TOWARDS DESIGN PRINCIPLES FOR DATA-DRIVEN DECISION MAKING – AN ACTION DESIGN RESEARCH PROJECT IN THE MARITIME INDUSTRY

Research paper

Tiemo Thiess, IT University of Copenhagen, Copenhagen, Denmark, tith@itu.dk

Oliver Müller, IT University of Copenhagen, Copenhagen, Denmark, olmy@itu.dk

Abstract

Data-driven decision making (DDD) refers to organizational decision-making practices that emphasize the use of data and statistical analysis instead of relying on human judgment only. Various empirical studies provide evidence for the value of DDD, both on individual decision maker level and the organizational level. Yet, the path from data to value is not always an easy one and various organizational and psychological factors mediate and moderate the translation of data-driven insights into better decisions and, subsequently, effective business actions. The current body of academic literature on DDD lacks prescriptive knowledge on how to successfully employ DDD in complex organizational settings. Against this background, this paper reports on an action design research study aimed at designing and implementing IT artifacts for DDD at one of the largest ship engine manufacturers in the world. Our main contribution is a set of design principles highlighting, besides decision quality, the importance of model comprehensibility, domain knowledge, and actionability of results.

Keywords: Data-Driven Decision Making, Design Principles, Action Design Research

1 Data-driven Decision Making and its Business Value

Data-driven decision making (DDD) describes organizational decision-making practices that emphasize the use of data and statistical analysis instead of human judgment only (Brynjolfsson et al., 2011). Provost and Fawcett (2013) understand DDD as the outcome of data science, which they define as follows: “Data science involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data” (p.53). Moreover, they state that data science relies on (big) data processing and engineering. So, following Provost and Fawcett (2013), DDD is the outcome of data science, data processing, and data engineering processes.

DDD is rooted in different technical disciplines, such as, decision support systems (Arnott and Pervan, 2008; Shim et al., 2002), business intelligence (Chen et al., 2012), data mining and knowledge discovery (Fayyad et al., 1996), and machine learning (Bishop, 2006; Samuel, 1959). But to turn data into value, it is equally important to also consider behavioral aspects of human judgment and decision making (Kahneman, 2003; Thaler, 1980; Tversky and Kahneman, 1992). Human judgment can, for example, be affected by cognitive biases. Due to their limited information processing capacities, humans often apply simplifying heuristics for making decisions, especially in situations characterized by high uncertainty (Tversky and Kahneman, 1974). Consequently, human judgments tend to be inferior to formal or algorithmic predictions in terms of predictive accuracy (Grove et al., 2000). Yet, at the same time there is a growing number of critical voices arguing that algorithmic decisions can be subject to biases too (Boyd and Crawford, 2012); for example, because they enact simplistic approaches to knowledge creation, are built on an uncritical use of black-boxed assumptions, or lack accountability and transparency (see Winner, 1980; Latour, 1987; Suchman, 2002).

From an economic perspective, there is a growing body of literature suggesting that DDD generates business value. Davenport and Harris (2007), for instance, found a positive correlation between the adoption of analytics in organizations and their annual growth rates (based on a survey amongst 32 companies). A survey research study by (Brynjolfsson et al., 2011) supported this finding by showing that, amongst 179 surveyed companies, the adoption of DDD leads to an increase in firm productivity of 5-6 percent. Likewise, a recent study by (Müller, Fay and vom Brocke, 2018) examining more than 800 firms over a period of seven years showed that the use of big data and analytics is associated with an average increase in firm productivity of about 4 percent, with some industries reaching returns on BDA of more than 7 percent. A similar positive impact of DDD on firm productivity was reported by Wu and Hitt (2016), but they also found that the value of DDD is mainly in enabling continuous process improvements (exploitation) and not in sparking disruptive product or service innovation (exploration). These quantitative studies are further backed up by a large number of qualitative case studies that generally report positive findings about the relationship between DDD implementation and business value (e.g., Manyika et al., 2011; vom Brocke, Debortoli, Müller and Reuter, 2014; Sodenkamp, Kozlovskiy and Staake, 2015; Someh and Shanks, 2015; Côrte-Real, Oliveira and Ruivo, 2017).

To sum up, existing research strongly suggests that DDD generates business value. However, the current body of knowledge on DDD mainly focuses on descriptive and explanatory studies. What is lacking, so far, is prescriptive knowledge on how to design and implement DDD in complex organizational settings. Moreover, there is a lack of research that investigates the role of decision-making processes and human judgment on the outcome of DDD implementations (Sharma et al., 2014).

Against this background, this paper reports the results of an Action Design Research project that was aimed at designing and implementing IT artifacts for DDD at one of the largest ship engine manufacturers in the world. Besides presenting the design of the artifact itself, we formulate a set of nascent design principles for DDD, that can help other researchers and practitioners to implement DDD in comparable settings.

The remainder of this paper is structured as follows: We first provide a theoretical background on the challenges of implementing DDD in organizational settings. We then describe the action design research method in general, before we report on the process and outcome of applying it in our case. In the main part of the paper, we present four proposed design principles and explain their theoretical and empirical justification. The paper concludes with a short summary and outlook.

2 Challenges of Implementing DDD

To discuss common challenges of implementing DDD, we use the data-to-insight-to-decision-to-value conceptualization by (Sharma et al., 2014) as a framework. Even though Sharma et al. (2014) use it to elaborate on a research agenda for creating value from business analytics, we find it particularly suitable as a framework as, in distinction to more established DDD concepts (for instance Shearer, 2000), it acknowledges the importance of human judgment and decision-making processes to creating value with DDD. Moreover, it supports our diagnostic that prescriptive knowledge about how to implement DDD to create value is lacking and that without appropriately considering human judgment and decision-making processes already in the design of DDD artifacts, DDD cannot unfold its potential, or even fail in some cases.

Data to Insight

Nowadays, organizations have technologies at hand that enable them to collect, store, manage, analyze, and visualize large volumes of data of varying formats and at increased velocity (Müller et al., 2016; Watson, 2014). Nevertheless, as Sharma et al. (2014) point out, “despite the hopes of many, insights do not emerge automatically out of mechanically applying analytical tools to data. Rather, insights emerge out of an active process of engagement between analysts and business managers using the data and analytic tools to uncover new knowledge” (p. 435). One of the most common mistakes in generating new knowledge from data is to start with the wrong initial question, or not having a clear question at all (Leek and Peng, 2015). For example, inferential questions are often confused with causal questions, leading to confusion between spurious correlations and real cause-and-effect relationships. Or analysts may confuse exploratory questions with inferential questions, also called “data dredging”, or exploratory questions with predictive questions, leading to “overfitting”. A way to overcome such pitfalls is to compose multi-disciplinary data science teams that possess not only the required statistical and computational skills but also the necessary domain knowledge to formulate the right questions and draw valid conclusions from analysis results (Sharma et al. 2014).

Lycett (2013) emphasizes the involvement of human “sense-making” in the process of turning data to insights. Following Lycett (2013), designers of DDD solutions take important decisions regarding what data is selected and what inferences are drawn from the data. Moreover, as designers are human, they are also prone to human biases (Tversky and Kahneman, 1974), which affects the insights that are generated and the decisions taken based on them.

Insight to Decision

Research on judgment and decision-making provides strong empirical and theoretical arguments that favor algorithmic or statistical decision making over human judgments, particularly when it comes to complex decisions (Evans, 2006; Tversky and Kahneman, 1974). For example, a meta-analysis of 136 empirical studies that compared statistical predictions and human judgments in fields ranging from clinical decision-making to economics showed that statistical techniques lead on average to a 10 percent higher accuracy than human judgments (Grove et al., 2000). The superiority of statistical methods over human judgments holds for trained, untrained, experienced, and inexperienced judges (Grove and Meehl, 1996). Theoretical explanations for these findings include human biases (e.g., ignoring base rates, failure to take regression toward the mean into account, over-weighting individual factors) and judgment heuristics (e.g., representativeness, availability, and anchoring and adjustment) (Tversky and Kahneman, 1974).

Yet, despite the overwhelming evidence for the benefits of using insights generated from data to inform decision making, practice shows that new insights are not automatically translated into good decisions. Instead, the conversion of insights into decisions is influenced by a host of psychological and contextual factors (Sharma et al. 2014). For example, due to humans' limited information processing capacities, decision makers tend to satisfice, that is, select a course of action that will satisfy the minimum requirements needed to achieve a particular goal, but which is not necessarily the optimal alternative (Simon, 1956).

In addition, organizational decision-making processes and practices can have a strong influence on translating insights into decisions, such as management's inertia in moving towards a data-driven culture or a fragmented use of analytics in single departments instead of enterprise-wide adoption (SAS, 2012). Adding to this, survey results by LaValle et al. (2011) and Ransbotham, Kiron and Prentice (2015) suggest that a "lack of understanding of how to use analytics to improve the business" and "turning analytical insights into business actions" are among the top challenges hindering a successful implementation of DDD.

Decision to Value

As mentioned earlier, there exists a growing body of empirical evidence that the implementation of DDD leads to increased organizational performance. However, these benefits are not evenly distributed across all industries and business functions. Müller et al. (2018) showed, for example, that only companies in certain types of industries are able to extract measurable productivity improvements from the use of big data and analytics, and according to Wu and Hitt´s findings (2016), the value created by DDD is mainly exploitative and gained via process optimizations.

One obstacle for turning better decisions into higher value is the observation that it is by no means certain that effective decisions will also be successfully implemented (Sharma et al. 2014). Besides decision "quality" (effectiveness), another important criterion of good decisions is decision "acceptance", that is, the likelihood that stakeholders responsible for the successful implementation of the decision commit to it (Sharma et al. 2014). Prior research suggests, amongst others, that the level of stakeholders' participation in the decision-making process (Vroom and Yetton, 1973) and the comprehensibility of the underlying decision model (Kayande et al., 2009) are factors impacting on decision acceptance – both of which are often not always given in automated DDD processes.

Furthermore, Sharma et al. (2014) argue that even when self-optimizing machine learning algorithms are applied, the outcome of those algorithms still needs to be accepted by human decision makers regarding its validity and usefulness, for instance: "in 'deciding' to deploy them to run operations in an unguided manner, and in 'accepting' the refinements to the algorithms generated via machine learning as being valid" (p. 436).

3 Action Design Research

To develop design principles for how to design and implement DDD in complex organizational settings, we employed Action Design Research (ADR) as a research method. ADR is "a research method for generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting" (Sein et al., 2011, p. 40). The motivation for ADR is to better serve the "dual mission" of Information System Research, that is, to "make theoretical contributions and assist in solving the current and anticipated problems of practitioners" (Benbasat and Zmud, 1999; Iivari, 2003; Rosemann and Vessey, 2008 as referenced in Sein et al., 2011, p. 38). Compared to more traditional design science research methods (e.g., Hevner, March, Park and Ram, 2004; Peffers, Tuunanen, Rothenberger and Chatterjee, 2007), which are often conducted in the form of stage-gate processes leading to a disconnect between the development of artifacts and their actual application in organizational settings, ADR fully recognizes the role of organizational context in shaping the design process as well as the deployed artifact.

The actual process of ADR consists of four stages, which build on different principles and tasks. The first stage, "Problem Formulation", is based on two principles: "Practice-Inspired Research" and "Theory-Ingrained Artifact". The first principle emphasizes that problems from the field can be knowledge-creation opportunities. Following this, the researcher's intent should not only be to solve a specific instance of an encountered problem, as a software engineer or consultant might do, but to generate general prescriptive knowledge that can be applied to solve the class of problems that the specific problem instance exemplifies. The second principle of the first stage acknowledges that the design and evaluation of artifacts should be informed by existing theory, rather than solely driven by the designer's creativity. In particular, there are three ways of using prior theory in ADR: (1) to structure the problem (2) to identify solution possibilities (3) to guide the actual design. (Sein et al., 2011) This reflects the assumption behind ADR that 'the action design researcher actively inscribes theoretical elements in the ensemble artifact, thus manifesting the theory "in a socially recognizable form"'(Orlikowski and Iacono, 2001, p. 121 as cited in Sein et al., 2011). This, however, constitutes just the first stage of ADR: "[The artifact] is then subjected to organizational practice, providing the basis for cycles of intervention, evaluation, and further reshaping" (Sein et al., 2011, p. 41).

The second stage of ADR, "Building, Intervention, and Evaluation" (BIE), builds upon three principles: "Reciprocal Shaping", "Mutually Influential Roles", and "Authentic and Concurrent Evaluation". (Sein et al., 2011) Reciprocal shaping refers to the complex relations and mutual influences between the designed artifact and its organizational context. The researcher may, for example, use the artifact to gain a better understanding of the organizational environment and then use this increased understanding to refine the selection of design constructs. The principle of mutually influential roles emphasizes the need for mutual learning between the involved roles, being the researcher(s), practitioners, and end-users. These roles, however, can overlap. The principle of authentic and concurrent evaluation points to the key characteristic of ADR that building and evaluation are not conducted in separated stages, but are rather ongoing activities that also involve practitioners and end-users into the design process: "Consequently, authenticity is a more important ingredient for ADR than controlled settings" (Sein et al., 2011, p. 44).

In the third stage, "Reflection and Learning", the researcher moves from building a solution for an instance of a problem to applying that learning to a broader class of problems. The principle "Guided Emergence" describes that the artifact is not just a result of the initial theory-informed design (Stage 1), but of multiple cycles of complex and continuous shaping in the context of the organization (Stage 2), e.g., due to new upcoming requirements or refinements based on insights from user involvement and empirical evaluations. Those refinements to the initial design of the artifact "provide an opportunity for the ADR team to generate and evolve design principles throughout the process" (Sein et al., 2011, p. 44).

The fourth stage, "Formalization of Learning", is based on the principle of "Generalized Outcomes". Because of the described aspect of situated learning, including aspects of organizational change together with the actual implementation of an artifact, the generalization of ADR outcomes can be tricky. However, to address this issue, it is suggested to generalize the generated knowledge, this is possible on different levels: (1) generalization of the problem of an instance, (2) generalization of the solution instance, and (3) derivation of design principles from the design research outcomes. (Sein et al., 2011)

4 Data-Driven Lead Generation in the Maritime Industry

In the following sections, we report from our ADR project of data-driven lead generation in the maritime industry following (Sein et al., 2011) and their suggested ADR steps and principles.

4.1 Problem Formulation

4.1.1 Practice-Inspired Research

We got the opportunity to work with one of the biggest international engine manufacturers in the maritime industry. The particular department that we worked with is supporting the global aftersales business with data analytics, process, and project capabilities. The department´s technical core is a mature enterprise data warehouse that extracts, transforms, and loads data from multiple sources into a common format and location for analysis by enterprise users. Moreover, the department is responsible for several digitalization projects, amongst those, the implementation of a company-wide CRM system that enables the company to support and optimize sales processes, to store important customer data at one shared location, and finally to become more customer-centric (one face to the customer).

An interesting first diagnostic that informed our conceptualization of a research opportunity is that from the department´s comprehensive portfolio of analytical apps, the apps with the highest usage are those that support and improve an existing business process. In contrast, more explorative apps, which are not embedded in a current or new business process, are those with the lowest usage, even though in the long run they might be much more promising than others. On the one hand, this supports the finding of Wu and Hitt (2016) that the value generated from DDD is mostly exploitative, on the other hand, it shows a need for developing business processes around DDD solutions and, thus, to shift the focus in DDD away from the data-to-insight process alone to the holistic data-to-insight-to-decision-to-value process (Sharma et al., 2014) in order to increase user adoption and value creation of DDD artifacts.

Furthermore, we found that the company-wide CRM is perceived as a promising and necessary tool to make the company more customer-centric. However, many sales processes are still key-account-driven and not well aligned with the pro-active approach that the new CRM system supports. So, there is a situation in which the system is ready for pro-active sales processes, but the organization needs still time to adapt to this new pro-active approach, especially because the users are partly lacking business processes surrounding the new system and its affordances. Those diagnostics led us to formulate the field problems as follows:

- lack of business process embeddedness for low-usage DDD applications
- under-utilization of CRM system due to lacking pro-active business processes surrounding it

The resulting initial research opportunity and question was:

- How to enable pro-active CRM processes via DDD?

Following the suggestion from Sein et al. (2011, p. 40) to “generate knowledge that can be applied to the class of problems that the specific problem exemplifies”, we abstracted the formulated field problems to the class of DDD-value-creation-problems.

4.1.2 Theory-Ingrained Artifact

Sein et al. (2011, p. 41) suggest three ways of using theory in the initial design of an artifact: “to structure the problem (…), to identify solution possibilities (…), and to guide the design”. In accordance, we choose the conceptualization of (Sharma et al., 2014) as a structural framework for discussing and utilizing theory regarding challenges of implementing DDD into the solution (artifact). Moreover, Shearer´s (2000) cross-industry standard process for data mining was chosen for guiding the design of data science sub-artifacts. Furthermore, Dearden´s, (2001) conceptual information-decision-insights-supervision framework (IDA-S) was chosen as a design theory to guide partially automated characteristics of the artifact.

The organizational support for the project was secured by managing expectations and involving stakeholders such as the application manager of the CRM system and a business manager into the design process from the beginning on. Moreover, one of the researchers is working as an industrial PhD at the host company, which helped to anchor the project in the organization.

4.2 Building, Intervention, and Evaluation (BIE)

4.2.1 Reciprocal Shaping

The general solution understanding was informed by initial design principles of pro-activity, embeddedness, partial automation, and data-drivenness that were derived from the diagnosed field problems and selected theory. The main design objective was to design a DDD artifact that creates a new data-driven and pro-active lead generation process within the CRM system.

In the first iteration, we developed a concept to generate lead events based on predicting upcoming major overhaul events for engines using machine learning algorithms trained on transactional data of spare parts sales. However, we lacked historical data regarding major overhaul events. The reason for this is that in the maritime manufacturing industry, in general, large amounts of data are available, however, on a product or event level, correctly labeled transactional data can be sparse.

In the next instance, we found an alternative approach to generate leads from transactional data. In particular, we found that certain events in the life cycle of ships, such as changes in ownership or upcoming dry dockings of ships, constitute lead events. However, this knowledge is usually not available in digital form but gained through implicit and informal key-account management activities or other forms of direct customer contact, e.g., during a service visit. Yet, we were able to identify an external database of ship registrations, which could be repurposed to extract the required information about lifecycle events by applying several business rules to transform the data. After a successful proof of concept, we worked closely together with practitioners to develop the right business rules and integrate them into the production version of the department's data warehouse.

At this point, we were able to generate initial sales leads based on relevant events in the life-cycle of ships. However, in many cases, the event-driven approach generated simply too many leads to follow up on all of them. As a result, we proposed to prioritize and segment the customer base so that leads can be selected according to metrics of (future) customer behavior, such as their customer-lifetime value (CLV), purchasing patterns, and probabilities to churn in a given future period (see Fader, 2012). One of the theoretical ideas behind calculating CLV is "customer centricity", which suggests focusing efforts on the customers with the highest future CLV. The assumption is that it is more rewarding to focus on already strong customer relationships than to try to (re-)launch weak customer relationships (Fader, 2012). After exploring different modeling approaches on the transactional customer data at hand in combination with an extensive literature search, we decided on using so-called Bayesian Buy-Till-You-Die probability models for estimating the customer metrics of interest. Amongst the reasons for this choice was that the company operates in a non-contractual market setting, which means that it is not clearly observable when a customer relationship ends and the next transaction occurs (Fader and Hardie, 2009), in contrast to, e.g., cellphone subscriptions. Another reason was that hierarchical Bayesian probability models allow for estimating individual-level parameters (Abe, 2008; Rossi and Allenby, 2003) and can utilize cohort level information when individual-level data is lacking (Efron and Morris, 1977).

After developing a working prototype, we contacted one of the company's regional sales organizations to introduce the initiative and run a pilot of the developed method. The resulting meetings were very insightful especially regarding how to enrich the generated leads with further customer, ship, and engine information, so that they can be represented in the CRM system in a way that the sales responsible can directly take action to follow-up, without having to seek for information elsewhere. In particular, we attached a slide deck to the leads that explains the lead generation campaign, e.g., dry dockings or owner changes, in detail. Moreover, we attached excel reports with further customer and ship insights.

Eventually, we developed a method of a data-driven lead generation that contains five steps. **First**, by applying look-up algorithms to compare the current version of the external ship registration database with the version from the month before, we create a change log of ship information to identify changes in the ships´ life cycle stages and, in turn, generate initial leads. **Second**, as there can be situations in which there are too many leads in a given month, we calculate CLVs and other behavioral customer metrics for the customers of interest to, for instance, identify the leads for customers with the highest future CLV, or leads for customers that are at risk to churn. **Third**, we enrich the leads with further customer and ship information from the company's data warehouse. **Fourth**, we use a lead uploading template to create and assign the generated leads directly in the CRM system. **Fifth**, we evaluate the performance of the generated data-driven leads via feedback meetings with the sales organizations and via quantitative analysis of CRM data to learn about and improve the quality of the generated leads. The first four steps of the method can be fully automated and implemented into extract transform and load processes (ETL).

4.2.2 Mutually Influential Roles

We conducted the BIE cycles following an IT-dominant schema (Sein et al., 2011) in which we were the researchers but also the leading designers and engineers of the artifact. Therefore, we were responsible for the formulation and technical implementation of design principles to ultimately create user-utility via an artifact for data-driven decision support. In this process, we were supported by a design team that consisted of a senior data warehouse engineer and student workers from the aftersales data analytics department. In addition, a wider group of business professionals and test-users from the company was supporting the team with valuable domain knowledge throughout the entire design process.

4.2.3 Authentic and Concurrent Evaluation

After the different design instances, the artifact was evaluated with regard to changes to the problem understanding, design principles, the need for further design cycles, and organizational effects. So far, 288 data-driven leads were created in the CRM system from which 73% have been worked on. We also got very positive feedback from the application manager of the CRM system, as one of the key stakeholders:

> "It's very interesting to see what scientific theories applied on our data sources can be used for. It has been important for us to include some of the receivers/end-users of the data-driven leads in the process to make it tangible for them and gain from their real-life expertise and not end up with a bunch of leads that only looked promising on paper. Having their stamp of approval is the first step towards a more pro-active sales process and thereby creating additional value. The data-driven leads will be an addition to their work and will save them some time when looking for new leads in the market, these leads come out of the box, being our CRM system."

Also the research question could be addressed with designing and implementing an working DDD artifact that creates pro-active business process by enabling sales responsible to take action without a prior customer inquiry: "We have to search for leads wherever we can, and using the data sources available is a natural next step in a more pro-active sales approach. It's important that we setup an automated process around it and analyze on the outcome of the data-driven leads, to optimize the process over time." (Application Manager CRM System)

5 Reflection, Learning, and Formalization of Design Principles

After multiple cycles of building, intervention, and evaluation, we successfully designed and implemented a DDD artifact for data-driven lead generation. This artifact can be used as a tool to develop similar artifacts in many different DDD context. Thus, we abstracted the solution artifact from the maritime industry to the higher class of DDD solutions.

Moreover, we reflected on the changes in problem and solution understanding as well as on design decisions taken and the feedback received from the practitioners. The aim of this phase was to abstract from the specific problems and solutions encountered in the case in order to generate more generic prescriptive knowledge about the design and implementation of DDD. We formulated this knowledge in the form of design principles, following the template proposed by Chandra et al. (2015).

DP 1: Given a lack of proof-of-concept, use theory-based models instead of data-driven machine learning algorithms in order to achieve concrete results.

DP 1 is based on the initial design principle of data-drivenness that was derived from the problem formulation stage. The principle was further shaped throughout the different BIE iterations towards its current formulation. A major design problem that arrived was the choice of a DDD modeling approach that could utilize transactional data for the data-driven lead generation artifact.

Broadly speaking, there are two cultures of using statistical models to gain insights from data (Breiman, 2001). The first tries to reconstruct and model the “true” relationships between data inputs and outputs in the form of some mathematical function. Typically, these input-output relations are deductively derived from extant theory, attempt to represent cause-and-effect relationships, and should be interpretable for humans. The second culture treats the process that has generated the data at hand as a complex and unknown black box. Instead of trying to discover the true inner workings of this black box, researchers simply build an algorithm that is able to predict the process’ output, given its inputs. The resulting model emerges in a purely inductive fashion, is often based on correlations instead of causation, and is typically incomprehensible to humans.

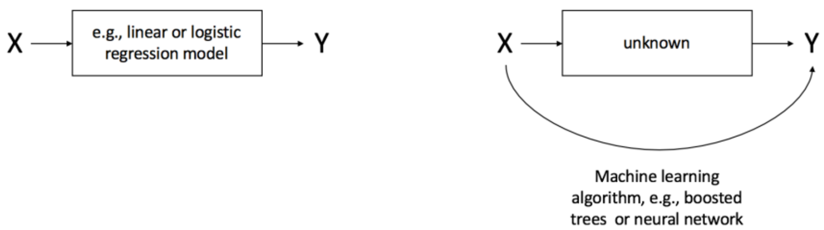

Figure 1: Two cultures of using statistical models

While traditionally the majority of researchers and practitioners followed the theory-based modeling culture (Breiman, 2001; Shmueli, 2010; Shmueli and Koppius, 2011), the data-driven algorithmic culture became more and more popular with the emergence of big data and the increasing adoption of machine learning in practice. Over the last years, black-box machine learning algorithms such as gradient boosted trees or neural networks have proved their usefulness when working with large and high-dimensional datasets and outperformed more traditional methods like linear or logistic regression in many of the recent classification and regression competitions. However, according to the no-free-lunch theorem (Wolpert and Macready, 1997), there is not one algorithm that fits all problems.

As described earlier, following the trend towards prediction with data-driven machine learning algorithms, we started the project with collecting a dataset comprising engine maintenance events and variables that are potentially correlated with this event. The goal was to create an algorithm that is able to predict maintenance events based on early signals like quotes, orders, or runtimes of engine parts as well as basic customer characteristics (e.g., industry, size). Customers for which the algorithm predicts high probabilities for an upcoming maintenance event are classified as leads and would be assigned to a sales representative for follow up.

However, we soon realized that there were not enough historical observations available in order to train an algorithmic model to the high-dimensional dataset, resulting in overfitting of the model and poor predictive accuracy on test data. Machine learning algorithms have been very useful in the last years for predicting customer behavior in B2C industries characterized by a high volume of transactions (e.g., retail, telecommunications, e-commerce). In contrast, our project is situated in a B2B industry with extremely durable products and a relatively small customer base. Moreover, the transactional data is used as secondary data, only repurposed for analysis. Eventually, the complexity of using machine learning algorithms was too high, as we lacked enough observations of correctly labeled occurrences of major overhauls. After having invested a lot of work and time in this first approach, we learned that it would have been better to have started with a less innovative, but more established and theory-based approach to generating insights from transactional data. This way, stakeholder engagement can be secured by presenting concrete results already at the beginning of a DDD project (quick-wins).

In the following, we decided to utilize an external database of ship registration to create leads based on changes in the life-cycle of ships. To prioritize and segment the leads, we were looking again for a suitable DDD modeling approach. Based on the learnings from our first DDD modeling iteration and informed by the theorem of Occam´s razor (Blumer et al., 1987), we searched for a DDD approach that constitutes a good trade-off between predictive accuracy and implementation complexity. We then turned to the marketing literature to search for alternative approaches for predicting customers' future purchasing behavior based on customer transaction data. Buy-Till-You-Die models (BTYD; e.g., Schmittlein, Morrison and Colombo, 1987), an example of theory-based statistical models, and especially those using hierarchical Bayesian models (Abe, 2008; Ma and Liu, 2007; Platzer and Reutterer, 2016), seemed to be particularly suited for our context, because they have been developed for predicting non-contractual customer purchasing (like in our setting), allow individual level parameter estimations, and require surprisingly simple data to be estimated. Only three variables are required for each customer: how many transactions a customer has made in the past (frequency), the date of the transaction (recency), and the monetary value of these transactions. Moreover, due to the possibility of using informative priors, and the utilization of cohort-level information when individual-level data is sparse, hierarchical Bayesian models do not necessarily require big amounts of data to produce good predictive performance (Efron and Morris, 1977; van de Schoot et al., 2015). In addition, BTYD models are based on sound behavioral theory, which enables them to provide useful managerial diagnostics, and have shown excellent empirical performance in the past (Fader et al., 2005).

Eventually, we got the best results with the Pareto/GGG model (Platzer and Reutterer, 2016). The dataset was an aggregated version of approximately 500,000 aftersales transactions. To benchmark the model, we predicted the number of future customer transactions one year ahead. Overall, with a mean absolute error (MAE) of 1.2, we got satisfying results. Especially when predicting future transaction for the whole customer cohort, the accuracy was with 93% very good (see Table 1; frequency as target variable had a minimum value of 0.0, a mean value of 1.6 and a maximum value of 93.0 in the validation dataset).

Model	**Actuals / Prediction**	**MAE**
Pareto/GGG	93%	1.2
Pareto/NBD (HB)	78%	1.4

Table 1: Predictive Performance of Pareto/GGG compared to Pareto/NBD (HB)

DP 2: Limit the complexity of models in order to gain support by managers.

Besides predictive accuracy and implementation complexity, comprehensibility is another important feature of any decision support system (DSS), as it increases the trust users put in the outputs of the system and, thereby, drives user acceptance of the system itself (Gregor and Benbasat, 1999). Kayande´s et al. (2009) 3-Gaps framework conceptualizes this idea in more detail. It proposes that the sizes of the gaps between the model implemented in the DSS, reality, and managers' mental models influence the performance of the DSS, it's acceptance by managers, and managers' performance.

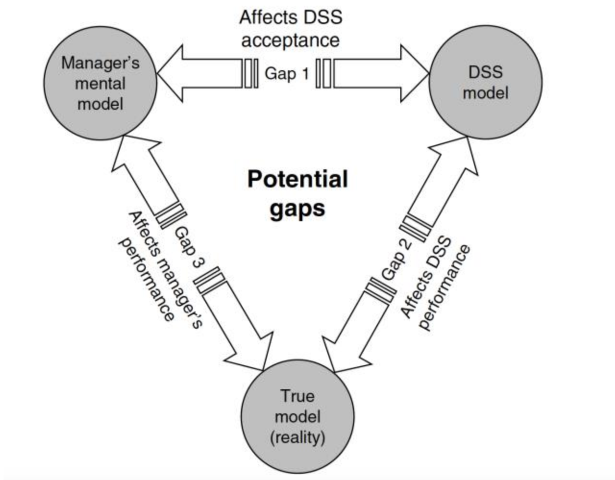

Figure 2: 3-Gaps Framework

As discussed in DP 1, to provide high predictive accuracy the DSS model must match the true but unknown process that generated the underlying data as good as possible (Gap 2). Likewise, a manager's mental model should be as close as possible to the true model (Gap 3) in order for the manager to make correct decisions (irrespective of using a DSS or not). Our main interest is in Gap 1. When the manager using a DSS does not understand the logic of the system, the gap between DSS model and the user's mental model is large. Consequently, the predictions provided by the system and the manager's experience and intuition are likely to be in conflict. In such situations, risk-averse decision makers tend to rely on their "gut feeling" instead on the DSS, even if following the advice of the DSS would objectively increase decision quality (Kayande et al., 2009).

The experiences we made over the course of the project support the above outlined theoretical arguments. As described earlier, we started out with using machine learning to classify leads for after-sales service from data about customers' spare parts and consumption profiles. This approach involved black-box algorithms like boosted decision trees and high-dimensional datasets. Although all project members and stakeholders had a strong analytical background, it was difficult for many to comprehend both the inner workings of the algorithms and the meaningfulness of the processed data. Moreover, when working with practitioners from different business domains such as engineering, the mentioned issues became more apparent. For instance, we needed to get detailed technical information regarding labeling major overhaul events from the past, to be able to predict future occurrences of such events. However, as the practitioners were not familiar with machine learning concepts and especially how the many input variables relate to the admired outcome, it was difficult for us to communicate and for them to comprehend what the model was supposed to do.

In contrast, the Buy-Till-You-Die models that we used later in the project were much better received. Although the mathematics behind these models are also complex and were unfamiliar to most project members, they are easier to conceptually understand as they require only three pieces of input information about each customer: their recency (i.e., the time of the last transaction) and frequency (i.e., the number of transactions made in a specified time period), plus the monetary value of the transaction for calculating CLVs. Nevertheless, the limited complexity of the input data creates the need for strong assumptions (e.g., purchase process follows a Poisson process). However, the format of an event log of historical customer purchase transactions and the marketing theory informed assumptions of, e.g., the purchasing and churn process, were very much in line with the experience and intuition of the involved domain experts, minimizing the gap between the DSS model and managers' mental models.

DP 3: Incorporate domain knowledge into the data-driven decision-making process in order to foster acceptance by managers.

Statistical models, like the Buy-Till-You-Die models we used, induce predictive models of customer purchasing behavior from historical data about past transactions. Apart from this source of information, there, of course, exist human experts who have developed expertise through years of experience in marketing and selling services to customers. In contrast to the statistical models, their domain knowledge tends to be implicit and heuristic in nature, for example, in the form of best practices or rules of thumbs. This knowledge, although it might be difficult to formalize, can still hold valuable information for predicting future customer behavior.

A small but growing stream of research is investigating how human domain knowledge and data-driven predictive models can be combined in order to construct better decision support systems (see, e.g., Dybowski, Laskey, Myers and Parsons, 2003; Sinha and Zhao, 2008 for an overview). Sinha and Zhao (2008), for instance, systematically compared the performance of data mining algorithms for credit risk scoring with and without incorporating experts' domain knowledge in the form of rules of thumbs and found that considering domain knowledge significantly improves predictive accuracy. Other researchers improved decision quality by using Bayesian approaches to incorporate prior believes derived from expert judgments into the model estimation process (e.g., Druzdzel and Díez, 2003; Langseth and Nielsen, 2003).

In our project, we integrated domain knowledge in the form of simple rules capturing experts' experience and intuition into the data-driven decision-making process. More concretely, we interviewed industry experts from the case organization to elicit what types of events at the customer side may lead to a demand for spare parts or maintenance service. We learned, for example, that events such as an upcoming dry-dockings or change in the ownership of a ship increase the likelihood that the owner will order specific spare parts or services. Moreover, we involved regional sales organizations into the development of the method, so that they could tell us how a lead needs to be represented in the CRM system and what additional customer and ship information is required to take immediate action. This way, eliciting and incorporating experts' knowledge into the artifact, e.g., from the decision makers that the method is targeted towards, also increased their level of participation in and influence on the final design of the DDD process, a key success factor for increasing the acceptance of the final artifact (Vroom and Yetton, 1973).

DP 4: Provide actionable insights instead of quantitative reports in order to increase use by decision makers.

DP 4 is based on the initial theory informed and practice inspired design principles of pro-activity, embeddedness, and partial automation.

Even if a DDD system produces decisions of high accuracy and acceptance, it is not given that end users will follow those decision proposals and take action. In their survey study on "Big Data, Analytics and the Path From Insight to Value," LaValle et al. (2011) highlight that many organizations fail to translate insights into actions because their analytics is too much focused on describing past and current situations and fails to provide actionable and prescriptive information. The authors recommend embedding analytics into operational business processes and users' daily workflows, instead of isolating it in standardized reports that are not accessed on a regular basis. Such a strategy “makes it harder for decision makers to avoid using analytics - which is usually a good thing” (Davenport, 2013).

As mentioned earlier, the practitioners of the department made the general observation that analytical applications without a business process surrounding it are used less often, and also that the CRM system was still lacking business processes to further foster usage and value creation. Based on those observations, we decided to push the leads generated by our DDD method directly into sales representatives’ daily newsfeed inside the CRM system, instead of building extra reports or dashboards that have to be pulled by them. The processes were designed so that every lead is created as a separate item accompanied with additional information regarding what to do in the form of a clear naming and description text, as well as via an attached slide presentation regarding the specific campaign. Moreover, based on the meetings with the regional sales organizations, we decided to enrich the leads with further ship and customer transaction information, so that the sales responsibles have all the information that they need for their regular lead follow-up at their fingertips.

The above-described design decisions were based upon the distinction between descriptive (“What has happened in the past”), predictive (“What will happen in the future?”), and prescriptive analytics (“How can we make it happen?”) (Watson, 2014). By enriching leads identified from ship life cycle events with predictions about future purchasing behavior and descriptive information about campaigns, customers, and ships, we generate prescriptive knowledge that sales responsible can translate into actions.

6 Conclusion

Existing research on DDD provides compelling arguments for its value, both on the level of individual decision makers (Grove et al., 2000) and on an organizational level (Brynjolfsson et al., 2011; Müller et al., 2018; Wu and Hitt, 2016). Yet, despite recent calls for research, there is a lack of research on how organizational decision-making processes and human judgment shape DDD and on how to implement DDD in complex organizational settings (Sharma et al. 2014). Hence, the goal of this ADR study was to develop practice-inspired, theory-grounded, and field-tested design principles for implementing DDD in the maritime industry, which can help other researchers and practitioners to implement DDD in comparable settings. Besides providing high decision quality, the formulated design principles acknowledge that systems supporting DDD need to be accepted by the involved stakeholders. Hence, our design principles highlight the importance of model comprehensibility, domain knowledge, and actionability of results. Although the proposed principles are inspired by diagnosed problems and grounded in theory and empirical data, due to the situated nature of ADR, we cannot claim that our list of design principles is complete or optimal. Nonetheless, we firmly believe that they represent a valid starting point and can provide the foundations for further research on how to design and implement DDD in complex organizational settings.

Next to the presented design principles, we contribute by abstracting the artifact from a specific data-driven lead generation instance to a tool for generating data-driven leads in many different contexts, thus, we abstract from a specific solution instance to the broader class of DDD solutions.

In future research, we attemp to further deepen the analysis of the impact that our designed artifact has on the process from data-to-value. Moreover, we attemp to further shape the designed artifact towards a generalizable tool for creating value with data-driven lead generation.

References

Abe, M., 2008. “Counting Your Customers” One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model. Marketing Science 28, 541–553. https://doi.org/10.1287/mksc.1090.0502

Arnott, D., Pervan, G., 2008. Eight key issues for the decision support systems discipline. Decision Support Systems 44, 657–672.

Benbasat, I., Zmud, R.W., 1999. Empirical research in information systems: the practice of relevance. MIS quarterly 3–16.

Bishop, C.M., 2006. Pattern recognition and machine learning. springer.

Blumer, A., Ehrenfeucht, A., Haussler, D., Warmuth, M.K., 1987. Occam’s razor. Information processing letters 24, 377–380.

Boyd, D., Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society 15, 662–679.

Breiman, L., 2001. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statist. Sci. 16, 199–231. https://doi.org/10.1214/ss/1009213726

Brynjolfsson, E., Hitt, L.M., Kim, H.H., 2011. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? (SSRN Scholarly Paper No. ID 1819486). Social Science Research Network, Rochester, NY.

Chandra, L., Seidel, S., Gregor, S., 2015. Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions, in: System Sciences (HICSS), 2015 48th Hawaii International Conference On. IEEE, pp. 4039–4048.

Chen, H., Chiang, R.H., Storey, V.C., 2012. Business intelligence and analytics: From big data to big impact. MIS quarterly 36.

Côrte-Real, N., Oliveira, T., Ruivo, P., 2017. Assessing business value of Big Data Analytics in European firms. Journal of Business Research 70, 379–390.

Davenport, T.H., 2013. Analytics 3.0 [WWW Document]. Harvard Business Review. URL https://hbr.org/2013/12/analytics-30 (accessed 11.24.17).

Davenport, T.H., Harris, J.G., 2007. Competing on analytics: The new science of winning. Harvard Business Press.

Dearden, A., 2001. IDA-S: A Conceptual Framework for Partial Automation, in: People and Computers XV—Interaction without Frontiers. Springer, London, pp. 213–228. https://doi.org/10.1007/978-1-4471-0353-0_13

Druzdzel, M.J., Díez, F.J., 2003. Combining Knowledge from Different Sources in Causal Probabilistic Models. Journal of Machine Learning Research 4, 295–316.

Dybowski, R., Laskey, K.B., Myers, J.W., Parsons, S., 2003. Introduction to the Special Issue on the Fusion of Domain Knowledge with Data for Decision Support. J. Mach. Learn. Res. 4, 293–294.

Efron, B., Morris, C., 1977. Stein’s paradox in statistics. Scientific American 236, 119–127.

Evans, J.S.B.T., 2006. The heuristic-analytic theory of reasoning: Extension and evaluation. Psychonomic Bulletin & Review 13, 378–395. https://doi.org/10.3758/BF03193858

Fader, P., 2012. Customer centricity: focus on the right customers for strategic advantage.

Fader, P.S., Hardie, B.G.S., 2009. Probability Models for Customer-Base Analysis. Journal of Interactive Marketing, Anniversary Issue 23, 61–69. https://doi.org/10.1016/j.intmar.2008.11.003

Fader, P.S., Hardie, B.G.S., Lee, K.L., 2005. "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model. Marketing Science 24, 275–284. https://doi.org/10.1287/mksc.1040.0098

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. AI magazine 17, 37.

Gregor, S., Benbasat, I., 1999. Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. MIS Quarterly 23, 497–530. https://doi.org/10.2307/249487

Grove, W.M., Meehl, P.E., 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. American Psychological Association.

Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., Nelson, C., 2000. Clinical versus mechanical prediction: a meta-analysis. Psychological assessment 12, 19.

Hevner, A., March, S.T., Park, J., Ram, S., 2004. Design science in information systems research. MIS quarterly 28, 75–105.

Iivari, J., 2003. The IS core-VII: Towards information systems as a science of meta-artifacts. Communications of the Association for Information Systems 12, 37.

Kahneman, D., 2003. Maps of bounded rationality: Psychology for behavioral economics. The American economic review 93, 1449–1475.

Kayande, U., De Bruyn, A., Lilien, G.L., Rangaswamy, A., Van Bruggen, G.H., 2009. How incorporating feedback mechanisms in a DSS affects DSS evaluations. Information Systems Research 20, 527–546.

Langseth, H., Nielsen, T.D., 2003. Fusion of Domain Knowledge with Data for Structural Learning in Object Oriented Domains. Journal of Machine Learning Research 4, 339–368.

Latour, B., 1987. Science in action: How to follow scientists and engineers through society. Harvard university press.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. MIT sloan management review 52, 21.

Leek, J.T., Peng, R.D., 2015. What is the question? Science 347, 1314–1315.

Lycett, M., 2013. 'Datafication': Making sense of (big) data in a complex world. Springer.

Ma, S.-H., Liu, J.-L., 2007. The MCMC approach for solving the Pareto/NBD model and possible extensions, in: Natural Computation, 2007. ICNC 2007. Third International Conference On. IEEE, pp. 505–512.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A., 2011. Big data: The next frontier for innovation, competition, and productivity.

Müller, O., Fay, M., vom Brocke, J., 2018. The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics.. Forthcoming. Journal of Management Information Systems.

Müller, O., Junglas, I., Brocke, J. vom, Debortoli, S., 2016. Utilizing big data analytics for information systems research: challenges, promises and guidelines. Eur J Inf Syst 25, 289–302. https://doi.org/10.1057/ejis.2016.2

Orlikowski, W.J., Iacono, C.S., 2001. Research commentary: Desperately seeking the “IT” in IT research—A call to theorizing the IT artifact. Information systems research 12, 121–134.

Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S., 2007. A design science research methodology for information systems research. Journal of management information systems 24, 45–77.

Peter E. Rossi, author, Greg M. Allenby, author, 2003. Bayesian Statistics and Marketing. Marketing Science 304.

Platzer, M., Reutterer, T., 2016. Ticking away the moments: Timing regularity helps to better predict customer activity. Marketing Science 35, 779–799.

Provost, F., Fawcett, T., 2013. Data science and its relationship to big data and data-driven decision making. Big Data 1, 51–59.

Ransbotham, S., Kiron, D., Prentice, P.K., 2015. The talent dividend. MIT Sloan Management Review 56, 1.

Rosemann, M., Vessey, I., 2008. Toward improving the relevance of information systems research to practice: the role of applicability checks. Mis Quarterly 1–22.

Samuel, A.L., 1959. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development 3, 210–229. https://doi.org/10.1147/rd.33.0210

SAS, 2012. The Evolution of Decision Making: How Leading Organizations Are Developing a Data-Driven Culture - SPONSOR CONTENT FROM SAS.

Schmittlein, D.C., Morrison, D.G., Colombo, R., 1987. Counting Your Customers: Who-Are They and What Will They Do Next? Management science 33, 1–24.

Sein, M., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R., 2011. Action Design Research. MIS Quarterly 35, 37–56.

Sharma, R., Mithas, S., Kankanhalli, A., 2014a. Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. European Journal of Information Systems 23, 433–441. https://doi.org/10.1057/ejis.2014.17

Sharma, R., Mithas, S., Kankanhalli, A., 2014b. Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. European Journal of Information Systems 23, 433–441.

Shearer, C., 2000. The CRISP-DM model: the new blueprint for data mining. Journal of data warehousing 5, 13–22.

Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., Carlsson, C., 2002. Past, present, and future of decision support technology. Decision support systems 33, 111–126.

Shmueli, G., 2010. To Explain or to Predict? Statist. Sci. 25, 289–310. https://doi.org/10.1214/10-STS330

Shmueli, G., Koppius, O.R., 2011. Predictive Analytics in Information Systems Research. MIS Quarterly 35, 553–572. https://doi.org/10.2307/23042796

Simon, H.A., 1956. Rational choice and the structure of the environment. Psychological review 63, 129.

Sinha, A.P., Zhao, H., 2008. Incorporating domain knowledge into data mining classifiers: An application in indirect lending. Decision Support Systems 46, 287–299. https://doi.org/10.1016/j.dss.2008.06.013

Sodenkamp, M., Kozlovskiy, I., Staake, T., 2015. Gaining IS Business Value through Big Data Analytics: A Case Study of the Energy Sector. ICIS 2015 Proceedings.

Someh, I.A., Shanks, G.G., 2015. How Business Analytics Systems Provide Benefits and Contribute to Firm Performance?, in: ECIS.

Suchman, L., 2002. Located accountabilities in technology production. Scandinavian journal of information systems 14, 7.

Thaler, R., 1980. Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization 1, 39–60.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty 5, 297–323.

Tversky, A., Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases. Science 185, 1124–1131. https://doi.org/10.1126/science.185.4157.1124

van de Schoot, R., Broere, J.J., Perryck, K.H., Zondervan-Zwijnenburg, M., van Loey, N.E., 2015. Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. Eur J Psychotraumatol 6. https://doi.org/10.3402/ejpt.v6.25216

vom Brocke, J., Debortoli, S., Müller, O., Reuter, N., 2014. How In-memory Technology Can Create Business Value: Insights from the Hilti Case. CAIS 34, 7.

Vroom, V.H., Yetton, P.W., 1973. Leadership and decision-making. University of Pittsburgh Pre.

Watson, H., 2014. Tutorial: Big Data Analytics: Concepts, Technologies, and Applications. Communications of the Association for Information Systems 34.

Winner, L., 1980. Do artifacts have politics? Daedalus 121–136.

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE transactions on evolutionary computation 1, 67–82.

Wu, L., Hitt, L.M., 2016. How Do Data Skills Affect Firm Productivity: Evidence from Process-driven vs. Innovation-driven Practices.