MERLIN

Modelling the Economic Reactions Linking Individual Networks

Milestone 5

Data Integration Challenges

By **Alan McMorran**

21/01/2021

Open Grid Systems

This report describes the additional enhancements made to the data and the challenges that were identified concerning the availability and integration of multiple data sources to create an accurate model. This report covers the solution to each of these challenges and summarises the data sources now used for the integration and conversion process.

Contents

1. Executive Summary

This is the third 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 could provide historical measurement data. These sources were described in the Milestone 0 report "SSEN Network Data Analysis".

The second stage described the process for building the distribution system analysis model from the 33kV planning data and the geographical datasets exported from the Geographical Information System.

This stage describes the additional enhancements made to the data and the challenges that were identified concerning the availability and integration of multiple data sources to create an accurate model. This report covers the solution to each of these challenges and summarises the data sources now used for the integration and conversion process.

1 https://project-merlin.co.uk/wp-content/uploads/2020/05/SSEN_Open-Grid-Systems-M0_v4a_pages.pdf

2 https://project-merlin.co.uk/wp-content/uploads/2020/11/SSEN_MS2_OGS_V2_pages.pdf

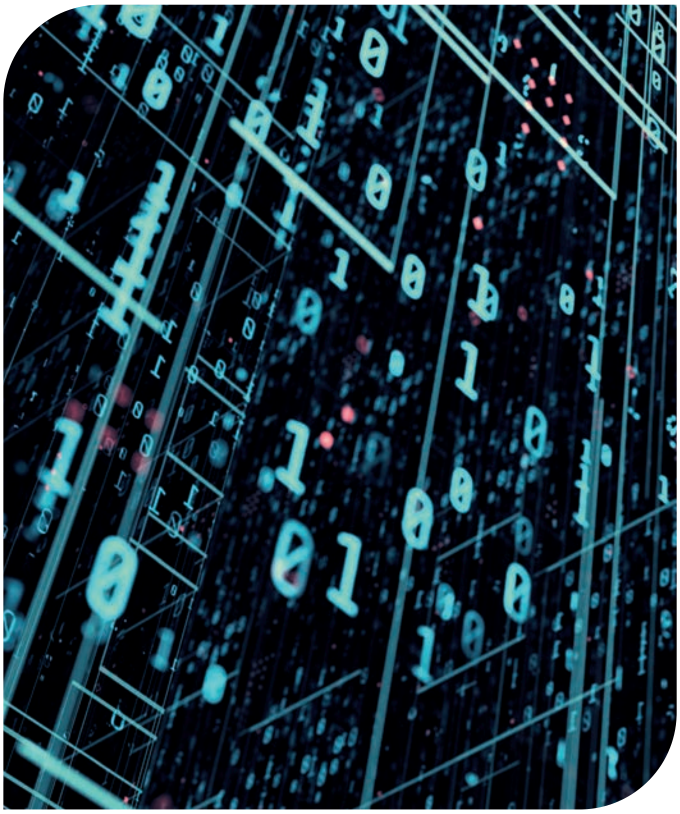
2. Introduction

The data conversion process developed as part of the MERLIN project has been discussed in two previous reports:

- **MERLIN Milestone 0**: SSEN Network Data Analysis[1]
- **MERLIN Milestone 2**: Creating an Integrated Distribution System Analytical Model from Planning and Geographical Data[2]

As the project has progressed, the data landscape has grown, with additional sources identified along with the challenges of integrating them into the Distribution System Analytical Model.

This report describes these additional sources, the challenges identified, and the solutions applied.



3. Data Overlap

3.1 Background

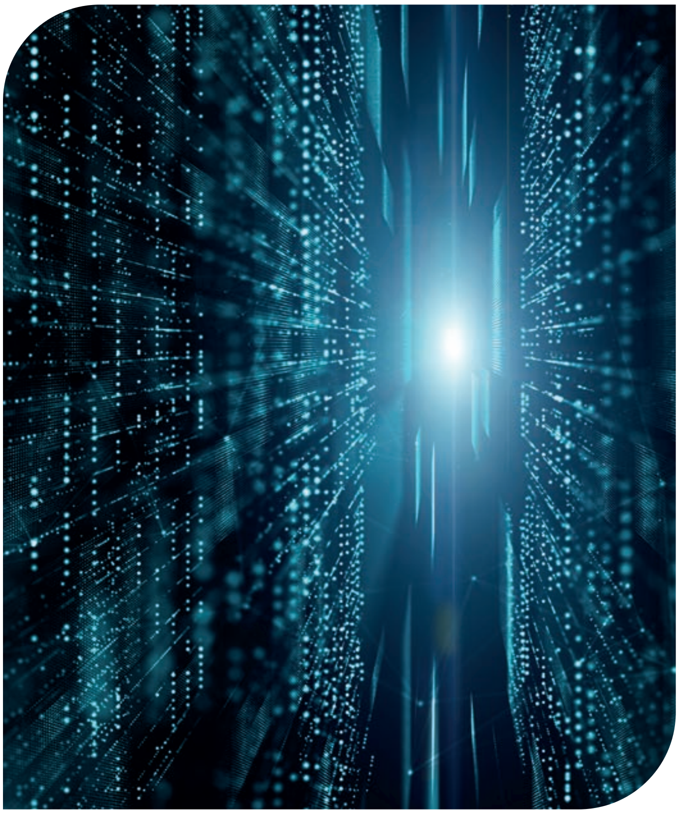
The approach defined in the previous reports covers how the core 33kV Extra High Voltage (EHV) network is extracted from the Long Term Development Statement (LTDS) data and then integrated with 11kV High Voltage (HV) data from GIS. This used the most accurate electrical models from the LTDS planning data to provide the core EHV network, with the HV data extracted from its system of record in the GIS. The overlap for EHV involved the merging of the line coordinate data from the GIS into the LTDS electrical model but did not create any conflicts. Currently the LTDS data does not cover HV feeders and as such there was a clear split between the two at the EHV/HV boundary. This approach allowed for loads to be accurately distributed across feeders, and successfully applied to the areas being modelled for MERLIN.

3.2 Generation Connection Overlap

1. An issue was identified as part of the data validation process, where generators at a hydro station in the Scottish network were modelled in the LTDS data and connected via an 11kV line to the primary station. In the LTDS data each generator has a single branch running from the hydro station bus to the bus in the primary station.

2. The GIS extraction of HV data modelled these same 11kV lines, however, it reflected the additional detail of the as-built network with all secondary substations, transformers, and connected loads, as well as small scale-generation connected to the feeder along its length. As such this single line in LTDS was multiple line segments with connection points in the GIS.

3. The single point-to-point branch in the LTDS data is a simplified abstraction of this detailed feeder from GIS, with the secondary loads aggregated at the primary station. The feeders instead had multiple points of connection along its length, with the termination point at the same hydro generation stations. By processing both sets of data then merging, parallel lines were defined: one as the detailed feeder from GIS; the other as a direct line from the primary station to the generation from the LTDS.



4. The two data sets were therefore modelling the same components in very different ways and there was no way to automatically identify if they were equivalent from the two data sets. This resulted in the erroneous creation of two parallel lines where there was only a single physical line. The solution was to exclude all the 11kV data from the LTDS conversion, and an additional mapping file was manually created that indicated which feeder from GIS the 11kV generators were to be connected to. This required additional manual mapping of the data, a one-off process, which created a mapping table that became part of the overall data conversion/ integration process.

4. Network Reference Numbers

4.1 Definition

The Network Reference Number (NRN) is a hierarchical, numeric numbering system used to identify the electrical system hierarchy. The system takes the form of:

- Primary Station: 3 or 4 digits e.g. 724 or 4623
- HV Feeder: 6 or 7 digits, the first 3/4 being the primary station, the next 3 being the feeder ID e.g. 724001 or 4623010
- HV/LV Transformer: 9 or 10 digits, the first 6/7 being the HV feeder, the next 3 being the transformer ID e.g. 724001010 or 4623010015
- LV Feeder: 11 or 12 digits, the first 9/10 being the HV/LV Transformer, the next 2 being the LV feeder ID e.g. 72400101001 or 462301001505

The NRN data allows for a nominal connectivity model to be determined for the HV/LV network, which is typically operated as a radial system (i.e. it does not contain closed loops). This could be used to augment the connectivity analysis of the GIS data to create more accurate HV (and potentially LV) datasets.

4.2 GIS NRN Data Alignment

The GIS data contains a number of NRN references, however, it is not complete and many secondary transformers and substations do not contain an NRN reference. The system of record for NRNs was determined to be another internal SSEN system used for storing customer/connection data. This data was extracted into CSV files and made available for integration with the LTDS and GIS data.

Through an analysis of this data, comparing names and postcode (location) data from the system of record with the names and GIS location coordinates, it was determined that in some cases the GIS NRN references did not always align with those of the NRN system of record. This can reflect secondary stations changing their feeder if there is a network reconfiguration, or a change between planning and implementation that are not "looped back" to the original design. This required the NRNs to be updated to reflect the system of record.

4.3 Derived Customer Numbers

The LV feeders are not modelled for MERLIN, however, it was noted that because the LV feeder data contains the customer numbers, the total number of customers supplied by the HV/LV transformer (the "end point" for the system analytical model being generated) could be derived from the downstream feeder data. Previously the customer data was taken from GIS, but as with the NRN data this was not always populated, and when it was, it was unclear if it was regularly updated when the system of record was changed. The customer numbers at the secondary substation load point for the system analytical model were therefore derived by summing the customer numbers on all LV feeders fed by the transformer.

4.4 Improved Connectivity Extraction

The connectivity extraction process described in the previous report uses the GIS data along with adaptive algorithms to derive the connections between points and stations. Issues were identified when substations on different feeders were shown to be directly electrically connected in the derived model.

During the conversion, when multiple lines passed through a common point, it was not always possible to determine when these points were electrically connected and when they simply happened to be electrically isolated but in the same place. In situations where feeder lines that ran close to (or on top of) each other, the algorithm would incorrectly consider them to be connected as they shared common termination points. Programmatically distinguishing between the two situations required more data than was available in the GIS.

The access to the hierarchical NRN data allowed this process to be adapted and updated. The NRN hierarchy was used to enhance the feeder extraction and correctly distinguish between circuits that are connected, and those which overlap geographically without being interconnected. This process used the NRNs to identify the individual circuits/feeders being fed from a primary station.

4.5 Hierarchical Extraction

The NRN number system also provided additional hints, as the secondary substation identifier generally increased in value the further the station was from the primary. For example, station 724001010 was closer to the primary station 724 on feeder 001 than station 724001020. The routing process was updated with some additional steps and criteria:

- Each circuit (e.g. 724001, 724002 etc.) was run independently to calculate the routes from each secondary to the primary
- The secondary routing order was determined by the station NRN number, running in ascending numerical order.
- Instead of using the shortest route back to the primary, the shortest route to either the primary or another secondary substation on the same feeder was used.

This significantly improved the data quality, resulting in multiple circuits that are electrically interconnected only at the primary station. The use of the NRN system of record over the GIS data also corrected some anomalous routing issues that had been identified previously, where stations were mis-identified as being part of the wrong primary station.

5. Distributed/Embedded Generation

The generation data within the initial conversion was all taken from the LTDS source. This included large-scale generators, such as hydro-electric power plants connected on the distribution network, with distributed (<11kV) and embedded (>11kV) generation represented as aggregated generators at the primary substation.

There is an addition internal system of record for distributed and embedded generation, with NRN data available for the distribution generation and geographical coordinate data for the embedded generators. The distributed generation data was extracted from this system as a CSV file, while the embedded generation data had already been extracted and is publicly available from SSEN's website as an Excel spreadsheet.

The conversion processes was updated to use these additional data sources, creating CIM DER generators for each entry, then electrically connecting them to the network. The process for identifying the connection point varied for each data set:

- For distributed generation, the extraction included the LV Feeder NRN (cross-referenced with unique meter number as part of the extraction process before the meter number was redacted for security purposes). As the network model being generated did not include the LV network, the generators were connected directly to the secondary substation derived from the LV Feeder NRN.
- For the embedded generation, the geographical location data and primary substation name was provided. For generators connected at 33kV the process was to identify the primary station based on its name, then the connection was placed directly at the primary station bus. For generators connected at 11kV the geographical coordinates were used to find the nearest secondary station(s), confirm they were downstream of the primary station for that generator, then the generator was connected at the secondary station bus.

The omission of NRN data for the embedded generation and use of spatial querying to identify the secondary station, raises the possibility that a generator would be incorrectly connected to a station that is geographically closer but not the station it is physically connected to. As the publicly available Embedded Capacity Register does contain the unique meter (MPAN) number, it should be possible to correlate the MPAN with the NRNs. This possibility is being explored with SSEN to enhance the available data.

6. Summary

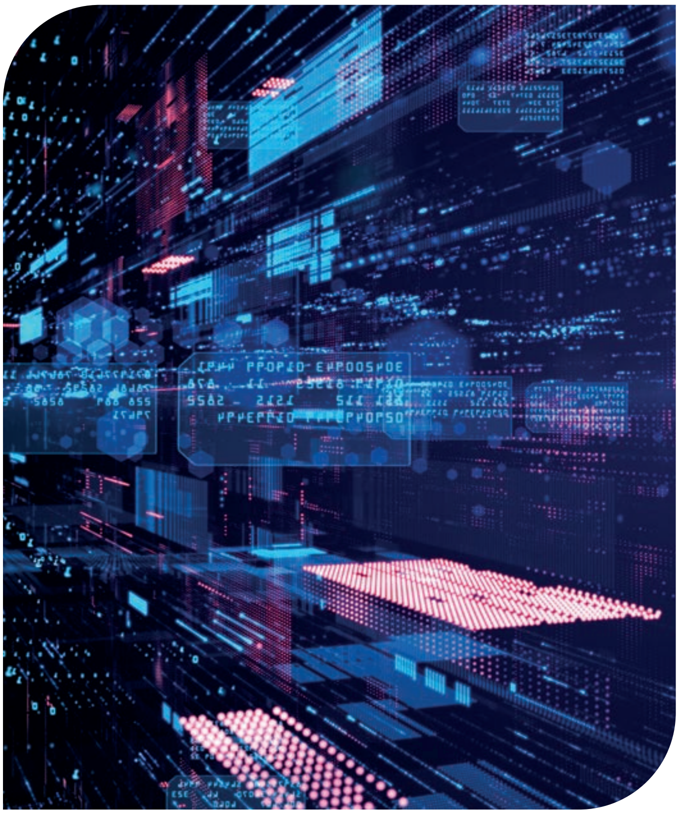
The initial data extraction process focussed on the merger of LTDS and GIS data to create the system analytical model. As this work has progressed, additional data sources were identified and integrated into the process. The process now involves multiple data sources and steps, all automated into a single consistent conversion process. The differences in data sets between the two SSEN distribution license areas, Scottish Hydro Electric Power Distribution (SHEPD) and Southern Electric Power Distribution (SEPD), also requires different configurations for each.

The data sources currently used are:

- Geographical Shape Files from the GIS
- LTDS planning data
    - Excel spreadsheets for SHEPD including public data sheet and internal master spreadsheet
    - PSS/E network analysis file for SEPD in proprietary RAW text format
- HV connected generator mapping data CSV
- Electrical line impedance data (manually extracted from PDF into CSV)
- Electrical line rating data (manually extracted from PDF into CSV)
- Authorised capacity data for secondary stations Excel spreadsheet
- NRN Primary Substation Data CSV
- NRN Secondary Transformer/Substation Data CSV
- NRN HV Feeder Data CSV
- NRN LV Feeder Data CSV
- Distributed Generation register CSV
- Embedded Capacity Register Excel spreadsheet
- UK Postcode data for location matching
- Load data extraction for feeders from PI Historian as Excel spreadsheet

In addition to the above sources, the process uses a registry to track the assignment of universally unique identifiers (UUIDs) for persistence between runs.

The process has highlighted how data used to create detailed system analytical models, is stored in multiple systems requiring integration, correlation and conversion. As more data sources are identified the overall process has been improved, resulting in better quality data and more accurate analysis to be conducted.



Contact us:

futurenetworks@sse.com
www.ssen.co.uk/Innovation
@ssen_fn

MERLIN
Modelling the Economic Reactions Linking Individual Networks

Scottish & Southern Electricity Networks