Big data

Where is it and is it useful?

ABSTRACT

A great deal has been made of *big data* and its potential usefulness in many applications in the electric power industry. It certainly is useful in many domains, such as banking and logistics, giving insights to previously 'hidden' data aspects that can and do escape the human eye. Data for assets in substations is required to facilitate condition assessments of equipment and to develop the subsequent action plans for maintaining a safe and secure electrical infrastructure. Often when one seeks to retrieve this data, it can be difficult to find. If it is found, the question arises: how accurate is the data and is it current? For those engaged in the development of condition (or health) indices, this can prove to be a major challenge. Data quality and timeliness determines the confidence one can place on the evaluations made.

KEYWORDS

data, big data, data quality, timeliness

1. Introduction

In my last column [1], the discussion centred on the essential requirement for communications of data, information and alarms to a place where it can be acted upon. This carries the implication of the movement of data and access to previous test records, inspection reports, etc. indicating that the realm of *big data* is with us now.

The application of *big data* tools and data analytics by electric power utilities should consider the specific features of the world of managing of major assets such as transformers, switchgear and battery systems. Challenges arise regarding not only the quality and timeliness of the data records of the equipment's past life, but also the amount of data that has gone missing. An IEC White Paper published in 2015 [2] points out the concerns of the age distribution of these assets. From my point of view, this includes many other transformers outside of the utility realm, such as industrial and renewable generation applications, as well as concerns such as safety hazards (explosion and fire), environmental issues (oil spills), loss of

Specific features of managing major assets such as transformers should be taken into account when considering the application of *big data* tools and data analytics

revenue, obsolescence, spare parts, and the *lack of knowledgeable and qualified manpower*.

A recent article on 'expert systems' vs. human expert written by Georg Daemisch and published in Transformers Magazine [3] brings to light a concern I have had for many years, regarding the need for better control of data being processed by not only 'expert systems' but by humans as well.

The output of HV equipment condition assessment indexing (manually or automated) is widely used as a tool for planning, refurbishment or replacement of these critical assets. A condition index can be a weighted value, or some other statistical combination of a range of factors and parameters influencing their condition, which in turn can be derived from a combination of inspection, off-line testing and on-line continuous monitoring data.

These approaches generally assume that the data presents a reliable picture of current transformer condition, meaning that the test record is complete and up to date. In an operational context, the data may not be as reliable as is often assumed, and yet asset managers must still make decisions about maintenance and replacement based on it.

2. Data quality

Other industries such as banking and logistics have reached the realm of *big data* faster than the power industry. Specific processing platforms and technologies have been developed for handling *big data*. But before utilities adopt these platforms, the application of domain and unique requirements of the power industry should be considered. While large historical data sets may be available for analysis, the quality of this data will have a significant impact on the quality of decisions that will be made.

There are five dimensions of data quality [4]:

- Completeness: Are all the records or fields present?
- Timeliness: Is the data up to date?
- Validity: Does the data conform to formatting and domain rules? For example, age cannot be a negative number.
- Consistency: Can related records be compared without conflict?
- Accuracy: Is the data a true reflection of the situation being recorded?

In some cases, a further dimension is added [5]:

- Uniqueness: Is the data recorded without repetition? For example, is the measured and/or calculated value from an IED or sensor constant regardless of changing operating conditions?

With respect to transformer data, there is scope for poor data quality along a number of these axes. Manual data entry and record keeping introduces opportunities for invalid or inaccurate results to be stored. The skill level of the technician performing the test or noting gauge readings can also impact accuracy of the result. Sending DGA samples to different labs for analysis may lead to inconsistences due to differences in calibration and/or processing of the samples, which make it more difficult to compare results against each other.

In the following, I will present cases that I have selected from my experience, and I believe readers have had similar experiences regarding this issue. I bring these forward to those who may not have been involved in such experiences.

Table 1. Record of monthly manual instrument readings

Bank 02	40 MVA			115-24.9 kV											
Capacity	Summer, w/o fans			24 MVA											
	Summer, with fans			40 MVA											
	Winter, w/o fans			33 MVA											
	Winter, with fans			54 MVA											
Month	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	**3**
Year	**2011**												**2012**		
Top oil temperature [°C]	35	33	55	45	45	65	55	55	60	48	38	40	40	35	39
Hot spot WTI [°C]	40	35	35	38	38	50	60	65	62	50	38	35	35	35	35
With fans running	0	0	0	0	4	4	8	8	8	8	0	0	0	0	0
With fans NOT running	4	8	8	0	4	4	0	0	0	0	8	8	8	8	8

While large historical data sets may be available for analysis, the quality of this data will have a significant impact on the quality of decisions that will be made

3. Case 1

Table 1 details the record of manual monthly inspections of a substation transformer with a top rating of 40 MVA (with all fans running). One will note immediately (looking at numbers in red boxes), that the record of top oil temperatures (OTI) and winding hot spot temperatures (WTI) do not reflect reality. It is physically impossible for the WTI to read lower than the OTI.

The transformer failed catastrophically (there was no fire) with a load of approximately 28 MVA. No fans were operating. Upon inspection it was found that the OTI at some point in time had 'peaked' a 110 °C, while the WTI was no more than 65 °C. The cooling fans were controlled by the WTI setting of 75 °C. The tank split, and 20,000 litres of oil were spilled in the area.

It is obvious that the person taking the readings as well as the person doing the data entry into this log had no appreciation for this physically impossible situation to appear to be normal.

So, the question becomes: How many more records are in the same state? Obviously, the information contained in this case is useless, unless a system is in place to compare these two values and raise the flag when logic dictates it is an obvious error. It could have averted a costly failure.

4. Case 2

Another example of poor data is detailed in Table 2, showing the record for one of a fleet of GSU transformers with an independent power producer (IPP). The engineer on site drew the conclusion that this transformer was in perfect health, as their training indicated that "no gas" is a good result. However, this record immediately suggests poor data quality, and should not be considered a reliable indicator of the transformer condition.

5. Case 3

In this case, off-line bushing test results over a five-year period, presented in Table 3, reveal some troubling issues and curious results and/or testing protocols.

Table 2. Example DGA record showing poor accuracy and uniqueness

Customer:	XYZ	Substation:	GT-GSU	Rating	212 MVA
		Make		Duty:	GSU
		Serial No.	623	kV rating:	400

Date	**Lab**	**Oil Temp**	**Max Oil Temp**	**Hydrogen**	**Methane**	**Ethane**	**Ethylene**
13/03/2008	A			0	0	0	0
11/06/2008	A			0	0	0	0
20/09/2008	A			0	0	2	0
Date	**Acetylene**	**CO**	**CO_2**	**Nitrogen**	**Oxygen**	**Propane**	**TDCG**
13/03/2008	0	0	31	11747		0	0
11/06/2008	0	0	7	11687		0	0
20/09/2008	0	0	60	14468		2	2

Table 3. Unit number TR 603: Off-line bushing test results on 29th May 2012 and 10th June 2014

Bushing serial No.	Rating plate C [pF]	Rating plate % PF	Test date	Measured C [pF]	Corrected % PF	Condition
789/15	313	0,24	29/05/2012	313,6	0,5	Yellow
789/14	318	0,26	29/05/2012	314,1	0,31	Green
789/13	316	0,24	29/05/2012	310,3	1,26	Red
789/15	313	0,24	10/06/2014	315	0,94	Red
789/14	318	0,26	10/06/2014	315,9	0,39	Green
789/13	316	0,24	10/06/2014	311	3,41	Red

Table 4. Unit number 604 (?): Off-line bushing test results on 21st May 2015

Bushing serial No.	Rating plate C [pF]	Rating plate % PF	Test date	Measured C [pF]	Corrected % PF	Condition
789/15	313	0,24	21/05/2015	306,9	0,21	???
789/16	318	0,26	21/05/2015	315,9	0,39	Green
789/17	316	0,24	21/05/2015	311	3,41	Red

It is obvious that bushing serial numbers 789/15 and 789/13 are in very bad condition and should be removed from service and replaced.

The last test data for this unit, shown in Table 4, leads to many questions, such as the unit number. Are the bushing serial numbers correct on the report? It can be observed that the N/P data – C[pF] and % PF – for these three bushings match exactly with the bushings reported on unit number 603, save for the serial numbers. IF the serial numbers are correct, why has the bushing number 789/15 suddenly tested below the N/P values? It should be noted that the tested values for 789/16 match with 789/14, and those for 789/17 match with 789/13.

The real situation is that the serial number 789/13, after removal from the transformer, had obvious oil leaks between the upper porcelain housing and flange. These leaks were not noted on any inspection report.

The test report of 21st May 2015 is highly suspicious of improper reporting, testing, OR maybe the tests were never carried out!

6. Case 4

An arc furnace transformer 76 MVA 53.15 kV – 595 V, built in 1965, a sealed unit with pressurized nitrogen system, had some troubling test results from a qualified laboratory, which suggested there was a need for a more thorough condition assessment and inspection. That inspection revealed issues that Georg Daemisch alluded to in his article [3] – that numbers by themselves do not tell the whole story, Table 5.

The testing of January 2010 did not raise any concerns, but subsequent testing conducted in September 2010 saw a sharp rise in moisture content as well as in all five Furanic compounds, and the resulting advice of the laboratory was as follows:

Calculated DP <215 estimated life; remaining lifetime <5 %

Recommendation: Investigate

Retest confirms damage to the insulation system that may place this unit in danger of failure.

Recommend this unit be investigated as soon as possible for cause and either repair, rebuild or replacement.

For those engaged in the development of health indices a major challenge is how accurate the data is and whether it is current

Table 5. Water in oil and furan testing, arc furnace transformer

Date	Average Temp. [C]	Water [ppm]	Saturation [%]	Est. moisture content in paper [%]	5H2F [ppb]	2FOL [ppb]	2FAL [ppb]	2ACF [ppb]	5M2F [ppb]	Total Furans [ppb]
22/01/2009	10	8	22,4	2,64	ND	ND	590	ND	10	600
14/01/2010	10	7	18,2	2,15	ND	ND	585	ND	6	591
13/09/2010	37	12	10,6	1,02	143	235	2658	44	47	3127
17/09/2010	22	17	28,4	3,32	80	164	2103	30	37	2414
23/09/2010	25	16	23,6	2,72	61	120	2750	34	31	2996
15/10/2010	20	15	26,3	3,12	41	116	2400	27	39	2623

Widely used tools generally assume that the data presents a reliable picture of the current transformer condition, meaning that the test record is complete and up to date

The DGA record in Table 6 did not raise any concerns, based on the standard 'levels' associated with Table 1 of IEEE C57.104-1991 [6].

The lab recommended (a) a retest in six months, and (b) that the analysis of this (last) sample showed no significant increase in the combustible gas volume.

The real clue to the condition assessment and why the numbers "fall out as they did" is contained in a comment at the bottom of the report from the site:

"You will recall that the transformer does not have an adequate seal around the secondary bus as it passes thru the top of the tank. A nitrogen monitoring system has been installed that is holding about ½ (.5) PSI, but they're running thru a tank of nitrogen in about a week."

Thus, this transformer was NOT sealed, gases were escaping with the nitrogen gas being lost to the ambient and ambient moisture entering via the two-way highway created by the leak.

I fully agree with Georg Daemisch [3]: automated systems cannot and do not use "transformer expert logic" to understand what is happening. The case number 4 is an example of "automation" making recommendations on individual sets of data without understanding the context, or having further anecdotal, yet vital observations contributing to the seriousness of this situation.

The owner of this transformer was concerned about how many more "melts" he could achieve before taking an outage to replace it. I never did offer an answer to that question as I walked away from the unit.

7. The impact of ISO 55000 on data quality

ISO 55000 is a series of standards governing asset management processes (including data collection and retention) within organizations [7].

Within the standard, the requirements on documented information such as test results are discussed. Regarding data quality, three key sections of the standard require that [8]:

- "The organization shall include consideration off [...] the impact of quality, availability and management of information on organizational decision making."
- "The organization shall determine [...] how and when information is to be collected, analyzed and evaluated."
- "When determining its information requirements, the organization should consider [...] its ability to maintain the appropriate quality and timeliness of the information." [8]

Fundamentally, ISO 55000 places a duty on the organization to consider what data is needed for the decision-making process. It does not specify details such as the level of quality or timeliness that is appropriate. The responsibility is placed entirely on the company to justify that their data collection strategy is appropriate for their operational needs. Therefore, if data is potentially out of date, it is not indicative of a poor asset management strategy or non-conformance to ISO 55000, as there can be strong operational reasons why particular tests may not be performed at the usual intervals. From this it can be concluded that there is nothing in ISO 55000 which provides explicit guidance on the handling of data quality and timeliness.

8. Data timeliness is a facet of data management

Timeliness of data can be presented along with the results, so that the user can make a judgment about the reliability of the output. The timeliness can be presented in a few diverse ways, for example using icons that indicate the accuracy and timeliness of each piece of data, such as those presented in Figure 1.

For transformers, this would mean mapping out the condition index calculation and highlighting which inputs are of suspect timeliness.

The recent work done (and soon to be published as a Cigre Technical Brochure), from Cigre WG A2.49, *Guide for Transformer Condition Assessment* addresses the issues of data timeliness AND data quality, with examples of mitigations that can help when information is not available or is obsolete.

Conclusion

This aspect of developing confidence in any health or condition assessment of

Table 6. DGA results of the arc furnace transformer (gas values are cited in ppm)

Date	H_2	O_2	N_2	CH_4	CO	CO_2	C_2H_6	C_2H_4	C_2H_2
11/16/09	ND	25,110	54,428	3	72	832	ND	4	ND
01/14/10	ND	16,275	50,424	3	36	812	ND	2	ND
04/15/10	5	27,798	61,331	6	115	1,490	Trace	5	ND
07/29/10	13	15,736	49,768	5	161	1,853	ND	7	ND
09/13/10	19	10,334	59,017	3	172	8,494	ND	4	ND
09/17/10	14	11,833	67,611	3	154	9,985	ND	4	ND
09/23/10	4	11,833	75,701	2	80	9,763	ND	5	ND
10/15/10	6	9,405	62,431	1	43	4,261	ND	2	ND

Figure 1. Examples of icons indicating the accuracy and timeliness of each piece of data

equipment is a concern to those involved. As such, thought needs to be put into the assessment criteria to include not only a current test and visual observations, but also previous older test data that may not be obsolete.

This requires the use of expert judgment to identify a data timeliness period or criteria for each test or condition parameter. Any data outside of these criteria is considered potentially obsolete and should be treated with caution or ignored.

In an operational context, the data may not be as reliable as is often assumed, and yet asset managers must still make decisions based on it

Industry and standards groups' works in progress and/or published:

Cigre WG A2.49, *Guide for Transformer Condition Assessment*, planned publication in Dec 2018

Cigre TB 732, *Advanced Utility Data Management and Analytics for Improved Situational Awareness of EPU Operations*, published in June 2018

Cigre TB 725, *Ageing High Voltage Substation Equipment and Possible Mitigation Techniques*, published in May 2018

IEC has created (February 2018) Technical Committee 123, on Management of Assets in Power Systems

Bibliography

[1] B. Sparling, *Future of Substation Monitoring: It is not Just the Software*, Transformers Magazine, Vol. 5 Issue 2, April 2018, pp. 64-68

[2] IEC White Paper, *Strategic asset management of power networks*, 2015

[3] G. Daemisch, *Expert systems vs. human expert*, Transformers Magazine, Vol. 5 Issue 3, July 2018, pp. 104-111

[4] K. Yin, S. Wang, Z. Liu, Q. Yu, and B. Zhou, *Research and development on data quality assessment management system*, in 2nd International Conference on Systems and Informatics (ICSAI), November 2014, pp. 992–997

[5] N. Askham, D. Cook, M. Doyle, H. Fereday, M. Gibson, U. Landbeck, R. Lee, C. Maynard, G. Palmer, and J. Schwarzenbach, *The Six Primary Dimensions for Data Quality Assessment: Defining Data Quality Dimensions*, The Data Management Association (DAMA) UK, Tech. Rep., October 2013

[6] IEEE C57.104-1991 – IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers

[7] ISO 55000 series on Asset Management, 2014 (ISO 55000: Overview Principles and Terminology; ISO 55001: Management Systems Requirements; ISO 55002: Guidelines for the application of Management System Requirements)

[8] B. Sparling and V. Catterson, *Big data … where is all the data and is it useful?*, TechCon North America 2017, Houston TX, February 2017

Author

Brian D. Sparling, a Senior member of IEEE, is the Senior Technical Advisor with Dynamic Ratings Inc. Brian has over 20 years of experience in the field of power and distribution transformers. For the past 25 years, he has been involved in all aspects of monitoring, diagnostics and condition assessment of power transformers. He has authored and co-authored more than 27 technical papers on various topics dealing with monitoring and diagnostics of transformers. He has contributed to many guides and standards with the Canadian Electricity Association, IEEE Transformers Committee and the Cigré A2 Transformers Committee.