WESTERN POWER DISTRIBUTION

INNOVATION

NEXT GENERATION NETWORKS

TIME SERIES DATA TOOL FEASIBILITY

CLOSE DOWN REPORT

LCN Fund
Low Carbon Networks

Report Title	:	Time Series Data Tool Feasibility Close Down Report
Report Status	:	Final
Project Ref	:	WPD_NIA_024
Date	:	30/11/2017

Document Control

	Name	Date
Prepared by:	James Bennett, Paul Charleton	31/08/17
Reviewed by:	Roger Hey	
Approved (WPD):	Roger Hey	

Revision History

Date	Issue	Status
31/08/2017	v0.01	Draft
	v.0.02	Review
30/11/2017	v1.00	Final

Contents

Glossary

Abbreviation	Term
CSV	Comma Separated Values. A file format where data elements are separated by commas on each row and structured in multiple rows for each new record. CSV data is easily imported into inspection and analysis programs such as MS EXCEL.
CE	Calculation Engine (PowerON element)
CT	Current Transformer
DNO	Distribution Network Operator
DNV	UKPN Project, Distribution Network Visibility
ENW	Electricity North West
HH	Half Hour (or Half Hourly) data points, being averages over this interval
HISTAN	Analogue History Database in PowerON
INM	Integrated Network Model, a next generation database tool derived from existing diverse sources of data
PowerON	Power on (Fusion), the WPD NMS system. Also referred to as ENMAC for historic reasons and also sometimes as PoF.
PSD	Primary System Design (a WPD engineering analysis dept)
RTU	Remote Terminal Unit
SCADA	Supervisory Control And Data Acquisition. Systems for remote monitoring and control operations via communications channel via which analogue values are returned from the remote sites to the centre (PowerON).
UKPN	UK Power Networks
VBA	Visual Basic for Applications, Embedded Excel programming language
VT	Voltage Transformer
WPD	Western Power Distribution

Reference Documents

1. Time Series Data Quality NIA Project 011, Close Down Report Version 1.0 31/08/2016
2. User Requirements Survey Report
3. User Requirements Specification
4. UKPN Distribution Network Visibility (DNV, Close Down Report), December 2013.

The following Reference Documents relate to the initial Data Analysis Work item

5. IBM Initial Data Analytics Engagement Final Report. Final report v1.0.pptx. Jonathan Batson, Kelly Freeman, Erwin Frank-Schultz. March 2016 Version
6. Company Directive STANDARD TECHNIQUE: TP6F, Power Measurement Conventions, February 2016
7. BSEN 60044-1:1999 Instrument transformers – Part 1: Current transformers
8. Company Directive STANDARD TECHNIQUE: OC12B/3, Numbering and Labelling Switchgear and other Apparatus within Substations, November 2004
9. Company Directive STANDARD TECHNIQUE: GE23/1, Relating to Data Assurance of Regulatory Submissions, October 2015
10. Analogue Data Quality Analysis Report – Cross Reference Report, July 2016
11. Analogue Data Quality Analysis Report – Heat Map Report, July 2016
12. Analogue Data Quality Analysis Report – Seasonal Analysis Report, July 2016

1 Executive Summary

It is a fundamental requirement for the DNO transition towards DSO that there is improved accuracy and visibility of the network representation available to the IT systems. This implies a much more holistic end-to-end data management capability which can deliver a greater understanding for engineering users. It is also highly likely that there will be benefit in DSOs (and even other organisations) using common systems or ones with standard interfaces.

The project followed on from a 2016 NIA project, which analysed time series data quality. The 2016 project covered a number of elements, developed in pilot form, for handling analogue data. This included tools to identify and rectify poor quality data, bulk data plotting, curve fitting and electrical analysis. We were looking to build on this initial work by integrating these tools (and more) into a user friendly data analysis utility, and this further project was initiated to assess the requirement for and feasibility of a new dedicated system for in-depth time series data analysis.

The project was to look specifically at the development of an integrated time series database and associated analysis system to be used for mainly analogue data gathered by the WPD systems. The two projects would therefore have prepared the ground in this important area. At the outset, the new solution envisaged the provision of data visualisation capabilities, analysis tools and reporting capabilities with the further possibility of allowing for permanent data rectification for known errors as well as corrected values to be available on demand.

The project team worked with a variety of user roles to gather and analyse requirements, formulate a system requirements specification and then moved on to assess supplier offerings. The project initially had three distinct work packages to address the project objectives:

- User engagement and production of a requirements specification for a new time series database based on research, including learnings from Time Series Data Quality project together with internal business input and wider DNO input;
- Review of the current solutions and technologies available and how they compare against our identified anticipated requirements; and
- Tender process for a time series database and integrated analysis tool. Costs from this to be used to inform a subsequent time series database development project.

1.1 Previous and Related Projects

All DNOs conduct the same or very similar data gathering and management activities and will have similar ongoing and future data related requirements. WPD have conducted analogue data gathering and data management investigations and projects including:

- Time Series Data Quality Project (February 2016 to July 2016) – the immediate precursor project which has investigated analogue data quality in WPD and lead to and understanding of, and initiated a rectification programme for, identified deficiencies;
- Project FALCON 2011 – 2015. Created a normalised, centralised database (Authoritative Network Model or ANM) drawing on major feeder data repositories such as PowerON, EMU and Crown and now being used as the basis for a new

Integrated Network Model (INM) project. Also looked at IP communications systems based on WiMAX radio as the physical carrier; and

- As mentioned above, convergence on the Common Information Model (CIM) standard for data is highly desirable and another WPD project (INM/CIM) was starting up at the same time as this project. INM is a Large pilot implementation covering the South West region of WPD initially as a proof of concept, itself building on the FALCON data abstraction layer (named the Authorised Network Model) which had been derived to support the FALCON network evolution modelling tool (Called the SIM). It quickly became clear that there was some considerable overlap in objectives and work elements between INM and the Time Series Data Management system projects and this view was reinforced by UKPN review of our findings.

2 Statement of Problem

It is apparent that analogue data management facilities have been a neglected area in the DNOs, perhaps not advancing quickly enough as data management tools have emerged and become commercially available. In WPD, Analogue data is currently acquired by the three separate PowerON (ex ENMAC) control systems covering the four operating regions of the WPD licence area[1]. PowerON operation also includes derivation of half-hourly average data derived from the source analogues mainly used for off-line analysis. A separate system called Datalogger takes an overnight feed of HH analogue data and makes this available to offline users such as engineers, along with metering data. Datalogger also provides a number of "derived parameters" (non-source data "derived" from other inputs). Engineers either access Datalogger to export data for analysis, or else liaise with the PowerON support team to obtain direct data exports from that system. This data may then be imported into systems such as MATLAB, Excel or bespoke analysis systems for further manipulation.

The above process is quite manual, disjointed, subject to lead time delays (when special requests for data have to be made) and are error prone (in particular there are issues between PowerON and Datalogger cross-references). In addition, unless the Engineer has access to such additional third party tools as MATLAB, along with the necessary skills needed to use them, the data management/visualisation and analysis facilities available can be very primitive.

A single, integrated, functional, performant and intuitive tool is therefore seen as offering a major step forward for WPD Planners and Engineers dealing with analogue data as part of their responsibilities.

This research project will seek to answer three questions;

1. What time series data requirements does the DNO have for a DSO future?

---

[1] East and West Midlands are currently a combined PowerON region, this may change however to split out the two regions.

This question includes raw data, visualisation and processing requirements together with functionality questions such as the requirement for smartphone apps or centralised vs. decentralised data repositories.

2. What time series data analysis tools and visualisation technologies are currently available and how do they compare to our current and future requirements?

3. What would be the cost to develop a tool satisfying these identified requirements?

3 Method

As described above, the project was to form three distinct work packages in order meet its objectives as set out in this document;

- Work Package 1 - Production of a requirements specification for a new time series database based on research including learnings from Time Series Data Quality project together with internal business input and wider DNO input;
- Work Package 2 – Review of the current solutions and technologies available and how they compare against our identified anticipated requirements; and
- Work Package 3 – Tender process for a time series database and integrated analysis tool. Costs from this will then inform a subsequent time series database development project.

This project was essentially a feasibility study with the aim to identify future industry requirements and the cost of developing existing offerings to meet these requirements. From this it will be determined whether a prototype demonstration project is appropriate and how much this is likely to cost. Any implementation would take place as a new Project.

3.1 Approach

The project followed an established IT project lifecycle which drives the overall timescale in a process limited schedule of activities. In the initial phase, user engagement and previous pilot project conclusions input was followed by requirements definition against which a next phase commercial tools review might then be conducted using a solutions assessment compliance matrix. From this a tender process might then follow for the top scoring systems offerings. These various activities have a natural duration based around the need for engaging with participants (including obtaining feedback from other DNOs).

4 Scope & Objectives

The project was principally concerned with value add in respect of analogue data management and rationalisation of business processes associated with offline data analysis. The project looked at how best to derive information from such analogue data and how to correct for errors, visualise and manage analogue data and export bespoke reports for export to other downstream utilities using *best-of-breed* data management and analysis systems to later develop a solution for improved collection, analysis and rectification of existing SCADA data used for network planning and control purposes. (This data includes half hourly recorded Volts, Amps, MW, MVAR, MVA. The incorporation of metering data and connectivity was to be considered for analysis purposes).

The conclusions of the project were intended to show how to deal with such data in the future, the limits of what can be done using current technologies and the likely cost of these systems.

The project was also expecting to look at moving some of the island data repositories into centralised and more functional and readily accessed database facilities better suited to meeting the needs of engineers and corporate communications personnel working in the new DSO environment. It would be desirable for users not to have to carry out disconnected operations such as one-off data set creation and creation of bespoke analysis tools likely to be used once then thrown away. On the way to achieving this, careful consideration would be given to the users data management and analysis requirements and how to align this with potential solutions. This managed approach attempts to formalise what has hitherto been a rather haphazard arrangement from the Users point of view. Getting the most out of the available data and maximising the return on investment made in analogue measurement systems is clearly highly desirable at a time when deriving ***information*** and hence value from input ***data*** has been identified as a key route to successful operations.

As the project progressed the project found itself duplicating elements of the INM/CIM project which was starting at around the same time after a previous delay. Given the scale and objectives of that project, this cross over was considered to have significant impact (see below).

5 Success Criteria

The Project includes the following main identifiable steps which can be assessed for completeness and an indication of successful conclusion:

- Requirements gathering, user interaction allowing processing to next step;
- Requirements specification – production of a complete specification against which the existing market can be compared and potential suppliers may quote;
- Tender process resulting in identifiable solutions and costs;
- Selection of a preferred solution; and
- Follow on instructions/details for following implementation project.

The progression to the next stage is predicated on the success of the previous steps in each case, so the ability to follow on in this way is a key measure of ongoing success. A final identified solution outcome is the ultimate measure of Project completeness and success.

6 Details of the Work Carried Out

The project was conducted as planned up to and including the production and review of the Requirements Specification. However considering the progress with the related INM project, the decision was taken to postpone some of the work at this point and await the delivery of the INM database and CIM capability expected around June 2018. In consequence the second work package was only partially completed whilst the third was not progressed at this stage as it was determined that given expected input in 2018 from INM, such a process conducted in early 2017 would be premature.

6.1 User Requirements Gathering

As part of the project, a 17 point questionnaire was created and distributed to some 64 users falling into the main target user group with the returns being assimilated and analysed in a User Requirements Response Review Document. A starting framework of system requirements, along with the specific requirements obtained from the user survey were merged into a System Requirements Specification document. Careful attention was paid to the actual tasks that the users are carrying out, rather than just assembling a series of data processing requirements, as it is important that the system is targeted at supporting the users in the execution of their work.

The user group was subdivided by role with these coming from six main groups:

- Planner
- Future Networks Team Engineer
- Primary System Design Engineer
- Operations Support
- Data Analyst
- Projects Staff

Not all of those who were sent a questionnaire responded, the return rate being 45.3%. An illustration of the breakdown of returns by user type is given below.

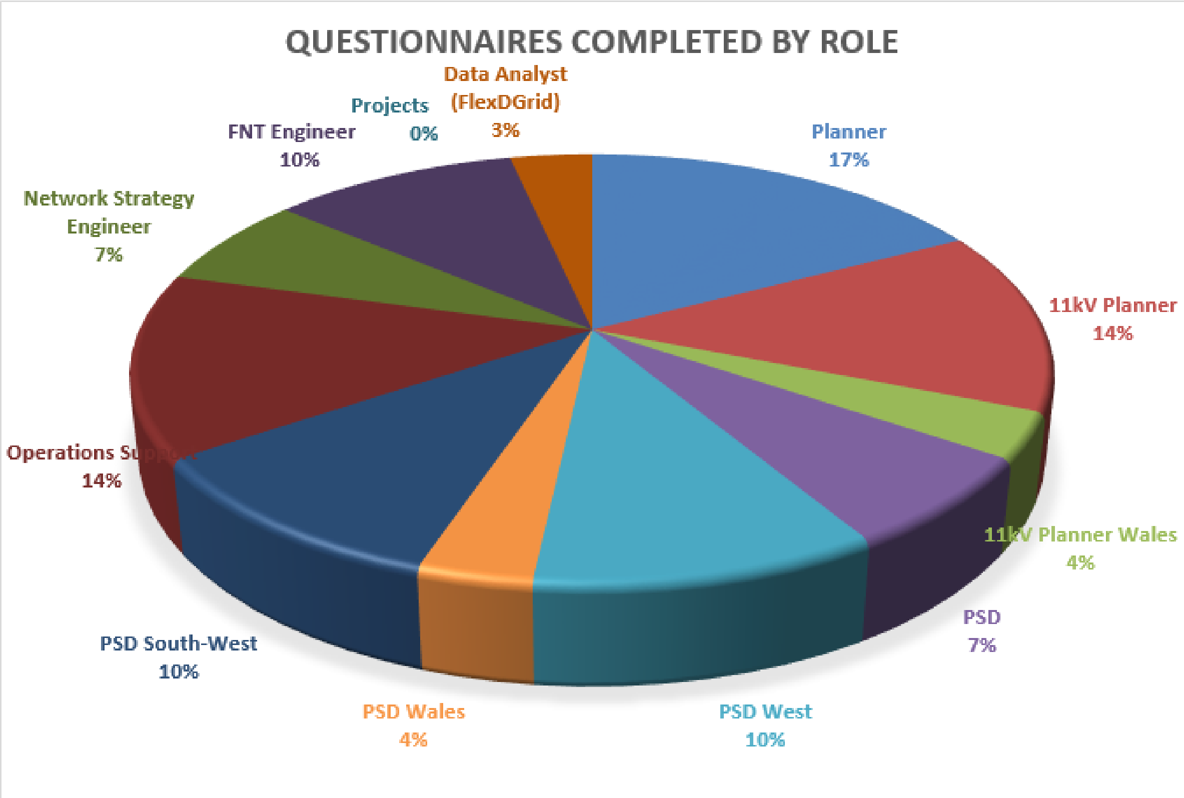


Figure 1 - Questionnaires Completed by User Role

The results of the Requirements gathering exercise are available in [Ref 2].

6.2 User Requirements Analysis

This task followed on immediately from the User Requirements Gathering activity and involved collating the results and conducting a simple analysis of the responses. For each of the main questions a simple percentage breakdown was derived allowing display in a pie chart format for rapid assimilation.

Based on the analysis of the user supplied inputs, the following were the main findings.

1. Over half of respondents use the current WPD data inspection mechanism *Datalogger* on a Daily/Weekly or Regular basis. For the rest usage is at a Monthly or Occasional level. For a replacement system the figures would be broadly the same (as expected);
2. The users are dissatisfied with the current toolset (60% of respondents stated that the system does not meet their current needs, 80% experience issues with the data while 80% do not expect that it will meet their future needs) as it is unreliable and data is incomplete. However they also appreciate that some of the main issues are with the source data coming in from faulty or incorrectly set-up analogues in the field. One respondent signalled significant issues with Excel crashing on him to the point where he relied on bitmaps to create his reports;

3. A major functional gap in the present data/data tools system concerns the direction of power flow given the high levels of generation now present on the network. This is a recurring theme with the users;
4. Users regularly have to supplement their Datalogger analysis with data from other systems either due to missing or inaccurate data. This requires further work for them and consequent loss of efficiency. Over 80% have had to conduct additional analysis using PowerON data due to Datalogger issues, but they are also creative in seeking such alternative paths and use a variety of supplementary inputs including Durabill, Rustrac monitor, PQBox 200, inference from previous analysis of other generators etc;
5. Data is used for a number of purposes including:
    a. Feeding into reporting processes (such as the annual load survey and CM1 Report) and including regulatory processes;
    b. Conducting more in depth analysis such as demand forecasting, load sanitisation process and assessing utilisation of a substation (i.e. demand vs capacity), with 70% of respondents stating that they use spreadsheets to conduct more analysis and 20% doing network modelling using tools such as IPSA and DINIS.
    c. For generation requests to look at profiles when deciding options.
6. The users readily identified a number of missing features for the current system (see next point for their thoughts on what to include in a new system). These inputs are further expanded in Section 6 below;
7. The users readily identified a number of additional features that they would like to see in a new Data Management system, including ready computation of minimum and maximum values, noise removal, event analysis, seasonal and trend analysis, directional information (corrected analogues). Three quarters of the respondents envisage the need for such additional processing capabilities; and
8. Most users remained unaware of how to get support for the existing data facilities, relying instead on personal links to people like Steve Quinn. In fact around 85% of respondents have never sought assistance with the application. This must be making many of them inefficient in the execution of their work when issues with the current toolset are encountered.

The user engagement also helped to define the various onward usage for the data as the different roles went about their work as the responses gathered from the user survey indicated a number of processes, tools and workflows with which the users are engaged when they are working mainly with the historic analogue data. These use cases provide a deeper understanding of how a new system would need to fit into the current Engineer's toolset.

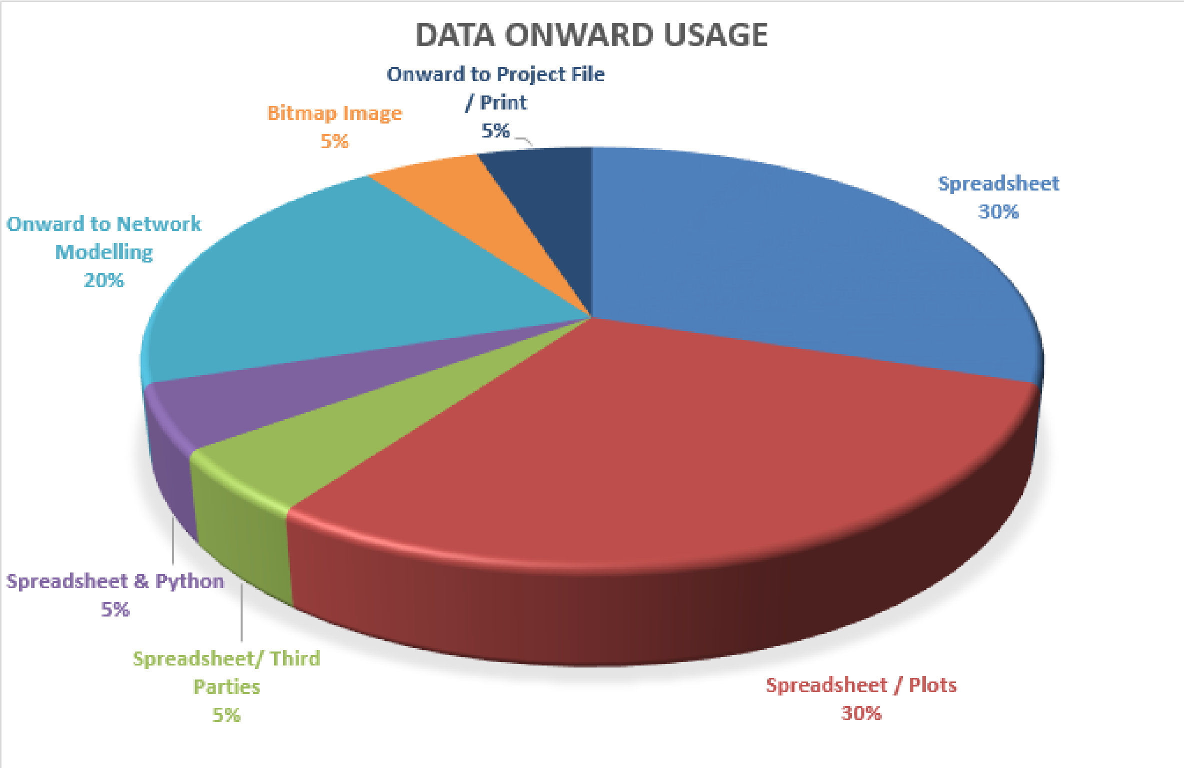


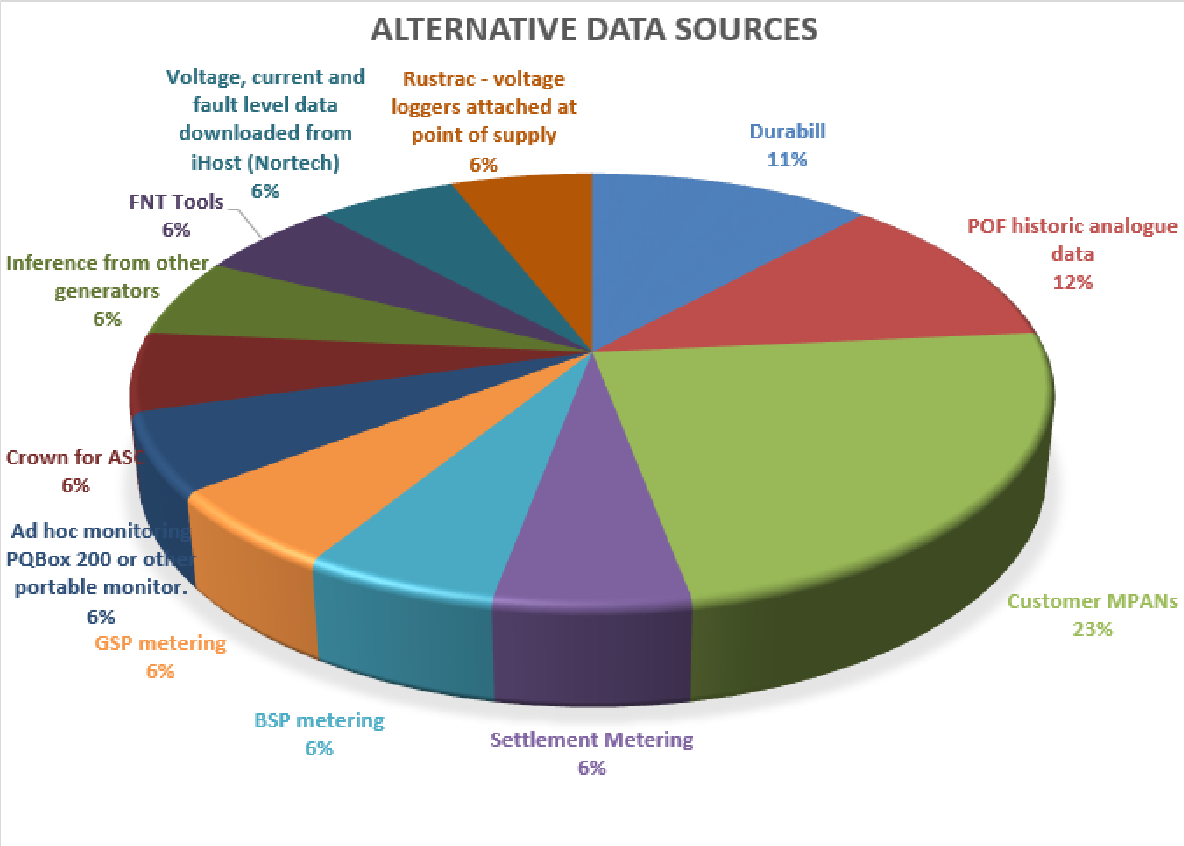
Figure 2 - What Users Do Next with Data

6.2.1 Input Data Sources

Users were asked about which data sources provided inputs to their standard workflows. The following breakdown shows the returns. The information derived from this shows that the main data comes from three main sources:

- Control system;
- Metering inputs;
- Dedicated tools/units and ad-hoc monitoring.

Secondary inputs are also needed to make sense of or otherwise support analysis of the data coming from the asset system (Crown) and bespoked analysis systems.



6.2.2 Substation and Customer Analysis

Users carry out a number of processes in support of their regular work tasks/activities. Tasks are mostly very ad-hoc, with engineers being 'reactive' to situations much of the time as customers chase the progress of their schemes, but work also includes regular reporting activities.

The following activities (and quotes relating to these) were noted:

- Regular load review – based on historic view (plots of gathered HH data). This may be as part of the annual load survey, carrying out analysis to determine minimum/maximum loads (over extended periods) or analysis of abnormal network events. Load studies are also conducted for outage requests and generation connection requests;
    - "Mostly I use historic analogue data to find system loadings and generator outputs to allow more accurate system modelling".
    - "We use historic analogue data 3 or 4 times a week for modelling feeder loads and customer consumption".
    - "My main use for historic analogue data is new business schemes and my work on them is mostly reactive, a regular ad-hoc use".
- Conduct review on a daily basis for metering data and substation current analogue data at 11KV. Further analyse the data, sometimes accumulating data from various data sets to find trends, peaks or troughs, diversified values (such as times when a group of generation was at its maximum rather than individually);
- Perform assessment of suitability/capability of a substation to support generation requests. All of the generation sites are managed in logger form in order to aggregate them for analysis, and monitoring generation;

- For the production of the load survey and CM1 report, both being regular data intensive activities within the business;
- The Engineers reported a need to understand the customers import / export requirements in more detail;
- There is an increasing requirement to know what power flows we are seeing in the network to determine how necessary reinforcement will be and if and when overloads could be anticipated; and
- A regular activity is in dealing with capital reinforcement projects, where loading is an issue or network reconfiguration is required.

6.2.3 Strategic Studies and Innovation Projects

For various strategic studies users detect minimum and maximum values - for DG, load and site demand and perform seasonal averaging and finding of extremes.

For innovation studies the data need can be particularly diverse and some quite elaborate processing is conducted by engineers, consultants and third parties using supplied data.

6.2.4 Load and Network Modelling

Feeder logger data and HH data for individual HV metered customers is used to feed power system analysis software to do load studies using these third party power flow / modelling tools.

6.2.5 Load Sanitisation Process

This is a process of correction of observed load by removal of the loads arising from abnormal events which skew observed figures on the occasions when the abnormal event is present.

In the WPD Primary System Design (PSD) Data team, users perform such a load sanitisation process, forecasting, and assessing utilisation of substations (i.e. demand vs capacity) and the PSD planning function requires clarity of demand in each individual substation, during normal and abnormal running.

6.2.6 Annual Load Survey and other Reports

This is an annual summary of load profiles at the detailed level, by site, which feeds into regulatory reporting. Users also reported doing “annual reports” and regular “ad-hoc studies and analysis”.

6.2.7 Regulatory Reporting

Regulatory Reporting is carried out for individual substations. To do this the user extracts a full year of data in order to forecast demand.

Note that any changes to the outputs would need to be strictly assessed as it may have significant impact on OfGEM reporting / RIIO ED1 investment impact. Changes to the methodology/process would therefore require reporting to OfGEM as it may change the RIIO ED1 mechanisms.

6.2.8 Data Processing Tools and Methods

The users were asked about and specifics forms of special processing that they needed to carry out on data. The spread of returns confirmed the view that the following are either mandatory or of interest to the users.

- Noise removal;
- curve fitting and long term trend analysis;
- detection of maximum/minimum values;
- analysis of form of response;
- seasonal or other averaging;
- Determination of capacity factor (margin);
- Visualisation.

6.2.9 Data Processing Tools – Expected Frequency of Use

The users were asked about their expected frequency of use of a data management tool.

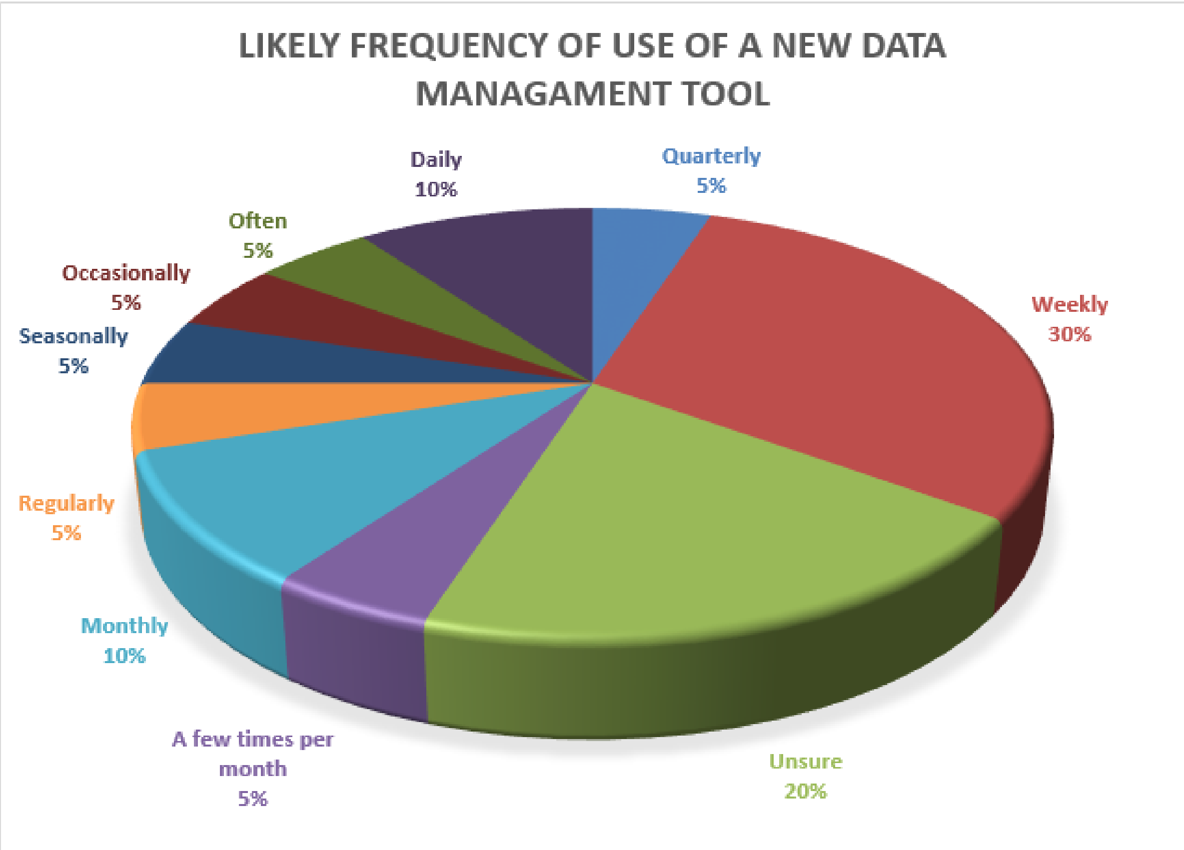


Figure 3 - Data Management Tools, Expected Usage

6.3 System Requirements Specification

The work on analysis of user requirements resulted in the production of a System Requirements Specification [Ref 3]. The suitability of all candidate systems may be assessed against this specification and judged according to how well such systems meet the requirements.

Requirements were classified under the following headings to provide structure:

- General Requirements;
- Data Input Requirements;
- Data Management Requirements;
- Data Analysis Requirements;
- Data Output Requirements;
- Linkage to Other Tools;
- User Interface Requirements.

6.4 DNO Requirements Review

The DNOs were invited to comment on the initial conclusions from the requirements survey and the Requirements Specification derived from it. Responses were received from UKPN and ENW. Both were broadly in agreement with the project conclusions and saw parallels with some of their own system developments. They also had advice for future systems developments.

The UKPN review in particular reinforced our emerging view that the ongoing parallel development of the WPD INM/CIM project was effectively providing the same level of data integration that UKPN had concluded to be most valuable for improving network visibility.

6.4.1 UKPN Review

From UKPN input it was clear that there was some crossover between the project and their own DNV (Distribution Network Visibility) initiative (Tier 1 Project, see Ref 4). UKPN observe that while most DNOs have limited monitoring at lower voltage levels, the UKPN London network is different having extensive and widespread monitoring capability, with Remote Terminal Units (RTUs) deployed in approximately 60% of the distribution substations. Network visualisation is therefore a very keen issue for UKPN and a major objective of the scope of the DNV project was to "Identify and integrate other data sources: Identifying other data sources that could be used in conjunction with network monitoring data to ensure the visualisation tools developed and trialled provide valuable information to users." This corresponds almost exactly to the main objectives of our project and drove a very similar project profile having the following major elements:

- Definition of business benefits and user requirements;
- Assessment of data requirements and data collection;
- Implementation of visualisation and power flow tools.

The UKPN review of our findings suggested that most requirements that they had also envisaged "were present in our analysis", while at the same time cautioning that it is important to maintain dialogue with the business while developing the platform to make sure it is fit for purpose, and designed in a way that meets the needs of the Engineers and Planners. More detail on the Visualisation Requirements was indicated as needed. The view was that the document mentioned a few points around visualisation but that it is important to decide early what type of visualisation suites would be most useful for users, with different preferences from different teams to be expected. This had been a major element of the DNV project.

The review suggested that frequency of updating certain data sources was missing (this will be based on user needs). RTU analogues get automatically updated but for the UKPN DNV, for example, some of the data sources require manual updating. UKPN had to get business commitment on the frequency of updating these so that the tool can be kept current and reliable. The frequency of updating each source was agreed with business users at UKPN.

As for WPD, the UKPN users work mainly with half hourly data. Feedback confirmed that while this is sufficient at the moment, this should be made futureproof. A flexible system can therefore be foreseen, one that would be able to adapt quickly to a more challenging future.

UKPN experience has also shown that data quality is a big issue to address. DNV has reports checking the health of RTUs, which not only alerts re: any non-communicating RTUs, but also checks if any RTUs give false readings. It was suggested by the review that it is not worth gathering all these data if they can't be trusted. In summary UKPN agreed that having a function in the platform to check for data validity is very important. Even better if some of the algorithms behind the analytical tools do the checks while calculating different factors and return the measured value along with a calculated value to compare (DNV does this for some of the information provided).

Data rationalisation was also agreed to be important, consistency between various sub systems, particularly around object naming, can be expected to increase the complexity of developing future data management systems.

As to specific tools, UKPNs report (Ref 4) stated that there is work still to be done to ensure the available load flow tools deliver the requirements of DNOs at a cost that is affordable. Other useful functionalities have also been identified for development such as predicting future load flows based on historic data for similar time of day, network conditions, running arrangements.

6.4.2 ENW Review

The ENW comments were provided from the perspective of the learning from their ATLAS NIA project. This project takes a particular interest in time-series data quality/processing to system normal as it relates to deriving half hourly P and Q true/latent/measured demand at G&P substations, in order to deliver new prototype tools/methodologies for long-term load scenarios.

The view was that our project scope was for a much wider system – extending to included feeder analogue data, interaction with switching logs for identifying abnormal running, and trying to provide a system to multiple users for 4 network areas.

At the top level ENW questioned whether one single system would be able to meet all these user requirements/solve all the actual and potential problems. This is a consideration for how we should translate from user requirements to implementation. This may be related to the capabilities of existing systems, and where the data is held. Even in the smaller scope of ENW, their solutions were split into several components. As a result ENW have MPAN metering data in one existing billing system (DADS), SCADA data in other existing system (FLA), and they then combine these analogues/metering data in an SQL MI Cube (a so-called Dashboard) to reveal HH P and Q true/latent/measured demand at G&P substations. In ATLAS, ENW have developed data-processing prototype tools/methods for data-processing to system-normal for all the spikes/zeroes/network switching. To achieve this they have used Matlab as a holding solution.   Some engineers have also created bespoke Python code to provide a per substation weather correction of HH true demand from historical data, thus providing a baseline for load scenarios.

ENW noted that there are both positives and negatives to having worked in these multiple steps, though the work is still at a prototype stage. Because ATLAS is focused on valid P and Q outputs at G&P substations, ENW may go further in that area than the scope of our WPD proposals.

They continue to work with averages over a half hour, but recognise this can hide some of the behaviour during a half hour e.g. at low values of P and Q there can be both positive and negative values within a half hour.

ENW agreed that autonomous error checking/validation application would be useful, but noted that there are lots of types of errors, so there would be a need to have different types of applications tailored different types of data and errors – with outputs needing to be reviewed by a human expert to resolve them. It is thus necessary to consider both the checking/validation approach, and how it is used. In ATLAS ENW have developed automatic data-processing of substation P and Q, but there is still an important element of manual review and sense-check.

It has been identified that consistency between various sub systems, particularly around object naming, will increase the complexity of developing such a data management system. Data rationalisation is clearly a feature of the immediate data management task. ENW agreed that consistency of naming would enable reduction in complexity of how to use and combine data – but addressing inconsistencies in the current systems potentially adds significant short-term investment of resources to see these fixed. In ATLAS, ENW chose to devote time to establishing naming consistency in some systems/databases, but not others.

6.5 Systems Survey and Procurement Aspects

A trawl of supplier websites was carried out using the corporate suppliers list for initial input for the list of Technology providers (IT).   An evaluation matrix was set up to match solutions/suppliers on to a set of general evaluation criteria.

	Functionality	Efficiency	Alignment of overall solution with the WPD operational paradigm	Availability of appropriate analytics	Competing visions in the organisation	Support of stakeholders	Price CAPEX	Price OPEX	Lead times for solution	Nature of host	Mechanisms for User access	Management of platform and application	Application development	Connection and distribution	In-house capability/skills
Hadoop/ HDFS /Spark															
KairosDB															
OpenTSDB															
Prometheus															
Gridkey															
Influxdata															
IBM Informix															
Affinity Meterflow															
SQLServer Streaminsight															
Rolta															
SAP Hana															
OSISoft PI															
Utilidex															
Informatica															
Nortech Ihost															
MapR															
Oracle Analytics															
Apache Cassandra															
Bespoke System															

A similar evaluation matrix can also be used on a per requirement basis to assess the technical aspects of each offering.

7 The Outcomes of the Project

The project successfully delivered the User Feedback Report and Requirements Specification and the latter was subject to DNO peer review. These two documents do not change as a result of the solutions chosen for adoption, are available for review and will be used as inputs in the next phase of data tool assessment work. We have concluded however that further investigation of such data management and visualisation tools is best conducted after the related WPD Project called INM/CIM concludes its current work packages in around June 2018. Some preparatory investigations will need to be carried out during later INM phases to prepare the ground for this.

8 Performance against Project Aims, Objectives and Success Criteria

The project met its first phase objectives of user engagement, requirements definition and DNO requirements review as noted above, with a full solution to be assessed once the INM project has also delivered. A number of data analysis systems were identified and preliminary qualifying investigation carried out, but we concluded that the activity was premature until the data abstraction layer facilitated by the INM/CIM project had delivered its own findings and solution.

9 Required modifications to the planned approach during the course of the project

The project proceeded as expected up to the point of carrying out full solutions investigations and competitive tendering process. The project was therefore truncated in respect of scope pending input from the INM/CIM project.

10 Significant variance in expected costs and benefits

The project remained within cost throughout.

Regarding benefits – these were in line with initial expectation although as noted the project was truncated in this phase due to pending linkage to another ongoing related project.

11 Lessons learnt for future projects

There is significant engineering user interest in the availability of improved tools for data analysis. However it is important not to rush into solutions, rather a coordinated approach to data correction and rationalisation should be initiated to provide a sound basis to build upon. A data abstraction layer (such as that implemented by the INM/CIM project) provides such a basis for the positioning of a data management solution, while initiatives designed to improve analogue data reliability are necessary to ensure that data is usable. Such initiatives are already underway in WPD and are already resulting in improved monitoring data collection.

There are a number of proprietary data processing solutions available to fit the DNO/DSO data analysis space.

12 Facilitate Replication

12.1 Knowledge Required

The main knowledge areas for the project are:

- Familiarity with the existing analogue management systems and field deployed equipment;
- Data analysis techniques;
- IT aspects of data processing (including familiarity with the tools developed by the Project; and
- Provision of engineering insight to guide development of tools and interpret results (user input with users of various types).

12.2 Products/Services Required

The key elements to allow for successful replication of this project were:

1. Stakeholder support;
2. User Support (Various engineering users);
3. Procurement support;
4. DNO Review availability.