

Data Collection Strategy

How can data collection be optimized to allow decisions to be taken, and then governed to ensure success?

The key is to understand up-front the range of decision support needed, and the way different stakeholders would use the data. If the complete mix of needs allows for very aggregated data for each role that uses it, then the selection and use of tools where the cost/capability mix is lighter can reflect that. If you have some users needing very fine data, then you would look to see whether other stakeholders' needs can be addressed by rolling up some of the fine-grained information.

Generally you would look for a data platform that is open – that can be accessed by visualization and reporting tools appropriate to each type of user. Where possible, it's best to avoid different data collection approaches that effectively overlap in terms of what they capture, and that limit the exposure of that data to purpose-specific user environments.

Ultimately you need to ID the areas of your business that need attention – is it cost, quality or delivery you need to query? To get started, pick one of your problem statements – and perform 5 Whys to get you to a rough theme or area which then you should focus your data collection activities. Example:

- Problem Statement: I think we have the right capacity, but we only deliver on time 60% of the time
- 5 Whys drill down to: We have material, but we think we have a bottleneck around testing
- Optimized data collection: Direct labour time capture / machine utilization will help you understand if you can better use existing capacity – or if your capacity consumption is wrong and you need to invest in that area

Running a similar exercise for other key areas of opportunity will give you a clear picture of where data collected once can be re-used, and it will help illustrate whether you have to allow for significantly different data capture capabilities to solve for the whole mix.

What would be considered a “Minimum viable data collection product” for smaller manufacturers

We see “viable” being a reflection of cost of ownership considerations as well as “being properly fit to the nature of the source of the data”. On the first front we'd suggest looking for a solution that can piggyback any existing equipment automation or sensors via common connectivity software built on standards such as OPC-DA, OPC-UA or MQTT, rather than requiring its own dedicated source-to-datastore hardware. As above, we think the key is that the data then be accessible using standard query and integration methods, rather than being bound to product-specific displays and reports. That way you can be sure that data collected once is easily available for many stakeholders to use.

Beyond those basics, it's important to assess the nature of the data you intend to use – if you're needing to capture information at subsecond intervals across thousands of sources, then that will mean looking at solutions built for OT environments, like process/machine data “historians”, rather than enterprise-style database platforms.

How to develop a strategy that allows for old and new equipment?

The main distinction here is often that older equipment is either less connectible, or less data rich. The approach we recommend to customers is to consider assets and processes independently of what data is easily available from a control system. Once you've defined the data you'd ideally make available for a mix of tracking, KPI's and detailed analysis, you can set a data collection and visualization strategy that allows for differences in what's collected, but has a common definition for each type of asset or process (or material, for that matter). Assuming you get richer data from newer equipment, you may find that the value from better reliability or quality performance that you're able to achieve on the newer plant helps you quantify the value of overhaul/upgrade or replacement of older assets.

And you can take simple steps to get some useful information flowing off legacy assets - old tooling can be modified, for example with IP-connected power taps to determine if it is **actually running** or not. So you can absolutely take those 'less granular' datasets and pull them into a larger dashboard to at least visualize current process flow and historic utilization.

How much data is too much data? Should a data collection strategy be proactive (collect in case needed) or reactive (collect once we know not having data creates risk)?

There is a chicken/egg debate here: some kinds of data platform are extremely cost-efficient at collecting and storing large volumes of data (machine/process historians), whereas the cost-efficiency for a more enterprise style of system may be much lower for a similar volume of data collected and stored (and as mentioned before the rate of collection necessary may outstrip some kinds of solutions). So we always recommend working "backwards" from the foreseeable mix of uses and users of data. If you can't reasonably foresee better tracking, maintenance, quality or risk management from holding a piece of data, then maybe it can be left out of the bigger strategy.

Another nuance is that the phrase "too much" is often used when maybe it would be more accurate to say "not well correlated in ways that make it easy to use". Consider how you would relate a stream of data to the asset where it was born, as well as to things like the category of asset that machine belongs to, the materials run through it, the specifications that define acceptable quality and performance, and so on. If you can see how a data element offers value with those relationships easy to see and follow, then that data probably has a useful place in your repository.

All that said, we do believe that where cost is not prohibitive, it is always better to be proactive and capture as much as possible. No one knows the future needs of the business - and generally, data is cheap to store and not do anything with until needed.

Would like to understand where others have found value in data

There are a lot of specific examples available from product vendors, consultants and peers. You'll have seen case studies addressing capacity utilization, quality improvements, even labour savings... So we think it's more useful to say that where you will find value will be in areas where you can take action if the data tells you something useful. That may sound obvious, but we see many cases where customers analyse the heck out of something that's not easily changed (the configuration of a long, complex packaging line, or the variation in chemical reactions in fixed vessels of different sizes), when more actionable problems are in play – are recipes consistently executed; are my suppliers sending material that's well within specification; do I have a consistent energy usage profile across runs of like products...? If you can identify problems where the fix is controllable, you can create a series of wins that will make it easier to get support for acting on insights derived into more entrenched areas.

Contact details

If you would like any further information on what is covered in this document please contact us using the details below:

Sean Robinson – Novotek UK & Ireland

E: sean.robinson@novotek.com