



Why is data quality important?

There has been a lot of emphasis on data quality lately. It's becoming obvious (yet again!) that having poor quality data is bad for business. Garbage in, garbage out.

Then why is it a hot topic?

I had several recent debates with the data practitioners about it. It appears we forego best data management practices for the speed of delivery. It's more important to produce insights and data products fast than spend effort on production grade data quality management. It's seen as bureaucratic and taxing, unnecessary for many data projects.

What are the key drivers of this?

- Desire to get results fast
 - The initiatives start off as an experiment, something that the business owners and the analysts put together to prove a concept. It works. The managers love it and demand it to be made live asap. To productionise the PoC would take months. "But it works fine now, right? Let's run with it"
- Lack of self-discipline
 - Some organisations lack the rigour to get on top of their data issues. They would do nothing until it becomes impossible to ignore them. Kicking the can down the road
- Technical limitations
 - We have plenty of tech to do production grade data quality management. However, not everyone has the full suite of such functionality available to them. Or they lack the tech expertise to develop an in-house solution. Interestingly, it was seen quicker to do data quality manually as and when than invest in automation!

None of the above should be an excuse to avoid embedding data quality management from the outset. If you do not address it early on, you WILL get into tech debt.

Data quality issues will inevitably overpower your organisation's ability to manage them. I've been around long enough to see poor data quality scupper many data products/projects. It gets difficult to implement production grade data quality management retroactively.

What can we do about it efficiently, so it's not seen as bureaucratic and taxing?

1. Start with self-discipline. Recognise that dedicating time and resources to managing data quality is crucial for the success of your organisation. Don't find excuses to avoid doing it. Make data quality a key part of your culture
2. Build in as much automation and standardisation as possible. Embed data quality management processes into ALL your pipelines. The less you do manually, the easier it will be for you to adopt the data quality culture. You will deliver QUALITY data products faster too
3. Streamline your architecture, processes and tools. Make it simple to work with your data estate. Implementing and enforcing data quality policies should be easy. The fewer barriers (technical or procedural) you put in the way, the wider the adoption will be
4. Put data contracts in place. Ensure your data quality issues are addressed at the source as much as possible

Remember folks, the more effort you put up front to build solid foundations for data quality, the bigger the benefits you will gain. If you want to have a chat about the ways of implementing data quality efficiently, drop me a note on val.goldine@ioblend.com. I'll be happy to show you how to do it with IOblend simply and efficiently.