21, rue d'Artois, F-75008 PARIS
http://www.cigre.org

CIGRE US National Committee
2019 Grid of the Future Symposium

# Challenges of Data Management & Analytics in the Future Grid

**T. MCGRAIL, B. GWYN**
**Doble Engineering**
**USA**

**J. BEARDSALL, M. ROWBOTTOM**
**Drax Power**
**UK**

## SUMMARY

With a growth in interest in 'big data' as electric grids evolve and data sources become more common and more productive, there needs to be a discussion of the management of data in a secure manner, and the role of analytics to provide information and have 'meaning'. This paper looks at a number of challenges that are beginning to be faced, and opportunities to ensure that the Future Grid is secure. Challenge 1 is the management of 'big data', which may provide value if appropriately viewed and analyzed; Challenge 2 is the management of security, for both data and systems which use the data; Challenge 3 is the need for appropriate urgency in analysis and action; Challenge 4 is to understand the meaning of the data and associated analyses, but also to understand the limits of our understanding.

## KEYWORDS

Big data, Analytics, Sense-Making

tmcgrail@doble.com

## CHALLENGE 1: DATA MANAGEMENT

There has been over the last few years a growth in interest in the idea of 'big data' [1]. Occasionally the characteristics of 'big data' have been summarized in terms of the 'V's of 'big data, some of which are collated in Table 1.

| Characteristic | Implication |
|---|---|
| Velocity | High rates of data generation, collation and analysis |
| Volume | 'incredible amounts' of data to be managed |
| Variety | Structured v unstructured: think photo content |
| Value | The value may be virtual: unrealizable |
| Veracity | Trustworthiness - verisimilitude |

**Table 1 Some of the various V's of 'big data'**

We collect data to support decisions – through analytics such as level checking, presence/absence and other, often more exotic approaches [2][3]. With increasing pressure on assets to perform continuously, to reduce outages for maintenance or through unplanned issues, and within the context of Industry 4.0 bringing condition/operation sensor ubiquity and autonomy, the *status quo* is insufficient: we need to know not only where we are now, but where we will be, and when [4][5]. A CIGRE Technical brochure indicates some of the key approaches to improving knowledge of substation asset condition, and optimizing asset usage – the approaches it describes are 'sensible', from Partial Discharge (PD) and Electro-Magnetic Interference (EMI) surveys to increased visual inspection and appropriate condition monitoring [6].

Analysis of data brings challenges – not least in understanding new and exotic techniques related to Machine Learning (ML), deep-ML, and Artificial Intelligence (AI) – but also in terms of how to apply them: what is needed to make the techniques valuable? Google provides a glossary of terms, but not everyone will agree with the Google definitions [7]. It has also been noted that the application of exotic analytics requires very 'clean' and well presented data – to the point where 95% of the possible value can be derived from data through basic statistics and meaningful organization of the data [8].

The application of condition monitoring is, at its core, an asset management activity [9][10][11]. Many organizations are embracing the approaches and structures described in the international standard for asset management, ISO 55000 [12]. Originally published as three documents in the 55000 range, including management system requirements and application, further documents are in development, including, for example, 'alignment of asset management, finance and accounting' in ISO 55010 [13]. The development and evolution of the Common Information Model (CIM) has been instrumental in enabling organizations to both structure their own data and to allow communication of data in a standard format with other organizations [14]. Even here there is possible cause for confusion: does data relate to an asset or to a location in a grid system, or both? The load on a transformer is of interest to the grid system management, but also for the ageing of the individual asset.

A big focus of an ISO 55000 asset management approach has been the balance of cost, risk and performance [13][15]. These have been augmented with the concept of ‘sustainability’ in revised versions – indicating a desire for an asset management which plans for the future. In terms of risk management, condition monitoring is a decision support tool to reduce the likelihood of catastrophic events and thus reduce risk – condition monitoring addresses the cause of hazards, rather than the consequence. Analysis of risk is a key element of asset management, and thus also must be addressed at the condition monitoring level: how does the data generated help us review the asset condition and the consequent probability of failure?

## CHALLENGE 2: SECURITY

Cyber security is a ‘hot topic’ in the industry at present, and rightly so. The consequences of cyber attacks can be devastating. Security applies not only to the data itself, and access to the data, but to management of the software systems and applications which generate, analyze and present the data and resulting information. For example, a Dissolved Gas Analysis (DGA) detector on a generator station auxiliary transformer may indicate the presence and severity of a possible problem, but a diagnosis of possible cause requires further analysis tools or sensors. Figure 1 shows a DGA detector in action, rising from a nominal 30 ppm to almost 80 ppm in about a week. The question is: do we believe the data and the resulting need for action? This is a very simple case, but if the plan is a de-energization of a power transformer as a result of such a rise, that has an impact on the grid as a whole: the recent events in the UK show the effects of a single lightning strike, which could be exacerbated by deliberate weakening of the grid’s resilience through cyber attacks on monitoring to necessitate the removal of highly inertial equipment, as discussed in the Financial Times of London [16].

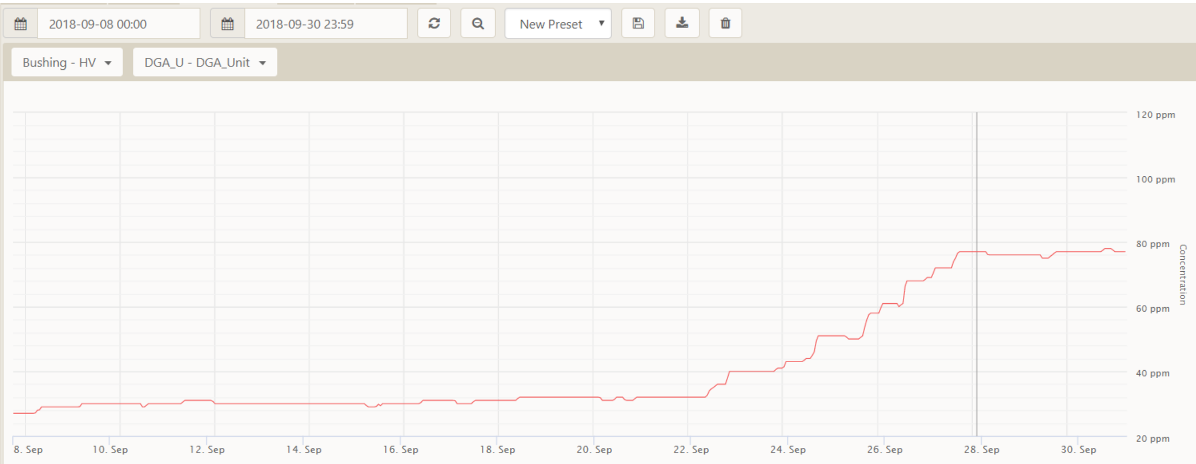


**Figure 1 Detector Type Monitoring of DGA Level**

## CHALLENGE 3: TIME IS OF THE ESSENCE

If assets did not deteriorate over time, they would not need maintenance, except to address external influences, and they would not need replacing except to address failures caused by external agents [17] [18]. But deterioration takes place through use/application and through interaction with the working environment over time: transformer paper insulation ages and becomes brittle with temperature, say, leading to replacement of a transformer before failure through mechanical forces.

We need to understand timescales of operation for failure modes – and the determination of parameters that indicate the presence and possible severity of a failure mode so that we can plan intervention in a timely manner, as per Figure 2 [18]. An analogy, using an automobile (car) can be useful: tire pressure reduces over time, eventually leading to the failure of the tire, and possible requirement for a new car.

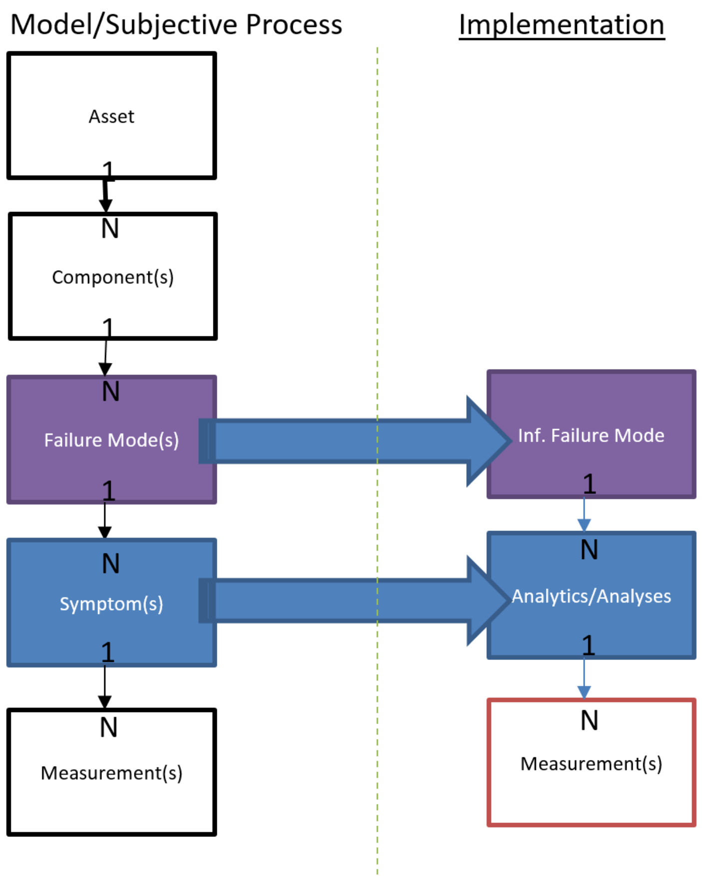


**Figure 2 Modeling the Failure Mode Analysis Process**

The process may take weeks to reach a conclusion. Detection of low pressure can therefore be an indication of a problem to be addressed within a number of weeks - we know that as the pressure reduces the likelihood/probability of tire failure increases, even if we do not have exact values: we understand the *urgency* [18]. Power transformers are much more complex assets than tires, and with many more failure modes and more parameters to indicate symptoms, the overall condition assessment is much more complex [19].

In a world of 'big data' more and more operational and condition data may be available, but the collation of that data is complex and machine learning and artificial intelligence only provide clues as to what may be a true cause for concern: a true failure mode in operation. Combining data from multiple measurements, each of which may be relevant to multiple failure modes, and then combining failure modes to assess a power transformer is not a simple process. To ensure that a *sense of urgency is not lost*, assessments must be calibrated for time and combined to reflect the true condition and intervention requirements of the asset. The means of reducing data to allow for a concise assessment for comparative purposes, includes the process of generating an Asset Health Index: and a health index which loses the sense of urgency is of little value in practice [20][21][22].

## CHALLENGE 4: ANALYTICS AND ACTION

An Asset Health Index is a useful place to start discussion of 'meaning' in data. Often a numeric value, an AHI is apparently simple to understand. But what does the number *mean*? We believe we understand numbers – we know which are bigger, and which are smaller and have some grasp of relative size – so how hard can it be? We are at risk of succumbing to the Dunning-Kruger Effect [23]: our estimation of our capability is not in proportion to our actual knowledge and capability. Hence the old saying: "*A little learning is a dangerous thing.*" The health index may take a vast amount of data, from the asset itself, from sensors, from maintenance history and produce a number which should be calibrated for time, monotonic in scale and urgency, auditable in terms of having a clear path from raw data to condition/health, justifiable in terms of the analysis used [19]. And yet many indices are none of those things – they just seem to be a way to collate a lot of data and produce a useful number: and we understand numbers, so we think we understand the analysis and the derivation… the result is often a health index which is disappointing and an asset management approach which is unjustified.

Machine Learning is growing in popularity [7][8]. But we may suffer from the same Dunning-Kruger effect: the analysis of data leads to a conclusion, but is the conclusion valid? The application of machine learning to PD/EMI spectra has been beneficial – even to the point of surpassing experts in their analyses of raw data [24], and as has been noted, much of the benefit of Machine Learning can be derived from good data clean up and standard statistical approaches [8].

It is important to have a plan to respond to anomalous or unusual data: in condition monitoring a response plan is a 'best practice': a response to a fire alarm should be to follow the plan in place at the particular location, not have a discussion and work out an approach 'on the fly'. But plans need to address the variability in data sources and applications – in monitoring a human heart rate, for example, a value of 140 may be considered very high, unless there was strenuous exercise being undertaken, in which case we may be more concerned about the duration at this level.

## CONCLUSION

Ultimately, we must be aware of the '*meaning´* inherent in the data we use for decisions: and also be aware of our own limitations in giving meaning to data, or to resulting parameters used for decision.

## ACKNOWLEDGMENTS

The authors would like to thank their many colleagues for conversations and discussion on matters related to the content of this paper; you know who you are.

## BIBLIOGRAPHY

[1] https://www.xsnet.com/blog/bid/205405/the-v-s-of-big-data-velocity-volume-value-variety-and-veracity

[2] T. McGrail, D. Angell & K. Elkinson, "*Asset Management - Data, Analysis & Decisions*", Electric Light & Power magazine, Volume 18, Issue 7, July 2013

[3] Tony McGrail, "*Asset Management: Data and Decisions*", IEEE International Smart Grid Conference, Perth Western Australia, 2011

[4] "Machine Condition Monitoring Market", B2B Market Research, www.marketsandmarkets.com
[5] "Industry 4.0: Opportunities and Challenges of the Industrial Internet", Price Waterhouse Coopers, https://www.pwc.com/gx/en/technology/pdf/industrial-internet-of-things.pdf

[6] CIGRE Technical Brochure 660 "Saving Through Optimized Maintenance in Air Insulated Substations", 2016
[7] https://developers.google.com/machine-learning/glossary/
[8] https://www.kdnuggets.com/2018/04/dirty-little-secret-data-scientist.html
[9] T. McGrail, R. Heywood, "*Intelligent Transformers: Monitoring, smart grid integration and asset management*", Coil Winding Insulation and Electrical Manufacturing Exhibition, Berlin, Germany, 2014
[10] T. McGrail, K. Elkinson, M. Lawrence *et al*, "*Aspects of Power Transformer Asset Management*", IEEE PES T&D Conference, Orlando, USA, 2012
[11] K. Wyper, G. MacKay, TransGrid Australia & T. McGrail "*Condition Monitoring in the Real World*" 80th International Conference of Doble Clients, Boston, USA, 2013
[12] ISO 55000 "Asset Management – Overview, Principles and Terminology", 2014
[13] https://en.wikipedia.org/wiki/ISO_55000
[14] Common Information Model: https://www.dmtf.org/standards/cim
[15] T. McGrail, "*Practical Steps to Improve Asset Management Applications*", TrafoTech Conference, New Delhi, India, 2016
[16] https://www.ft.com/content/bcb9f3b2-c352-11e9-a8e9-296ca66511c9
[17] https://www.through-life-engineering-services.org/downloads/27302_White_paper_amended.pdf

[18] "Condition Monitoring: Data Driven & Knowledge Based Analyses", T. McGrail & K. Elkinson, CIGRE Canada, Vancouver, 2016
[19] "Developing & Using Justifiable Asset Health Indices For Tactical & Strategic Risk Management", McGrail, Rhoads, White, CEPSI Conference, 2018, Kuala Lumpur, Malaysia
[20] "Transformer Health Indices & Probability of Failure", McGrail, Kennedy, Lawrence, TechCon Australia/NZ, Sydney, Australia, 2017
[21] "Turning an Asset Health Index into a Probability of Failure", McGrail, 85th International Conference of Doble Clients, Boston, USA, 2018
[22] "Translating an Asset Health Index into a Probability of Failure", CEATI report 30113, 2018

[23] "Unskilled and Unaware of It", Kruger & Dunning, Journal of Personality and Social Psychology. 77 (6): 1121–1134, 1999
[24] "AI approaches to interpretation of electrical CM data", Dr. I. Mitiche, BINDT 16th International Conference on Condition Monitoring & Asset Management, Glasgow, UK, 2019