Designing a collaboration process for big data analytics projects

Janine Adina Hagen
LMU Munich, hagen@bwl.lmu.de

Follow this and additional works at: https://aisel.aisnet.org/ecis2021_rip

Recommended Citation
https://aisel.aisnet.org/ecis2021_rip/5

This material is brought to you by the ECIS 2021 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2021 Research-in-Progress Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
DESIGNING A COLLABORATION PROCESS FOR BIG DATA ANALYTICS PROJECTS

Research in Progress

Janine Adina Hagen, LMU Munich, Munich, Germany, hagen@bwl.lmu.de

Abstract
Although there is broad agreement that big data analytics (BDA) creates business value, knowledge about how organizations should proceed to extract value remains limited. Scholars have found that utilizing BDA insights in business greatly contributes to BDA value, but existing BDA processes tend to focus more on data science activities and do not adequately outline collaboration activities with business users. However, this relationship seems troubled and may even impede BDA success. To alleviate this discrepancy, we conduct an action research project with a tool manufacturer. The goal is to design a collaboration process for BDA. We apply social capital theory to evaluate the process, as social capital can improve collaboration. This paper gives the first glimpse of the collaboration process which intends to contribute to literature and practice by outlining critical collaboration activities that improve the relationship between business and data science and increase BDA use in business.

Keywords: Big data analytics, collaboration process, relationship, social capital theory.

1 Introduction

Big data analytics (BDA) is increasingly transforming business and society by enhancing agility, innovation, and competitive performance (Loebbecke and Picot, 2015; Mikalef et al., 2020). However, companies seem to be struggling to leverage the potential of BDA. Several surveys report that most BDA projects fail to deliver business value. Gartner, for example, found a success rate of only 15% for BDA projects and forecasts that only 20% of analytics insights will deliver business outcomes through 2022 (Asay, 2017; White, 2019). In fact, the understanding of how organizations should proceed to extract value from BDA remains limited (Côrte-Real et al., 2019). What is known is that the utilization of BDA insights in business is the most critical contributor to BDA value (Côrte-Real et al., 2019; Kamioka and Tapanainen, 2014). BDA utilization means the degree to which BDA applications are used to support business activities such as operations, product development, marketing, and sales, as well as customer and supplier relations (Côrte-Real et al., 2019; Tallon and Kraemer, 2007). This indicates that it is not only the much-discussed data scientist (e.g., Davenport and Patil, 2012) who is important for extracting value from big data, but likewise, the business user who has to leverage the BDA solutions, for example for process improvement, product innovation, and customer experience enhancement (Grover et al., 2018; Gupta and George, 2016).

Information systems (IS) scholars have developed several process models that outline the activities needed to extract value from big data (e.g., Abbasi et al., 2016; Jagadish, 2015; Philipps-Wren et al., 2015). Despite several minor differences, these process models all adhere to the overall BDA value path – from information over insight to decision and finally action (Sivarajah et al., 2017). However, they tend to focus on the analytical and technical activities associated with BDA (e.g., data preparation and analysis) and append the business perspective as the final step in the process, called, for example, “data interpretation” or “data usage” (Jagadish, 2015, p. 3; Philipps-Wren et al., 2015, p. 23). Thus, business users serve as “recipients” of the BDA process results but are not assigned to actively
contribute to the full process within the scope of distinct collaboration activities with data science professionals. Although the need for collaboration and a good working relationship between data experts and business users is underlined (Abbasi et al., 2016; Jagadish, 2015; Gupta and George, 2016), the literature does not provide concrete solutions of what this can look like in terms of distinctive collaboration activities. However, these collaboration activities are an important component when designing for collaboration (Briggs et al., 2009). Moreover, they take on importance against the background that the relationship between data science and business professionals appears troubled due to a gap in the representation and interpretation of BDA applications in business context, and the absence of network ties, which has the potential to imperil collaboration (Hagen and Hess 2021).

Eventually, we find a discrepancy in the IS literature between the high importance of BDA utilization in business units and the insufficient consideration of business users and their collaboration with data science professionals in BDA process depictions. To improve collaboration and increase BDA use in business, we argue that it is necessary to refine our BDA process understandings. Against this backdrop, we pose the following research question to guide our research:

**How can a collaboration process between data science experts and business users be designed for joint BDA projects?**

To answer this question, we develop a collaboration process together with an international tool manufacturing and trading company that likewise considers the activities of data science and business professionals, and their collaboration. The tool company has faced several collaboration challenges between data science and business professionals on prior BDA projects (e.g., a lack of mutual understanding) and aims to develop a collaboration process blueprint to improve the groups’ relationship. A supervised machine learning project serves as an environment to develop the process. To evaluate the quality of the process and its potential to improve the relationship, we apply social capital theory (Nahapiet and Ghoshal, 1998), as the presence of social capital has been shown to improve inter-team collaboration. Thus, social capital theory offers a suitable framework to guide our study. As we intend to solve a practical organizational problem through intervention while simultaneously contributing to IS knowledge, we conduct a canonical action research project (Davison et al., 2004). Eventually, by designing a collaboration process for BDA, we aim to offer a collaboration-oriented vehicle that yields a satisfactory working relationship between data science and business professionals and increases business utilization by appropriately incorporating the business perspective and collaboration activities into the process.

This research in progress contains preliminary results, namely an initial version of the high-level BDA collaboration process and an exemplary drill-down of the collaboration activities in the model development phase. Moreover, we introduce social capital theory and our approach to evaluating the final process design, which will be iteratively developed and improved in the future course of this research.

## 2 Related Work and Theoretical Background

As for the conceptual background, we first summarize insights from related work on BDA process frameworks and derive the research gap we want to address. Then, we present social capital theory and how we intend to apply it to evaluate the potential for our process to improve the working relationship between data science and business professionals.

### 2.1 Big Data Analytics Process and Collaboration

The IS literature provides various definitions for big data analytics (BDA) ranging from a sole focus on the specifics of the data itself to necessary tools and processes up to the associated business value (Mikalef et al. 2018). As our study focuses on the procedural perspective of BDA to increase business value and the necessary collaboration activities, we follow Ghasemaghaei et al. (2015) in understanding BDA as “tools and processes often applied to large and disperse datasets for obtaining meaningful insights.” In essence, BDA comprises the application of analytical skills to analyze the
data, as well as functional skills to deduce business-relevant insights (Russom, 2011). This requires a good working relationship between data science experts and business users and collaboration among them (Gupta and George, 2016). In general, we understand collaboration as “the process through which a specific outcome, such as a product or desired performance, is achieved through group effort” (Kotlarsky and Oshri, 2005, p. 40). Specifically, we define collaboration for BDA as the process through which business value based on BDA is achieved through a joint effort between data science experts (e.g., data scientists, data engineers) and business users (e.g., functional business professionals from the marketing or supply chain department).

Existing BDA processes in the IS field offer an initial understanding of the major activities and the stakeholders involved. First, Jagadish et al. (2014) and Jagadish (2015) point out that the big data lifecycle is more than an analysis of big data and comprises the following steps: data acquisition, information extracting and cleaning, data integration/aggregation/representation, modeling and analysis, and interpretation. The authors state that human collaboration is a BDA characteristic that makes these steps challenging, but they fail to provide concrete recommendations on how to shape this collaboration. Moreover, the focus is on analytical activities associated with BDA, and business activities are only considered in the final step. Second, Philipp-Wren et al. (2015) suggest in their BDA framework that the activities needed to proceed with data sources are: data preparation, data storage, analysis, and data access and usage. Their proposal involves data science experts in the first three phases (preparation, storage, analysis) and business users only in the final stage (usage). Third, according to Abbasi et al. (2016), the big data information value chain comprises data, information, knowledge, decisions, and actions. Again, data experts are assigned to the first part of the value chain (data, information), and managers to the second part (decisions, actions). As for the knowledge phase, the authors suggest both parties be involved.

In summary, we find that, although business utilization has been identified as the most critical contributor to unlocking the value of BDA (Córte-Real et al., 2019), IS literature does not provide specific solutions on how business utilization can benefit from actively involving business users in the entire process. BDA frameworks tend to remain at an aggregated level regarding the activity description and take primarily an analytical/technical point of view. It seems the responsibility for the majority of the BDA process is assigned to data science experts, whereas business users are primarily involved at the end. They appear to be the “recipients” of what the process yields without having a real opportunity to integrate their perspective and requirements right from the beginning. Although the need for collaboration and a good working relationship between data experts and business users is underlined (Abbasi et al., 2016; Gupta and George, 2016; Jagadish, 2015), no concrete solutions are provided for how this can look like in terms of specific collaboration activities. However, this collaboration may require special attention, as the relationship between the groups seems troubled (Hagen and Hess 2021). Our research aims to reduce this discrepancy between the high importance of BDA utilization in business and the poor consideration of business users in the BDA process by developing a collaboration process for BDA that incorporates all the relevant interactions from both data science experts and business users.

### 2.2 Social Capital Theory

Scholars have found that the higher the social capital within a group (i.e., the better the relationships between the group members), the better its performance (Aquino and Serva, 2005). This tends to happen because the presence of social capital can reduce transaction costs, enhance mutual commitment, and facilitate collaboration (van den Hooff and de Winter, 2011). As the presence of social capital has a positive impact on collaboration, this theory is well suited to evaluate the quality of our process to facilitate collaboration for BDA. Specifically, we aim to measure how the social capital develops throughout the process and identify which collaboration activities are capable of having an impact on the relationship in some way.

Social capital is “the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships” (Nahapiet and Ghoshal, 1998, p. 243). It facilitates the
activities of actors within a social structure (Coleman, 1990). Nahapiet and Ghoshal (1998) distinguish three interrelated dimensions of social capital: structural, relational, and cognitive. The structural dimension refers to the network of relations as a whole and the overall connection patterns between actors (e.g., network ties, identifiable pattern of linkages). The relational dimension refers to the quality of relationships between actors that influences the behavior (e.g., trust, expectations, friendship). The cognitive dimension refers to shared representations and interpretations, and systems of meaning among groups (e.g., shared language, codes, and narratives).

The theory of social capital has been applied to several topics where humans and groups interact, including education, public health, and governance (Jackman and Miller, 1998; Portes and Sensenbrenner, 1993; Woolcock 1998). In IS, the theory has been used, for example, to examine the relationship between business and IT departments (van den Hooff and de Winter, 2011; Wagner et al., 2014). We apply social capital theory to guide our process design and to evaluate the process concerning its potential to nurture the relationship between data science and business professionals. We postulate that a process that yields high social capital has a positive impact on collaboration. Our research model follows a common format that has been proved to apply to action research projects (Davison et al., 2004): In a specific situation (joint BDA project) that has salient features (weak relationship and collaboration), explicit outcomes (strengthened relationship and collaboration) are expected from certain actions (collaboration process activities). Our research model is summarized in Figure 1, which also indicates how the practitioners’ and the researchers’ perspectives are interwoven in the research project.

![Figure 1. Research model.](image)

3 Methodology

We first introduce canonical action research as the chosen research method and briefly describe the company with which we are conducting the project. We then outline how our research design ensures the necessary systematic rigor.

3.1 Research Method and Case Introduction

To answer our research question, we opt for action research as this approach is well suited to solve a practical organizational problem and at the same time contributes to knowledge (Baskerville and Myers, 2004; Davison et al., 2004). These two imperatives are met in our research: First, the partner company suffers from the practical organizational problem of weak collaboration on BDA and needs a new collaboration process standard. Second, the BDA field can be enriched by a precise understanding of how data science experts and business users should collaborate when deriving business value from big data. Action research is acknowledged to be entirely appropriate for the IS field, as this field is very practice-oriented (Baskerville and Wood-Harper, 1998). Specifically, we follow the canonical action research (CAR) method and adhere to the principles and associated criteria to ensure both the rigor and the relevance of our research (Davison et al., 2004; Susman and Evered, 1978). CAR is characterized by its iterative, rigorous, and collaborative approach, which differentiates it from other action research approaches, e.g., IS prototyping, action science, and action learning (Baskerville and Wood-Harper, 1998). The CAR project is executed together with an international manufacturer and retailer of B2B tools (e.g., drills, saws) and related services, referred to as “ToolCo” in the following.
Headquartered in Germany, ToolCo has almost 4,000 employees across 50 global locations and sees over €1 billion in annual revenue. The company is currently transforming into a data-driven organization. To date, the focus has been the case-based development of BDA prototypes (e.g., for sales forecasting). The data science team’s collaboration with the business units has been rather sporadic and does not follow any standards for collaboration. This resulted in an unsatisfactory situation characterized by multiple challenges, including a lack of mutual understanding (see section 4.1 for all the collaboration challenges). Moreover, ToolCo feels that business does not really use the developed applications after the project is finished. ToolCo is seeking a collaboration standard that can be applied to BDA projects and that entails concrete activities to overcome the current collaboration challenges. After careful examination of three planned BDA projects at ToolCo, we jointly decided to use a project called “content automation” as the environment to develop and evaluate the collaboration process for BDA. The goal of this project is the development of a supervised machine learning tool to accelerate the onboarding process for supplier content in the online shop. This project can be seen as a typical BDA process improvement project (Grover et al., 2018). It started in March 2020 and is expected to last at least one year. This period is necessary to build the relationship with ToolCo, analyze the situation, and plan, execute, and evaluate the activities. In this way, the involved researcher has a facilitative involvement (Baskerville and Wood-Harper, 2008) and helps ToolCo with procedural knowledge and an independent viewpoint while also studying the situation.

3.2 Research Design

We adhere to the principles of CAR to ensure systematic rigor (Davison et al., 2004) and describe in the following how these principles are reflected in our overall research design. Our study design and the CAR principles are depicted in Figure 2 below.

- **Principle of the researcher–client agreement (RCA):** This agreement guarantees mutual behavior (Davison et al., 2004). First, functional project objectives were documented in a business case. We agreed that the development of a collaboration process for BDA is the key CAR project objective and that mitigating the collaboration challenges should be the major evaluation criterion. We approved CAR as an appropriate approach. As for data collection, we agreed that process development workshops, meeting observations, interviews, and supplementary surveys would be central data sources. A non-disclosure agreement for research projects was signed.

- **Principle of the cyclical process model:** Our project consists of five phases: diagnosis, action planning, intervention, evaluation, and reflection (Davison et al., 2004; Susman and Evered 1978). Proceeding through these steps ensures systematic rigor. Given this cyclical nature, we anticipate several iterations, especially for the action planning and intervention phase. We expect the process to evolve during the project, as the mitigation of the collaboration challenges and the satisfaction of the stakeholders involved will be measured continuously.

- **Principle of theory:** We agree with McKay and Marshall (2001) that a clearly articulated theoretical framework must be applied to the phenomenon of interest. As described in section 2.2, we apply social capital theory to evaluate the intervention outcome, namely the collaboration process. Specifically, we assigned the collaboration challenges contested by ToolCo (see section 4.1) to the respective dimension of the social capital that they impede. For example, the challenge of “lacking mutual understanding” represents a lack in cognitive capital, as neither of the groups has shared narratives (Hagen and Hess 2021; Nahapiet and Ghoshal, 1998). To measure how the challenges may be mitigated during the process, we monitor these items in regular surveys and interviews, together with overarching KPIs like satisfaction with the collaboration. Thus, we explore which activities in the process have the potential to mitigate the collaboration challenges and how. Moreover, in the course of our research, we aim to explore what we can learn from a theoretical perspective from the practical outcome of our action (Baskerville and Myers, 2004).

- **Principle of change through action:** Intervening in an organizationally unfavorable situation to induce change is a core aspect of CAR. This is achieved in the project as the results from the first
conceptual iteration are directly applied to the content automation project to improve collaboration. The insights regarding the collaboration challenges and mitigation approaches from the diagnosis phase went directly into the first process concept.

- **Principle of learning through reflection**: We are aware of our dual responsibility as action researchers. Details on our expected practical and academic contribution can be found in section 5.

![Research design of the canonical action research project.](image)

### 4 Preliminary Findings

The following preliminary results stem from the diagnosis and action planning phase. As for the diagnosis, we executed an independent analysis of the situation by conducting several interviews with the core team, which consists of the head of data science, the scrum master for data science, the product owner for data science, the director of content management, and a content engineer. Moreover, we screened the first project documents, such as the business case. These results directly informed the action planning.

#### 4.1 Results from the Diagnosis Phase

During the diagnosis phase, we analyzed several aspects to get a holistic understanding of the situation (Davison et al., 2004). The goal of ToolCo’s content automation project is the development of a supervised machine learning tool which is supposed to accelerate the onboarding process for supplier content in the online shop and, thus, reduce the manual effort for data consolidation, validation, and classification in business operations. That is why the director of content management approached the data science team, which delivers BDA solutions as a centralized unit to the company, to solve this business need on a collaboration project. As indicated above, the major motivation for ToolCo to join the CAR project was to develop a collaboration process blueprint that has the potential to mitigate prior unsatisfactory collaboration experiences with business units. The final content automation application is supposed to be programmed by external software developers. However, ToolCo did not experience significant issues at this interface before, which is why the group is not considered in our research as another potential stakeholder group. Specifically, the dissatisfaction resulted from several challenges during the collaboration and a weak relationship between the data science team and the business units. We found six factors that have continued to impede collaboration for BDA and that represent a lack of social capital between the teams (see Table 1 below for an overview). We assigned these challenges to the respective dimension of social capital that they jeopardize and brainstormed
collaboration activities that could mitigate them. These collaboration activities were further incorporated into our process design.

<table>
<thead>
<tr>
<th>Collaboration challenge (raised by)</th>
<th>Social capital</th>
<th>Mitigating collaboration activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrealistic expectations from the business sponsor (data science)</td>
<td>Relational</td>
<td>Actively manage expectations regarding BDA potentials and limits</td>
</tr>
<tr>
<td>Lacking effort of data science in building bridges to business (business)</td>
<td>Relational</td>
<td>Proactively utilize product owner as bridge builder</td>
</tr>
<tr>
<td>Lacking mutual understanding, with data science having limited clue about business processes and business having limited clue about statistical methods (both groups)</td>
<td>Cognitive</td>
<td>Conduct basic enabling session at the beginning and occasional ones during the process</td>
</tr>
<tr>
<td>Incongruent mindsets, with data science accepting only 100% solutions, and business aiming for “80/20” solutions (both)</td>
<td>Cognitive</td>
<td>Agree on clear definitions of done that suit both parties</td>
</tr>
<tr>
<td>The data is owned by business, leading to lost data access when project portfolio changes (data science)</td>
<td>Structural</td>
<td>Jointly improve technical infrastructure (e.g. data lake)</td>
</tr>
<tr>
<td>No space for “real data science” in projects due to time pressure (data science)</td>
<td>Structural</td>
<td>Install “data science boot camps” quarterly aside from project business</td>
</tr>
</tbody>
</table>

**Table 1. Collaboration challenges and mitigation activities (work in progress).**

### 4.2 Results from the Action Planning Phase

The action planning phase started by screening and discussing available BDA process frameworks in IS literature (e.g., Abbasi et al., 2016; Jagadish et al., 2014; Hirt et al., 2017; Phillips-Wren et al., 2015; Sakr and Elgammal, 2016). As a foundation for our collaboration process, we decided to use a combination of the BDA framework by Phillips-Wren et al. (2015) and the end-to-end supervised machine learning process by Hirt et al. (2017), as this suited ToolCo’s project situation best. Doing so yielded an adapted high-level BDA collaboration process (Figure 3).

![High-level BDA collaboration process](image)

**Figure 3. High-level BDA collaboration process**

We find that the existing BDA frameworks do not showcase the business activities sufficiently throughout the process, which is why we had to make several extensions to integrate the business perspective and collaboration activities. Specifically, we drilled down on all phases and designed business involvement and collaboration activities. This proved to be especially important in those phases that focus on the analytical/technical steps, which are usually only assigned to data science. In this way, we ensure that data science activities do not remain a black box for business and that mutual understanding is increased. Moreover, we find that responsibilities continuously change during the process; this leads to a high number of interfaces, which could potentially spell trouble. Table 2 illustrates the collaboration activities of the model development phase, the stakeholders involved, and the expected impact on the relationship.
<table>
<thead>
<tr>
<th>Collaboration activities in the model development phase</th>
<th>Involved stakeholders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educate business users regarding classification models (high-level)</td>
<td>Product owner, business user</td>
</tr>
<tr>
<td>Define most important objectives (e.g. false positives more problematic than false negatives?)</td>
<td>Business user, product owner</td>
</tr>
<tr>
<td>Explore the data and make sense of it</td>
<td>Data scientist, business user</td>
</tr>
<tr>
<td>Build multiple supervised models based on existing data (e.g. PIM)</td>
<td>Data scientist</td>
</tr>
<tr>
<td>Benchmark all models on performance metrics (e.g. confusion matrix)</td>
<td>Data scientist</td>
</tr>
<tr>
<td>Provide business feedback and implications on algorithm performance</td>
<td>Business user, product owner</td>
</tr>
<tr>
<td>Select final supervised model</td>
<td>Data scientist, business user</td>
</tr>
</tbody>
</table>

**Expected impact on relationship:** Hands-on education mitigates wrong business expectations and nurtures a common language; real-world business experience improves algorithm, which in turn increases trust in the application

**Table 2. Collaboration activities in the model development phase (work in progress).**

### 5 Expected Contribution and Outlook

With the following expected contributions, we meet the twofold requirements of action researchers, who are solving organizational problems by intervening while simultaneously contributing to knowledge (Baskerville and Myers, 2004; Davison et al., 2004):

- **Contribution to practice:** We intend to deliver a practical outcome to ToolCo, which is a proven and reusable collaboration process for BDA. This process will showcase how business users can be integrated throughout the process, including the analytical phases, which is supposed to increase the probability that the final BDA solution really fits business needs and, thus, increases their use in business. Moreover, ToolCo will benefit from a clear understanding of how certain collaboration activities have a positive impact on the relationship between data science and business measured by their potential to mitigate collaboration challenges. Thus, ToolCo can ensure that it executes these activities in future projects, which will improve overall satisfaction with BDA projects in the company. The process can also serve as a blueprint for other companies facing similar problems.

- **Contribution to research:** We intend to alleviate the discrepancy between the high importance of business utilization for BDA projects and the inferior consideration of business users and collaboration activities in BDA process models. Therefore, our intended theoretical contribution is threefold: First, we extend BDA literature and our understanding of BDA processes (e.g., Abbasi et al., 2016; Phillips-Wren et al., 2015). We drill down on the phases and add activities to provide a better understanding of the collaboration that is required and how the stakeholders have to interact. Second, our collaboration process contributes to the literature on BDA value (e.g., Côrte-Real et al., 2019; Mikalef et al., 2020, Vidgen et al., 2017) by offering a collaboration-oriented process to increase business use of BDA, which, in turn, affects BDA value. Third, we exploit the mechanics of social capital theory as an evaluation measure for our collaboration process.

The first version of the collaboration process illustrated above is currently being applied to the content automation project (intervention phase). As for the targeted results, the final depiction of the collaboration process will include decisive collaboration activities and explanations of how they positively impact the relationship, collaboration, and business use. The process quality, defined in terms of its potential to overcome collaboration challenges and to create social capital between the teams, is measured with a comprehensive approach. It consists of ongoing discussions of the project progress, interviews, and complementary surveys (evaluation phase). Weekly sprint meetings offer the possibility to implement necessary changes to the process right away (reflection phase). The final business use of the tool is planned to be monitored with a Google Analytics application. Lastly, to inform theory with our practical action, we aim to examine to what extent social capital theory is capable of evaluating the collaboration process quality and what we can learn from a theoretical perspective according to the practical outcome of our action (Baskerville and Myers, 2004).
References


