Integrity of *Data*

Grid Data Quality – The Ultimate Checklist

Intro

High data quality is a default precondition for digitizing and automating many processes that are currently seen as a bottleneck for a fast energy transition. If you have all relevant grid data at your fingertips, you can identify *new grid connection points* more quickly, make *more accurate calculations* around grid connection process, better pinpoint the areas where *grid expansion or reinforcement* measures are most urgently needed, and *make more sustainable and reasoned decisions* for strategic grid planning.

Ultimately, you will also be able to introduce intelligent controlling of the power consumption or power generation at peak time, using the smart meter data.

Yet many DSOs we speak to wish their data had much higher quality than it is now. At the heart of this issue lies the problem of "Shit in = Shit out". When data has been input incorrectly, it affects analytics, insights, planning and investment decisions.

One of main issues with data quality assessment and improvement that distribution system operators have is caused by a heterogeneous system landscape that has grown over many years. Which means, knowing that your data is not at an optimal level of accuracy is not enough. You also need to know where the need for improvement exactly is, and considering the massive amount of data scattered across various software systems and databases, that's not an easy task.
So, how can this issue be solved? By aggregating and linking data together in a shared context. This way, you can assess its quality in a more structured way and be laser-focused and consistent about improving it.

Because the crux of the matter is:

If you *don't do anything* about the quality of your data, it won't get better *on its own*.

The Checklist to *Improve* the *Quality* of your

Grid Data

Checklist

STEP#1

Make an inventory of all relevant data sources

Questions to ask

What internal data sources contain the information that allows you to understand what happens in the grid?

How

Make an inventory of all the systems that provide geospatial data, information on individual grid assets, data on the grid participants (households, PVs, heat pumps, etc.), etc.

The most common systems are:

1. Geographic information system (GIS)
2. Distribution management system (DMS)
3. Network information system (NIS)
4. Asset management system
5. Energy data management (EDM)
6. Enterprise Resource Planning (ERP)
7. Supervisory Control and Data Acquisition (SCADA)
8. Meter data management system (MDM)

Considering that in many EU countries, more and more households are equipped with smart meters, you'll also need to include the systems that record and provide smart meter data, which will enable a more granular analysis of the electricity consumption and generation behavior.

STEP#2

Designate your single-source-of-truth system

Questions to ask

Which system can be the hub where all data will come together?

How

You're looking for a system that will enable you to create grid models and run complex calculations in order to validate data. The results of these calculations will indicate information gaps and inconsistencies which you will use in order to correct data in the source systems.

As long as your data sits in silos, you can’t know what needs to be improved and where.

To identify the information gaps and errors, you need to aggregate all that data in order to be able to view it holistically and in the right context.

In addition to that, the results, provided they are based on correct data, can be used for further purposes, such as optimizing investment decisions for strategic grid expansion.

STEP#3

Aggregate data from all relevant sources

Questions to ask

1. Is the data I work with up-to-date?
2. Do I have flexible integration options?
3. Has my data been maintained consistently?

How

1. Our distribution grids are constantly changing because of the massive adoption of renewable power systems in the private sector. This is why keeping data up-to-date both in your single source of truth as well as in the source systems has the number one priority. This is crucial so that you don't perform important calculations on the basis of obsolete data.

2 It is generally recommended to use standardized formats for data exchange between various systems. However, some standardized formats such as CIM or CGMES allow a high level of customization, which often makes it hard for a plug-and-play data exchange. Therefore, it's better to work with configurable, when possible API-based connectors (aka interfaces) that are flexible and scalable enough to enable faster and easier data aggregation and transformation.

3 Ensuring data compatibility upfront or at least as soon as possible is important to make data aggregation as less time and efforts-consuming as possible. Ideally, data has been maintained consistently all along through the assignment of a unique identifier for individual assets or facilities that is shared between all relevant systems. If this hasn't been the case, then this practice should be adopted at least as soon as you start on the evaluation and improvement of your grid data quality.

i

A final evaluation of data quality should take place ***AFTER*** data is aggregated. Each single dataset can seem error-free in itself; however, it is only when you put all datasets in the context of a power grid model that all inconsistencies and errors can be detected.

Good to *KNOW*

How up-to-date exactly data must be, depends on two factors:

1. How fast these data are updated in the first place
2. Whether you're going to use these data for planning or operational tasks

Usage for planning tasks

When it comes to planning-related tasks – e.g. for strategic grid planning or evaluation of new connection requests – there is generally no need for having the grid status in real time. It's enough to know its general, usual state. This also sets the approach you'd choose for data integration: Depending on how fast data changes happen, even manual exports and imports may be sufficient in individual cases (although it's generally recommended to automate data exchange whenever possible to avoid unnecessary waste of time).

Usage for operational tasks

In the case of operational tasks – for instance, for grid congestion management or incident detection – you'd generally want to get alerted to errors and problems in real time. Therefore, data must be updated every minute or maybe, even every few seconds. For the approach to data integration this means – automation of data exchange is unavoidable, and your integrations must be robust to handle large amounts of data while keeping the latency low.

STEP#4

Check the quality of individual datasets

Questions to ask

Do you have all pieces of information and are they accurate?

How

Ideally, you have a tool that allows you to run automatic checks on single datasets – e.g. information related to a specific transformer or a substation – and detect information gaps and inconsistencies such as false KVa value on a fly.

STEP#5

Run contextual quality checks in a grid model

Questions to ask

1. Are all grid areas actually connected with each other in the grid model?
2. Is all data we have aggregated in a grid model is correct in a given context?

How

1. Examine the grid topology by basically performing the so-called power grid tracing. This will help you detect the areas of the grid that are not connected with each other in the model, because, for example, the switches on both sides of the line have been wrongly recorded as open. Then you can correct the gaps and errors in the model bit by bit and in a consistent way.

It can be done manually, but ideally, you can deploy a tool that allows to detect such affected areas ***automatically***.

2 Run load flow simulations to check all data for accuracy and plausibility in the context of a specific distribution grid. If, for example, the iterative process of the load flow simulation isn't converging or delivers unrealistic results (say, the load factor of 300% or more), it is a clear sign that some data is erroneous and requires correction. Using this approach, you can fairly easily trace the origin of the errors and correct data in the responsible source system.

Real-life examples

The following examples come from one of envelio's customers, a German distribution system operator FairNetz. The single source of truth system that FairNetz has deployed is the Intelligent Grid Platform (IGP). The internal data source systems include NIS, EDM, GIS, ERP.

Data enrichment

The grid calculation for a specific low-voltage area showed a high load, even though this area had only a few electricity consumers according to the data provided from NIS. Upon closer examination of the consumer data that came from the EDM, FairNetz was able to determine in the grid model that the high load was caused by a large number of heat pumps recorded in the system. Yet, the technical information about these heat pumps was missing in the NIS entirely. As a result, FairNetz enriched the data in the NIS by adding approximately a thousand heat pumps.

Data correction

Although the IGP indicated a high calculated load for a specific area, the actual station load based on measurement data was found to be low. The data investigation revealed very high capacity coming from storage heaters that had been recorded in the NIS and therefore, transported into the grid model where they were used for calculations. At the same time, FairNetz could see in the EDM data that many end-customers who presumably were in possession of storage heaters had relatively low electricity consumption. As a result, FairNetz conducted a customer survey and was able to remove 25 MW of storage heating capacity from the NIS. These 25 MW reflect approximately 6% of the substation capacity in FairNetz's grid.