REALISING NETWORK INTELLIGENCE THROUGH MASTER DATA EXPLOITATION AND DYNAMIC DATA MODELLING

Jim WHYTE
SP Energy Networks – UK
Jim.whyte@SPenergynetworks.co.uk

Fiona FULTON
SP Energy Networks - UK
Ffulton@SPenergynetworks.co.uk

Alex WHITE
CGI- UK
Alexander.white@CGI.com

Adrian PUTLEY
CGI- UK
Adrian.putley@CGI.com

Bruce PAISLEY
SP Energy Networks – UK
Bruce.paisley@SPenergynetworks.co.uk

Brian ROBINSON
CGI - UK
Brian.robinson@CGI.com

ABSTRACT

With the coming tsunami of 'Smart Meter' and 'Smart Grid' data SP Energy Networks (SPEN) considered that it needed to explore 'Dynamic Data Modelling' within 'Master Data' structures. This step is considered essential in determining the future approach to Data Management within this predicted 'Big Data' environment. A project called Distribution Intelligence for Network Operators (DINO) was formulated to facilitate phase 1 of this exploration. This initial investigation consists of enhancement of operational insight from a high volume 'Alarm' data output that is generated within our existing Network Management System (NMS). The dynamic data source is a high volume, bespoke relationship, alarm event historian system called PSAlerts.

Novel dynamic data modelling techniques were explored to relate this dynamic data source to a specific integrated network, communication and protection master data model. This 'Integrated Network Model' (INM) was crafted from a snapshot of the overall NMS enterprise data relating to the network automation business area; along with a locally held, informally managed, semi-structured radio communication dataset. Daily experience with the 'Automation' team business process led to specific use cases that could utilise this dynamic data stream analysis, but are currently masked in the noise due to the volumes of data. Complex Event Processing (CEP) translated the dynamic data streams into the required network intelligence. Final integration of both these into a Proof of Concept (PoC) dynamic 'Smart Network Model' (SNM) was completed to provide the final output visualisation and follow-on action capability.

INTRODUCTION

Data Management

There are 3 fundamental siloes of enterprise data that are 'Data Managed' within Distribution Network Operators (DNO's). These are Asset Management, Network Management Systems (NMS) and Geospatial Information Systems (GIS). Each of these fundamental systems has different business users and business functions. Hence, each System holds specific business information. However, they also rely on cross-system information built into data models whose individual silo requirements can vary significantly between the systems in structure, level, connectivity and granularity. Business requirements have therefore created complex point-to-point integration of data across the 3 siloes but only where they intersect operationally. For example the cross-system maintenance of network plant information for asset management, design & planning and network operation. Dynamic network updates and complex integration requirements are predominantly managed manually by a variety of 'Data Management' teams along with suitable IT and system admin support.

Within each enterprise system there are also a number of specialised 'enterprise' level datasets that are required to be integrated and maintained within the underlying data models. These data sets are used to manage increasingly complex network management and regulatory reporting requirements, e.g. customer databases for CI/CML regulatory reporting, customer quote data for connection performance reporting, plant health indices for asset management.

Underneath both of these levels of 'enterprise' data requirements there are also a number of supporting semi-structured data sets and dynamic data streams that have been generated within the business or come from new innovation projects. Some of this data has not yet been allocated within the enterprise systems. These are therefore usually managed within local Excel, Access and bespoke data collection systems. Though they always start as a reflection of 'fit for purpose' data quality and alignment their locally managed, ad-hoc nature mean they eventually degrade as a result of the dynamic nature of the network and this increases risk and impacts operational efficiency.

Last but not least is the predicted dramatic increase in dynamic data volumes which will come from increased levels of network monitoring and control. This requirement is derived from analysis of the change requirement in the industry to cater for future Low Carbon Technology (LCT) scenarios. These change requirements over 2015-23 are anticipated in the SPEN RIIO Business plan[1] and recently reinforced by a published Distribution System Operator (DSO) vision document[2]. The major innovation research drive within DNO's, including the DINO project, is the need to explore methods to cater for change to this more dynamic system operation and management.

Network Automation DINO use case

Traditional network management cannot be considered particularly dynamic, with the main areas of automation coming from specific requirements like protection fault clearance or voltage regulation at primary substation. These systems did not generate much data and any data signalling was traditionally transported over copper pilots between required sites. Things became more complex with the introduction of primary substation Remote Telemetry Units (RTUs) that started to feed increasing volumes of network monitoring and alarm data to central control systems over SCADA (Supervisory Control and Data Acquisition). This started the journey to increasingly complex NMS IT systems for integration of this data into the existing central operational network 'Control' processes.

Increasing complexity was then driven by regulatory incentive focus on outage management and Customer Minutes Lost (CML). This led to the industry to invest in the first wave of systems to actively manage the reconfiguration of the network based on reaction to fault indication. The resultant output of these systems was to improve CML through maintaining continued availability of existing supply after transient fault or automatic reconfiguration of the network to restore supplies for those not affected by the 'isolated' permanent faulted component. This requirement is termed in the industry as 'Automation' of the network and has now widely been installed in many rural and urban applications.

This can be argued as being the original of what is currently being termed within Smart Grid innovation as 'Active Network Management' (ANM) techniques. It has to be noted though that this would be a non-persistent ANM requirement as it is only activated based on the presence of fault conditions. There is also a relatively low consequence of failure as the result of non-operation only resets to previous non-automation levels which fails to deliver the improved CML benefits.

This increasingly automated network complexity, on top of the growing number of protection alarms and network monitoring requirements are already producing resultant data management challenges in the NMS system:-

- **'**<u>Too much Data'</u> – 5,500 automation sites and potentially 6-8 million alarms per year (10-15 per minute), leading to real-time priority "signal to noise" issues and overall comprehension difficulty
- Lack of <u>'Data Integration'</u> between Alarm data and multiple network and communication models hampering holistic problem interpretation and situational awareness

It also has to be noted that 'ANM' techniques are a key pillar in the Smart Grid and DSO vision where the requirements will become increasingly persistent to manage network constraint and of higher consequence where things like thermal, voltage or fault level limits are being maintained by them. This future requirement will inevitably lead to much greater volumes of feedback data, control signalling and communication management.

APPROACH

Therefore for this initial introduction of the DINO concept we had the combination of this developing dynamic network vision along with support from within the existing business 'Automation' team. This team had been built with the vision of a very progressive engineer who had been the driving force for this requirement into the business. His team are at the forefront of the issues these growing dynamic network requirements create. He clearly foresaw the challenges that increasing volumes and complexity of the communication & control system, along with associated data management requirements, were bringing into the business. Recognition of the benefit that this project could bring to his team now and from the predicted increased in ANM requirements was key to the success of this project.

Our chosen partners to deliver this investigation of improved Data Management possibility, within a clearly defined business use case, were CGI. It was clear on discussion with them that they had previous experience of the growing complexity of data management within the data cultures of DNOs and the problems this creates when trying to drive data driven innovation. They had also given this problem some considerable thought and brought to the table a coherent path forward within the concepts of 'Master Data Management' and 'Dynamic Data Modelling'

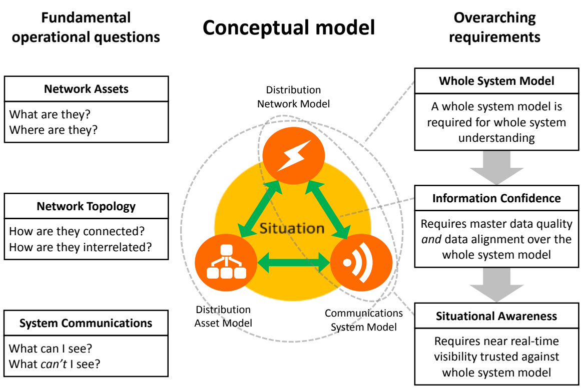

Fig 1: Situational awareness requirement

The challenges we set CGI were,

- The issue of too much data
- The issue of data exchange/discovery

All within the context of the automation PSAlerts business use case.

We also wanted to explore benefits from modelling the solution in the SGAM framework and understand the potential use of CIM to help with eventual data discovery and exchange requirements.

Integrated Network Model

In this use case the Master Data modelling requirements were in the main derived from the enterprise NMS system. This is GE PowerOn Fusion (POF) in SPEN. An interesting observation on the connectivity within this system is that it is built as a connected networks model but without electrical/thermal characteristic values.

Complication in creation of the INM came from the complex internal data modelling structures in POF. This is managed within the Control room environment by a specialized Operational IT team. This team manages the overall POF data model including naming and classification of all the electrical objects. POF also held 3 related but separate data models that were relevant to the use case,

- Plant Model
- RTU & Alarm Model
- Automation model

Analysis of this data model created additional modelling consideration for the project team. After careful consideration a pragmatic approach was chosen to create a more manageable INM. Of the various choices available it was decided to go down the route of not incorporating the SCADA configuration data held by POF which links the actual electrical components to the physical RTUs. Instead, linkages between network asset SCADA points and RTUs are inferred on the basis of these being co-located at the same network sites.

The communications architecture of the secondary automation (SA) network also causes the data points on this network to appear to be grouped within large (virtual) RTUs which appear, to POF, to be at the data concentrator sites. The presence of the secondary RTUs therefore had to be inferred when building the INM, though this was easily achievable by combining the POF data with a database of automation radio communications topology maintained by the automation delivery team. Some simple fuzzy matching was needed to match the latter's sites names to the substation sites in POF, which achieved a hit rate in the 9x% range. This was fine for the PoC but would need further refinement or automated inferencing plus a limited amount of manual data clean-up work to deliver 100% matching within practical BaU systems. To complete the INM data map, the derivation of geospatial co-ordinates from GIS for final visualisation was also required.

CGI took the approach below after consideration of the INM modelling requirement and the fact that this initial exploration was at PoC scale,

- Base INM on CIM, keep it as simple as possible:-
  - ➢ Used a Graph database which closely mirrors CIM classes (vertices) and associations (edges)
  - ➢ Use techniques to avoid need for voluminous class hierarchies that can occur in pure CIM
- Connectivity modelled as per CIM using Terminals and Connectivity Nodes
- Geometry fields included for spatial coordinates and (future) schematic positions
- Data Quality Issues linked to affected data records and (future) corresponding Data Dictionary items.

This approach provides the interchangeability of CIM without being tied to it in every tiny detail.

Project data learnings from this process included,

- Some inconsistencies were present in the POF data (mostly historical and known to the team)
- The Automation radio communication local spreadsheet database, which was built out of necessity by the original project management team, is a good illustration of an inappropriate repository for structured business data. Its continued use though is due to a lack of a clear 'enterprise' data integration pathway.
- Fuzzy matching was needed between substation sites and RTU/concentrator/radio nodes, currently only related by their site names rather than by appropriately managed common IDs.

Complex Event Processing

Complex Event Processing of high volume data streams derives the valuable network intelligence to present to the business users. This is used within decision support analysis and drives proactive or remedial actions. This is achieved by correlating the relevant dynamic data events against the network master data (the INM) and filtering out non-relevant events. This capability can derive logical, temporal and spatial data relationships and patterns, and for DINO the rules have been derived from requirements for the chosen use cases.

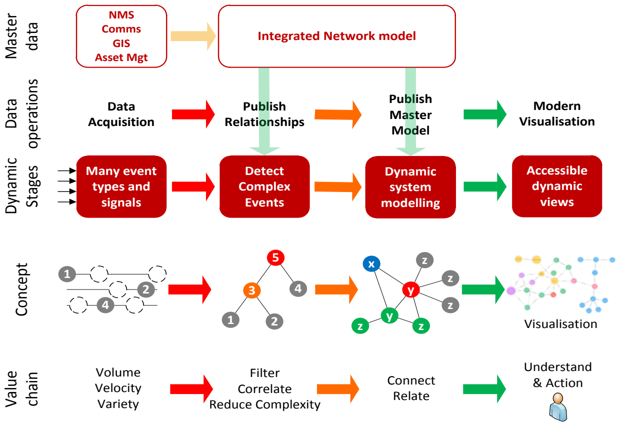

Fig 2: DINO concept overview

At the start of the project a number of business use cases were presented to CGI for conversion to CEP algorithms by the project sponsor within the Automation team. There was a requirement through the course of the project to develop and adapt this list. The final use cases that were chosen to demonstrate the business value of the CEP concept were as follows:-

- RTU loss of volts – manage escalation to prevent equipment failure
- Slow control execution – detect "false" control failures
- Standing communications fails – re-annunciate periodically as a reminder
- Identify potential point of failure in communications mesh

It will help to present the alarm progress and business process risk of one of these in detail (Fig 3),

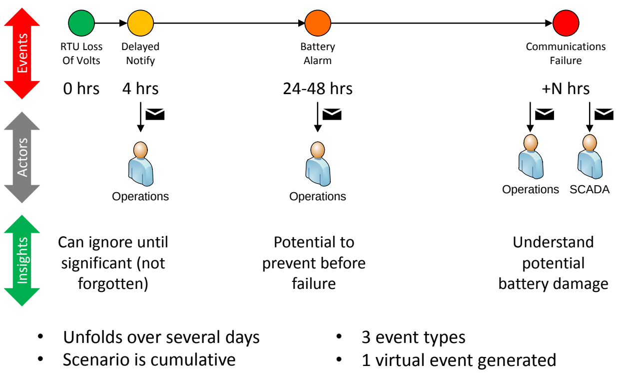

Fig 3: RTU loss of volts use case

Within this use case, loss of site volts can, over the course of days, see total loss of site alarm functionality through total discharge of and damage to the battery at an RTU. With the number of automation sites and alarms that were triggering and clearing this overall pattern was very difficult to spot until the loss of this vital functionality was eventually reported explicitly, not until after the event. This CEP gave the opportunity to action this pathology before it reached critical levels.

APPLICATION

Within the DINO project CGI delivered a Dynamic data model which presented the output of the integration of the business use case CEP algorithms with the INM. This they defined as the Smart Network Model (SNM). A geospatial and dashboard presentation of the resultant network intelligence was created. Google Maps was used to provide the geospatial insight due to its ease of integration and low cost. Geospatial locations were derived from existing substation locations available from the enterprise GIS system. CEP Event streams could be interrogated within the geospatial location point (Fig 4) and were tabulated and gathered within a main dashboard presentation (Fig 5).

An additional automation radio communication map with point to point mapping was installed for the identification of potential points of failure in the communication mesh CEP (Fig 6):

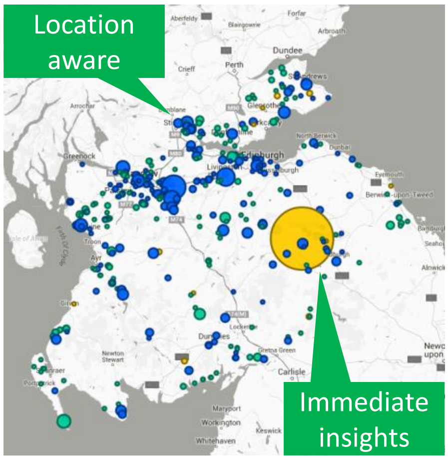

Fig 4: CEP spatial visualisation

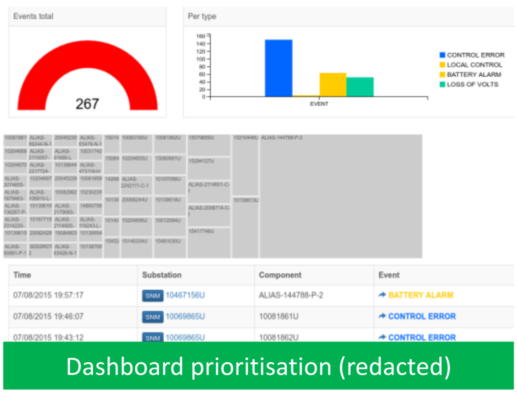

Fig 5: CEP dashboard

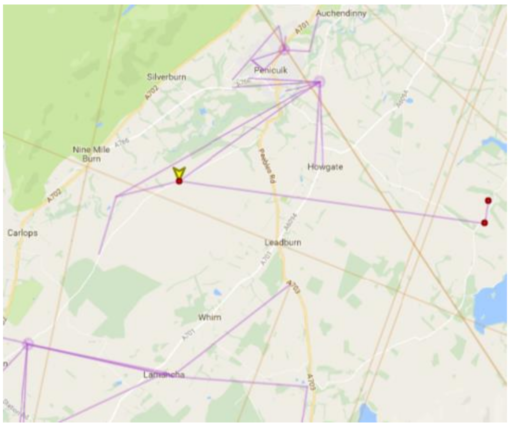

Fig 6: Automation Radio Comms mesh

Visibility of INM data was also provided within the system presenting a use case specific view using the SNM. The Graph database showed the full complexity of the relationship between plant, protection and communications within any network site. The presented picture (Fig 7) is an overhead line autorecloser; larger primary sites are of considerable complexity! Network intelligence was also built into the SNM views with indication of event alarm and consequence:

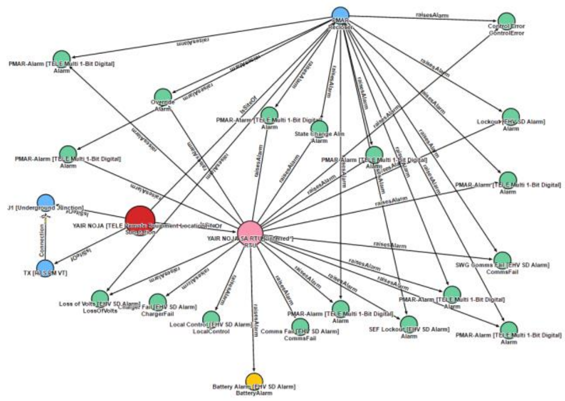

Fig 7: Autorecloser representation in Smart Network Model (Active Battery Alarm)

The level of delivery of this SNM system was at PoC level (Fig 8) and decisions to keep costs low were made for the technical architecture of the overall system. The use of Cloud technology was a major innovation point within SPEN as this led to additional security requirements with a very keen eye on Cyber-security and the critical nature of network assets.

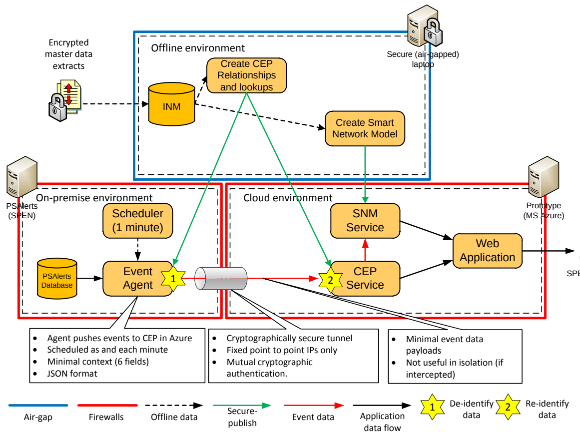

Fig 8: High level technical architecture of DINO PoC

CONCLUSIONS AND NEXT STEPS

This investigation into Master Data Modelling, even at this PoC scale, has provided valuable insights on the path forward for future SPEN Data Management requirements. Final lessons learnt have still to be fully clarified but initial thoughts are:-

- SGAM is useful in helping take repeatable approach in info/data handling – this will be helpful in looking at other data areas
- CIM is not a panacea: it is an excellent template but should not be beholden to its every detail particularly for internal solutions (however this does not diminish the benefits of CIM interoperability).
- An INM is possible and rules to allow matching of data can make it reasonably automated with manually handled exceptions. An INM may be a better way to serve data to multiple processes rather than enforcing ubiquitous CIM compliance over all applications and burdening operational systems with data requests.
- A whole-system-model approach is essential to deriving insight from an increasingly complicated distribution system.
- Metadata management and data mastering are key to forming a whole system model and delivering the future DNO/DSO information management strategy

The final requirement for Dynamic Data Modelling has also received intriguing insight into the possibilities from the application of CEP within this business use case:-

- Demonstrated a significant decrease in events needing human attention using CEP (separate high priority events from background)
- Demonstrated that a quick build and quick change to business cases can be achieved with this approach – i.e. try first then invest

It was also clear from all stakeholder engagement that they understood and supported the need for this dynamic modelling from the predicted increase in data flows emerging from all Smart requirements.

The deliberate thin-slice approach within the PoC was both a positive and a negative. It provided the quick demonstration of the possible but also generated the response ‘if only it could do this’. SPEN plan to build on this feedback to prove that the initial use cases demonstrated the feasibility of the platform but the core platform has applicability to many more areas. Business acceptance of new approaches always takes time and goes beyond the technical. Though you can use CEP to provide valuable insight the ultimate reaction to this by business teams is reliant on other business factors. For example, those with deep expert knowledge of previous failure mechanisms will see less value as are in essence already doing CEP in their head! However, it is the long term repeatability and consistency in approach where benefits will materialise.

There is currently a clamour for the use of data analytics for delivery of increased network intelligence in all aspects of network management. It was interesting that during the project other small scale local data management systems within SPEN were analysed and that these too could be broken down into the concepts within DINO. This analysis helped in the insight that careful consideration is required on the presentation and use of the DINO approach in helping to make final business decisions. It also reinforced the belief that Dynamic Data Modelling and Complex Event Processing is currently and will increasingly prove its worth within SPEN network management requirements.

The next step is therefore to utilise further project based business use case exploration of the DINO concept but moving much more towards an operationally ready system. This will refine the need for better and more consistent network data modelling structures like the interim suggested graph database or on to full CIM compliant data exchange. It is essential that this exploration will drive the much needed ‘single source of truth’ and access to consistent overall network connectivity within the data infrastructure that will be needed for the levels of dynamic data the new Smart world will deliver.

ACKNOWLEDGMENTS

We would like to acknowledge the support and guidance of John Kirkwood who as the Automation team manager delivered the use case for the DINO project but who unfortunately passed away during the course of the project.

REFERENCES

[1] SP Energy Networks, 2014, ‘SP Energy Networks 2015-2023 Business Plan’

[2] SP Energy Networks, 2016, ‘SPEN DSO Vision’