

SETUP RISKS

d4t4science.com | Philipp M. Diesinger, November 10th, 2016

1 INTRODUCTION

The term *data science* describes the utilization of scientific problem solving in combination with recent advances in data technologies to solve business problems. Often times, new analytical techniques like machine learning, causal inference or scientific modelling can solve business problems in new or more efficient ways when reasonable data is available.

Setting up data science in corporate environments is a challenging task. Many companies struggle or even fail at it and end up with sub-optimal setups that significantly limit the potential positive impact data science can have on their organization. Some risks and setup models will be discussed in the following.

1.1 STRUCTURAL SETUP

There are two opposing types of data science setups with regards to distribution of talent:

- a) *Distributed setups* attempt to integrate data science into existing business units such as business analytics, business intelligence or IT teams. These units naturally exist within different functions across a company.
- b) On the other hand, *capability-centric setups* have a central data science capability which can be utilized by functions across the company.

The distributed setup is business function centric and thus follows a traditional business structure. Support capabilities center around a function to enable it. The capability-centric setup on the other hand focusses on centralizing the data science capability and thus on optimizing its impact on the business functions around it. Compromising hybrid models between these two extremes exist as well but will not be further discussed in the following.

The capability-centric model introduces advantages over the distributed model. However, higher effort is required to set it up successfully. More decision makers may be involved in the process and achieving consent can lead to significant delays which may come at a potentially high opportunity cost. A capability-centric setup is also much harder to grow organically which might be challenging in risk averse environments. However, the long-term benefits of the capability-centric model often times outweigh the initial

setup challenges. Companies who eventually adopted a capability-centric setup for data science often times followed a different approach initially.

2 CAPABILITY-CENTRICITY

There are numerous criteria to be carefully considered when implementing a data science capability. Three selected key factors qualifying a capability-centric setup over a distributed one will be described in the following.

2.1 AGILITY

Data science requires agile setups to thrive. Prototyping, fast-forward approaches and research are essential aspects of data science. Scientific problem-solving is a key driver for the outcome quality of data science projects but at the same time a highly non-linear process that requires a significant amount of flexibility and adaptability.

In corporate environments, projects will always be carried out under strong constraints such as limited financial resources and time or other boundaries like data limitations. To not let constraints build up to a level at which data science cannot function anymore a supportive and enabling environment needs to be established.

In a capability-centric setup, almost academic levels of agility can be established by shielding data science from the firm corporate network of surrounding business processes (*bubble of agility*). Similar dedicated agile environments can be found in startups and are the reason why garage setups are very successful in data science. In distributed setups, data scientists are typically integrated in larger pre-existing teams, and therefore are likely to be subject to the same business processes that the existing team is already embedded in. This reduces the agility of the data scientist's work environment and thus their ability to succeed.

2.2 SYNERGY

In capability-centric setups project knowledge can be built up and shared among a group of data scientists more easily. Furthermore, a comprehensive understanding of a company's data landscape can be established when working across functions. Centralizing knowledge can enable strong

synergetic effects (*connecting the dots*) that are lost in distributed setups where knowledge remains fragmented.

Many successful data science projects that made it into the press in the past years were fostered by this effect. A team that works on a sales project one month and on a finance project the next month might realize that customer insights and market elasticity data from the sales project can be utilized for more precise decision making on financial investments across the company's product portfolio.

2.3 TALENT AND SCALABILITY

Strong talent is crucial for data science but unfortunately a very sparse resource. Following recent studies, data science talent can be expected to remain for about two to three years within a corporation before seeking new opportunities. Any measures to extend this period by increasing the attractiveness of a position should be taken into consideration.

Centralizing talent enables data scientists to grow and synergize with their peers and therefore develop the company's capability. A capability-centric setup will furthermore establish interesting career development opportunities as the larger team size facilitates a more diverse set of seniority levels. A centralized team can create an attractive work environment for data scientists. Distributed setups can turn into isolated experiences which may lead to frustration.

Furthermore, distributed setups come with the risk of exposing data science talent to mere analytical support roles which will significantly hinder unlocking their full potential.

Last but not least, a capability-centric setup can often times be scaled more easily than a distributed one including support from external partners to adapt to fluctuating internal demand.

3 CAPABILITY-CENTRIC MODELS

A specific collaboration model for a capability-centric setup is an internal consulting approach. In this model, a central data science capability engages temporarily on a project basis with functions across a (potentially global) corporate environment. A consulting setup can become a strong internal driver towards utilization of data and scientific problem solving for smart decision making.

Another example for a capability-centric model is a garage setup. This setup focusses even more on establishing an agile fast paced environment by deliberately setting up data science outside of the corporate environment.

There is no single best solution for establishing data science capabilities in corporate environments. Existing structures and processes need to be considered carefully to identify the optimal setup for each corporation individually. People and mind sets play an important role which needs to be considered as well.

