



Data quality – reduce to the min or push to the max? Keep the balance to achieve your goals in Predictive Maintenance.

Data is key – especially in the field of Industry 4.0. No data, no prediction! However, finding the right balance between the min and the max data quality possible can be a challenge. Many companies are collecting an incredible amount of data (hello big data!). But to be honest: the same also struggle with handling it to extract the right information based on pre-defined KPIs. This can be a real pain in your neck, especially when you want to achieve the right findings and argumentation, to point out a certain problem based on Predictive Maintenance. In this article, you learn what really matters when it comes to collecting the right data quality to realize a valuable Predictive Maintenance Solution.

Fairytale or true story? Write your success story based on the right measurement

Imagine you have a problem with the high temperature in your office. To document your problem, you measure the room temperature every time you feel the temperature in your room is too hot and capture the results in an Excel spreadsheet. With this, you walk to your boss presenting the data to persuade him from your problem by suffering hot temperature – confident that he will understand your suffering now by proven data. Surprisingly he doubts the need for an air conditioner claiming that your measurement is not representative and even not trustworthy. “Measuring the temperature only at one point in time is not creating the right image of the situation!”, he says. Instead, he suggests recording the temperature every 30 minutes to get a full understanding of the temperature development throughout the day. To ensure uninfluenced recording, he will use three data loggers, recording the temperature in the office for a full year!

This example illustrates some of the aspects which must be considered working with measurements and data.

Before you start setting up a measurement and data recording, you should define first what you want to achieve in Predictive Maintenance. Based on that, you will afterwards define the specific requirements to find out the right information to achieve your goals.

The three musketeers of data quality: granularity, completeness, timeliness

In order not to end up in getting the fairytale stamp, you will need to present your boss representative, reliable and consistently collected data which support your arguments being acknowledged as facts. Follow the familiar argumentation structure: Assertion, justification, proof! Properly collected data are your proof and gateway to your success story. Write it and consider the three musketeers of data quality: granularity, completeness, timeliness (short: GCT).



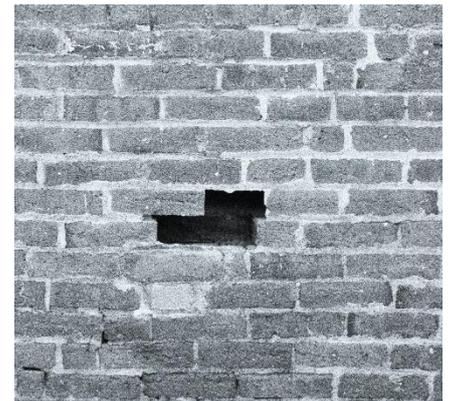
As this example shows, you don't always need the highest data granularity to visualize the big picture. We can still recognize the sea and the clouds despite lower data granularity.

Data granularity – the right level of detail

This is the level of detail the data provides. For time series of data like the temperature measurement, this is the frequency of measurements (e. g. every 30 minutes). The granularity defines which information you can extract from the data (e. g. min, max, avg, ...). Decisive for the decision is the characteristic of the measured signal and the expected result (e. g. slow moving signal like room temperature or fast oscillating signal like vibration). Data granularity has as a huge impact on data capturing, storage, transmission, analysis, and visualization. Obviously, the more data you get, the higher the granularity. This turns into a cost issue in case you are using data roaming and cloud storage, as every bit must be paid for. Consequently, you should look for the minimum required.

Data completeness – the courage to the gap

This is the wrongly missing of data if there should be data available. In my above-mentioned example, you would expect one measurement per day. If one is missing, you do not know if it wasn't hot or if recording was skipped. In data analytics, it may lead to wrong conclusions and will require some pre-processing of data, e. g. automatic filling of missing data. The importance of data incompleteness must be evaluated, and counter measured to be installed accordingly. E. g. it could be required to have the entire data acquisition system powered by uninterrupted power supply (UPS) or you may need to have data storage / buffering on the edge. The more important data completeness is for your application, the more effort you must invest to make it reliable, the more expensive it will get. Doing as much as possible is often not required!



Pay attention to the nature of the gap - sometimes a gap has no negative impact, except for the fact that it might look visually strange – like this stone wall.



Think about the flow of your data! It needs to be available when you really need it in the right frequency. Otherwise, you get stuck like in a traffic jam.

Data timeliness – the flow to go

Or also called "latency", is the degree to which the data is available at the right time. E. g. your boss didn't want to see the measurement immediately after recording but he wants to see it at the end of the week. Specifically for measurements which are transmitting data via mobile network, or which are battery based, the timeliness has an impact. If you need continuous and immediate update of your data, the solution will probably not go with battery powered sensors as continuous measures and data transmission is energy intense. The same is true if your roaming cost need to be low. In that case, sending data packages every hour or once a day may be the solution to optimize costs.

Go for the best fitting signal processing technology to achieve your goals in Predictive Maintenance

Today's technologies for signal processing can satisfy nearly all requirements. It is rarely necessary to apply the latest available technology but more cost – and resource effective, and thus the best fitting technology. If you consider the three aspects above for designing your solution (GCT), you are on the right track to achieve the required data quality to achieve your goals in Predictive Maintenance.

How did the story end? We are sure you will know it already after reading this article 😊

After collecting the data based on the three aspects (GCT), it turned out that the temperature was only in 15 days a year, in total 60 hours, above 30 degrees Celsius at the office. This means 60 hours of sweating (out of 1760 hours). Unfortunately, your boss concluded not to buy an aircon based on this data. However, he offered free ice cream during these days.

Author: Michael Bruckhaus, Managing Director, PREMAS AG

Pictures: shutterstock / unsplash.com

Want to know more about Predictive Maintenance? Contact us.

PREMAS

Preventive Maintenance Service AG

Haldenstrasse 1 · 6340 Baar · Switzerland

Phone +41 41 766 82 01

info@premas.ch · www.premas.ch