



Get machine-readable data for industrial AI



Roger Herger

4 min read · Mar 20



Listen



Share

More

```
{  
  "options": {},  
  "indexes": [  
    {  
      "v": {  
        "$numberInt": "7"  
      },  
      "key": {  
        "_id": {  
          "$numberInt": "3"  
        }  
      },  
      "name": "_id",  
      "ns": "test.model"  
    }  
  ],  
  "uuid": "fef3g943f024984844gd7db4969b907"  
}
```

Exemplary metadata of a model.

Modern machine learning (ML) algorithms have one thing in common: they are data-hungry beasts.

Or to put it more amiably: They are your teenage offspring, regularly munching away three-quarters of the family meal without blinking twice. As a loving mom or dad, you naturally want to see them grow and thrive. That's why you enjoy feeding and raising them. And your professional self knows: they are the future money makers in your family.

This is where the problem on your industrial AI journey starts: Your company must be able to generate new data regularly and reproducibly. Personally, I believe that this is one of two big challenges of implementing AI in an industrial setting. And it is one that is very closely tied to your corporate culture.

Machine-readable data is key to industrial AI

The solution can be named simply: Get machine-readable data; from production, engineering, and labs. However, the path to this goal is paved with a few stumbling block. Let's address some of them.

Machine-readable data are much more than electronically stored data. They are the products of processes that were thought digitally from the bottom up. For you as an executive, this means that you must be willing to align your processes digitally. Meaning also: manage the change.

My previous experience on large IT implementation projects with multinational companies showed that you should attribute about 30 % of your time and money budget to change management. Save money here and you will fail later. Some of your employees will love the change, some will be strongly against it, many will have fears we need to address. So it's better to take this very seriously from the beginning. Also keep in mind that there are excellent companies that can help you with that.

Modularity and simplicity save you money

Second, I think it's best to rethink your processes as a string of repeatable modules. As a manufacturing company you probably have excellent work instructions that have been proven in practice. You are — per definition — repeatable. Unify these instructions in modules that are as simple as possible. The crucial keyword here is: Simplicity. Remember that you need to digitally implement each of these modules at end. The simpler, the cheaper.

Maybe you need expert support to do that in your operational technology stack of your manufacturing environment. Some of my co-entrepreneurs are building their startups with this very purpose in mind: Getting the relevant data from production and testing machines for further processing.

Metadata is the grease for your data motor

Third — and this is often the most difficult step — , record the metadata about your process, i.e., “data about your data”, simply put. You know your product, you know your manufacturing process, all the dos and don'ts. Your computer does not. We've seen a few

times where data was basically available, but the metadata was missing. That ended up making the data useless for ML.

When starting up our own company [maXerial](#), we were aware of these caveats. We had the advantage of starting with a greenfield approach. If you cannot do that, start in a simple and small field, and learn. However, keep in mind that you want to transform your entire company over time.

Find all your data — in seconds

We used the first year to build our database system — in addition to building up the financing for our industrial computed tomography system. I set a very clear goal: I do not want to lose a single byte on a sample on its way through our lab. To this day, we stick to that credo.

We defined a data model that represents our first business case — X-ray computed tomography of an industrial component — in our database. We always know, for example, who the sample came from or who made the measurement. But more importantly, not only we know it, but our database knows it. Big difference. Every measurement we generate is fed into the database.

We have established processes that literally force us to store the data in this way in our database. We think this will be an incredible treasure for our customers later on. The longer we are in the market measuring for our customers, the more we can get out of their data for them. We can train models on it or analyze trends over time. We can retrieve any 3D volume ever measured for them. How cool is that?

In our next article, we will cover how to build a suitable environment for industrial AI.

Further reading

This is the second article in our series on industrial artificial intelligence (AI). More articles in this series (list updated on release):

- (1) [How to bring AI to your manufacturing company](#)
- (2) [Get machine-readable data for industrial AI](#)
- (3) [Build sandboxes and let them play](#)
- (4) [Problems you can solve with ML in your company](#)
- (5) [Your route to success in industrial AI: Think big, start simple](#)

(6) [From pilot to maintainable AI technology stack](#)

(7) [What you can learn from your smartphone for industrial AI](#)

Machine Learning

Data

Artificial Intelligence

Manufacturing

Change Management



Edit profile


Written by Roger Herger

3 Followers

Roger is an entrepreneur in artificial intelligence and X-ray technology. He develops data-driven materials for high-tech industry.

More from Roger Herger



 Roger Herger

Your route to success in industrial AI: Think big, start simple