therefore undertaken to:

- Perform additional connectivity analysis of the data to identify components interconnected via zero-length and zero-impedance branches, transformers and switches/breakers that indicate they are within the same substation

- Identify common naming prefixes for buses that can be used for initial naming of these substations

Create the CIM containment hierarchy structure of:

- Geographical Region
- Sub Geographical Region
- Substation
- Voltage Level

based on a pattern of:

- Network (SHEPD for this data)
- Grid Supply Point
- Derived Substation
- Defined voltage



#### 4.2.2 Extracting the Target Area

The target area for MERLIN was defined as being the network supplied by the Fort William GSP. The LTDS data lists the GSPs for buses, branches and generation, but for further verification of this, the extent of the network was calculated using the complete network with the edges defined by the “open points” between networks. These open points are switches on the network that are normally in an open state, but can be closed to allow networks to be supplied via alternative GSPs in the event of abnormal network conditions.

It was confirmed by SSEN planning engineers that the lowest numbered bus defined for a GSP area is the bus that is attached to the transmission system. Tracing from this point outwards identifies the extents of each GSP territory. When run against the full SHEPD CIM network generated from the LTDS spreadsheets, this identified the network for each GSP, from which the network for Fort William was then isolated and extracted.

Running the full data transformation algorithm identified 58 substations for the Fort William area and a further 51 Junctions that were defined as buses between stations. These junctions define either points where multiple lines interconnect (e.g. T points) or where two lines connect. An example of this is where a line transitions from overhead to underground. Each segment has different electrical characteristics and as such is modelled as a separate segment with the junction marking the point of interconnection.

The resulting fort William network model comprised 133 nodes, 109 lines and 19 transformers that was capable of solving a standard three phase balanced power flow.

#### 4.2.3 Process Validation

As the models produced through this exercise are used as an input to further stages in the MERLIN project, it is recognised that errors at this stage could magnify and be difficult to remediate. As such, several validation steps were undertaken to satisfy that the data had been correctly imported.

- The data was validated using Open Grid Systems own internal CIM validation engines to check that the output formats conformed with IEC 61970-552 and the data attributes, references and types were in compliance with IEC 61970-301 and IEC 61968-11.
- The elements generated were compared with the list from the LTDS spreadsheet to confirm the number of buses, branches, and transformers in the CIM data matched those in the source spreadsheet.
- Additional data validation was run against the core CGMES 2.4.15 profiles to check that the core required attributes and references were present.
- Topological processing was executed to validate that there was a single interconnected island.
- A balanced three-phase Newton-Raphson power flow was run using Open Grid Systems’ own power flow engine. The power flow successfully converged with bus voltage results within the nominal 0.95-1.05 range.
- The network was exported into a PSS/E RAW format and provided to the SSEN planning engineers for validation. SSEN planning engineers confirmed that the model was an accurate representation of the network and that using the default loading and generation data within the LTDS data, the results were in line with expectations.

### 4.3 Southern Electric Power Distribution

#### 4.3.1 PSS/E Data

The SEPD data is maintained independently to the SHEPD data with the master data stored and maintained within PSS/E, an analysis application from Siemens PTI. The PSS/E standard file format is known as RAW and is a text-based, white-space delimited file format that defines the core components of the electrical network.

As was discussed in SSEN Network Data Analysis paper PSS/E RAW is widely supported by third-party applications but remains a proprietary format owned and defined by Siemens PTI. The PSS/E data for SEPD is equivalent to the internal LTDS data for SHEPD, containing commercially sensitive data on generators and loads, and as such is not publicly available.

Conversion of this data into CIM was undertaken using Open Grid Systems' PSS/E RAW to CIM convertor, creating a CIM electrical network model conforming to the ENTSO-E CGMES standard. The data created at this stage was similar to that for SHEPD in that:

- Substations are not defined in PSS/E RAW, but as with the SHEPD LTDS data there are bus naming patterns that correlate with real-world substations.
- Buses are joined with zero-impedances lines/ switches/breakers indicating intra-substation connections.
- Not all buses with common prefixes are within the same location.
- Loads and embedded generation on the 11kV network are aggregated to individual loads and generators on 11kV buses.

The same process for calculating substations was followed for SEPD as was undertaken for the SHEPD data. Additional data was available for SEPD where the public LTDS spreadsheet had a map for the bus name to the corresponding substation name. As such the SEPD data conversion process contained an additional step that matched the bus prefix (e.g. ACTL) with the corresponding substation name (ACTON LANE) or provided a warning in the logs where no match was found.

#### 4.3.2 Extracting the Target Area

As with the SHEPD data, the SEPD data contained the full network, but also contained the 132kV sub-transmission system that for SHEPD is part of the transmission network, not distribution. The target areas for MERLIN were identified as BSPs, all of which came under the COWLEY GSP. This allowed the conversion extents to be calculated via connectivity analysis of the network from the interconnection point defined within the COWLEY GSP. This analysis traced the network extents covering the 132kV and 33kV networks with the edges again defined as the "normal open points" on the network.

As GSPs in the English and Welsh network are defined at the boundary to the 132kV network, while in Scotland they are to the 33kV network, the area covered by a GSP in SEPD is larger than in SHEPD. This is reflected in the resulting network which covers 87 substations, 297 nodes and 190 lines. Individual BSP networks can be extracted from this by performing connectivity extent analysis at each 132/33kV boundary.

#### 4.3.3 Process Validation

With the data in PSS/E format, there was no requirement for additional validation of import scripts, as the PSS/E converter is an off-the-shelf component that Open Grid Systems provides to other customers and is used by other clients for their daily submissions to the ENTSO-E Day Ahead market.

As with the SHEPD data the validation involved multiple steps:

- The data was validated using Open Grid Systems own internal CIM validation engines to check that the output formats conformed with IEC 61970-552 and the data attributes, references and types were in compliance with IEC 61970-301 and IEC 61968-11.
- The elements generated were compared with the original PSS/E RAW file to confirm the number of buses, branches, and transformers in the CIM data matched those in the source data.
- Additional data validation was run against the core CGMES 2.4.15 profiles to check that the core required attributes and references were present.
- Topological processing was executed to validate that there was a single interconnected island.
- A balanced three-phase Newton-Raphson power flow was run using Open Grid Systems' own power flow engine. The power flow successfully converged with bus voltage results within the nominal 0.95-1.05 range.

The Fort William area was the initial target for the project, and as such the validation with SSEN engineers for the SEPD data will be undertaken at a later date.

## 4.4 Geographical Information System Data Integration

The conversion of the electrical connectivity data into an IEC CIM structure is the first step in the creation of the overall system model. The second data source to be integrated is the Geographical Information System (GIS). In SSEN this is a GE Smallworld system. The different data tables for both SEPD and SHEPD were described in the SSEN Network Data Analysis paper.

### 4.4.1 Conversion to CIM

For both SHEPD and SEPD a detailed mapping definition was constructed for each data table and its attributes into the corresponding CIM classes, attributes and relationships. This required separate configurations for each network, with the resulting CIM data reflecting the data within the GIS as electrical components (e.g. overhead lines, cables) and related structures (e.g. substations, poles, towers). The different GIS schema were mapped to a common CIM definition for both SHEPD and SEPD.

Open Grid Systems GIS to CIM conversion software was used to produce the CIM data, serialised as a CIM RDF XML file in compliance with IEC 61970-552. The conversion was restricted to a geographical boundary covering the areas of the network being analysed for MERLIN (Fort William in SHEPD and the Oxford are supplied by the Cowley GSP for SEPD).

### 4.4.2 Correlating Planning and Geographical Data

The next challenge was to correlate the two distinct data sets by identifying the geographical routing for lines and cables between the substations and junction points created from the conversion of the LTDS spreadsheet and PSS/E RAW data into CIM (the planning data).

This required three steps:

- Match the substations identified in the planning data with the substations defined in the GIS data.
- Identify where the junction points in the planning data were in the GIS data.
- Correlate the electrical lines and cables between substations and junctions in the planning data with the overhead lines and cables in the GIS data.

### 4.4.3 Substation and Junction Mapping

#### 4.4.3.1 Scottish Hydro Electric Power Distribution

The distinct GIS schema for SHEPD and SEPD required different approaches to mapping substations from GIS and planning. The GIS data used for SHEPD identified the primary stations by their Network Reference Number (NRN), a unique numeric code used to identify stations and transformers. The planning data did not use this NRN and so an external mapping table was defined that correlated the name derived from the planning data with the NRN. An extract of this is shown in Table 2.

Name	NRN	ID	Type	Latitude/Longitude
FWIL	715			
GARVAN	715302144			
GLENF_	715302174			
GLENSA	714			
GLENSA3B		133	Pole	56.68193048727568, -5.384493122375858
GLENTTR	715305100			

Table 2 SHEPD Substation Mapping Table Extract

In some cases there was no direct substation NRN and instead the point was identified as being another GIS component. In the extract above a station is marked as being at the location of Pole 133 and the latitude and longitude specified.



The identification of these points required the manual identification of points in the GIS data. In some cases they directly mapped to components such as transformers or poles, in two cases there were not components at the points identified and so an explicit location was added.

These cases were identified as:

- Missing data within the GIS extract where a substation was visible in the satellite view but had no corresponding GIS entry. This could reflect an error with the export or that the dataset being used had some data missing.
- An ambiguity where a substation was really an aggregation of lower voltage data so in the context of a planning model there was a collection of loads and generation at a single point, but the reality the components were spread across a geographical area .

The challenges of aligning “real world” built assets with electrical analysis models is common for utilities. Network analysis tools perform complex mathematical calculations on data designed to represent the characteristics and behaviour of the electrical network. Modelling and analysing every single component is unnecessary when aggregated equivalents can provide the accuracy required.

The same process was undertaken for the junctions, the components identified as being points of interconnection between lines and cables but that did not have any other additional components such as load or generation attached. This required manual matching of the 51 junctions identified as being part of the Fort William network created from the planning data.

The junctions matched to different GIS components with a mapping table defining these. An extract is shown in Table 3.

Name	ID	Type	Latitude/Longitude
CARNOC3A	225	Pole	56.69055112269584, -5.529290933906907
CIAA3A	ARKAIG H418_L	Switch	
DUBHH3A to ACHEA3A : A i1	ALLT DUBH	Power Transformer	56.960794721620516, -5.040451507422176
FORW- 1L5 to INVERL3C : A i1		Junction	56.82237399679491, -5.080094106662102

**Table 3** SHEPD Junction Mapping Table Extract

For some components where the ID was not null but also not unique within the data the latitude and longitude were also provided. For example the Pole ID was often a sequential number indicating its number for a circuit so there were multiple Poles with ID 225 in the GIS data. The inclusion of the latitude/longitude allowed the path finding algorithm to correctly identify the correct pole to be used.

#### 4.4.3.2 Southern Electric Power Distribution

The conversion of the SEPD data for the Oxfordshire area supplied by the Cowley GSP took a similar approach. An additional mapping spreadsheet was available for SEPD that identified the substation a bus ID was within. Using this data the process became:

- Planning Data derived substation IDs mapped to full names from the mapping spreadsheet.
- Substation names from planning mapping spreadsheet mapped to GIS substation names.

The mapping spreadsheet successfully matched 95% of buses from planning data, the remaining 5% were manually identified and added to the spreadsheet.

The majority of the substation names aligned between planning and GIS, however some outliers were found that required manual matching. An extract of the mappings that required manual entries is shown in Table 4.

Planning Data Substations	GIS Substations
AIR PRODUCTS	AIR PRODUCTS DIDCOT
ARDLEYGEN	ARDLEY LANDFILL GENERATION
MANOR FARM PV GENERATION	MANOR FARM PV FARM
CULHAM U.K.A.E.A.	CULHAM JET
SUTTON COURTNEY	SUTTON COURTENAY

**Table 4** SEPD Planning to GIS Substation Mapping Table Extract

As can be seen in the above, some are down to minor differences in nomenclature, others are down to the station being named for the organisation (UK Atomic Energy Authority) and the other using the facility (Joint European Torus). Sometimes the inability to match is simply due to a misspelling in one system. *Fuzzy* matching helps to identify potential matches, with human input to then set explicit mappings allowing the conversion to be run automatically and repeatedly.

The Cowley SEPD data contained far fewer Junction points than Fort William with only eight identified as part of the automatic connectivity analysis of the planning data. Each junction was then mapped to either a GIS element such as a tower, pole or substation, or given an explicit coordinate identified from the GIS data.

#### 4.4.4 Line Routing

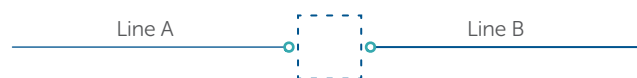
Once the substations and junctions were matched to their GIS equivalents the next stage was to correlate the branches between substations defined in the planning data with the overhead lines and cables within the GIS data. This is not a simple one-to-one mapping, and the goal of this conversion process was to automatically match data where possible so that the processes could be made repeatable and support future changes to the data. Manually mapping thousands of lines to the equivalent GIS components is not practical so an automated process was developed.

##### 4.4.4.1 Connectivity Generation

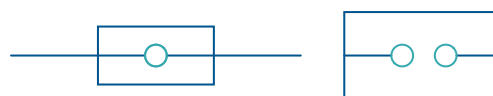
The GIS data was provided as Shapefiles, a standard GIS export format that defines locations as points, lines and polygons. The challenge for building electrical networks from GIS data is that the open formats such as Shapefiles and GeoJSON do not define explicitly connectivity. Instead, connectivity must be derived by finding find points that have common locations and *joining* these points with a node.

This challenge was discussed in the SSEN Network Data Analysis paper, and examination of both the SEPD and SHEPD data identified a number of challenges to mapping GIS line/cable data into electrical components:

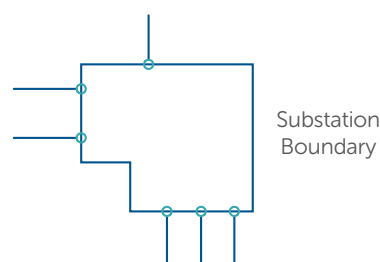
- The GIS had another component (e.g. a tower, pole, switchgear) between the lines. As such the two end points of the line are not at the same coordinate.



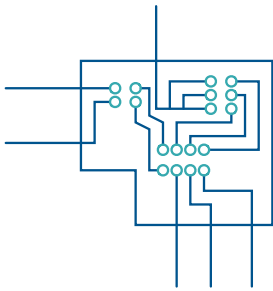
- The lines appear to be at the same coordinate when viewed at normal zoom levels, but upon closer examination they finish close to each other but not at the exact same point. This can happen when lines are manually drawn without a snapping function, or they are imported from different sources requiring coordinate conversion:



- The internals of a substation are not included for SHEPD data within the GIS and a substation is defined as the boundary of the station. The internal connectivity of the station is therefore not included and lines terminate at the edge of the station:



- For SEPD data the primary substations did contain internal cable layouts, resulting in a large number of lines, cables and connection points within a substation:



- As the line and cable data is used by field crews for locating, maintaining and servicing equipment, this can result in details that are not required for electrical connectivity models. An example of this was locations where the three individual conductors for each phase of a line or cable were represented independently. This resulted in GIS representations where sections had three parallel lines before going back to a single line:



- When the data is examined within the context and understanding of how electrical networks are constructed and maintained this is logical, however it poses challenges when trying to build electrical networks from the GIS data.

Lines are often split into multiple segments within the exported data:



There can be multiple reasons for this:

- Individual segments reflect the changes in construction where a line's conductor type has altered.
- They represent the individual spans between towers or poles.
- The internal storage structure of the source GIS has split lines to accommodate its own data organisation.

Creation of a line from station to station therefore has to merge these into a single route from station to station (or station to junction)

#### 4.4.4.2 Path Finding

The initial approach to creating a full network model was to build an electrical connectivity model from the GIS data that could also be used for analysis. Examination of the data however indicated that this approach was not practical due to the ambiguity in the connectivity generated. The approach taken was to instead use the connectivity already defined and validated within the planning data as the core electrical network, then to integrate the location data from the GIS data. This required the system to automatically calculate the potential routes within the GIS data between the stations/ junctions connected in the planning data then integrate the location and routing data with the electrical data.

At a high level this can be illustrated with some examples of the challenges that were identified during this work.





Figure 3 Example Planning Network Topology

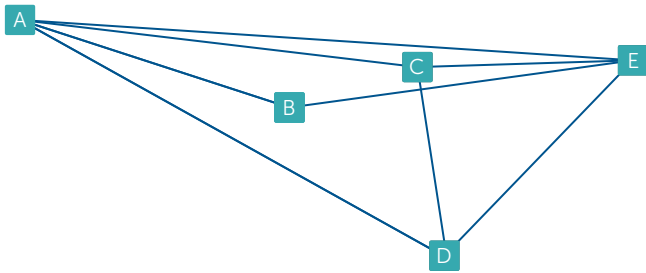
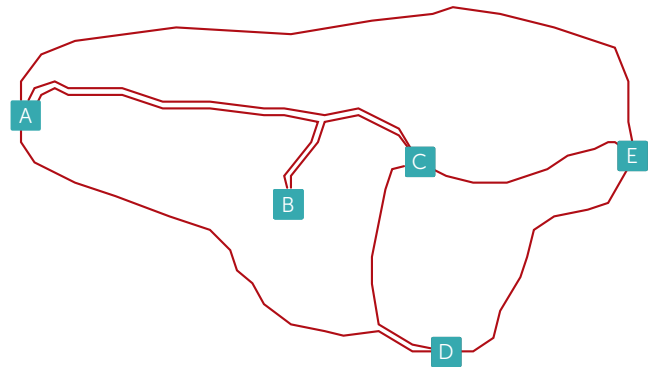


Figure 4 Example GIS Network Topology



#### 4.4.5 Automatic Geographical Integration Challenges

##### 4.4.5.1 Example Network

Figure 3 shows an example of a simple network extracted from planning data. There are five stations, with the lines between the stations identified as the direct straight lines shown in blue.

The same substations are shown in Figure 4 with the geographical routes for lines and cables shown in red. The process for merging the datasets involves taking each branch and finding the potential corresponding routes in the geographical data.



Figure 5 Route Identified from A to B

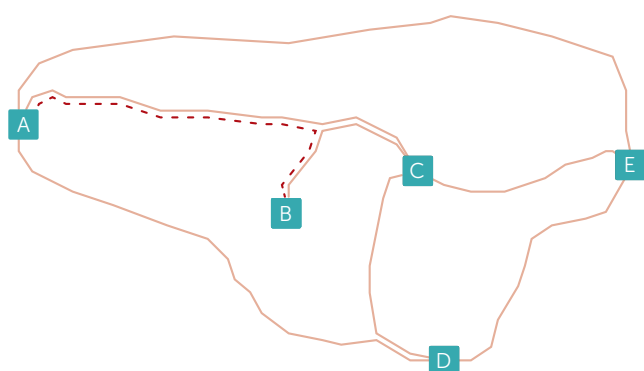
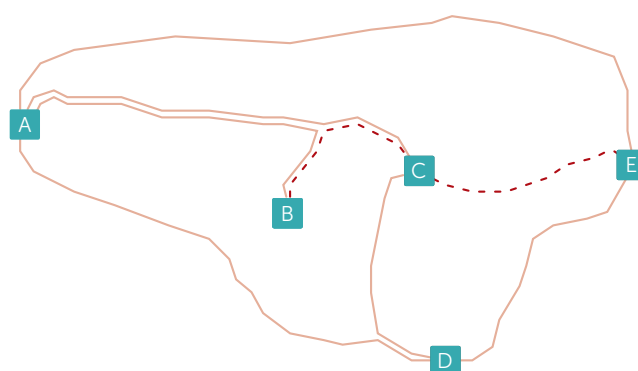


Figure 6 Route Identified from B to E



#### 4.4.5.2 Route Finding

Dijkstra's algorithm was used to find all paths between the source and target stations, so for example when looking for routes from A to B the result is shown in Figure 5.

An initial challenge is illustrated when looking for the route from B to E. There is no direct geographical line between B and E, and when the internal cabling of the stations is not included, each line terminates at the boundary of the station. As such there is no single node at a station as would be expected for a route-finding algorithm based on vertices and edges. Instead there were multiple nodes at the end of each line either at the edge of the station or inside (as was illustrated in 4.4.4.1).

To compensate for this, a processing step was introduced to connect the nodes at substations so that paths could then traverse through the stations. When this was applied for finding the routes from B to E the shortest path was then correctly identified as shown in Figure 6.

In some cases the shortest path was not the correct path. An example here would be to find the path from A to E.

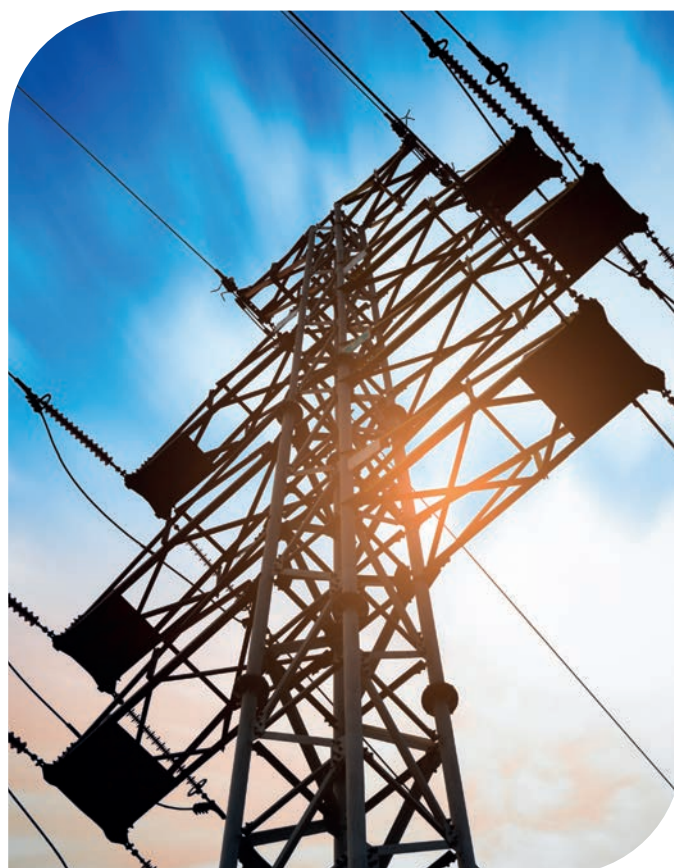


Figure 7 Routes Identified from A to E

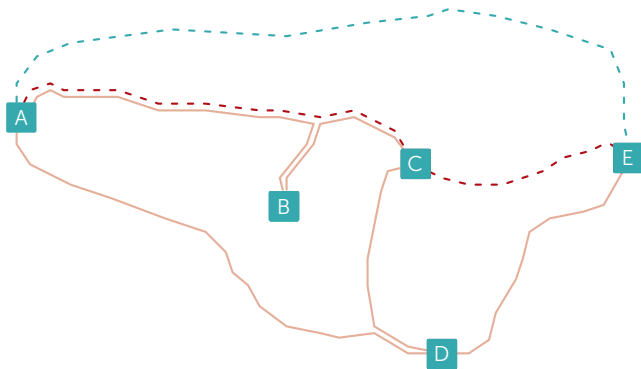


Figure 8 Initial Route Identified from C to D



The shortest path is A to C to E as shown in dark red in Figure 7. However the correct path was the direct line from A to E highlighted in green. In this case the reason was that the line from A to E was a 132kV line while the A to C to E lines were all 33kV. As such the algorithm was updated to use both the routing and the electrical parameters. Both the GIS and the planning data identified the voltage of a line and using this data allowed the algorithm to correctly identify the route.

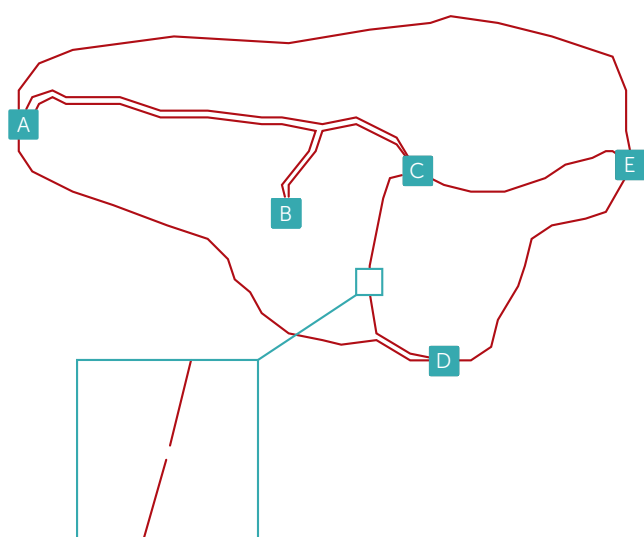
Running this process on the approximately 300 lines within the EHV data for SHEPD and SEPD quickly identified other anomalies where lines were unable to find routes or the routes identified were illogical.

An example of this was that when finding the route from C to D in the example above, the shortest route was identified as the path highlighted in dark red in Figure 8.





Figure 9 Illustration of Gap Identified in Route from C to C



Logically the route should have been the direct connection from C to D. Examination of the data and manual tracing of the route from C to D found that there was a gap between the sections as shown in Figure 9.

This required adjustments to the tolerances within the connectivity analysis. There were a number of cases where this issue arose in both SHEPD and SEPD data. This required the algorithm's configuration to be updated as part of an iterative processing-analysis-correction cycle that was run for both SHEPD and SEPD. Each dataset required different parameters, settings and tolerances to successfully locate the location data. This resulted in a multi-stage matching and processing system to create a repeatable conversion system able to deal with the challenges that arose including:

- Identifying parallel lines so nodes that were geographically close (e.g. on the same tower) were not mistakenly joined.
- Correcting for situations where the change between overhead and underground cabling omitted the connecting cable resulting in a gap.

- Accommodating data containing lines that terminate outside, on and inside the defined boundary.

The result was network data for both SHEPD and SEPD that included line routing, substation and junction placements taken from GIS with the electrical parameters taken from the existing electrical planning data for 33kV and 132kV.



#### 4.4.6 Process Validation

Validation of the GIS to CIM conversion was undertaken in several stages.

- The initial process outlined in the previous section describes the steps undertaken to build a process that worked for the SSEN data, but has to be proven to be providing correct data. This was undertaken by checking the results of the conversion processes by rendering the data with the Cimphony product, overlaying calculated line routings upon satellite imagery.
- Once the process was completing without warning or errors from the data conversion and validation steps, a "sense-check" of the data was performed by looking at the drawn shapes from the CIM data and the GIS data and confirming that there were no obvious errors in the data.
- The SSEN GIS data was loaded into the open source QGIS geographic information tool so the data could be rendered. This allowed for manual, visual verification of the routing in the geographical display within Cimphony using the CIM data, and the original Shapefiles visualised in QGIS.
- Again, overlaying all data upon satellite imagery then provides a further check as any issues with coordinate conversions can be spotted where plotted elements do not align with the corresponding imagery.
- Automated validation was also performed by checking for buses/lines in the GIS that did not have corresponding LTDS data, and where lines within the LTDS data had not matched with the GIS data. This then either resulted in an update to the configuration and mapping parameters, or could be ignored if the lines were internal substation routes that did not require GIS data.
- Finally, a detailed check was undertaken of core stations by verifying routes from these stations and comparing them with the data shown in QGIS. Manually verifying the routes for cables and lines by tracing them manually within QGIS and comparing the results with the geographical routing on the satellite imagery.



## 5. High Voltage Data

The next phase in the data conversion was to integrate the HV 11kV data into the network data. Unlike the 33/132kV data, the HV data was not maintained and managed for the complete network, but was instead managed within the GIS and then electrical networks created on ad-hoc basis as required.

When studies are undertaken on the HV data within SSEN they are focussed on individual HV circuits, however for MERLIN the desire was for the HV data to be integrated with the EHV planning data to create full network connectivity from the GSP down to the secondary substations and transformers. The approach taken for EHV of creating the CIM network data from an existing planning source then integrating it with the GIS data location from another source was not possible as there was no separate, maintained planning data source.

Instead the aim was to replicate and automate the manual conversion processes undertaken by SSEN planning engineers when building HV circuits from GIS.

### 5.1 Geographical Information System Data Conversion

The GIS data, as discussed in the previous sections, does not contain explicit connectivity data, this is instead derived based on proximity of lines and equipment. An understanding of the naming conventions and topology of the SSEN distribution systems allows for a similar approach to the EHV data extraction to be taken.

#### 5.1.1 Network Reference Number (NRN) Hierarchy

The NRN system used by GB DNOs has a hierarchical structure, assigning unique numeric identifiers to substations and transformers. Primary Substations have a three number code, (for example 734 is the NRN for the Pinegrove substation in the Fort William Area). Each feeder that is supplied by the primary station then has a six digit code, the first three of which are taken from the primary station (e.g. 734013).

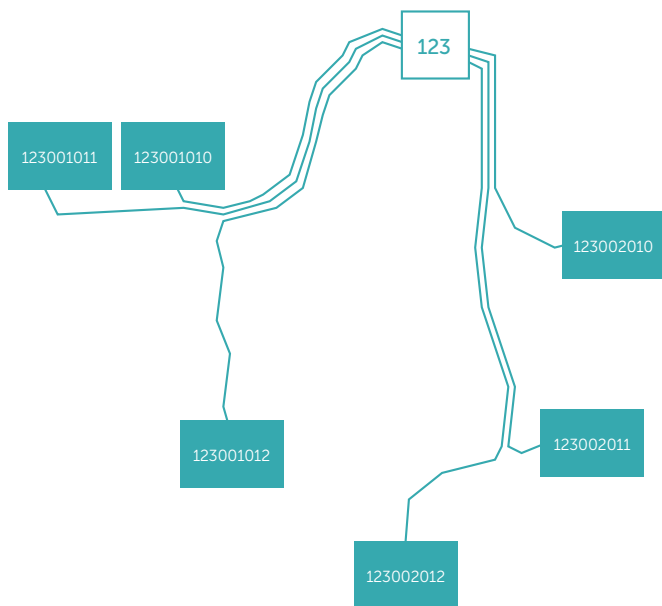
Finally the transformer on each feeder has a nine digit code (e.g. 734013010). Where a transformer is within a secondary substation, the substation is marked with the same NRN.

This NRN hierarchy allows the nominal connectivity from the GIS data to be determined by working back from the substations and pole-mounted transformers. The cable/line route is then determined using a similar approach to that taken for determining the geographical routing between substations on the EHV network. Instead of finding the route between two primary stations, instead each secondary station and pole mounted transformer is analysed to determine the route to the corresponding primary substation.

#### 5.1.2 Deriving Network Topology

The challenge with the HV data is that there is no separate electrical model defining the network topology in terms of buses and branches. For the EHV data the existing electrical model defines junction points where multiple lines are connected, so when line routing is derived it can be assumed there is a single line on this route. In some cases multiple lines find the same route, which is not surprising given towers carrying two circuits are common especially at higher voltages.

Figure 10 Secondary Substation Topology Routes

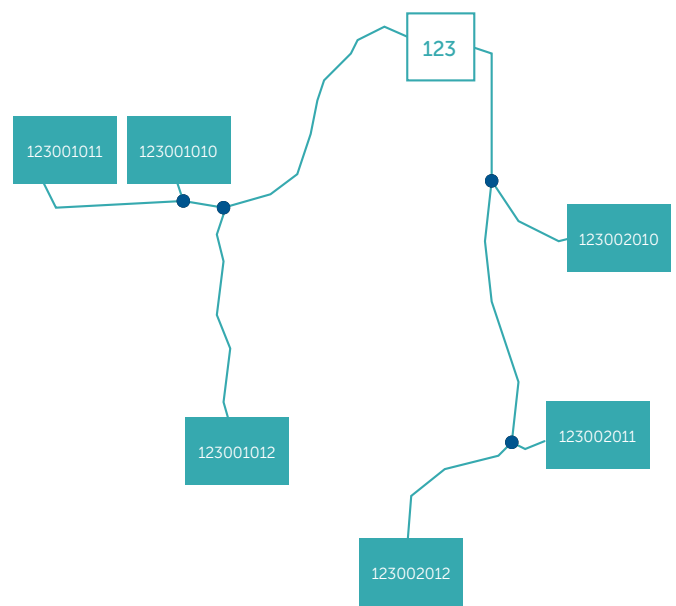


With the HV data deriving routes from secondary to primary station does not in itself derive the topology of the network. For example, on a first pass of deriving the connections from secondary stations to primary we end up with a result like that shown in Figure 10.

This is a simplified example but shows that using the same route-finding algorithm on the HV data, six paths are identified. In reality it is unlikely that each substation has its own dedicated line/cable route running to the primary station. Examining the GIS data for the HV connections does not generally show multiple lines in parallel at the HV level, so instead a second step identifies the points of connection where routes from the different secondary stations meet. At these points a node/junction is inserted, and then the single line continues.

Running this algorithm on the data above provides a topological map as shown in Figure 11.

Figure 11 Secondary Substation Topology from Connectivity Analysis



This process was repeated for the primary stations within the test network to calculate the HV topology for the secondary substations served for each primary station.





## 5.2 Line Parameter Data

For the EHV data the electrical parameters were defined with the planning data as explicit impedance, admittance and ratings data. For the HV data the line parameters in the GIS instead define the line type and voltage. The level of detail is often dependent on the age of the cabling, with recently installed equipment containing accurate line construction information and historical data (where the GIS information was digitised from printed maps), this information is either missing or marked as “unknown”.

External data sources were provided by SSEN in the form of documents listing tables of cable types and their parameters (maximum ratings and impedance/admittance per unit length). These documents were then digitised to create a standard catalogue of CIM wire information from which impedance data is then derived.

For lines with unknown types, defaults can be assigned based on the cable's voltage and whether it is an underground cable or overhead line. If data quality is improved in future GIS exports, the explicitly defined types can then automatically be used.

Line lengths can either be taken from those defined with the GIS or automatically calculated based on the line routing information. As part of data validation, where an existing length is provided it can be compared with the calculated value and any discrepancies identified. The assumption for this conversion was that the GIS lengths were more accurate and so were used where they existed. This was because they were more likely to reflect the actual length, including the additional length inherent from overhead line sagging, that is not automatically captured by a pure geographical-based distance calculation.

The impedance and admittance values were calculated for each branch by multiplying the length by the per-length values for the given type. These were then stored against each segment as explicit resistance, reactance and susceptance values.

The line ratings were taken from the same line parameter data and assigned to each line as a CIM Operational Limit Set with an Apparent Power Limit.

## 5.3 Distributing Load Data

### 5.3.1 Equalised Distribution

The planning data contains loads at each primary station representing the aggregated load from the downstream residential, commercial and industrial service points. For accurate modelling and analysis of the network the load data should be placed at the secondary substations so that the flows on each HV branch can be calculated (and thus the impact of DER placement more accurately identified).

The GIS data does not list the loads at each secondary substation or transformer, and as such the initial approach was to distribute the load equally across all secondary substations. This involved determining the number of feeders at each primary station then the total number of secondary substations and transformers being fed by the primary. Loads were added at each of these substations and transformers, the load split equally across each and the load directly connected at the primary station set to zero. This allowed for validation of the general connectivity and that the system still solved with the additional branches, nodes and distributed load.

### 5.3.2 Distributing by Customer Connections

This approach for load distribution was crude but useful for validating the initial HV networks. A more accurate load distribution required the use of data stored within the customer management system. This listed the number of customers connected to each secondary transformers and larger commercial/industrial loads that had connection agreements that must be factored into any planning studies.

An extract from this system was provided that listed the number of customer connections and/or the total minimum load required for larger customers at each secondary substation and transformer.

This allowed for the load to be split of the load from the planning model to be more accurately distributed, and for any changes in customer load profiles to be accurately reflected at the secondary substation level.

## 5.4 Limitations of HV Network Creation

The creation of the HV electrical data is limited to the quality of the GIS data. The EHV data starts with a separate electrical model that already supports planning activities, and augments this with geographical data. If no geographical route is found the electrical branch still exists and worst-case can be rendered as a direct line between points on a map (showing there is a connection even if the system cannot derive its route).

For HV data the electrical connectivity is derived from the naming conventions used by the NRNs. If stations are added without NRNs or with incorrect NRNs this will not impact the use of GIS data to show assets, but will impact deriving the connectivity hierarchy. By deriving this network from the source material, changes need only be made at source with the resulting network then automatically reflecting the changes. Similarly if a new station is added the next extract will automatically include this new section of the network.

Deriving accurate connectivity from GIS data is a challenge for all utilities. The iterative analytical approach taken for this work has allowed for automatic correction of some data quality issues, and SSEN, like other utilities, is constantly improving the quality of the GIS data as part of ongoing projects. The work undertaken for MERLIN can help this by not only correcting issues identified but feeding them back to the engineers maintaining the GIS data so as to improve the overall quality.



## 5.5 Process Validation

As for the previous GIS to CIM conversion, the resulting HV CIM data was validated using a number of steps:

- The data was validated using Open Grid Systems own internal CIM validation engines to check that the output formats conformed with IEC 61970-552 and the data attributes, references and types were in compliance with IEC 61970-301 and IEC 61968-11.
- Topological processing was executed to validate that there was a single interconnected island.
- A balanced three-phase Newton-Raphson power flow was run using Open Grid Systems' own power flow engine. The power flow successfully converged and the results were compared with the power flow simulation run on the EHV data of the same area confirming that the total system load was the same.
- A "sense-check" of the data was performed by looking at the drawn shapes from the CIM data and the GIS data and confirming that there were no obvious errors in the data.
- The SSEN GIS data was loaded into the open source QGIS geographic information tool so the data could be rendered. This allowed for manual, visual verification of the routing in the geographical display within Cimphony using the CIM data, and the original Shapefiles visualised in QGIS.
- Overlaying all data upon satellite imagery then provides a further check as any issues with coordinate conversions can be spotted where plotted elements do not align with the corresponding imagery.
- Finally, a detailed check was undertaken of the feeders by verifying routes from these stations and comparing them with the data shown in QGIS. Manually verifying the routes for cables and lines by tracing them manually within QGIS and comparing the results with the geographical routing on the satellite imagery.

## 6. Conclusions

The challenges of integrating network data is one faced by all electrical utilities. The approach described in this document covers the integration of multiple sources of data, and the challenges in deriving electrical models from GIS data. These challenges are not unique to SSEN and the quality of data is similar to that seen in other utilities in the UK, Europe and North America.

This process has highlighted a number of challenges when correlating abstract electrical models with real-world physical models but has also highlighted how advanced analytical techniques can combine with manually defined configurations to create repeatable, automatic processes for creating integrated, standardised data.

A key driver is to ensure that processes are repeatable so that updates in sources systems can quickly and automatically be reflected in the integrated analytical model. The work undertaken has shown that this is possible, but the discrepancies in naming conventions and topology can require manual updating of mapping tables and other configuration parameters when new components are added. The processes described in this document are undergoing continual refinement and validation as part of the MERLIN project's ongoing analysis and simulation work.

The creation of 11kV HV data from GIS data shows that a process previously undertaken manually by planning engineers can to a large extent be automated. This process still requires manual configuration to accommodate the missing data within the GIS extract, but as with the EHV data integration process, this configuration can be defined once and re-used for subsequent extractions.

The quality of the source GIS data and its adherence to naming conventions may require manual verification of data, but is also a driver for improving and updating the data quality.

The reliance on extracts from external systems, for example spreadsheets of line parameter data and customer counts for secondary stations has been identified an issue for long term maintainability. Integrating the conversion process to automatically pull data directly from source systems removes the reliance on extracts that themselves need to be kept current.

Further work will explore the integration of a third parties system's electrical model into the integration process, that of real-time control. This will map the real-time measurement data into the integrated planning/geographical model so that the planning applications can use real-time and historical data to better predict and analyse the network.

A further phase of the GIS connectivity analysis will explore whether the process used to build the HV data can be used with the LV data. This would extend the network to the level below the secondary, deriving the LV topology, deriving customer connection counts directly from the service points within the GIS and potentially distributing the load to individual residential, commercial and industrial premises. This would then support the further integration of additional data such as smart meter readings and the modelling of domestic DER and Electric Vehicle (EV) infrastructure.

Contact us:



**futurenetworks@sse.com**  
**www.ssen.co.uk/Innovation**  
**@ssen\_fn**



**Scottish & Southern**  
**Electricity Networks**

Scottish and Southern Electricity Networks is a trading name of: Scottish and Southern Energy Power Distribution Limited Registered in Scotland No. SC213459; Scottish Hydro Electric Transmission plc Registered in Scotland No. SC213461; Scottish Hydro Electric Power Distribution plc Registered in Scotland No. SC213460; (all having their Registered Offices at Inverlismond House 200 Dunkeld Road Perth PH1 3AQ); and Southern Electric Power Distribution plc Registered in England & Wales No. 04094290 having its Registered Office at Number One Forbury Place, 43 Forbury Road, Reading, Berkshire, RG1 3JH which are members of the SSE Group. [www.ssen.co.uk](http://www.ssen.co.uk)