# UIG Task Force Update

Thursday 25th October 2018

Dear Customers and Industry Colleagues,

In early October the Unidentified Gas (UIG) Task Force published an executive summary of their Sprint 2 findings. The team is now pleased to share the most recent findings from Sprint 3.

## Background

Since the implementation of Project Nexus in June 2017, gas shippers have experienced much higher than expected absolute levels and volatility of UIG. This is severely affecting their ability to predict demand and commercially manage their businesses from an immediate cash-flow perspective, because UIG is reconciled (corrected) over an extended and unknown future period. In July 2018 Ofgem approved the UNC Modification 0658 to drive a more centralised and focussed approach to the resolution of UIG, mandating Xoserve as the Central Data Service Provider to take on a leadership role on behalf of the industry. I'm pleased to confirm that the third sprint of the UIG Task Force completed earlier this week.

## Sprint 3 Findings

The priority focus for Sprint 3 was to carry out more detailed and complex modelling. This allowed us to delve further into the current components linked to historic UIG volatility, as well as introduce new components to the model, with a view to pinpoint the problem areas that are driving volatility.

To this end, applying the use of machine learning to the model confirmed that the Non-Daily Metered (NDM) algorithm correctly accounts for wind speed, day of week and holiday factors. However, overall, the use of machine learning showed that it wasn't possible to deliver materially better results based on the current inputs. This has indicated that as a next step we need to focus on the inputs to the model, rather than the mechanics of the model itself.

Having now procured additional weather data items (e.g. precipitation, solar radiation), we are using these in Sprint 4 to test whether their inclusion in NDM estimation processes would make a material difference to UIG.

A significant finding is that within the NDM demand sample dataset there are a small number of sites, which have a measured consumption far greater than the Annual Quantity (AQ) recorded on our systems. This can occur where sites have had a usage change and the shipper needs to increase the AQ to reflect the real consumption. There could also be erroneous historic meter readings or incorrect site set up data, which would need to be resolved. These outliers alone have been shown to have a marked impact on UIG levels and, depending on the End User Category (EUC) profile allocated on our systems, this impact would affect either UIG base, volatility or both.

Furthermore, we have identified a number of large consuming sites within Classes 3 or 4 (Non-Daily Metered), which all have AQs above the Class 1 (mandatory Daily Metered) AQ threshold of 58.6m kWh. In the absence of Xoserve receiving daily meter reads for these

sites, there is a likelihood that if these sites' actual usage patterns are different from the NDM allocation profile (assumed to be fairly flat for EUC Band 9), for example a large hospital that uses more gas in the winter and less in the summer, then this would likely contribute to daily UIG volatility. We are drilling down to understand the root causes behind each site, in order to work with customers to suggest correction actions, as well as drawing out longer-term industry recommendations over the coming weeks.

We can confirm that the review of the geographic clustering of the NDM sample sites shows that we have a good representation across the country.

Sprint 4 will continue to probe the NDM algorithm through the use of additional weather data inputs, and through further investigations into the NDM sample data set to determine how representative it is of the whole market. In relation to out-of-date or missing AQs, we are now simulating historic allocation variances caused by the AQ changes of the NDM EUC Band 9 sites, to see if the change in EUC profile or increased AQ level has an impact on UIG. Using the sample data set, we will simulate the AQ calculations using different read frequencies (monthly, bi-annually and annually) to see what the impact could be to both base and volatility UIG. This may support recommendations in regard to read frequencies and/or EUC bandings to help support the reduction of UIG. Our investigation into the accuracy of daily metered nominations has been carried forward from Sprint 3 and will be concluded in this sprint.

Building on the work of the Allocation of Unidentified Gas Expert (AUGE), which is investigating the impacts of the standard volume-to-energy conversion factor (1.02264, as prescribed in the Thermal Energy Regulations) on UIG, we have modelled the difference between actual and standard outside temperature on daily volume conversion, for a single colder than average Local Distribution Zone (LDZ). This analysis supported the AUGE's findings that standard temperature conversion contributes to positive UIG in winter and negative UIG in summer. The analysis also showed that the annual effect is non-zero, i.e. that winter under-recording of actual energy does not fully offset the summer over-recording of actual energy. On this basis, these investigations will continue into Sprint 4 to model the impact on warm and average LDZ's, and to assess national impact on AQs and therefore UIG.

Please see below for more detail on the full scope of Sprint 3.

We will continue to provide monthly updates at the DSC Change Management Committee. The Investigation Log on our website provides further detail on all Task Force activities. If you have any further questions or comments, please contact us uigtaskforce@xoserve.com.

Kind regards

Ranjit

Ranjit Patel – **Chief Customer Officer**
Ranjit.Patel@xoserve.com | 0771 7854 091

**Address:** Xoserve Limited, Lansdowne Gate, 65 New Road, Solihull, B91 3DL
**Company Website:** http://www.xoserve.com

## Sprint 3 Findings (Ref # relates to the UIG Investigation Log)

| Area | UIG Hypothesis | Findings to date | UIG Impact Peak Volatility % | UIG Impact Annual Average % | Confidence in % | Confidence rationale | Further Action to increase confidence/root cause (None / Sprint 3 / Later Sprint) | Associated UNC Modifications |
|---|---|---|---|---|---|---|---|---|
| Composite Weather Variable (Log Ref #13.2) | All NDM Sites Class 3 and 4 are assigned gas using a standard algorithm, on the basis of their assigned End User Category. Any actual and allocated usage would contribute to UIG. | The use of machine learning showed that it wasn't possible to deliver materially better results based on the current weather inputs. This has indicated that as a next step we need to focus on the inputs to the model, rather than the mechanics of the model itself. | 4% | TBC | MEDIUM | Still unclear whether the linkage to UIG is due to the CWV formula or the make-up of the NDM sample, or a combination of both | Sprint 4 (based on findings from Sprint 3 sample site ref 13.3):<br><br>Re-run machine learning as per Sprint 3 findings with the removal of large EUC bands 6 upwards to validate the potential correlation to UIG. (removing non-weather sensitive consumers)<br><br>Carry out further machine learning with additional weather data to assess if weather is a significant predictor of UIG. | UNC Mod 0659 (use of extra weather data items) |

| Area | UIG Hypothesis | Findings to date | UIG Impact Peak Volatility % | UIG Impact Annual Average % | Confidence in % | Confidence rationale | Further Action to increase confidence/root cause (None / Sprint 3 / Later Sprint) | Associated UNC Modifications |
|---|---|---|---|---|---|---|---|---|
| Accuracy/ out-of-date AQs for Class 3 & 4 sites (Log Ref #3.2) | The difference between the live AQ and a more accurate figure contributes to UIG each day. Base and volatility | Sprint 3 has identified corresponding trends of decreasing NDM AQ and increasing average 'back-cast' UIG levels starting in 2014<br><br>There are a number of sites in the NDM sample data which are consuming significantly different levels of energy than expected based on their sites and meters AQ and EUC profile. The analysis indicates that a small number of sites in the larger EUC's could be significant contributors to UIG. Depending on their profile this could be either base, volatility or both.<br><br>As per Sprint 1 - In Class 4, 32bn kWh of NDM AQ (7%) is overdue for a meter reading (against the UNC obligation of one read every 1 or 12 months, depending on AQ and nature of meter reading equipment). | TBC | TBC | TBC | Correct AQs are not known until meter readings are loaded, which could subsequently increase or decrease UI.G | <u>Sprint 4</u><br>Simulate historic allocation variances caused by the AQ changes of the NDM EUC9 sites, using the NDM sample set to investigate if the change in EUC profile or increased AQ level has an impact on UIG.<br><br>Commence a drill down investigation into the number of actual sites which have erroneous AQ's, considering the scale of AQ we believe is missing and the % UIG this would account for when scaling the sample data to represent the overall population. In addition to this we would also investigate the number of sites which are using more or less than 10% against their original AQ, and the % UIG this would account for when scaling the sample data to represent the overall population.<br><br>Using the sample data set, simulate the AQ calculations using different read frequencies (monthly, 6 monthly and yearly) to see what the impact could be to both base and volatility UIG. This may support recommendations in regard to read frequencies and or EUC bandings to help support the reduction of UIG. Results must be split by EUC and map to the timing of the AQ changes.<br><br>We will be modelling the changes in the contribution to total AQ of the different EUC groupings over time and the impact this has had on allocation.<br><br>In line with the Sprint 3 findings of corresponding trends of decreasing NDM AQ and increasing average 'back-cast' UIG levels, as part of a future sprint, we will be analysing legacy NDM AQ amendment trends compared with Rolling AQ performance to see if there is a relationship with UIG trends.<br><br><u>Customer Activity</u><br>Engage with individual customers to highlight the need to improve on the volume of actual reads for sites where the AQ is out of date. | UNC Mod 0672 (Incentivise Class 4 read performance) |

| Area | UIG Hypothesis | Findings to date | UIG Impact Peak Volatility % | UIG Impact Annual Average % | Confidence in % | Confidence rationale | Further Action to increase confidence/root cause (None / Sprint 3 / Later Sprint) | Associated UNC Modifications |
|---|---|---|---|---|---|---|---|---|
| DM Nomination Accuracy (Log Ref #9) | Inaccurate DM Nominations cause UIG in Nominations | Limited data set analysed thus far, (part of gas year 2017/2018).<br><br>Initial analysis suggests a general trend of nominations improvement throughout the day, the final run of nominations appears reasonably accurate (likely due to the scheduling charges). | TBC | TBC | TBC | | <u>Sprint 4</u><br>Increase data set to cover 2016/17 and 2017/18 full gas years, the aim to make visible a component of within day UIG volatility at nomination stage with a view of sharing our findings with the industry who may wish to make recommendations to adjust the current scheduling charges. | |
| Standard conversion factor (Log Ref 12.2) | Using a standard figure throughout the year could understate winter usage and overstate summer usage. | <u>Sprint 3 findings</u><br>AUGE (Allocation of Unidentified Gas Expert) also investigating this topic and findings so far are that temperature is the most sensitive element of the standard conversion factors. Our internal analysis of the impact of using actual temperatures instead of the standard 12.2 degrees in a colder than average LDZ indicates that the annual effect is non-zero, i.e. that summer over-recording of actual energy does not fully offset the winter under-recording of actual energy.<br><br><u>Sprint 2 findings</u><br>Results of previous 3rd party investigation mapped against %UIG for first 12months post Nexus – evidence is scattered, no correlation found. | TBC | TBC | TBC | | Carried forward to Sprint 4, to model warm and average LDZs against actual temperatures, and assess national impact on AQs and therefore UIG. | |

| Area | UIG Hypothesis | Findings to date | UIG Impact Peak Volatility % | UIG Impact Annual Average % | Confidence in % | Confidence rationale | Further Action to increase confidence/root cause (None / Sprint 3 / Later Sprint) | Associated UNC Modifications |
|---|---|---|---|---|---|---|---|---|
| Suitability of NDM Sample site data (Log Ref 13.3) | Any difference between actual NDM usage and allocation based on the sample is a cause of UIG | Sprint 3 clustering analysis demonstrates that we have a fair representation of sample sites across the country.<br><br>A possible relationship has been identified between sites in lower EUC bands and the potential correlation to UIG purely based on the data in the sample set of sites (investigation activities for this finding can be found under ref 13.3).<br><br>Outliers identified within the NDM sample set may have a significant impact on levels of UIG (investigation activities will continue under Ref 3.2). | TBC | TBC | TBC | | Sprint 4<br>Investigate how representative the NDM sample data set is when looking at EUC's. We aim to plot the distribution of sample users by EUC vs. the full data set (nationally and by LDZ). This is the enabler to identify specific EUC findings, or LDZ specific behaviours and will help indicate if the sample data is representative or not. This will also support the need to consider if the sample set can be extrapolated up in size to map to UIG levels nationally. We can potentially look for patterns on the EUC's across the country (this would be findings under 13.1) | UNC Mod 0654 (mandating the provision of NDM sample data) |

## TASK FORCE SPRINT 4 SCOPE AND RATIONALE

In addition to the next steps outlined above, sprint 4 of the UIG Task Force will investigate the following areas detailed in the table below:

| | | |
|---|---|---|
| *Investigation Log 13.1* | **Accuracy of NDM Algorithm including EUC definitions** | Utilising NDM sample data, investigate a 12month data set to help identify if demand estimation is influenced by geographic factors, such as proximity to weather stations, rural vs. urban to validate existing modelling to determine if improvements are required. This will be done by using the actual usage over the period v's the calculated usage from the UIG formula. |