

What and why - commonality and compatibility

“...we require a system which can act on the data, collate and send it onwards to a coding/modelling environment. **This is a data pipeline.**”

Daniel Riddoch

Date of release 24/01/2025

What do we mean?

With any large project, there will be a large amount of data, coming from a wide variety of sources. To best use the data we can collect, we must be able to use all of it, in any combination we wish. This allows data scientists to produce large and complex models, and will allow AI/ML models to develop a full picture of the system.

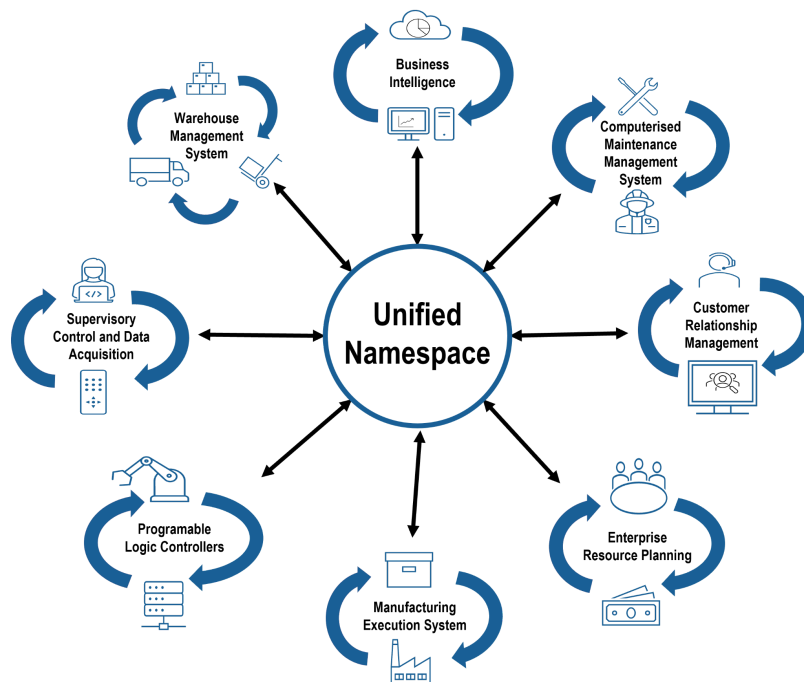


Figure 1: How a unified namespace can unite disparate business areas(Credit: Op-tec Systems)



The problem is that data is not always so neatly arranged. Data from different sources may well be in different formats, have different names or units. Converting everything can take time, and missing name or unit changes can lead to errors in coding and/or erroneous results. One goal is to create a unified namespace, which can allow all different business areas, from process and operations engineers, through storage and support to management standardise and communicate clearly and without error.

How does the problem arise?

When dealing with data from different sources it is common to have to connect different sensors and actuators in different bespoke ways. This is extremely inconvenient, but even once set up, this can lead to variables which represent the same thing being measured in different units (an accelerometer may measure time to the thousandth of a second, whereas a PLC may measure to the second), with respect to a different basis (time will be measured relative to when the sensor is turned on, or using an internet connection) or in incompatible data types.

Another, perhaps less obvious problem, is that of variable naming. Data from three different sources could label the time measured as: 'Time,' 'time,' and 'Elapsed time.' Each of these names is absolutely correct in isolation, but the subtle differences between these three would cause a problem if the data associated with them was to be combined.

This problem can also develop over time. With any process, machines will gradually be replaced as they become life-expired, or if significant changes make replacement machines more efficient or changes in process which require different machines. These changes can disrupt commonality, and will change how, and for how long, data has been collected.

So how can we approach solving these issues in a non-invasive, passive and streamlined manner? It is important that we create a separate system to deal with these issues before coding or modelling is done. This will ensure that all inferences gained from modelling are accurate and trusted and there is no chance of the confusion that could arise from having any references to non-standard units, names, types or bases.

How can we solve the problem?

To do this, we require a system which can act on the data, collate and send it onwards to a coding/modelling environment. This is a data pipeline. A data pipeline, with a microservice architecture allows us to perform these alignments, translations, conversions and renaming on the data, re-collate it, and send it on to another environment.

By performing these changes within a data pipeline, the calculations can be performed on the edge, meaning any inference done remotely (using cloud computing for example) can be isolated from the

non-unified data. This also means minimal latency and storage difficulties as the changes can happen on the edge and no data needs to be stored or sent elsewhere for pre-processing.

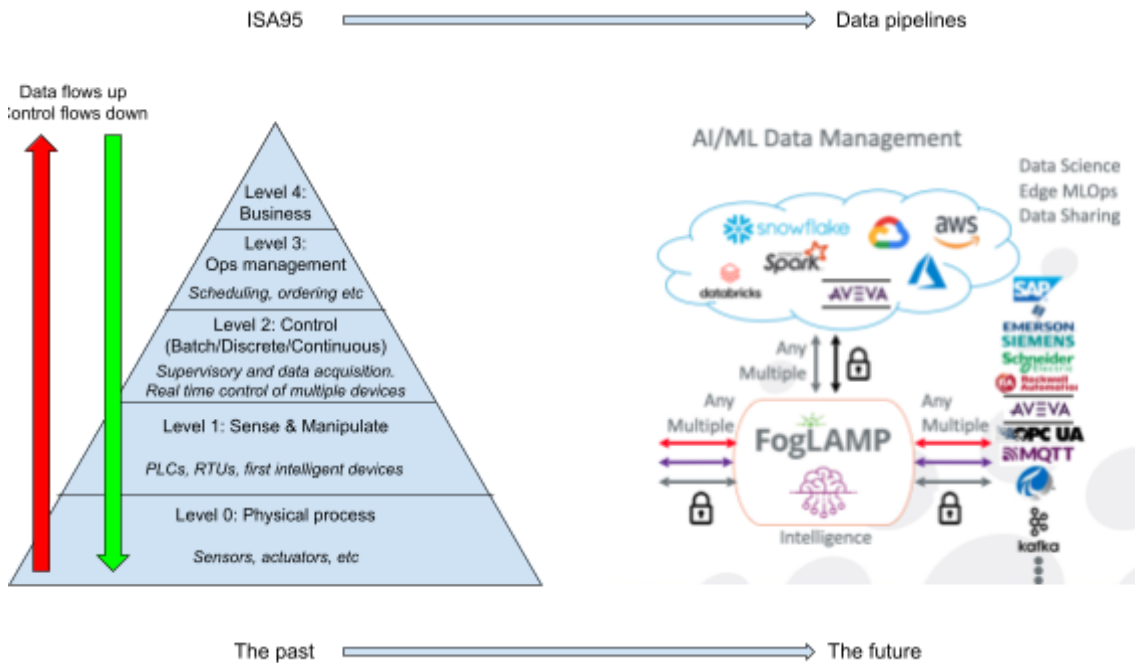


Figure 2: A sketch showing the increased flexibility to act on data in different ways using data pipelines

What other benefits arise?

Many pipelines replace the need for the bespoke connections between each sensor and a coding environment, giving a single system for collecting, aggregating, unifying and transferring data. This will reduce the system complexity but also the cost of purchasing and maintaining several different systems.

Data pipelines are also easily replicable and scalable. So building further capacity, whether by sending more data through the pipelines or building parallel or interconnected pipelines, is easier than with a bespoke approach.

Example

The following example comes from a real use case, where Réseau de Transport d'Électricité (RTE) teamed up with Dianomic and the Linux Foundation to update their existing substation infrastructure to standardise in line with IEC 61850.

<https://lfenergy.org/how-french-transmission-system-operator-rte-leverages-lf-energy-and-lf-edge-to-build-next-gen-substation-monitoring-and-controls/>

IEC 61850 is an internationally agreed standard for communication protocol of substation equipment. The objective is to standardise substation communication with external resources, to facilitate standardised semantic modelling of substations and optimise the working environment to allow for best practice condition monitoring and preventative maintenance, thus providing best case supply characteristics.

One of the problems faced by this project was that there was no need to upgrade infrastructure that predates IEC 61850, and to do so would be expensive and disruptive. As such, legacy systems had to be integrated into a framework compatible with IEC 61850. This required, among other things, a way for legacy equipment's communications to appear as if they were from an IEC 61850 compliant device.

This required data to be gathered from all the equipment within a substation, and transforms to occur on all the data gathered to make it IEC 61850 compliant. This had to occur near real time, at the edge and separated from any external involvement, and in a way that was scalable and could be deployed within many different substation sites and environments seamlessly.

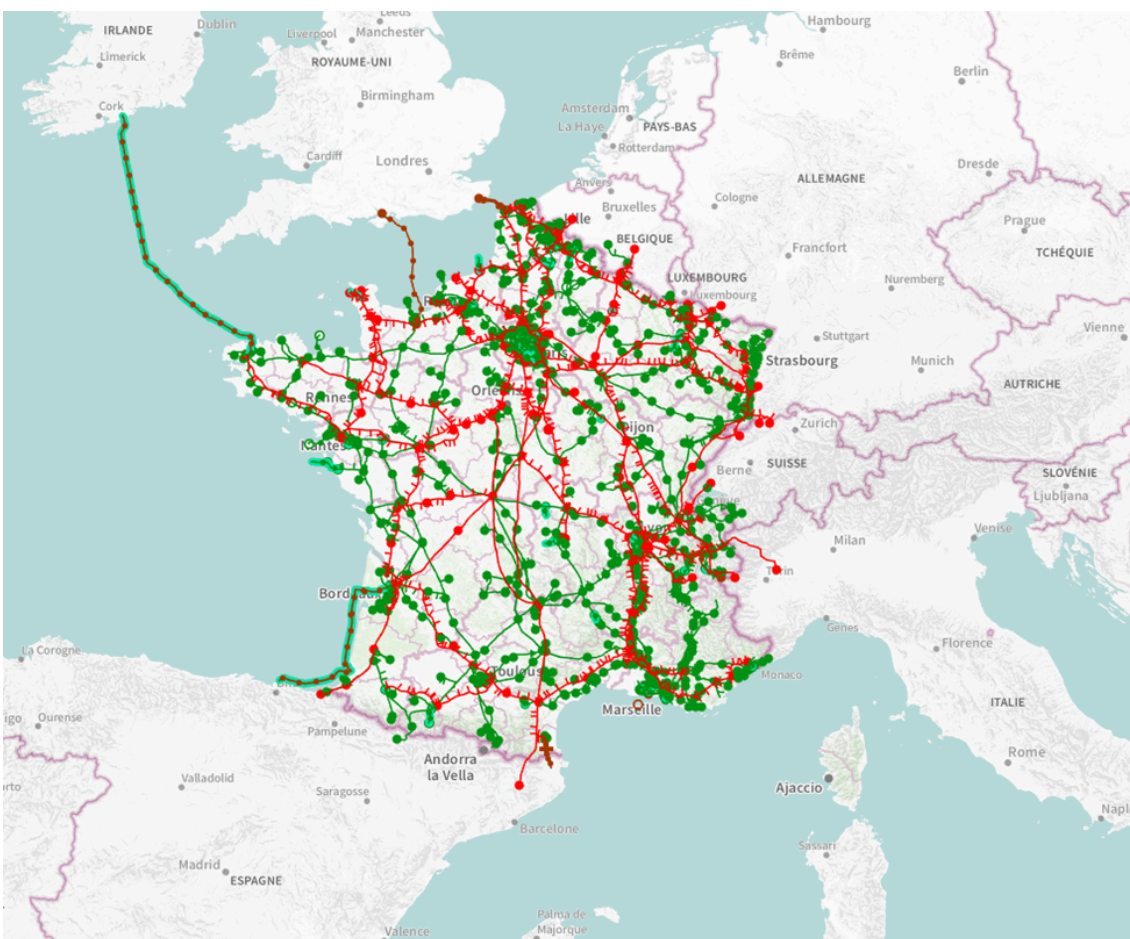


Figure 3: RTE’s transmission grid – Red are 400 kV substations and Green are 225 kV substations. 150, 90 and 63 kV substations are not displayed for readability(Credit: Linux Foundation)

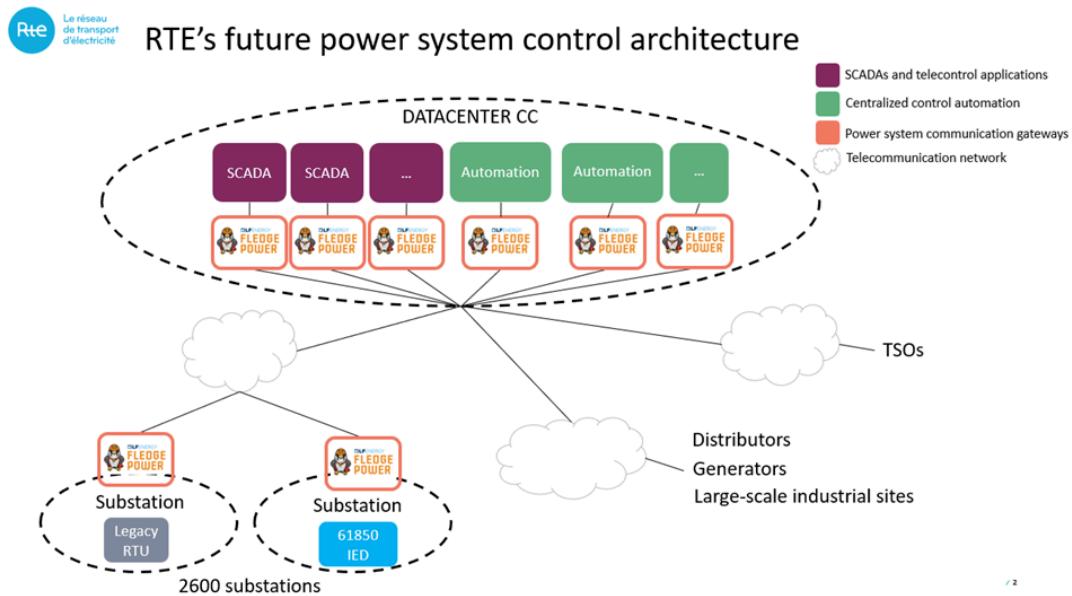


Figure 4: Sketch showing how RTE addressed the architecture requirements for converting legacy architecture to IEC 61850(Credit: Linux Foundation)

As such, RTE created the Linux Foundation Energy’s FledgePOWER gateway, working with Dianomic’s FogLAMP suite. The FogLAMP suite provides the flexibility, deployability and scalability to meet the challenges of multiple data transformations, different communications protocols and different running environments. The plug in architecture made the development and deployment of the solution seamless, allowing for easy scalability, and integration with future substations designed following the IEC 61850 protocols.

The result of this upgrade was a significant cost saving, compared to moving the devices over to new hardware, and a significant reduction in complexity too. For future-proofing, the solution is also bi-directional, meaning that set point control can be automated within the pipeline allowing for automatic machine control. Furthermore, the solution is also ML/AI ready so when more intelligent tools are developed they can be seamlessly integrated. Finally, the FogLAMP suite’s single management architecture makes it easy to manage multiple different machines within many different substations.