**Automaatiopäivät[22] 2017 Seminar:**

# BIG DATA APPLICATION TO COMBINE THE VALUE CHAIN DATA TO KNOWLEDGE

*Matti Häkkinen*
Savcor Oy, Insinöörinkatu 8, 50100, Mikkeli, Finland
Tel: +358 40 521 6068, E-mail: matti.hakkinen@savcor.com, https://www.savcor.com/

*Mika Suojärvi*
Savcor Oy, Insinöörinkatu 8, 50100, Mikkeli, Finland
Tel: +358 40 567 2447, E-mail: mika.suojarvi@savcor.com, https://www.savcor.com/

*Jari Suihkonen*
Savcor Oy, Tekniikantie 14, 02150, Espoo, Finland
Tel: +358 40 750 0572, E-mail: jari.suihkonen@savcor.com, https://www.savcor.com/

## ABSTRACT

Nowadays the amount of data collected from industrial processes are increasing with an accelerating speed. Modern computer systems enable to save data with much longer history length and much faster sampling rate than ever before. Digital information is generated in whole lifecycle of product, starting from raw material production to end product recycling. This raises a relevant and important question; how I can exploit all that data?

Many bigger companies who works with consumer goods and services have utilized big data type approach already for several years successfully with the data they collect from their customer transactions. The obvious aim is to find a competitive edge compared to their rivals. Industrial business to business companies are not in the same level yet, but the ever increasing competition forces also them to fine tune their manufacturing processes. Companies are forced to improve their efficiency by producing better quality in larger amounts and with decreased cost.

Efficiency can be improved by an approach that lets operators and process engineers find answers to questions about the manufacturing process based on measured time series data.

## INTRODUCTION

This paper presents how industrial companies can handle huge amount of process data in an efficient way and how they could achieve competitive edge with a better process diagnostics.

## BACKGROUND OF PROCESS DATA MANAGEMENT

During the last decade digital information, process control and business execution systems have become standard applications in process industry. At the same time the amounts of data collected and saved are increased a lot. This increasing trend will continue in the future also. Today huge amount of data is collected, but at the moment a minor part of that is utilized. Typically production related data are scattered in several organizations and systems, such as forestry department (FIMS, forest information management system), purchasing, maintenance, etc. (ERP, enterprise resource planning), production planning (MES, manufacturing execution system), process operation (DCS, distributed

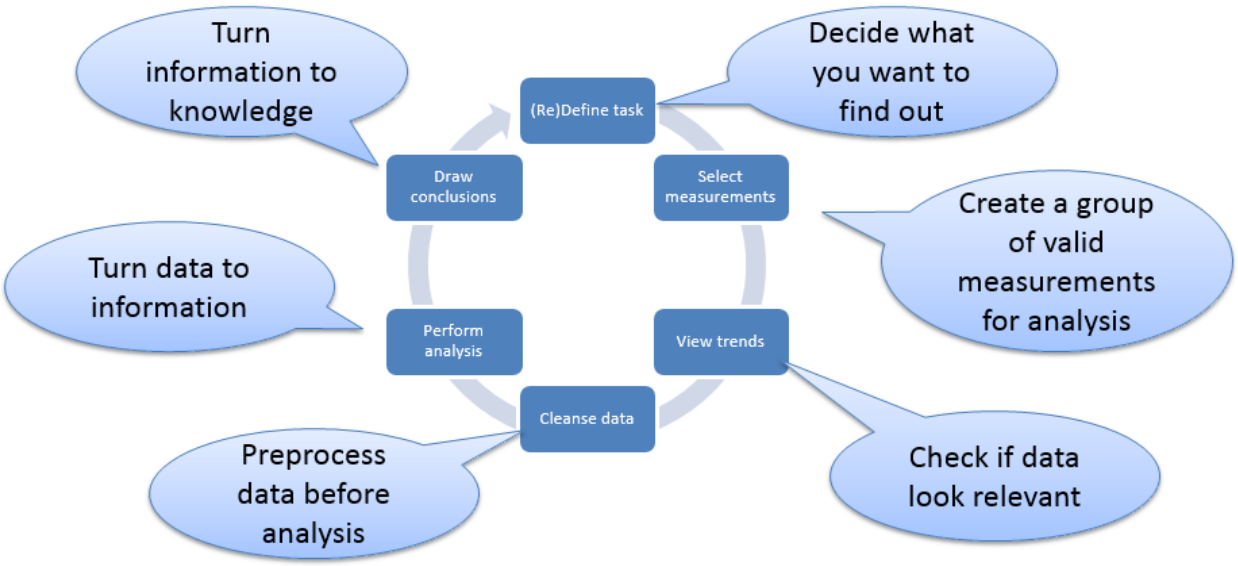
control system), quality management and customer related systems (ERP, CRM, customer relations management). A production plant or an entire company would achieve huge benefits if all this data would be combined seamlessly together and the users would get a total view to the process.

## BETTER PROFITABILITY WITH EFFICIENT PROCESS DATA MANAGEMENT

Collected data mass contains a lot of unused business and process improvement potential. Traditional way to utilize this potential has been modelling and simulation. Challenge in these approaches is how to keep complex process models up-to-date. Long production chains have also a lot of process evolution and typically process evolves faster that it's possible to update process models. To take this potential in use an efficient data handling system is needed. The system should be able to combine all data sources into same system and the user interface should be able to support the user to turn huge amounts of data to useful information. The user shouldn't need to spend lot of time to get useful data into data handling system but instead of that the user should just focus to troubleshoot the problem or improve the process and business efficiency.

Relationships between process phenomena are very complicated and several measurements affects together to another measurements. Therefore there are often a need to combine original measurements into a new and more useful measurement, e.g. to calculate dry solid flow based on original flow and consistency measurements. To enable this type of free data modification there must be an easy way to add own calculations into the data handling system.

When all relevant data is available in the data handling system it is possible to start process analysis. The workflow of process improvement or troubleshooting cycle is illustrated in the picture 1.




**Picture 1,** Process Analysis workflow

The first step is to define the target: what you want to find out?

The next step is to create a working set of measurements. In this point it's important that user can combine data from each different data sources to the same system and the data connection should be online to avoid any laborious and slow data importing to the analysis system.

Process data contain always some bad values e.g. huge outlier values, erroneous values, shutdown data values. This type of bad data affect a lot to the analysis results and therefore those values must be removed before analysis. To perform this kind of data cleansing efficiently the tool should support the user to utilize his/her process knowledge to judge which values are not relevant and then those values should be able to remove easily from the dataset.

When data is cleansed the user can perform process analysis, e.g. statistics and correlation calculations. In this step it's typical that the user already finds some useful information from the data but when there are lot of measurements this can be laborious. Computer power can be used to make this step much more efficient. A computer can check quickly huge number of process and quality measurements to find out those measurements which correlate best with the set target. The system should be able to rank the measurements and give a candidate list of measurements which seem to affect the most to the target measurement. In this step the data is turned into information.

It's important to remember that without any artificial intelligent system a good process knowledge from the user is required to get useful results from the studies. In this step the information is turned into new knowledge and the user can redefine the task and start the loop again.

## CONCLUSIONS

The amounts of data are increasing with an accelerating speed. The resources in the mills are decreasing. The competition between companies and mills is strong and the situation is not going to be easier in the future. In addition to all this the pressure to save natural resources is also raising.

All these things forces the companies to run the processes more efficiently and also in a sustainable manner. One of the most cost efficient way to achieve the targets is to utilize already existing data in more efficient way.

Demand for efficient data handling system is to utilize the potential of information collected from the whole production chain. For example, optimization end product production process and end product properties by taking into account raw material properties.

In this scope many smaller scale system demands are remaining and also some new challenges have to be solved. Cause–effect finding have to be fast, data retrieval has to be automatic and online, system have to be able to find relations over long delays and system have to operate reliably without process modelling. When change happens in process upstream, downstream process can be optimized for the coming situation proactively – and not only solve problems afterwards.

New generation process diagnostics enables further optimization of the current production chain compared to optimization department by department. Current production chains can utilize part of the potential, but wide scope process diagnostics reveals also a lot of potential which requires modifications in processes, business models and organizations. Customer case "Analysis of log data" in presentation