15th Maj 2017

Data Strategy Guidelines for Lean Startups

- Deriving Learnings Faster

Nicklas Millard

(nimi12ac@student.cbs.dk)

MSc In Business Administration and Information Systems (E-Business)

Copenhagen Business School, CBS

Supervisor	Jacob Nørbjerg, jno.itm@cbs.dk
No. of characters	136.255
No. of normal pages	90

COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Data Strategy Guidelines for Lean Startups

Nicklas Millard

MSc In Business Administration and Information Systems (E-Business)

Abstract

Lean Startup Methodology is an alternative to the conventional plan-and-execute business approach. New businesses following the lean startup methodology are inherently experimental and data driven. Emphasis is placed on deriving learnings to verify or reject value and growth hypotheses—this is, how value and growth is generated. This thesis claims current literature on lean startup methodology provides insufficient guidance on how to establish a data strategy, that is, the approach to collect, store, and use data to generate actionable insights and learnings.

This study is undertaken following design science research methodology. An IT artifact featuring three data strategy guidelines designed to facilitate learning in lean startups is developed, explicated, and evaluated. The artifact is designed using justificatory theories and concepts from authoritative authors on lean startup methodology, agile programming, and business intelligence. Three evaluation episodes are carried out to evaluate the artifact's performance in terms of utility to the entrepreneurial community and knowledge contribution to the body of research on lean startup methodology. First, a summative evaluation is conducted by using a self-completed survey whereby volunteer entrepreneurs are asked to assess if the artifact increases the entrepreneurs' knowledge on how to derive learnings. Second, a formative evaluation is performed by conducting a semi-structured interview with a business intelligence manager to identify areas of improvement. Last, a proof of concept is carried out to reveal its feasibility and whereby practical insights and pitfalls are documented.

Contribution to knowledge is derived from the novelty of the artifact and its attempt to bridge the research gap by offering a concrete approach to the challenge of collecting, storing, and using data in the context of lean startups.

Table of Contents

1 Introduction

This section introduces the reader to the problem field, its scope and the purpose of conducting research exploring the implementation and use of data strategies in lean startups.

1.1 Problem Field and Scope

A solid business plan and model has long been the hallmark of promise or indicator of starting a successful business. The plan-for-success paradigm is characterized by thorough analytical activities prior to product launch. However, this upfront planning may yield poorly performing startups due to the uncertainty embedded within the nature of startups. According to a press release from the European Commission (2013), 50% of newly started business fail with the first five years and about 98% of product launches fail (Bosch et al., 2013).

Startups often operate within new markets or by bringing new products to existing markets, with only a roughly sketched out business plan (if any). Thorough analysis prior to launch may not be feasible for startups, and, in the realization of this, a new paradigm called lean startup methodology started to emerge, popularized by Ries (2011) with his seminal work "The Lean Startup". This paradigm focuses on lean principles such as eliminating waste, deriving learnings, and verifying or rejecting *leaps of faith* assumptions – these may also be called creating validated learning in the LSM terminology, which is facilitated by the Build-Measure-Learn feedback loop, whereby a minimum viable product is created with the intend to collect data which in turn reveal if users find the product valuable and if growth is happening at the expected rate. However, within the literature of LSM, little research has been conducted on the process of deriving learnings – despite its prominent role in LSM.

Deriving learnings in lean web-based startups is an interdisciplinary activity which requires business understanding, programming proficiency, and quantitative and qualitative data collection and analysis techniques. In sum, this may be categorized as business analytics. A hot topic in contemporary business discourse due to enabling property in terms of competitive advantage. However, how do lean startups, that regularly lack manpower, time, money, and technical skills, make use of business analytics to drive validated learnings?

This research sets out to develop a set of guidelines that startups can adopt when creating and implementing a data strategy – i.e. the collection, storage, and use of data to generation actionable insights and learnings. The guidelines are evaluated by members of entrepreneurial community, an expert in business intelligence, and by conducting a proof-of-concept whereby insights and pitfalls are documented.

1.2 Purpose

This research project is motivated by the fact that lean startups are inherently experimental and data driven in terms of seeking validated learning through the Build-Measure-Learn feedback loop. However, such learnings are based on thorough data collection and analysis which in the first place may prove to be a resource-intensive endeavor. As resources are typically scarce in startups, such a commitment to data driven decision making is likely to prove difficult.

The challenge is trifold: First, the challenge lies in figuring out what data to collect. Second, the continuous development of software that tracks user behavior and metrics is time consuming. Last, extracting and analyzing stored data requires—at least to some degree—a specialized technical skillset depending on how advanced analyses that are needed. These three aspects (collection, storage, and usage) of data are what I perceive to comprise a *data strategy*.

This led me to the research question

"How to approach collection, storage, and use of data to generate actionable insights and learnings in the context of lean startups".

In the light for these challenges for lean startups, I strive to create a better understanding of how lean web based startups can create data strategies and what the imminent obstacles of implementing a data strategy might be. My goal is to provide an IT artifact that features a set of guidelines on how to implement a data strategy in lean startups.

2 Theoretical Foundation

To suggest guidelines and recommendations that will help startups implement data strategies, I first need to gain a comprehensive overview of the theories and concepts surrounding the related fields such as lean startup methodology, programming paradigms, and data analytics.

The theoretical foundation is structured in the following way:

1. **The Evolution of Entrepreneurship:** Get an overview of the different phases entrepreneurship has undergone and where it is heading, in terms of business planning and product development.
2. **Build-Measure-Learn feedback loop:** Explore one of the core models of lean startups.
3. **Technique, technology, and business intelligence in lean startups:** Thorough examination of literature on LSM, agile programming and dispersed but connected areas of analytics, such as data collection, data storage, data usage and presentation.

2.1 Evolution of Entrepreneurship

This section investigates the evolution of entrepreneurship, beginning at the conventional plan-for-success approach to the experimental, data-driven approach that has gained traction in recent years. This section serves the purpose of providing an overview of the problem environment and how it has progressed in recent years.

2.1.1 The Conventional Approach

Creating startups and businesses has for a long been carried out in a waterfall like fashion, where the entrepreneur would start out with having an idea and write down the business plan to be executed. To test if the idea holds water, a series of analytical initiatives and business planning would be conducted prior to launching the business, such as for example a market research, competitor analysis, go-to-market strategy, marketing plan (Toren, 2012), all of which may end up in an elaborate business plan to be executed. A solid business plan based on an array of analysis have long been the hallmark of promise or indicator for a successful business (Rich & Gumpert, 1985; Ries, 2011).

The business plan has typically been an important instrument for business owners and entrepreneurs, particularly in terms of attracting funding from investors as it is an aggregate of the thoughts (in terms of vision and mission statements), market analysis (environmental factors, competitor analysis, and go-to-market strategy), financials, etc. Especially the need for providing a solid financial plan showing break-even and return-on-investment are of special interest in the traditional school of thought, as these are typically the most interesting points for investors. Despite the rigid planning-for-success that the traditional waterfall approach business embrace, only approximately 60% of startups survive to the age of three and 10% survive past 10 years (Gage, 2012).

This conventional startup paradigm has been described along similar lines by Sarasvathy (2001, p. 246) using the term causation. This paradigm is characterized by time consuming analytical activities to understand an existing market and how to optimally bring a new product to market.

2.1.2 The Roots of Lean Startups

This elaborate pre-planning of a startup may however not be feasible as startups operate within mass amounts of uncertainty and unknowns, which essentially makes elaborate plans practically impossible write down. In the light of this realization a new type of startups began to emerge, called lean startups, popularized in the early 00's and 10's (Ries, 2011). Before trying to grasp the concept of lean startups, I briefly dive into the underpinning theories, concepts, and principles that the lean methodology builds upon to get a better understanding of the its roots from the operations management field.

The core of lean startup methodology (also known as LSM) is originating from lean production and thinking. The principles and application of lean thinking and production has in recent years had significant impact on academia and several industries and are making its way into formal educational (Blank, 2013). The origins of lean thinking have commonly been attributed to the

Japanese Motors industry and particularly the innovations at Toyota Motors and its Toyota Production System that was formed due to a scarcity of resources and intense domestic competition, resulting in methods as for example the Just-In-Time production system, Kanban, and a high level of employee problem-solving (Hines, Holweg, & Rich, 2004; Ries, 2011, p. 18).

The lean approach to operations is focused on eliminating waste (sometimes called “muda”) and excess. Later on, a critical point in lean thinking became the focus on value (Hines, Holweg, & Rich, 2004, p. 995). Typically, waste is defined as any activity that does not add value to the end-product or output. To eliminate waste, you need to identify the output that customers value and cut away anything that does not support the output (George, 2013, p. 26).

For the purpose of identifying value you need to listen to the customer, also known as voice of the customer (VoC), interpret the customers’ feedback and prioritize down to critical requirements. These requirements are often referred to as Critical Customer Requirements (CCRs) and for them to be useful, a CCR should: be specific and measurable, be related to a product/service attribute, not present alternatives, be unambiguous, and describe what – not how (George, 2013, p. 58). In a similar manner, Hines, Holweg, & Rich (2004, p. 997) has defined value enhancing as a focus on customer value that is created if internal waste is reduced or if additional customer valued services and features are offered.

Criticism has been given to the lean startup methodology for being too narrowly focused. Emphasizing agile product development on basis of customer feedback over other important areas of business models such as company’s value network, value delivery, and revenue model (Ojala, 2016).

2.1.3 Lean Business Plans

Applying these core principles of eliminating waste and focus on value adding activities to startups, we quickly see how elaborate business plans stand in contrast to the lean thinking methodology. The business plan is solely an internal document which does not directly add value to the customer. Ries (2011, p. 22) advocates that instead of making complex plans that are—essentially—based on assumptions, lean startups must make constant adjustments based on an iterative model called Build-Measure-Learn feedback model (which I explore in section 2.2).

Sarasvathy (2001) has developed the term *effectuation* that has similarities to the lean startup's feedback model. The effectuation process allows the entrepreneur to explore different business plans on-the-go by progressively learning what is valued by customers. Effectuation allows the change and construction of goals over time, as opposed to set out a strategic direction to be followed as proposed by the conventional (causation) paradigm. Sarasvathy (2001) notes that the effectuation paradigm is not regarded as being better than the conventional paradigm – it simple offers an alternative approach that may fit for some entrepreneurial situations better, where uncertainty is a key risk factor.

However, despite the lean startup's emphasis on a learn-on-the-go approach by adjusting according to feedback from customers, Berry (2012) takes another stance and proposes that the ideal business plan should grow organically in the same way a lean startup does. In that way, it is argued that business plans fit perfectly with lean startups as they develop organically with each Build-Measure-Learn iteration.

2.1.4 Product Development

The traditional product development process is based on the same waterfall school of thought as the traditional business plan is: that planning (often done at a distance to the market) is a way of achieving success. However, despite thorough planning, only one in four products were winners even though 50% of American firms' resources were spent on innovative new products that failed (Cooper, 1990), and there are even claims of as high as 98% of new products fail (Bosch et al., 2013). A traditional and popular product development process presented by Cooper (1990)—that is still being used—is stage-based development systems, where new product innovations go through different stages before being released to market as seen below.

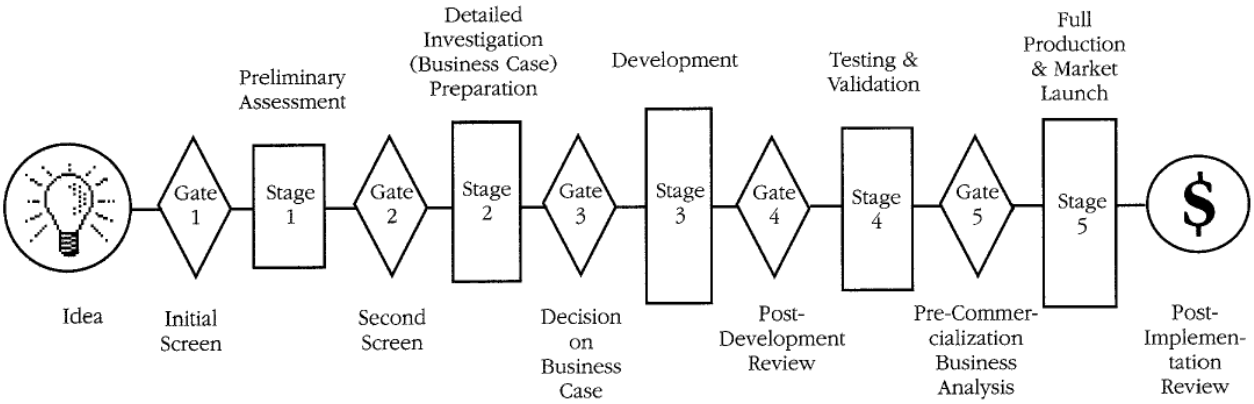

Figure 1 - Overview of a generic Stage-Gate system, Cooper, 1990, p. 46

The main appreciation of using a stage-gate system is the simple process overview and roadmap it provides. The approach lays out the activities for each step, or stage, of the process. Each gate has its own set of inputs, deliverables, exit criteria, and output. The gates' role includes a review of the input quality, assessment of quality from an economic and business standpoint that results in either a go/kill/hold/recycle decision. When a product is moving from one gate to another an action plan is formulated and necessary resources are allocated. I will provide a brief explanation of each step in the system, excluding the gates:

1. **Preliminary Assessment** – a new idea is submitted for screening. Determine technical and market merits of the idea, possibly involving focus groups and quick in-house appraisal of proposed product.
2. **Detailed Investigation** – project must be clearly defined. Market research, competitive and financial analysis is conducted.
3. **Development** – product development, marketing plan, and updated financial analysis is conducted, as well as detailed testing.
4. **Testing & Validation** – testing the product (in-house and field tests), production process, customer acceptance, and economics of the project.
5. **Market Launch** – marketing plan is executed and product is released to market.
6. **Post Launch Review** – product's performance (e.g. customer adoption and finances) in the market is reviewed and the overall project's strengths and weaknesses are assessed.

Evidently, this system provides an easy-to-follow approach that puts discipline into launching new innovations, where each step and gate is trying to ensure no critical activities have been forgotten, as well as no gaps in the process. Nevertheless, such elaborate planning without receiving actual market response throughout the process has been criticized by advocates of the lean startup methodology and agile principles (Ries, 2011; Blank, 2013; Cooper, 2016). Since this stage-gate system was proposed for managing innovation in firms, agile methodology has rapidly made its way into product development, both in startups and larger firms, and almost becoming omnipresent.

The critic is mainly due to the exploratory nature of startups, where there is commonly a lack of clear requirements, customer segment and business models (Bosch et al., 2013), and especially that long and rigid processes cannot keep up with quickly changing environments and customer needs.

Lean startup and agile methodology contrasts sharply with the traditional product development process. The principles of LSM are, as articulated in section 2.1.2, focused on value creation and validated learning where experimentation and short feedback cycles are praised. Additionally, the agile principles are in line with those of LSM, where uncertainty and unpredictability are dealt with by collaborating with people close to the process, rather than using a formal process framework such as a gate-based development system. Cooper (2016) has taken these critical views presented by advocates of LSM and agile into consideration in the newly proposed Agile-Stage-Gate system which is a hybrid model of the traditional stage-gate system and the agile software development principles of being close to market through user feedback.

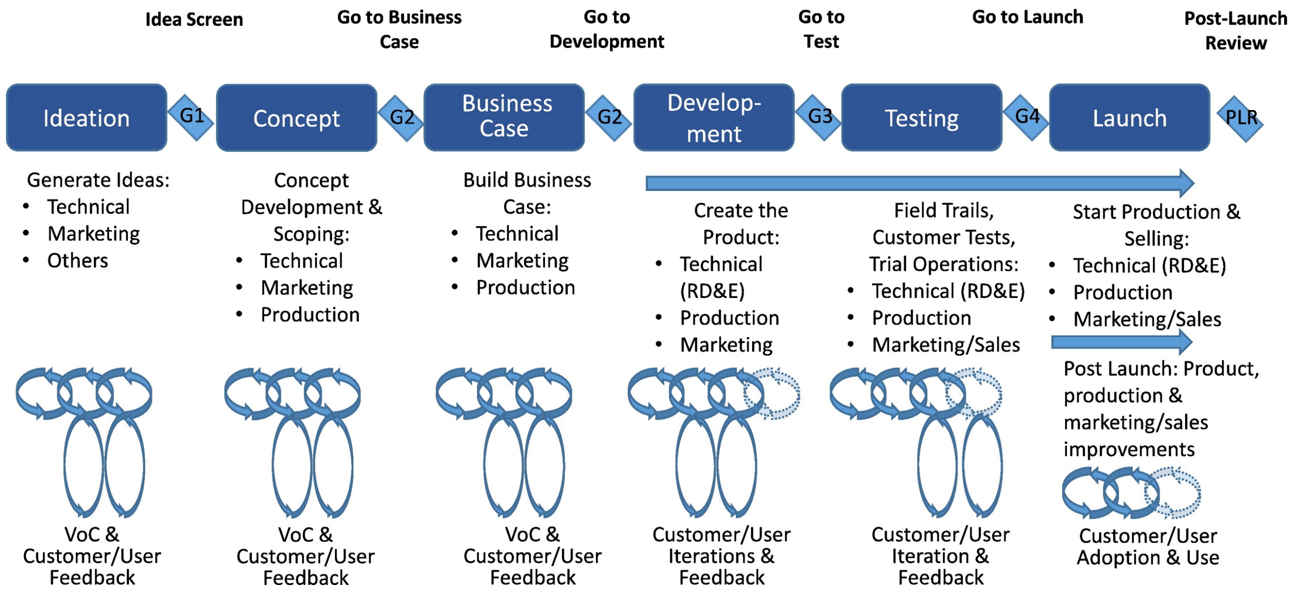

Figure 2 - New Stage-Gate system adopting agile principles, Cooper, 2016, p. 169

According to Cooper (2016, p. 167) the adoption of agile principles is possibly the most significant change to our thinking about new product development in thirty years. Even though this Agile-Stage-Gate model is designed for manufactured products, it has its roots in IT projects. Research studying three large European high-tech firms where Stage-Gate and agile were merged for IT projects showed that the integration posed several benefits, such as: better internal communication, more visually intuitive progress metrics, more efficient planning, improved customer feedback, and improved moral amongst team members (Cooper, 2016. 168).

Furthermore, in addition to the benefits of LSM and agile principles, the contrast from thorough pre-planning and execution of plans, it allows one to act upon unexpected results of user interaction. Parker, Van Alstyne & Choudary (2016, p. 58) argues that product design should always leave room for discovering unintended usage, resulting in design evolving based on actual customer use, which is articulated as a form of “anti-design”.

Regarding product development, the lean startup methodology advocates for creating a minimum viable product (Ries, 2011, p. 93), typically mentioned by its abbreviation “MVP”, which is basically a stripped-down version of the envisioned product, lacking many—perhaps even essential—features. The MVP’s purpose is to collect actual usage data to test hypotheses (explained further in section 2.2.1) and create validate learning. This helps entrepreneurs to start the learning process as quickly as possible. The MVP is the topic of section 2.2.2.

Moving from the planning intensive approach to a fluent, lean, and agile approach, the next phase of product development can be found within the literature of dispersed areas of data subjects such as analytics, statistics, and data engineering. As users often have a difficult time articulating their needs, their actions and behavior on for example a website or in an app are perhaps a more ideal way of investigating value adding features.

This rather new advance in product development has been recapitulated as evidence-based engineering by Bosch & Olsson (2016, p. 29), analytics, or data driven decision making by Fisher et al. (2012). This concept refers to the ambition of using captured usage data to validate new development and features' value it delivers. This thinking is certainly in keeping with the lean and agile principles of a customer centric focus in development as well as only commencing work on development projects and features based on previous iterations' learnings. The benefits of data driven engineering and data analytics and how these activities fit with LSM and agile approach will be explored in subsequent sections.

2.2 Build-Measure-Learn model

This learning-based and experimental approach to business and product development leads me to investigate the previously mentioned Build-Measure-Learn feedback model in further details. Later I will examine how this model is convergent with the recently increased interest in data and analytics. Eric Ries popularized the concept of lean thinking in startups with his book "The Lean Startup" (Ries, 2011), a widely famous work that has inspired numerous articles and books. Some of the most prominent ideas presented in Ries' (2011) work in relation to the research question of this thesis are the concepts of minimum viable product, leap-of-faith, validated learning, innovation accounting, and especially the Build-Measure-Learn as shown below.

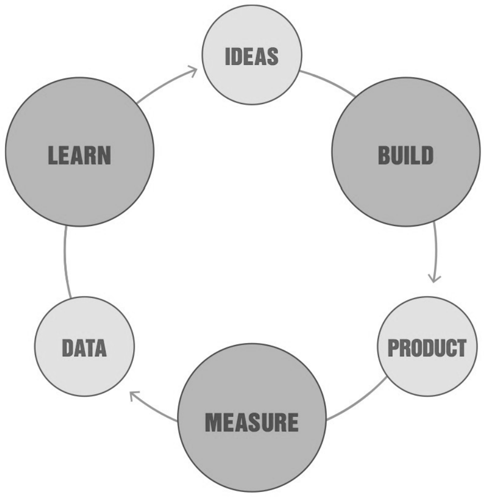

Figure 3 - Build-Measure-Learn Feedback Loop, Ries, 2011, p. 75

However, interesting as it is, before plunging straight into the components of the Build-Measure-Learn feedback loop, I will start by touching upon the concepts of leap-of-faith assumptions and minimum viable product as these are fundamental to the feedback loop. Having a glance at these concepts prior to the feedback loop itself will provide us with deeper understanding of each of loop's components and ultimately how they relate to creating successful data strategies.

2.2.1 Leaps of faith

Every business is started based on a set of assumptions—according to Ries (2011, p. 76) these are also called *leaps of faith assumptions*—The two most important assumptions are the value hypothesis and growth hypothesis. The value hypothesis denotes the company's assumption that their product or service is adding or creating value for customers, which relates to the value proposition that the product offers–this is for example how to resolve customer "pains" and provide customer gains. The growth hypothesis is concerned with the company's initial idea of how to attract customers or users to its product. Not only through the use of short term growth strategies such as PR stunts, but also the company's long term strategy articulated as its *growth engine*.

It is fundamental for lean startups to figure out (measure) if their initial hypotheses are correct or not, and whether to pivot or persevere based on results from testing the hypotheses. Pivoting or persevere are two essential concepts of LSM, however out of scope for this thesis.

Because the leaps of faith are highly error prone, the need for testing is vital to startups. Ries (2011, p. 81) proposes that entrepreneurs must first build a company that has the capabilities to test assumptions systematically, secondly, the testing must be done without drifting away from the company's overall vision. The go-to approach for testing hypotheses are to create a minimum viable product.

2.2.2 Minimum Viable Product

The centrepiece of Build-Measure-Learn is the minimum viable product. Here the entrepreneur takes the idea of the envisioned product and creates a minimum viable version of it. This is not to say that the entrepreneur should create a minimal product by itself nor to release a MVP filled with flaws. As illustrated by Pasanen (2014), the MVP should ideally be a functional and usable product. The ultimate purpose of the minimum viable product is to help the entrepreneur start the learning process in which the leaps of faith assumptions are tested to provide validated learning.

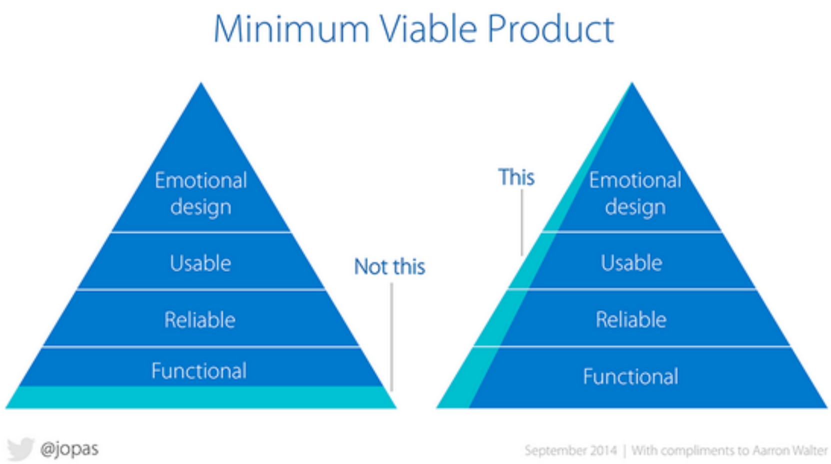

Figure 4 - Illustration of a MVP, Pasanen, 2014

Obviously, the initial product offering is not suited for everyone. Instead of focusing on the mainstream market, Blank (2010) and Ries (2011) argues that the MVP should be targeted towards possible early adopters who has a keen interest in trying out the product at an early stage (think of e.g. gamers buying alpha or beta access). The early adopters are ideally also people who wish to be a part of the product development by providing feedback because the product solves a pressing problem for them and they wish to see it succeed.

After collected data on users' perceived value of a feature has been analyzed, the findings will be implemented into the next product development iteration to create a more refined product in terms of providing actual value for customers. This now leads me back the feedback loop that I started out with in the previous section.

2.2.3 Components of the Build-Measure-Feedback loop

As illustrated by Ries (2011, p. 75 – fig 3.) the feedback loop is a three-step cyclic process with three "gates" in between, much in a similar way as the stage-gate system in traditional product development process. Before moving from one step to another, the current step must be completed. The steps and their purpose is briefly explained below.

1. **Idea** – The entrepreneur has an idea for a product or service which is based on a *value* and *growth hypothesis*.
2. **Build** – To test these hypotheses, the entrepreneur builds a *minimum viable product*. In this way hypotheses are quickly validated or rejected in contrast to start out with elaborate business planning. Rapid, validated learning is regarded the way to succeed in LSM.
3. **Product** – Release product to highly interested customers and potentially first-movers within the industry.
4. **Measure** – Set up measurements, targets, and overall goals. Collect real data to test value and growth hypotheses. Use the LSM concept of "*innovation accounting*" (described further in subsequent section).
5. **Data** – Analyze performance to examine if the company is on right track with targets.
6. **Learning** – Based on conclusions drawn from data analysis, it is time for the entrepreneur to decide whether to *pivot or persevere*. Which, respectively, means to change direction or keep on track and adjust the product.

According to Ries (2011, p. 75), a startup's product(s) are really a set of experiments used to create validated learning. The Build-Measure-Feedback loop is a way to steer the startup in the right direction. Focus should be placed on minimizing the time it takes to get through this loop. In this way, startups can minimize waste (one of the core principles of lean thinking) and build products that provide real value.

2.2.4 Measuring and data

Besides the down played focus on planning, one of the key aspects of LSM that contrasts to conventional businesses is the attention placed on measurements and data. As opposed to traditional accounting, the LSM has its own system called innovation accounting (Ries, 2011, p. 113). This is specifically designed for startups. It is argued that traditional accounting (for example forecasting revenue) is not suitable for startups due to their unpredictability. Innovation accounting is instead concerned with learning.

Innovation accounting provides startups with an objective accounting framework to assess the learnings. It begins with decomposing the leaps of faith assumptions into a quantitative financial model. Instead of revenue, data on a set of metrics related to the assumptions are collected and analyzed. Such (simple) metrics can for example be the number of visitors, signup rate and churn rate in a period. When selecting which metrics to measure, it is important to steer clear of the so-called *vanity* metrics. An example of a vanity metric is the gross number of signed up users on a site. Another is tracking the total number of user stories implemented. These numbers will only continue to grow, but does not provide any information of whether the product or service is providing value to customers or if real progress is happening. Contrary, *actionable* (actionable, accessible, and auditable) metrics that demonstrate cause and effect provide clear indications of value and how the company is performing.

The focus on learning by using innovation accounting provides entrepreneurs an alternative way to measure performance. However, this approached has received some criticism. Ladd (2016) who conducted research on 250 teams found that LSM may produce "false negatives". Good ideas may be rejected because a seemingly lack of customer demand due to the lack of rules revolving around innovation accounting. Ladd (2016) advices entrepreneurs to declare rules for go/no-go decisions upfront, additionally, his research also indicates that too much feedback from customers might lead startups to change its idea too frequently and thereby create confusion.

An additional caveat from critics of LSM, while useful, the methodology has at times been criticized for being *too* data focused (Croll & Yoskovitz, 2016). The criticism of this approach is pointing out that using data to optimize your business or product without having an eye on the bigger picture, can be dangerous and fatal. Also, data driven optimizations are good at optimizing an already known system, but cannot find new ones (Croll & Yoskovitz, 2016, p. 38). Finding new opportunities requires human ingenuity.

2.3 Technique, Technology, and Business Intelligence in Lean Startups

So far, the technical aspect has been kept at a minimum. However, as this thesis is concerned with lean *web-based* startups, understanding the technical and programming aspect of LSM is essential due to its data intensive focus. This section is focused on how agile methodology fits with LSM and which challenges that might occur as well as which technologies and strategies that supports data-driven initiatives in the context of lean startups.

Despite Ries' (2011) thorough work in developing and educating entrepreneurs about the lean startup methodology, one of the most crucial parts, learning, is still somewhat in the dark. How exactly to collect the data in the "measure" phase is only lightly touched upon, and no concrete strategies for this has been presented. The same goes for conducting data analysis. Furthermore, Steve Blank (2013), an often cited advocate of LSM, also fails or avoids to present tangible ways of implementing strategy for data collection and analysis. Perhaps this is purposely done not to scare off non-technical entrepreneurs. Another answer might simple be that explaining business analytics beyond saying "it creates better outcomes" is difficult as stated by Stubbs (2013).

Nevertheless, solely being aware of the importance of data and which metrics to watch out for only gets you so far. Without the understanding of the actual data process (collecting, storing, and using), you will not be able to conduct any actual data analysis yourself. Without this technical understanding, an entrepreneur striving to building a data driven lean startup might face tremendous issues when trying to validate his hypotheses due to either not being able to write the necessary analytics software, conduct the data analysis, or both.

The literature, theories, and concepts that do exist on data and business analytics are often targeted towards a more resourceful audience such as large corporations, and describes the creation of large data projects that likely require investments that lean startups cannot afford due to scares resources. For me to examine this "data issue" further, I will draw on literature from different fields of data related subjects and process frameworks.

2.3.1 Scrum framework for Lean Startups

There is a striking similarity between the agile manifesto (Beck et al., 2001) and LSM. With good reasons – both advocate for a close-to-market, customer centric approach with focus on value creating and mitigating risk by following an iterative process framework whereby feedback is provided by actual users or customers.

Agile principles (Beck, et la., 2001) are basically a set of values that other frameworks such as Scrum can build upon. Scrum, created by Schwaber & Sutherland (2013), is a widely recognized and utilized agile programming process framework due to its lightweight and simple to understand structure. The purpose of Scrum is not to provide a technique for building products, but rather to lay out a guiding framework in which you can employ different techniques and processes (Schwaber & Sutherland, 2013). The iterative development that Scrum promotes is organized into relatively short time-boxed iterations called *sprints* that typically lasts two to four weeks or less. These sprints contain: planning, daily meetings, coding, review, and evaluation. At the end of a sprint, a potentially shippable product has been developed. Successive sprints are undertaken to implement new use cases based on user feedback.

Product or system requirements are kept in a *Product Backlog* which is a Scrum artifact that contains all use cases (referred to as stories), features, bug-fixed etc. This backlog is dynamic and constantly changes according to accommodate new market conditions, user needs, and business requirements. Before a sprint is undertaken, items from the product backlog are placed in a *Sprint Backlog* – the items that needs to be finished at the end of a sprint. These sprint items are selected by the whole team, that consists of a Product Owner (responsible for the product backlog), Development Team (responsible for coding), and Scrum Master (responsible for the Scrum process and enactment).

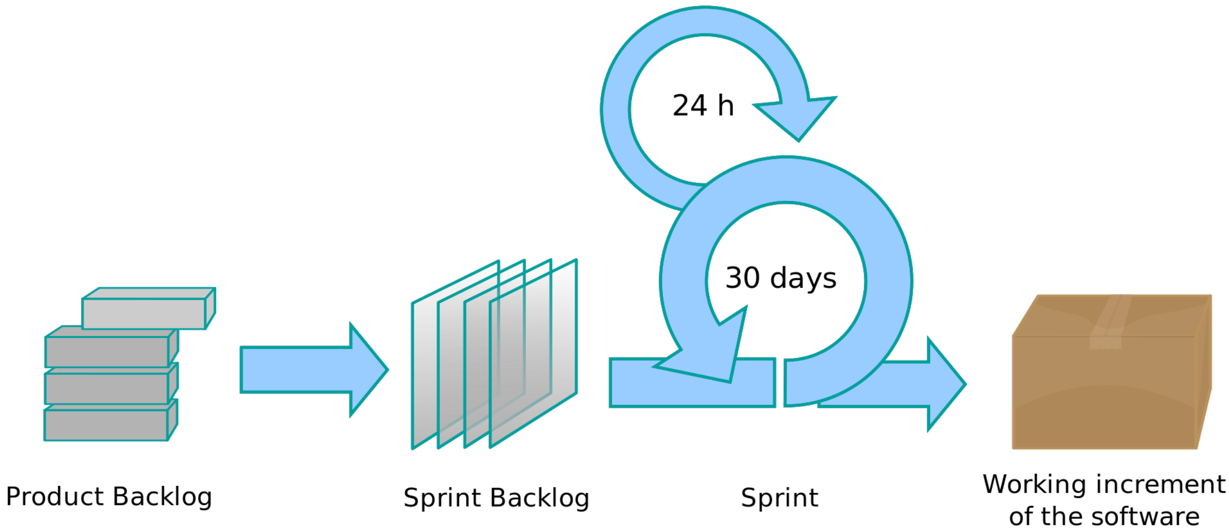

Figure 5 - generic overview of a scrum sprint from https://en.wikipedia.org/wiki/Scrum_(software_development)#/media/File:Scrum_process.svg

The above figure illustrates:

1. Product requirements are placed in the Product Backlog
2. Specific features and user stories are placed in the Sprint Backlog
3. The sprint itself starts – the sprint backlog is locked during this time to avoid scope and feature creep.
4. A new, working product version is created and possibly released to users (full market release or to a limited user base)

The sprint process as illustrated above has a resemblance to the Build-Measure-Learn model from LSM. Both are iteration-based processes that value high relevance for the users by incorporating actual user feedback (articulated or behavioral) and acknowledges that requirements are never set in stone, but will evolve over time. Also, both processes conclude with a usable product which may take the form of a minimum viable product or further product enhancements. It should be noted that the BML model's focus is not the product, but the learnings derived from usage of the product.

Even if Scrum matches the lean startup methodology, one of the criticisms of Scrum is its lack of innovative thinking. Cohn (2014), who teaches a certified ScrumMaster course, criticizes Scrum for having become more concerned with ticking off check-boxes rather than exploring new innovative ideas and solutions. This is potentially a side effect of moving to shorter sprints that leave less time to recover if promising but risky approaches fails. This box-ticking contrasts with the agile methodology that deals with unpredictability by relying on people and their ingenuity (Nerur, Mahapatra, & Mangalaraj, 2005).

Moreover, agile methodology and lean thinking somewhat collides with the lean startup methodology because of the overhead that is introduced in the Build-Measure-Learn model. Agile methodology encourages lean thinking by cutting down waste, by reducing for example over-production (i.e. developing non-value adding features) and documentation. Oppositely, the BML model emphasizes measurement and extensive data collection which requires activities that are likely time consuming and not directly value adding in respect to the customer – time that could have been spend on developing the product, and not analytics software and performing data analysis.

2.3.2 Analytics Strategy

While it is entirely possible to create an analytical strategy based on existing vendor solutions such as Google Analytics, this may not provide an optimal analytical strategy approach in Lean Startups due to lack of flexibility. The relatively low traffic new products and websites attract will most likely show large fluctuations in usage – this poses an analytical issue of having too little data to conduct traditional analytics (Parikh, 2014). To overcome this issue, Parikh (2014) suggests that startups' analytics software should be tailored to accommodate monitoring user activity at a granular, individualized level by for instance tracking individual sessions or cookies and the associated behavior of those. The development of an in-house analytics solution further allows for a higher degree of flexibility, control and data access (Periscope Data, n.d.; Oxley, 2014), but does require more time to develop. Despite the excess overhead presented by developing a proprietary business analytics platform, it is the approach that may yield the greatest results due to it is ability to support a startup's unique business model—or search for one—and strengths (Stubbs, 2013).

Firstround.com (n.d.) conducted an interview with Ben Porterfield, Vice President of Looker (a business analytics software company: https://looker.com) to explore how to establish an analytics infrastructure in startups. Porterfield touches upon subjects of how to best store data, common mistakes, and which metrics to measure. Some of Porterfield's advices are:

- **Start as early as possible** – Postponing data collection only delays finding out which features that provide actual value to customers.
- **Make data accessible** – Everyone in a startup benefits from easy data access. Data is not only for engineers – the people who interact directly with customers or develop features must be able to retrieve data by themselves without having it first translated by an engineer without direct knowledge of the work by e.g. the customer agent.
- **Provide Self-Service tools** - In keeping with advice above, providing a self-service tool to access data will eliminate the bottleneck of needing to have all data requests go through a data team. Further, Porterfield states "Game changing insights don't always come from the analysts or data science group, they often come from the users who are closest to the problem".

Additionally, Porterfield has provided some caveats of analytics, that are perceived by him to be common mistakes in startups:

- **Too much focus on product building** – Take a second to consult the collected data opposed to keep building new features. Understand engagement, how the product is being used, and why customers come back before moving ahead.

- **Not tracking enough** – More tracking allows the startup to see how granular changes to the product or in the market affects the sales and engagement.
- **Not thinking of who needs access to insights** – Build a self-serving platform that answers the questions asked by specific jobs in contrast to providing a simple overview of general metrics' performance.
- **Storing data the wrong places** – Making data inaccessible with SQL, one of the most popular query languages, will inevitably lead to lost value in terms of data analysis. Event and transactional data should be easily accessible to create a deep understanding of user behavior and product value.

2.3.3 KPIs and Metrics

Before starting to explore the data, you must first gain an ample understanding of the business problem that you are trying to solve and its strategic context (Hodeghatta & Nayak, 2017; Provost & Fawcett, 2013). In lean startups, this may denote the value and growth hypotheses – also known as leaps of faith. These hypotheses are decomposed into quantifiable metrics that can be tracked and by using innovation accounting lean startups determine if the initial hypotheses hold true or if they need to be changed. Not only do startups need to track the progress at a grand level, they also need to track individual product features. Every implemented feature must be instrumented with its own set of metrics to measure its behavior, performance, and usage (Bosch & Olsson, 2016).

Difficulty in measuring metrics in startups come from the uncertain nature embedded in them. Startups simply do not always know which metrics are key. Croll & Yozkovitz (2013) have proposed some rules of thumb for what makes a good metric:

- **Comparability** – A metric must help the startup determine which direction the company is heading. This is done by comparing to e.g. other time periods, competitors, user groups, etc.
- **Understandable** – Memorable and easy to understand its value to the company.
- **Be a ratio or rate** – Create a fundamental understanding of the startups' direction at a glance.

- **Changes behavior** – Knowing how your behavior will change based on changes in the metric is key. If a metric does not promote behavioral change, chances are it is not critical to success.

Innovation accounting has been criticized for lacking clarification of rules for go/no-go decisions. Croll & Yoskovitz (2013) and Ladd (2016) proposes that startups need to draw a line in the sand to foster a disciplined approach. Furthermore, the metrics must be directly aligned with the startups' goals.

Especially the metrics' alignment with a startup's goal is a critical point. Croll & Yoskovitz (2013) states that being able to decide which metrics to track, the entrepreneur must describe the startup's business model in terms of: acquisition channel, revenue source, product type, and the delivery model and then set up the metrics that matter the most for each of the components. This supports the claim made by Stubbs (2013, p. 12), stating that the application of business analytics should support the individual, unique business model and capitalize on the context of the specific business.

2.3.4 Data Storage and Architecture

On a more technical note, databases play a key role in data strategies. Database administration is a complete topic by itself, and will therefore not be examined to great extends. However, the understanding of database management systems and query languages are critical to developing a data strategy.

Relational database management systems (RDBMS) are central to modern applications (Schlossnagle, 2004). In general terms, a database is a collection of persistent data and RDBMS is a system for managing databases. RDBMSs provide the foundation to store and retrieve data collected by e.g. usage of a website or features. The "relational" part denotes that data is organized into tables that can be referenced by other tables—in this sense: have a relation to other data. The tables consist of headers and rows that can be accessed by e.g. different variations of SQL (such as MySQL) which are typically used as the de facto database languages in many organizations to interact and manipulate with relational databases (Clifton & Thuraisingham, 2001; Nayak et al., 2013). Relational databases require a predefined scheme to place the data into. An example of a scheme and data may look like the following:

id	sessionId	userId	browser	elementClicked	timeElapsed	url
1	*1*	*1*	*Safari 10.0.2*	*a – some link*	*12000*	*www.url.dk/test*

Each column has its own name and data type. For instance, id, sessionId, userId, and timeElapsed (in milliseconds) may be of type Int (integer) and browser, elementClicked, and url are of type Varchar (variable-length characters) or text. The columns sessionId and userId are referencing other tables. Rows in a database are referred to as records. Trying to insert data of the wrong data type will cause and error.

NoSQL—Not Only SQL—another type of database and query language is an emerging alternative to the traditional SQL-based RDBMS (Nayak et al., 2013). Advantages of NoSQL is the schema-less nature, flexibility that it provides, high speed, and scalability compared to RDBMS. However, NoSQL databases lack a standardized query language. Database providers of NoSQL has developed their own query language, e.g. Cassandra supports CQL, MongoDB uses mongo query language and Parse also has its own. This does pose an issue if a startup wishes to switch from one NoSQL database provider to another (Nayak et al., 2013).

Different database access patterns have been developed over time. These define the way you interact with a database, using a programming language such as PHP 7. The database access pattern determines where and how SQL appear in the code base (Schlossnagle, 2004, p. 306). Some ways of implementing SQL in the code is by Ad Hoc Queries, which essentially is not a pattern – here SQL is written directly into a particular spot to solve a specific problem. The issue that this approach poses is in terms of refactoring and reuse. To overcome this issue, you may deploy an Active Record Pattern instead. In this way, a class directly corresponds to a row in the database. All database access is encapsulated by the class itself. I have provided a simple example below, to illustrate a class that corresponds to the above database schema collecting click-stream data:

```
class SomeVeryBasicAnalytics {
    private $sessionId;
    private $userId;
    private $browser;
    private $elementClicked;
    private $timeElapsed;
    private $url;
}
```

Another useful pattern is the Mapper pattern which can deal with several tables at once. This pattern uses a separate class that knows how to save an object – without the object's own class having any of the database access itself.

Considering this topic of data storage–data stored in a database—the most important thing to keep in mind is the database's design to support operational record keeping and analytical decision making (Kimball & Ross, 2013). The operational part is concerned with the actual operation of e.g. a website, making sure that new users can sign up, etc. The analytical part has a different objective: to evaluate performance. For instance, this involves counting the number of new user signups and compare that number to last week or month and understanding why the users signed up in the first place. The analytical database can take the form of a *data warehouse* which are often located on a separate hardware system to avoid load on the operational system which slows it down.

Kimball & Ross' (2013) advice on how to design an analytics database have some resemblance to those of Porterfield's:

- **Understand the business user** – Understand their goals, objectives, and which decisions they need to make.
- **Deliver relevant and accessible information** – Make user interface simple and match the users' cognitive processes, monitor data accuracy, adapt to change.

One technique to make data accessible and deliver relevant data to business users, is dimensional modeling. Dimensional models can be placed in a DBMS and simplifies the schema in contrast to manually pulling data from various normalized tables (Kimball & Ross, 2013). Creating a dimensional model can be done using a star schema, which translates into a fact table derived using multiple dimension tables as visualized below:

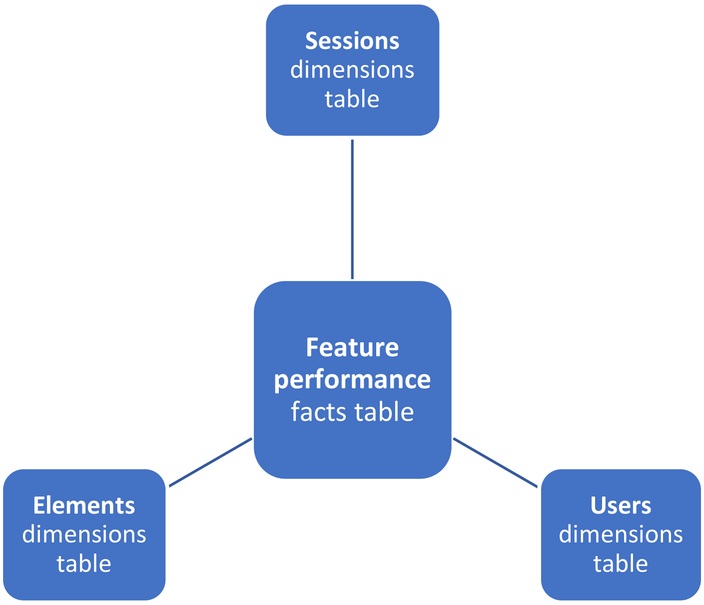

Figure 6 - Example of a feature performance facts table

2.3.4 Business Analytics

The popularity of being data-driven has increased during recent years. The concept of conducting data-driven or (informed) business is an immensely wide concept, ranging from relatively simple activities like observing website stats to more advanced data scientific methods. Businesses need to develop the right capabilities to perform business analytics. Some of the skills or knowledge that are required by a data analyst are such as: comprehensive understanding of the business and its problems, data analysis techniques, computer programming, data-storage, and statistical methods in data analysis (Hodeghatta & Nayak, 2017; Stubbs, 2013,).

The application of business analytics needs a strategic context – without it organizations cannot decide what data to focus on and not least what they are trying to achieve (Acito & Khatri, 2014). A typical example from marketing (my professional background) is to use data to determine which customer segments that has the highest probability of responding to an offer sent for example by e-mail.

According to Stubbs (2013) to successfully leverage business analytics requires understanding of 1) how to generate insight, 2) how to manage information, and 3) how to act upon the insights. These three activities lay the foundation of my articulation for what covers a data strategy. Business analytics comprises an array of analytical methods as for example *reporting*, *trending*, *segmentation*, and *advanced analytics* such as predictive modeling. What distinguishes common analytics from business analytics is the focus of being highly relevant to the business, generating actionable insight and providing performance measurement and value measurement (Stubbs, 2013, p. 6).

Stubbs (2013) and Hodeghatta & Nayak (2017) argue that businesses should be willing to sacrifice model accuracy for ease of implementation and execution. Provost & Fawcett (2013) notes that being able to conduct analytical activities is only one of the fundamental principles – the other is to collect the right data, and investing in data acquisition can generate great pay offs. Provost & Fawcett (2013, p. 11) states that data should be regarded as a strategic asset. In a similar manner, Stubbs (2013) reasons that knowledge derived from analytics is justifiably seen as a competitive advantage to the company that generates it. Some of the fundamental advanced analytics types from data science used to generate insight are *classification* and *class probability*, *regression*, *similarity matching*, and *Clustering* (Provost & Fawcett, 2013; Hodeghatta & Nayak, 2017). These analytical methods can help businesses uncover trends and create deeper understanding of consumer needs.

A commonly followed process for performing business analytics is the Cross Industry Standard Process for Data-Mining (CRISP-DM) as shown in the diagram below. CRISP-DM offers a general overview of the business analytics process.

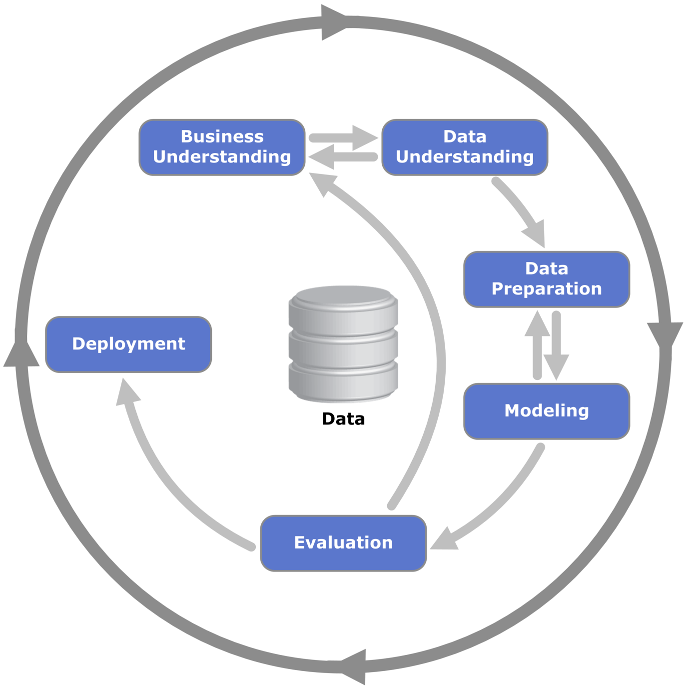

Figure 7 - CRISP-DM model from https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining#/media/File:CRISP-DM_Process_Diagram.png

This is an iterative process where the first iteration is about exploration of the data. The subsequent iterations build upon the findings from previous iterations to come to insights that promote action. An iteration is composed of the following steps described by Hodeghatta & Nayak (2017) and Provost & Fawcett (2013):

1. **Business understanding** – Initial focus must be placed on understanding the problem, objectives, and requirements from the business' perspective (Hodeghatta & Nayak, 2017, p. 91). At this stage creativity plays a large role and carefully considering the desired outcome is of high importance (Provost & Fawcett, 2013, p. 187).
2. **Data understanding** – Data is the raw material for data-driven decision making. Comprehensive understanding of its strengths and limitations is essential and often there is not an exact match between data and business problem (Provost & Fawcett, 2013, p. 28). If data is not available to solve the problem, a process for collecting data is required (Hodeghatta & Nayak, 2017, p. 92).
3. **Data preparation** – Often, data needs cleaning or to be converted before it can be utilized. Typical data preparation involves converting data to tabular form or inferring missing values,

as well as converting data to other data types (e.g. string to integer) (Provost & Fawcett, 2013, p. 30).

4. **Modeling** – Apply data mining techniques to data. Output is some sort of model or pattern recognition.
5. **Evaluation** – Examine the results from the model or pattern. Assess its validity and reliability before moving on. Evaluation should also serve to help ensure that the model is able to solve the problems from the original business goals (Provost & Fawcett, 2013, p. 31).
6. **Deploy –** Implement the model into for instance working software in order to realize returns on investment. Typically, this requires the model to be recoded to fit a production system (Provost & Fawcett, 2013, p. 33).

Applying the complete CRISP-DM cycle to startups may however be too resource intensive. Also, not all startups even have the required knowledge to fully apply this model. Instead, Stubbs (2013) argue that even relatively unsophisticated techniques may deliver quick wins. Sophistication can increase over time as the startup progresses and knowledge is accumulated. The need for flexibility is also key for startups as this allows innovation – but the execution phase does need a higher level of control. Being flexible without regard for control is likely leading to a situation that prevents execution. To mitigate this need for control but remain flexible, an agile process framework can be applied to the overall process.

2.3.5 Data Visualization

At some point, the outcomes of business analytics should be presented to business users such as customer service agents, managers, etc. The data visualization techniques (visual encoding) have a great impact on how we perceive results (Heer et al., 2010). Knaflic (2015) has identified some of the most commonly used visuals that are needed, some of these include: simple text, scatterplot, line, heatmap, vertical, horizontal, and stacked bar. Also, it is noted that we experience cognitive load whenever we try to understand some piece of information. As our cognitive abilities have limitations, it is important to avoid clutter when presenting data – anything that does not increase our understanding should be removed (Knaflic, 2015, p. 73).

3 Methodology

This section presents my approach to conducting research. First, I will present my research philosophy and design, including why these specific approaches have been selected. Second, methods of data collection and analysis are put forth. And last, the validity and reliability of my research is addressed.

3.1 Overview of Research Design

This research is undertaken to offer an understanding of how lean startups can develop and implement a strategy for collecting, storing, and using data to generate actionable insights – this is referred to as a data strategy. The research involves investigating the technical side of lean startups and best practices when writing and using analytics software that supports the data-driven approach the lean startup methodology acquires. At the end, the research concludes in an evaluated IT artifact.

The research process has been structured by first obtaining a comprehensive overview of key elements of the lean startup methodology, agile programming, and business analytics. Theories and concepts surrounding the technical side of LSM has been examined to help creating a preliminary set of data strategy guidelines. This includes literature on agile programming, database management, and data analytics. After the initial guidelines has been proposed, they will be evaluated by entrepreneurs by completing a survey questionnaire combined with an expert interview with a business intelligence manager and lastly a proof of concept.

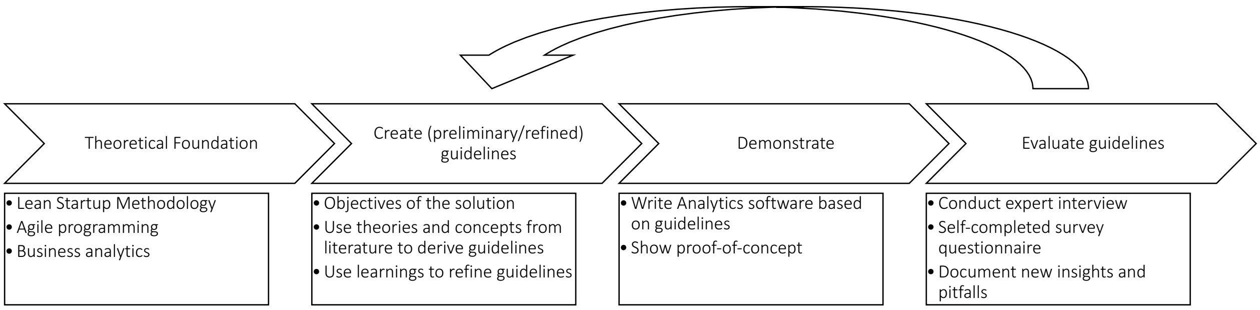

Figure 8 - Overview of Research Design.

The demonstration phase comprises an example of a data strategy by writing analytics code, the collection of data, data analysis. This phase acts as a proof-of-concept which shows the applicability of the guidelines.

As visible from the figure showing an overview of the research design, a cycle concludes with learnings derived from evaluating the data strategy guidelines, which are then used to refine the guidelines.

A detailed account of my approach is established in the upcoming sections that surrounds my research philosophy and design, and later describing how data is collected and analyzed.

3.2 Philosophy

This research is built on a functional pragmatic philosophy due to the practical orientation of the research question, that seeks to answer the question of how to approach the implementation of data strategies in lean startups. This is in line with the view on pragmatism described by Saunders, Lewis, and Thornhill (2012, p. 130), where pragmatism favors concepts that are relevant when they support action.

The purpose of this research is not to establish a universal, objective “truth” that solves all future data strategy implementation issues in lean startups. The purpose is purely to create a deeper understanding how lean startups may—successfully approach—implementing data strategies. The outcome is ideally a set of evaluated guidelines that will help entrepreneurs to better generate actionable insights by consulting collected data. This desired outcome is consistent with how pragmatism is concerned with the instrumental view on knowledge – that the outcome of research should be useful in action to make a deliberate change in practice (Goldkuhl, 2011, p. 140).

This research will not rely on large samples and statistical analysis as seen for example when adopting a positivistic approach to research (Saunders, Lewis, & Thornhill, 2012, p. 140).

3.3 Strategy

The pragmatist stance makes it appropriate to construct the research strategy in a way that intervention into the researched phenomena is suitable, in contrast to merely observing the phenomena (Goldkuhl, 2011) that is advocated by a positivist stance. Hence this research has been designed using an iterative-based research method: design science research (DSR) to build artifacts such as guidelines and recommendations for a data strategy. The benefit of design research is the dual orientation of the method: both contributing to existing knowledge and assist in solving a practical problem (Chatterjee & Hevner, 2010, p. 179).

Peffers et al. (2008) has provided a useful framework that involves six steps to create, improve, and evaluate IT artifacts. These involve problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and lastly communication. According to Peffers et al. (2008), the research can depart from any of the identified steps and move outward. This research follows a problem-centered approach where the basis of the research is founded in an observed problem.

Instead of following the framework rigorously I have merged it with the proposed guidelines presented by Hevner et al. in Chatterjee & Hevner (2010, p. 277). The merging leads me to the following strategy: First, *problem awareness* and motivation that comprises the relevance, motivation, and scope presented in section 1. Second, *problem solution* which incorporates the objectives of the solution (section 5.2) and design and development (section 5.3). Third, *demonstration* to provide proof-of-concept shown in section 6. Fourth, *evaluation of artifact* presented in section 7.

My reasoning is a combination of deductive and inductive reasoning. The research question is a result of investigating literature on data gathering and analysis in the context of lean startups. Surprisingly, my inquiry into this field reveal a lack of research on this specific subject. This is despite the mass amounts of literature on business analytics that have been published since 2012 – around 17.500 publications (according to Google Scholar) which equates roughly 9,5 new articles per day. I start out deductively by seeking out literature on lean startup methodology, agile programming, and an array of data subjects to develop a set of data strategy guidelines. Secondly, I develop the data strategy guidelines with the context of lean startup in mind. And lastly, the guidelines are evaluated in a series of evaluation episodes.

3.4 Approach

In applying design science research, I will elaborate on activities undertaken in each of the steps in my research strategy.

3.4.1 Problem awareness

On a personal level: by having a keen interest in entrepreneurship, software development and data, I searched for literature covering these subjects. Especially the creation of proprietary data analytics software for lean startups was an area of interest. However, the search reveals a lack of information and research in this particular area. In a general context: according to a press release by the European Commission (2013), investing in entrepreneurship is one of the highest return on investment that Europe can make, but about 50 % of newly started businesses fail within the first five years. A better use of IT can significantly increase the survivability of new businesses. Likewise, it is stated in the press release that web-based startups require tailored support measures. During this initial phase of the thesis a research question is formulated and search for related theories and concepts focusing on lean startup methodology, agile programming, and data subjects is conducted.

3.4.2 Problem Solution

When the problem is defined, a solution can be proposed, also referred to as the artifact. The artifact is created after consulting literature related to the problem, but not in the same context as the specific problem. Theories and concepts have been placed in the context of Lean Startup methodology. The initial solution corresponds to the proposed list of data strategy guidelines. Each guideline is informed by its own set of theories and concepts.

3.4.3 Demonstration

Succeeding the creation of artifact, a proof-of-concept is carried out to demonstrate its applicability in the specific context. The demonstration shows how a lean startup may consult the guidelines to lay out a data strategy finding a solution to a specific business problem.

3.4.4 Artifact Evaluation

Evaluating the artifact is a key activity in design science research. According to Venable et. al. (2016), who created the "Framework for Evaluating Design Science Research" (FEDS), the researcher must define *when to evaluate*, *for what purpose*, and *how*. First, when choosing an appropriate evaluation method, I consider two dimensions: the functional purpose (formative or summative) and the paradigm (naturalistic or artificial).

The functional purpose of the evaluation is neither purely formative nor summative. Formative evaluation deals with producing empirically based interpretations of the artifact in order to further improve it. The summative evaluation on the other hand is carried out to judge to what extend outcomes match expectations, e.g. how effective the artifact is. In my evaluation, I am interested in both assessing whether the guidelines are effective (at increasing entrepreneurs' knowledge), and which aspects of the guidelines that need improvement.

How the evaluation is performed can be either naturalistic or artificial. The distinction between the two paradigms is whether the evaluation is performed in the natural environment where the artifact is intended to be used, or if evaluation is conducted in for example a laboratory setting. My approach is mostly naturalistic, as I intend to have actors from the natural environment examine the artifact and then evaluate its performance. However, elements of an artificial paradigm are present, such as carrying out a proof-of-concept. Given I had more time to perform a longitudinal study, I would be able to evaluate its performance first-hand by observing the implementation and use within a startup.

3.5 Data Collection

Three underlying assumptions about practitioners (i.e. entrepreneurs) drive the initial data collection: 1) practitioners perceive data driven approach to be crucial for business success, 2) practitioners lack clarity of how to be data driven in practice, and 3) there is a general lack of knowledge on data strategies in the entrepreneurial community.

I employed a mixed methods approach to collecting primary, empirical data. Non-probability techniques for selecting the samples that received my questionnaires was used. I deem this the most appropriate way for my thesis, as I want to deliver guidelines to a specific target group (entrepreneurs with a lean startup). The first questionnaire was conducted to establish entrepreneurs' own self-assessed data capability – i.e. how capable a startup is in terms of carrying out a data strategy. This also provides me with a baseline for startups' capability. The second questionnaire's purpose was two-fold: 1) assessing if the entrepreneurs' knowledge on data strategies after receiving the guidelines increased and 2) assessing the individual guidelines' performance – for example which guidelines is the most difficult or easiest to understand and why.

The initial questionnaire was distributed in entrepreneurial interest groups on Facebook, and sent directly to my network, as well as people I know have connections to other entrepreneurs. People in my own personal network were asked to pass on the questionnaire to other relevant people as well. My approach is considered to be a mix of two volunteer sampling techniques: snowball sampling and self-selection sampling. When identifying specific entrepreneurs or cases is too difficult, time-consuming, etc., it is appropriate to employ this sampling technique (Saunders et al., 2012).

The sampling technique for the second questionnaire was based on availability. The questionnaire was sent to people who had provided their e-mail in the first questionnaire and wished to further partake in my research.

A qualitative semi-structured interview was conducted with a professional business intelligence (BI) manager in a mid-sized firm with approx. 300 employees. The BI manager has prior to the interview received a copy of the data strategy guidelines and asked to read them thoroughly to assess their performance in startup contexts and to identify areas of improvements. Even though the nature of the interview was mostly semi-structured, a predefined agenda had been put forth.

I consider quantitative data from the survey questionnaires alone to be insufficient due to the lack of deeper understanding that qualitative data from an interview can provide me with. Hence, a thorough evaluation of the constructed guidelines needs both quantitative data to assess by how much (quantitative survey) the guidelines created deeper understanding, as well as why this is (expert interview) – this approach may also be referred to as triangulation (Saunders et al., 2012).

3.6 Data Analysis

Before quantitative data is useful, it needs to be processed and analyzed to derive meaning. Quantitative analysis techniques such as graphs, charts and statistics help explore, present, and describe the data (Saunders et al., 2012).

Primary, quantitative data have been obtained using survey questionnaires. The dataset includes categorical, dichotomous, and ordinal data. Data have been prepared by inputting it into an excel sheet. The initial self-completed survey questionnaire captured data on variables to establish a baseline for startups' data strategy capability. Variables and how they were measured are show below.

Variable	***How it was measured***
Collects data on product performance	By selecting *Yes* or *No* (dichotomous)
Using software to collect or analyze data (3rd party)	By selecting *Yes* or *No* (dichotomous)

Has a proprietary analytics platform to collect or analyze data	By selecting *Yes* or *No* (dichotomous)
Performs data analysis	By selecting which types of data analyses are performed, if any (categorical)
Additional Variable (Not used to determine data strategy capability)	***How it was measured***
Own self-perceived data strategy capability	By selecting the degree of capability (ordinal)

I assess the degree of data strategy capability based on how many of the features above that a startup possesses.

In preparing the quantitative data for the assessment, I had to make a choice to clean answers for the "Performs data analysis" variable. I found that some respondents who selected some type of data analysis at times had also selected "Try to spot patterns by myself", as well as respondents who had selected "No data analysis" had also selected the spotting patterns option. The cleaning involved that everyone who was conducting data analysis and selected the spot patterns option was converted to "Conducting data analysis" – in a similar manner, everyone who had selected "No data analysis" and the spot patterns option, as well as respondents who had only selected the spot patterns option, were converted to "No data analysis" In effect, the categorical data were converted to dichotomous data.

Additionally, the preparation involved coding all variables using numerical codes. During the coding process, two new variables were created. The first has been described above, whereby the "Performs Data Analysis" variable was converted from categorical to dichotomous. The second variable is an ordinal type called "Data Strategy Capability Properties". This variable is comprised of examining each case for the number of data strategy capability properties that the case has.

I have followed the *exploratory data analysis* approach (Saunders et al., 2012) to explore and present data, which emphasizes the use of diagrams to explore and understand my data. This approach allows for flexibility to introduce unplanned analyses – but still with the research objectives in mind. For example, individual, categorical variables have been summarized to find the frequency of each category, and contingencies have been made visible using stacked and clustered bar charts.

Audio-recording qualitative data collected using a semi-structured interview has been transcribed to allow for further analysis. Units of data have been categorized using a concept driven approach whereby data is organized into two main categories: *suggested improvement* or *evaluation*. The two categories serve an internal and external aspect. First, they are meaningful in relation to the data, that being, the interviewee was asked to evaluate and suggest improvements. Second, they are meaningful to the functional purpose of evaluation as described in FEDS as being summative and formative.

3.7 Validity and Reliability

In this section, I will clarify what ensures the reliability and validity of my conclusions drawn from the research.

I will explicate how I designed the two web-based, self-completed survey questionnaires, which act as my main sources of data. When using this data collection method, it is important to remember that the likelihood of being able to contact the respondent again is slim. Therefore, emphasis is placed on constructing the questions in a way that the participants understand the questions in a way that I expect them to understand them (Saunders et al., 2012). For this, I provide some explanatory text below each question, to make sure what is meant by terms or phrases that might be somewhat unclear to the participant, perhaps due to lack of prior exposure to terms from business analytics literature.

To avoiding having answers that do not reflect the respondent's view, I allow for blank-answers (no answer) to all questions. To increase the participants' likelihood of completing the survey, I explain the purpose of the survey up front using a cover letter in combination with keeping the number of questions low.

In terms of validity, the initial questionnaire aims at establishing a baseline for startups data capability, and the second questionnaire aims at evaluating the performance of my designed artifact– i.e. exploring if reading and using them leads to greater understanding of how to conduct data-driven business to generate learnings. The semi-structured interview with the expert is conducted to assess the guidelines applicability and depth – i.e. if they cover the subject in-depth and in a useful way.

To create a baseline for data capability, I first need to clarify what data capability comprises. This process started out with first identifying reoccurring terms and concepts in the literature on data subjects. Often, I found that terms and concepts could be placed into one of three categories: collect data, store data, and use data. I have decomposed data capability into the following variables that I need to examine: 1) if the startup collects data on product performance or use, 2) if the startup has deployed some form of 3rd party data collection software, 3) if the startup has built its own proprietary analytics software, and 4) if the startup performs any kind of data analysis. Data on additional variables have been collect on 1) startups own perceived level of data capability, and 2) if data collection and analysis is perceived by the startup to be critical to business success.

The initial survey was conducted using the free version of SurveyMonkey.com, which only provides me with the option to ask 10 questions. At times, this limitation led me to combine questions that would be more appropriately asked separately.

4 Presentation of Findings

In this section I present the findings from the initial survey, establishing a baseline of startups' own perceived data strategy capability.

4.1 Initial Survey Findings

The initial survey was conducted to gain an understanding of startups' perceived data strategy capability. The degree of how capable a startup is perceived to be, is determined by how many of the following four properties it has: 1) if data on product performance and use is collected, 2) if 3rd

party software is used, 3) if a proprietary analytics platform has been developed, and 4) if data analysis is conducted.

First, I present some general data on respondents. I received 61 responses of which 50,8% had been in business for 0-1 years, 27,9% for 1-3 years, and 21% for 3+ years. Due to my survey questionnaire was distributed to entrepreneurial interest groups online, and a volunteer sampling technique was used, I found it necessary to find the distribution of web based vs. non-web based respondents. This is the number of people conducting business by purely digital means vs those where most business is conducted in e.g. a physical store. I found that the majority of respondents (72%) were web based. However, both types of startups have been included for further analyses because of data strategy capability is not based on the type of business.

4.1.1 Data Strategy Capability

Below I will present my findings which suggest a baseline for the data strategy capability of startups.

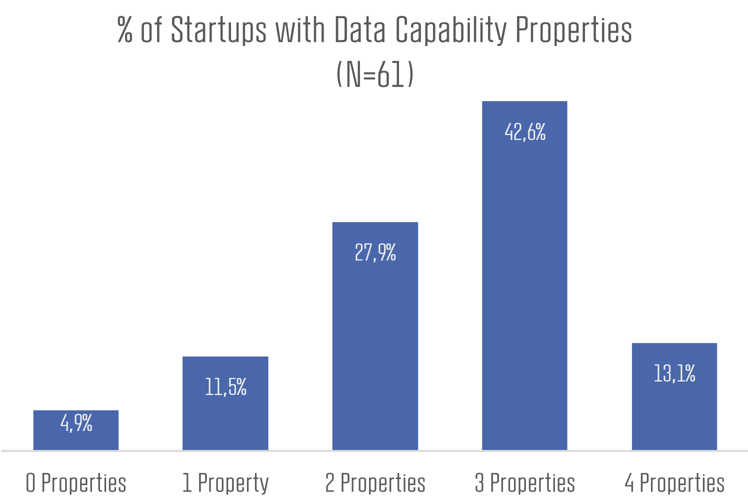

Chart 1 - No. of Data Capability Properties held by Startups, in %.

As mentioned previously, data strategy capability of startups is measured by examining how many of the properties that comprises a data strategy a startup has. As presented in chart 1, findings from the survey shows that close to half (42,6%) have three of the properties. Translating these findings into a degree of capability, I can suggest that 13,1% of respondents are completely capable of carrying out a data strategy, 42,6% are somewhat capable, 27,9% are neither capable or incapable, 11,5% are somewhat incapable, and lastly, 4,9% are completely incapable. Collectedly, startups with zero to two properties make up 44,3% - roughly speaking, only half of the respondents are capable regarding carrying out a data strategy.

A caveat of my approach to measuring data strategy capability: the degree to which each of the properties a startup has have not been examined. This means for example that a respondent who states data on product performance and use is being collect, can do so in a very limited way or do so extensively. My survey questionnaire design was limited due to restrictions from SurveyMonkey.com (only 10 questions were allowed). I will be cautious when drawing conclusions from my initial survey due to imposed limitations in this design.

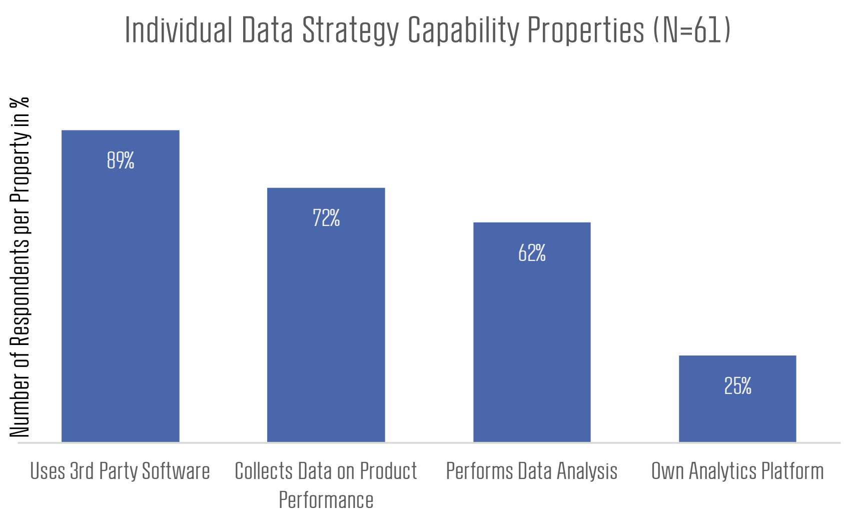

Chart 2 - Chart 2 - No. of Respondents per Property, in %.

Due to the high number of respondents stating having three of the properties, I further investigated which percentage of the properties that were held by the surveyed startups, depicted in chart 2.

I found that most (89%) of respondents are using 3rd party software contrasting to only a quarter using a proprietary analytics platform as well as 72% collects data on product performance. Use of 3rd party software and collecting data on product performance may be closely linked due to 3rd party software often has the capability to collect data on certain metrics (critical or not). The low percentage who has developed a proprietary analytics platform can perhaps be attributed to either lack of knowledge, lack of time, or sufficient value is derived from 3rd party software. Understanding why few have developed a proprietary platform would need further investigation.

Interestingly, my findings show that most (89%) are using 3rd party software, however, only 62% conduct any type of data analysis. This may indicate a lack of knowledge on how to conduct data analysis, collecting unnecessary data, or ineffective or insufficient data collection.

This led me to look into which types of data analyses (if any) were reported by the respondents. Chart 3 illustrates the percentage of respondents that have reported conducting different data analyses – as well as no data analysis, and try to spot patterns.

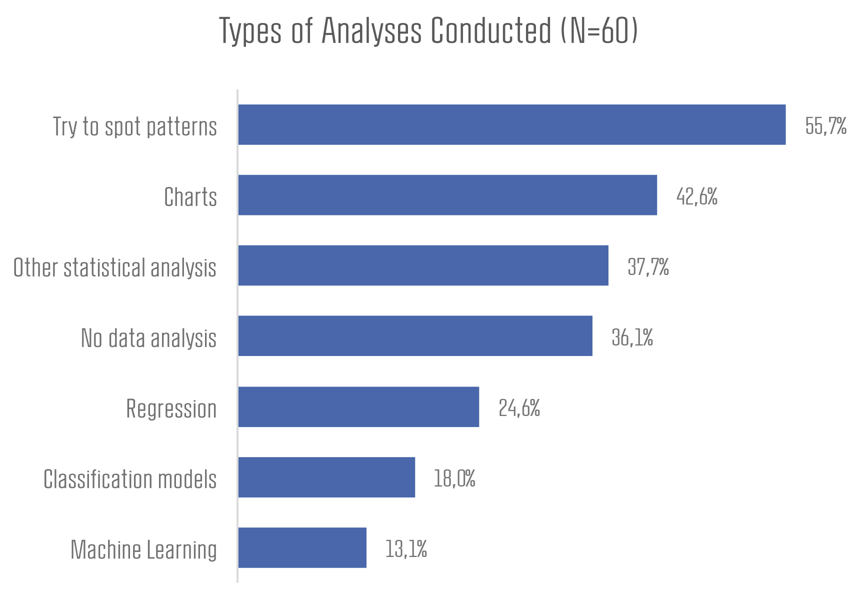

Chart 3 - No. of Respondents per Data Analysis Type in %

The most frequently reported technique is essentially not a type of data analysis. Secondly, a third of respondents reported no data analyses were conducted.

Additionally, 59% (13) of those who reported no data analyses were conduct, also reported trying to spot patterns by themselves. This possibly indicates a desire to become data-driven, but lacks knowledge on how to do so in practice. Supporting this claim, 89% of respondents reported that data collection and analysis is perceived to critical to business success.

4.2 Summary of Findings

Findings from the initial survey questionnaire shows that slightly more than half (55,7%) of respondents are somewhat to completely capable of carrying out a data strategy based on self-reported possessed properties.

The most frequently reported property is use of 3rd party software to either collect or perform data analysis and least reported property is use of proprietary analytics platform. Interestingly, 62% of respondents conduct data analyses whereas 89% reports that data collection and analysis is perceived to be critical to business success. Surprisingly, despite the perceived importance of using analyzing data, the most frequently reported method, which essentially is not a data analysis technique, is trying to spot patterns by themselves.

Overall, the results indicate an interest in data collection and analysis, but may also suggest a lack of knowledge on how to carry out a data strategy in practice, due to only 62% of respondents performs any type of data analysis and only 25% have developed a proprietary analytics platform.

5 Initial Data Strategy Guidelines

I have now examined literature on related topics of data strategy and the context of which the strategy is to be implemented in; lean startups. Based on the literature I will construct a set of guidelines that supports startups' endeavors to base decisions on data rather than pure intuition. In this section, I explicate the creation process, each guideline, its underlying theory, discuss its implication and to what purpose it serves.

5.1 Guideline Context

First, understanding the context that these guidelines have been developed for is vital; Lean startups are operating in uncertainty and rather unstable environments in terms of often lacking a defined target audience, market, and even sources of revenue. Startups frequently arise out of an urge, curiosity, or simply an aspiration to be a business owner and then having to actively search for a market or even create one. Limited monetary resources, time and know-how combined with an experimental approach to business are likewise traits commonly held by startups. Furthermore, products and services provided by lean startups may only attract a narrow, non-mainstream audience.

In startups, resources are likely scarce which does not leave much room for lengthy, money, and time intensive processes such as thorough market research to find the optimal audience or fine-tune a product before launch. Scarcity leaves the entrepreneur to "make-do" with and examine what resources are available to them.

Lean startups are surrounded by a great deal of unknowns and uncertainty. Contrasting to a plan-and-execute approach, lean startups mitigate the risk from uncertainty by experimenting. Ideally, this is done using an iterative process such as the Build-Measure-Learn loop, whereby a minimum viable product is built to verify or reject the entrepreneur's initial value and growth hypotheses. Learnings are derived from data about e.g. customer preferences and behavior. The voice of the customer—both articulated or behavioral—guides the direction of startups. Based on learnings derived from iterations, new and enhanced versions of the product are released to further generate new learnings or even lead to the conclusion that a pivot is needed.

Summarized, the context is lean startups that are characterized by:

- Scarce resources
  - Time, money, know-how, etc.
- Experimental approach to business
  - Create learnings by going through an iterative cycle
- Uncertainty
  - Lacking defined market, target audience, and revenue channel, etc.

5.2 Guideline Objectives

Deriving learnings from the Build-Measure-Learn iteration is a key activity in lean startup methodology, as it informs the focus of subsequent product development and enhancements. However, the lean startup methodology literature provides insufficient guidance on how to establish a strategy for generating learnings. To address this gap, this thesis develops, explicates, and provides data strategy guidelines with emphasis placed on a holistic approach in respect to the BML model. A data strategy considers how to determine critical metrics that are aligned with the startup's business model, how to set up appropriate measuring mechanisms, and the relevance of making data easily accessible and sharable.

In summary, the objective of the artifact is to increase the entrepreneurs' knowledge on how, in practice, to conduct data-driven business and generate actionable insights from data, by laying out a set of guidelines that can easily be understood and adopted. All in the pursuit of deriving learnings.

5.3 Data Strategy Guidelines

I regard the term data strategy to be comprised of the collection, storage, and usage of data to generate actionable insights. Theories and concepts related to each of these components have been consulted to establish a set of guidelines. The data strategy guidelines are generic due to the fact that a data strategy must be tailored for the individual company's business model to be effective. Certain parts of data strategies have been neglected as my focus is mainly on the technological aspect. Other, interesting, and important components such as the organization culture, data policy, and legal aspects – these facets have been neglected in favor of others.

Challenge and purpose	Theory/Concept	Guidelines summarized
Avoid excessive overhead Challenge: - Creating analytics software is time consuming. - Time wasted tracking unimportant metrics. - Poorly understood business objectives.	*Business Understanding* (Hodeghatta & Nayak, 2017; Provost & Fawcett, 2013) *Value and Growth Hypotheses* (Ries, 2011) *Good Metrics* (Croll & Yoskovitz, 2013) *Effectuation* (Sarasvathy, 2001)	**Guideline 1) Understand the Business Problem & Relevant Metrics** - Understand the strategic context. - Set KPIs/metrics according to hypotheses being tested. - Understand how each metric will help answer questions.
Gather necessary data Challenge: - Gathering data without clear purpose leads to unnecessary overhead. - Tracking metrics effectively requires collecting the right data.	*Tailored for individual level* (Parikh, 2014; Oxley, 2014) *Feature metrics* (Bosch & Olsson, 2016) *Sophistication can develop as startups progress* (Stubbs, 2013) *Data warehouse techniques, Star-Schema* (Kimball & Ross, 2013)	**Guideline 2) Appropriate Data Collection & Processing** - Select data sources. - Gather data that will help the startup answer its hypotheses. - Build appropriate database to store collected data. - Data cleaning.

	Business Analytics Life Cycle (Hodeghatta & Nayak, 2017; Provost & Fawcett, 2013)	
Data Accessibility <u>Challenge:</u> - Inaccessible data have limited value to the startup. - Inaccessible data slows the Build-Measure-Learn cycle. - Cognitive overload from presented data	*Sharable & accessible data* (Interview with Porterfield by Firstround.com; Kimball & Ross, 2013; Stubbs, 2013) *Avoid clutter* (Knaflic, 2015)	**Guideline 3) Make Data Easily Accessible & Sharable** - Stick to common data formats. - Allow team access to full dataset. - Create self-service analytics platform. - Visual encoding is important for understanding data.

The illustration below shows a data strategy based on the proposed guidelines applied to the Build-Measure-Learn feedback loop. The BML process is described as having to first complete one phase before moving on to the next – deploying a data strategy, it might at times be necessary to move back in the process to for instance fine-tune the analytics software. The first analytics software deployment may not yield accurate, and useful results and essential tracking parameters may first reveal themselves after consulting the data discovering a lack of variables.

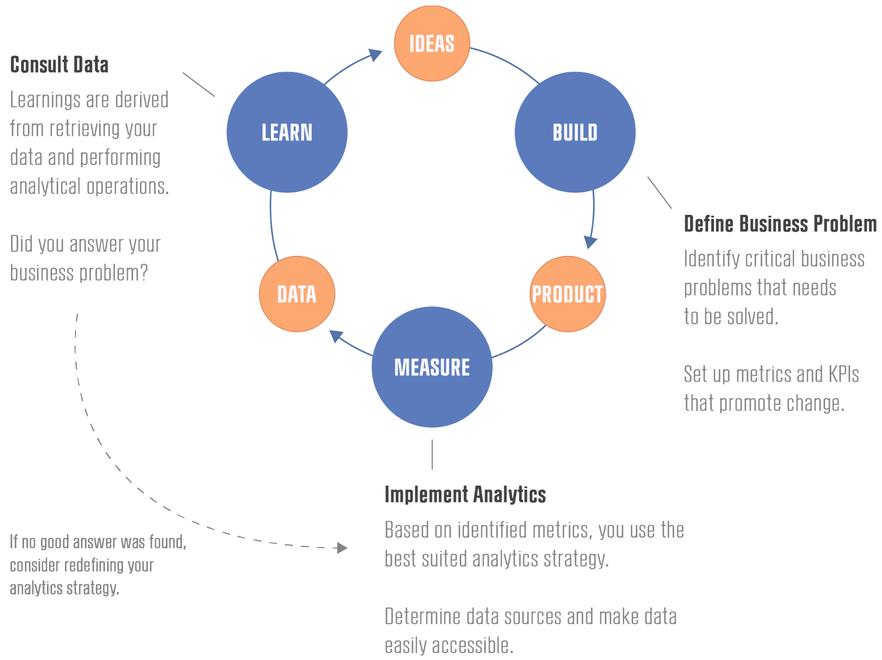

Figure 9 - Data Strategy Applied to BML model

Not all theories consulted in the theoretical foundation are directly contributing to the construction of the artifact, but instead contributes to the overall process in which the artifact was made by providing a thorough understanding of how to customize the theories to fit a Lean Startup methodology.

The final artifact can be accessed using this link: http://nickmillard.com/dsg-v1.pdf

5.3.1 Guideline 1: Understanding the Business & Relevant Metrics

Startups commonly operate in environments characterized by a high amount of uncertainty which is associated with a high level of risk. Changes in the external environment, such as customer preferences and behavior, have a high impact on startups and the need for tracking these changes are vital. Additionally, startups regularly have to discover customer preferences and behavior in the first (Ries, 2011). Typically, change and preference discovery can be recapitulated as having relevance to either the value hypothesis, being the hypothesis of the startup's product providing value to customers, or the growth hypothesis, concerned with how the startup attracts customers.

An outline of the process of getting to understand the business and derive critical metrics is provided: 1) Understanding the business starts with defining the value and growth hypotheses. Breaking down the value and growth hypotheses into quantifiable metrics is fundamental to answering the business problem and fully comprehend its level of complexity. 2) Upfront clarification of go/no-go decisions based on metric results is necessary to enforce a disciplined approach (Ladd, 2016; Croll & Yoskovitz, 2013). 3) Build a minimum viable product designed to test hypotheses using available resources (Sarasvathy, 2001; Ries, 2011). 4) Evaluating metrics after each Build-Measure-Learn cycle is important to ensure the metrics' relevance to business' success. Some metrics' importance might fade as startup progresses, therefore, the evaluation must be in respect to current business problems that need an answer or solution (Stubbs, 2013). 5) The end-of-cycle metrics evaluation must result in adjusted metrics to ensure their continued relevance to business progress.

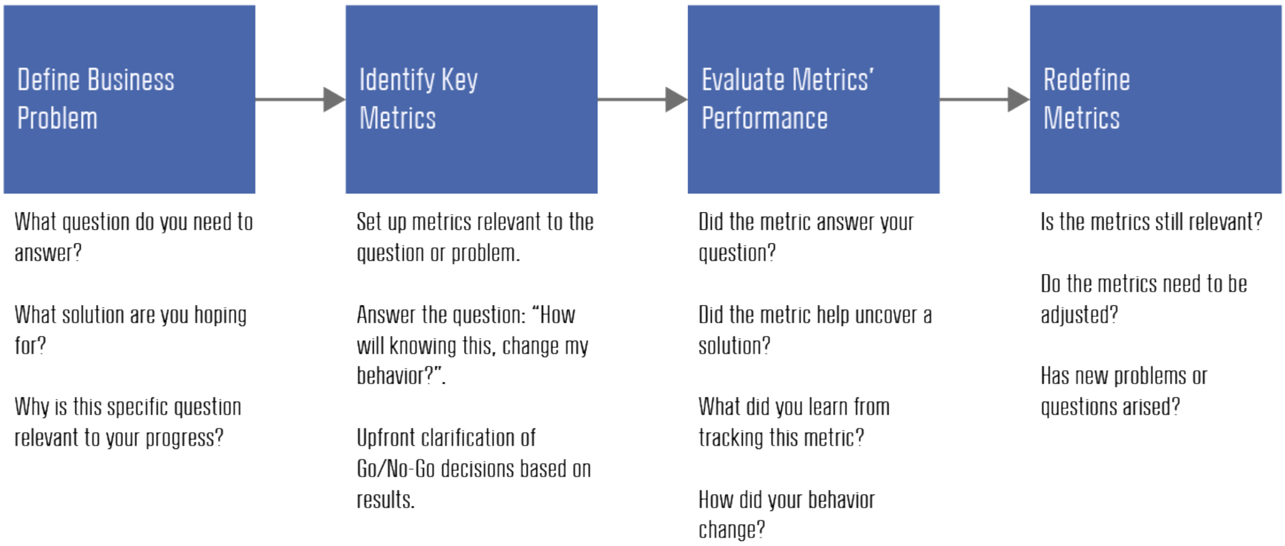

Figure 10 - Understanding the Business Problem & Relevant Metrics

The decomposition of the value and growth hypotheses is facilitated by laying out the components of a business (Croll & Yoskovitz, 2013). This also ensures the metrics' alignment with the startups' goal. Aspects of a business model comprises: acquisition channel, selling tactic, revenue source, product type, and delivery mode. Additionally, designing a product through the process of effectuation may likewise be facilitated by first gaining a comprehensive overview of the initial business model components.

Parallel to identifying key metrics, a startup must make sure that tracking these metrics result in actionable insights – if a metric does not promote any action, then it is likely not a key indicator of performance in terms of business success (Stubbs, 2013; Croll & Yoskovitz, 2013). A good metric is summarized as being: comparable, easily understandable, promotes change, and likely takes the form of a ratio or rate.

The overall purpose of this guideline is to help startups start the process of identifying what really matters – which results in metrics with high level of relevance.

I have the following expectations when applying this guideline:

- By first decomposing the *leaps of faith* assumptions, it becomes easier to identify critical metrics.
- Identifying and tracking critical metrics leads to more informed business decisions.
- Evaluating metrics after each BML loop, ensures only relevant metrics are tracked.

5.3.2 Guideline 2: Appropriate Data Collection & Processing

Once the metrics have been identified, it is of paramount importance to determine which data are needed to accurately measure the metrics. The choice of analytics strategy, e.g. level of granularity and flexibility informs the structure of the data infrastructure, and if proprietary analytics platform is needed (Parikh, 2014; Oxley, 2014). The uncertain environment lean startups find themselves in suggests that a high level of flexibility is required to accommodate rapid changes to the product. Additionally, startups must be willing to sacrifice accuracy in favor of speed and agility. A process that drains the startups' time and money, but yields highly accurate results might not be suitable in a startup context (Stubbs, 2013). Importance is placed on ability to respond quickly to changes in the market, customer preferences and behavior. Therefore, it is suggested that startups should start out with relatively unsophisticated techniques to deliver some quick wins. Sophistication can develop over time as the startup progresses and gains confidence in its data collection and analysis techniques (Stubbs, 2013).

The data collection and storage technique must be designed for the individual startup's strategy, goal, and business model. The development of an in-house analytics platform over a vendor solution allows for a higher level of flexibility, control, and data access. The platform should ideally be developed as early and track as many facets of the product's use as possible (Firstround.com, n.d). Every implemented feature ought to have its behavior, performance, and usage tracked (Bosch & Olsson, 2016).

Tracking metrics, features, and user activity at a granular and individualized level require the appropriate database construction. Tracking data is placed in a relational database whereby it is possible to reference other data tables to create a compounding table. Deploying a database warehouse technique called Star-Schema fits this purpose (Kimball & Ross, 2013). Using this technique involves setting up tables that store a set of dimensions of e.g. a feature, user, session, etc., and compounding selected dimensions from those tables into a single, facts table. In the code base, mapping object properties directly to table columns is facilitated by using an Active Record database access pattern (Schlossnagle, 2004).

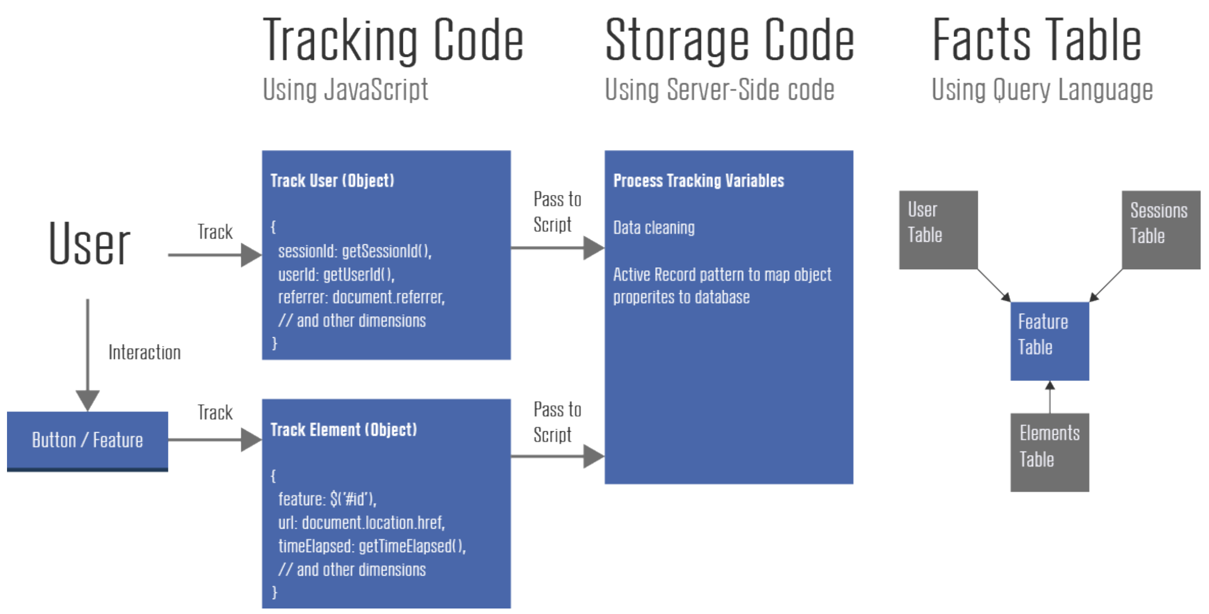

Figure 11 - Data Collection & Processing flow

The diagram above shows how user events are tracked, stored, and later processed in the database. Essentially, the collection and processing flow is broken down to three separate activities. The granularity and flexibility is dictated by the chosen analytics strategy. First, tracking code is implemented to monitor how users interact with the product. In this case, it is assumed that the product resides on a website, and therefore, JavaScript is an appropriate programming language for tracking click-stream events. Second, tracking data are send to a server-side script that processes the data and stores it in a database. Additional dimensions can be added in the server-side code. The database used for analytics will at this point hold dimensions tables. Employing the concept of Star-Schema, a compound table is constructed by selecting metric relevant dimensions from the previous tables.

Before value and validated learning is realized from data, purposive analysis must be performed. Data analysis for the sake of analysis itself without regarding the business problem or question is inconsequential (Stubbs, 2013). Actionable insights are derived when there is a clear purpose with data analysis, hence the following outline is proposed, inspired by Hodeghatta & Nayak's (2013) Business Analytics Life Cycle: 1) Identify business problem and associated metrics derived from steps provided in the first guideline. 2) Retrieve relevant data and clean it. Raw data likely require some preparation before used for further analysis. 3) Explore data to unveil its characteristics and relationships among variables. This leads to the first discovering of insights that can later be used to inform business decisions.

The purpose of this guideline is to provide an understanding of how in practice to select and track metrics, as well as how to store data in a way that user activity can be analyzed at a granular and individualized level.

I have the following expectations when applying this guideline:

- Starting with relatively unsophisticated techniques leads to learnings being derived faster.
- Designing an in-house analytics platform yield superior results compared to a vendor solution (e.g. Google Analytics) due to it allows for a higher level of flexibility.

- Granular and individualized tracking mitigate the risk of having too little data for traditional analytics.

5.3.3 Guideline 3: Make Data Easily Accessible and Sharable

Insights can only generate action if they are shared with the right people, at the right time – and often, innovative insights are derived from people closest to the problem. Everyone in a startup benefits from having easy access to data and results from analyses. Providing a self-service analytics platform may be one way to achieve this. Also, it eliminates the bottleneck of having all data requests going through one person or team (Oxley, 2014). Even if the startup only consists of one person, it is beneficial to create a platform in which the most important metrics, data analysis results, etc. are displayed: easy data access also results in faster production and better products by speeding up the Build-Measure-Learn iterations.

To build an analytics platform some key considerations must be addressed, such as: who needs the data, for what purpose do the person need data, when do the person need data, and how will the person access the data (Firstround.com, n.d.; Stubbs, 2013; Kimball & Ross, 2013). But first, it starts with storing data in an easily accessible manner. Hiding data away in hard-to-reach places is obstructing the value generating and learning process. Making data accessible by e.g. SQL is the bare minimum. Better yet, make the full dataset available on a proprietary website where everyone in the startup can access it. Next step is to provide a "report" overview of all metrics and how they are performing. Additionally, the self-service platform should support the need for customized queries without having to type SQL commands – it is likely that not everyone in the startup has the necessary technical skills to type advanced database queries.

When visually encoding data it is important to do so in a way that reduces clutter the most. Anything that does not add any value in terms of understanding should be avoided (Knaflic, 2015). Graphical perception experiments find that data are easiest to decode when presented as e.g. scatter plots and bar charts (Heer et al., 2010). In keeping with this notion, the analytics platform should present data in the simplest way possible (and not simpler) to enhance understanding.

Building an analytics platform may be a resource intensive endeavor and impose a great deal of overhead. However, it also provides flexibility in terms of the data collection and analysis (Periscope Data, n.d.; Oxley, 2014). A proprietary analytics platform is likely more capable of answering or solving problems than off-the-shelf software (Stubbs, 2013). The overhead may be mitigated by creating the platform over several iterations. In that way, the workload is spread out and learnings about how to best collect and analyze are found in the process as well.

I do have to point out that not all business problems or questions only can be answered by using a proprietary analytics platform. Vendor solutions do have a valid role in collecting data and delivering actionable insights. Especially at first, when focus is solely placed on building an MVP – later, the startup can progress towards their own platform. Here, emphasis is placed on sharing access to the vendor solution and at minimum teach people how to read the different metrics that are presented.

The purpose of this guideline is to emphasize the importance of making data easily available throughout the startup and providing an outline of considerations that must be addressed in order to create an effective analytics platform.

I have the following expectations when applying this guideline:

- Easy data accessibility and sharability leads to learnings being derived faster.
- Addressing key considerations results in a highly relevant analytics platform.
- Building an in-house analytics platform facilitates data access and shared insights.

5.3.4 Adaption of SCRUM to generate learnings

I have now set up three guidelines that focus on different aspects of a data strategy for lean startups. – but without a defined process for how to implement and use these, their value may only be limited to startups. The process of creating and deriving value from an analytics platform needs to support both flexibility that allows for innovation as well as a degree of standardization which enables structure and execution. I propose that balancing flexibility and standardization can be mitigated by using an adaption of Scrum as the process framework, as illustrated below.

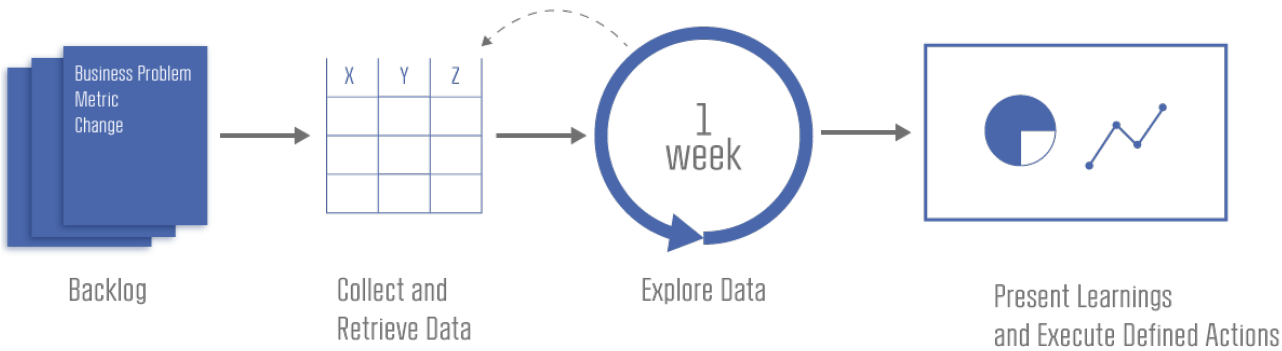

Figure 12 - Analytics Iterations

The above illustration is a proposed process framework to generate actionable insights. In a similar manner as Scrum, this framework does not provide defined techniques for how to identify business problems, retrieve data, explore data or how to present findings and associated actions. Rather, it provides an overall framework that guides the process.

In the same manner as the product backlog in software development contains features, user stories, etc., the backlog here contains the business questions and problems that need answers and solutions. Additionally, I suggest that each business problem in the backlog should contain the associated metric and what will change, knowing the answer to the question or problem. The next step is to collect and retrieve all necessary and related data that might help unearthing a solution. Data may be from an array of different sources – external and internal. As resources in startups are often scarce and emphasis is placed on execution over accuracy, I suggest that an iteration may last one week or less. Ideally, each iteration amounts to new learnings and actionable insights that are shared with the rest of the startup team.

6 Artifact Demonstration

The demonstration is conducted as a proof-of-concept to establish the artifact's feasibility and practical potential. First, a brief, theoretical e-commerce case is presented. Secondly, each guideline is applied to the case.

6.1 Brief Theoretical Case

A simple case should help reveal the feasibility, new insights, and shortcomings of my proposed guidelines. I have chosen an e-commerce startup as the type of business because I assume most have experience with an online store in one way or the other, and this allows for easily recognizing some of questions or issues that such businesses might have. The generic e-commerce business model is relatively simple. An e-commerce startup gets its revenue from charging for products, which they deliver, either electronically or physically. Costs are associated with cost of goods, hosting, payment fees, etc.

For this particular example, we imagine that the e-commerce startup's team consists of two members, one mostly responsible for the business aspect, i.e. accounting, marketing, sales, etc. and the other responsible for the technical aspect, such as website development. It has a small assortment as it is still figuring out which products to offer and uses three types of customer acquisition and sales strategies: ads and interaction on social networks, paid search engine advertisement, and e-mail newsletters that incentivizes purchase.

The theoretical e-commerce startup does currently not have any established data strategy, that is, a strategy for collecting the right data, storing data in an appropriate manner, and make use of the data to drive learnings and data informed decisions. To set up a data strategy for the—very simple—startup, I apply the proposed guidelines.

I have set up a very simple testing website that consists of an index, three product pages, and an analytics dashboard. Purchases are as well simple, they can only contain one product of a quantity between 1-3. The website has been created using PHP7 (mix of object-oriented and procedural approach), html5, CSS3 with the materializecss framework, JavaScript, and MySQL. The website can be accessed at http://nickmillard.com.

Link to the website has been distributed to my network to see if actual insights can be derived from users' interactions on the website, by e.g. simulating being referred from social network sites and search engines, as well as purchasing products.

6.2 Identify Critical Metrics

The first step to establish a data strategy is to gain an understanding of the business and identify critical metrics. According to the Lean Startup Methodology, a startup is based on two main assumptions (Ries, 2011): the value and growth hypotheses. First, these hypotheses must be decomposed into quantifiable metrics. I propose that the decomposition can be facilitated by laying out the components of the initial business model. The e-commerce startup's components look like this:

- **Acquisition channe**l: Advertisement and interaction on *social networks* to create brand awareness and drive sales. *Search engine marketing* used to drive in-bound traffic as well as using paid advertisement on search engines.
- **Selling tactics**: Discounts are used as incentives to convince customers to purchase. The discounts are advertised on social networks, through paid search engine results, and e-mail newsletters. Different campaigns are used to test alternative imagery, text, etc.
- **Revenue model**: Revenue is derived in a straightforward one-time transaction manner.
- **Product type**: Physical products. At this point, the startup is uncertain about which kind of product that will perform the best in terms of sales. Therefore, it only offers few products.
- **Delivery model**: Physical, packed and shipped.

Second, identifying metrics must be done so with respect to these components – but not necessarily all components. This ensures the metrics' relevance and ability to track real progress. Two important aspects must be addressed. First, the metric must promote change and second, go/no-go decisions should be established upfront. The list of identified metrics below has been greatly reduced to only containing few metrics for the sake of simplicity in later steps.

	Metric	**Go/No-go decision**
Acquisition channel	- % of sales per channel	Majority of marketing budget allocated to the best performing channel.
Selling tactic	- % of sales per discount campaign	Stop campaigns with less than 10% sales of all campaign sales.

Product type	-	% of sales per product	Discontinue products with less than 10% of all product sales.

Table 1 - Proof of Concept, identify critical metrics

Identifying the most critical metrics helps inform the subsequent data collection, processing, and analysis.

6.3 Setting Up Appropriate Data Collection & Processing

When metrics have been identified, the next step is to determine which data is needed and how to collect it in order to measure the metrics. It is proposed that ideally as many facets should be tracked. Due to sophistication can develop over time, I have chosen to only set up tracking methods for few dimensions.

Granular and individualized tracking has as well been proposed as necessary for startups to register user behavior. As all visitors will have a session ID, provided that the `session_start()` PHP function has been called, I use the session ID as the common field in all tables. Figure 13 shows a diagram of the database's design that I have created to track the identified metrics. It allows to track sessions, purchases, page views and events (such as button clicks), referrers and if discount codes have been used. Due to the common field provided by the `sessions` table, it is possible to run queries that returns a granular and individualized view of each sessions' behavior and if purchases have been made, as well as if the session came from another site, such as a social network, search engine, and so forth.

Obviously, this database could be normalized even further to split `sessions.referrer` and `transactions.discount_code` into their own tables. However, this example fits the purpose.

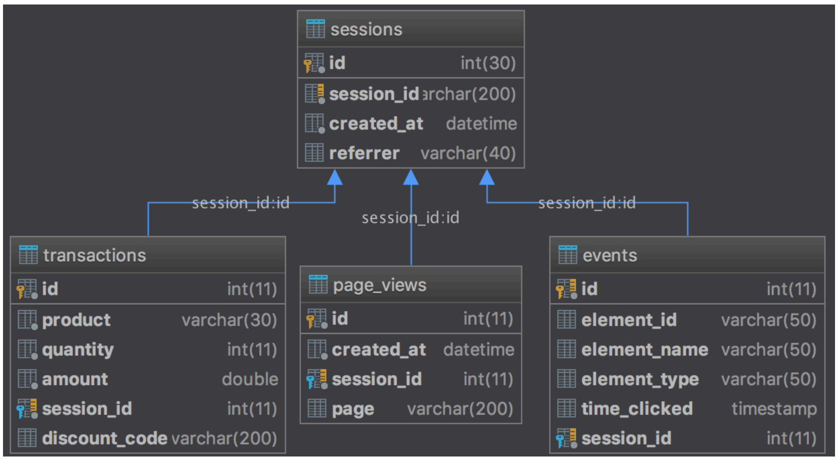

Figure 13 - Proof of Concept, database diagram

As an example, page views are registered using the following PHP and MySQL code:

```
// Update page_views table
$query = "INSERT INTO page_views (page, created_at, session_id)
          VALUES    (
                    '$page', '$date', ( SELECT id
                                        FROM sessions
                                        WHERE session_id = '$sessionId')
                    )";
// Send query to database
$db->query($query);
```

The code is run every time a new page is loaded. This allows me to run queries as the one below to get a granular view of which pages the specific user visited as shown in figure 14.

```
SELECT page, created_at, session_id FROM page_views WHERE session_id = 87;
```

	page	created_at	session_id
1	/thesisdev/index.php?_ijt=7qnnuvmtord528t…	2017-05-05 10:58:09	87
2	/thesisdev/index.php?ref=searchengineref=…	2017-05-05 10:58:15	87
3	/thesisdev/landingpage.php	2017-05-05 10:58:18	87
4	/thesisdev/index.php	2017-05-05 10:58:20	87
5	/thesisdev/product2.php	2017-05-05 10:58:22	87
6	/thesisdev/index.php	2017-05-05 10:58:23	87
7	/thesisdev/product1.php	2017-05-05 10:58:25	87
8	/thesisdev/thankyou.php?quantity=1&produc…	2017-05-05 10:58:28	87
9	/thesisdev/index.php	2017-05-05 10:58:30	87

Figure 14 - Proof-of-Concept, granular view query result

The database has been specifically designed in respect to granularity, individualized tracking, and the metrics. By running slightly more complex queries, it is possible to retrieve results used to track the identified metrics from guideline 1 – this is covered in the subsequent section.

6.4 Making Data Easily Accessible & Sharable

It is proposed that benefits can be derived from making data easily accessible and sharable. Some of the associated benefits are faster Build-Measure-Learn iterations and elimination of bottlenecks. The guideline suggests that an analytics dashboard should accommodate the following: easy access to data and insights, consider the needs of individual startup members, and make queries possible without the need for typing in database commands.

For this theoretical startup, I have developed one business/marketing dashboard that meets the needs of the startup member responsible for business operations. This dashboard shows the current state of the business in terms of identified metrics.

Figure 15 - Proof of Concept, simple business dashboard accessible at www.nickmillard.com/dashboard.php

Now, learnings may be derived faster in terms of which channel that generate most sales, which product has highest demand, and if a campaign is yielding results. If no dashboard was used, each time any of the metrics needed to be consulted, manual database queries would have to be written. Additionally, individual sessions' behavior can easily be explored by all team members.

7 Artifact Evaluation

This section is concerned with evaluating the designed artifact. First, the four-step evaluation process is explicated. Secondly the results from evaluation episodes are presented. This includes three types of evaluation methods: expert interview, survey, and artifact demonstration.

7.1 Evaluation Process

Following the evaluation strategy proposed in FEDS (Venable, Pries-Heje, & Baskerville, 2016, p. 83), evaluating an artifact is a four-step process: 1) explicate the goal of the evaluation, 2) choose evaluation strategy, 3) determine the properties to evaluate, and 4) design the evaluation episodes.

The evaluation is carried out ex-post and follows a combination of both naturalistic and artificial paradigms to assess its performance (summative evaluation) and how to improve the artifact (formative evaluation). First, the goal of the evaluation is to assess the guidelines' effectiveness and efficiency in the natural environment and identifying difficulties and areas for improvements. The choice of evaluation strategy leans mostly towards the Human Risk & Effectiveness Strategy – an appropriate strategy when the majority of risk is related whether the artifact fulfills a need or solves a problem.

Next, the general set of features and properties subject to evaluation have been identified as: 1) the artifact's ability to increase entrepreneurs' knowledge on how to create a data strategy, 2) if entrepreneurs find the guidelines sufficiently describe the process of creating a data strategy, and 3) identify which aspects of the guidelines that are unclear, ambiguous, or in otherwise need improvement.

Lastly, time constraints influence my choice of evaluation episodes. Due to limited time, I can only carry out one evaluation round, ex-post. I have chosen to first provide a proof-of-concept to establish the artifact's feasibility. Each guideline will be applied to a theoretical case and have its own set of success criteria derived from the expected results when applying the individual guidelines. Data on artifact performance and difficulties are collected using a self-completed survey questionnaire distributed to volunteer entrepreneurs. Lastly, areas of improvement are examined by an industry expert with whom a semi-structured interview is conducted. The guidelines are evaluated in the following way: First, a general assessment concerned with the overall impression of the guidelines' applicability and depth. Second, each guideline is evaluated based on its individual performance regarding knowledge creation, applicability, and areas of improvement.

7.2 Evaluating the Guidelines

Multiple evaluation episodes have been undertaken to ensure thorough artifact evaluation. A survey questionnaire was mainly used to assess the performance. An industry expert was interviewed to point out areas of improvement, and, a proof of concept (documentation) to reveal if it is possible, and more importantly, meaningful, to apply the guidelines, as well as finding new insights and shortcomings. During the application of guidelines, a log was kept documenting findings.

Compared to the initial survey, the number of respondents is drastically lower. This may be due to respondents having to read the guidelines prior to responding. Additionally, not all respondents have fully completed the survey.

Evaluation Survey Questionnaire (N = 20)

Position in the startup	Founder/Co-Founder (N = 14) CEO (N = 2) Early Employee (N = 4)
Programming proficiency	Advanced (N = 6) Intermediate (N = 6) Elementary (N = 1) Beginner (N = 4) None (N = 2)

7.2.1 Evaluation of Guideline 1

The first guideline is proposed to emphasize the importance of business understanding and identifying critical metrics that promote action, the foundation on which a data strategy is built. According to the guideline, identifying metrics is facilitated by layout out the components of the initial business model. During the proof of concept phase, I found this approach to metrics somewhat unilateral and simplified to the point that other important metrics may be overlooked. Focus is placed on relating metrics to specific components, and therefore lacks the ability to include metrics that are related to an overall strategy or goal – as pointed out by Verbossen, Business Intelligence (BI) manager at Plato Group:

“You approach it from a critical problem, but you can also approach it from business goals […] Most of the time the goals are about turnover or profit, or maybe cost prediction. Those goals can help you prioritizing what you are going to measure.” - Verbossen, Appendix 11.3

However, I did find the guideline to outline a process which enforces a disciplined approach to metrics, making sure every metric was relevant and had associated outcome in the form of Go/No-go decisions. Likewise, Verbossen pointed out this strength “*What I find very strong, is the example "If a metric does not promote action, it's probably not that important". That's a thing we are implementing right now as well.*“ (Verbossen, Appendix 11.3). In keeping with this, most survey respondents—13 out of 16—reported they learned more about how to break down the value and growth hypotheses and identify critical metrics.

For improving the guidelines, Verbossen suggested identifying metrics using *KPI trees*. In this way, you set up a strategic goal as the root and then add branches of KPIs that influence this goal. Additionally, it was suggested to “*strive to have a balance with metrics in quality, quantity, and timeliness […] if you only think of metric in terms of quantity, then you are not having a good touch on the quality.*” (Verbossen, Appendix 11.3). This point is especially useful in the context of reviewing if newly implemented features are adding real value to users, or if for example the service level is suffering because of too much focus on for instance sales or feature development.

Lastly, in terms of improvements, Verbossen suggested adding the concept of lacking and leading metrics and balancing them to have a better view of past performance and what can still be influenced.

"[…] your financial results, very important, but it's really lacking, it's a result of what we did last month maybe. So, try to balance lacking and leading. For instance, open quotations, that's a leading KPI, because that we can still influence." (Verbossen, Appendix 11.3)

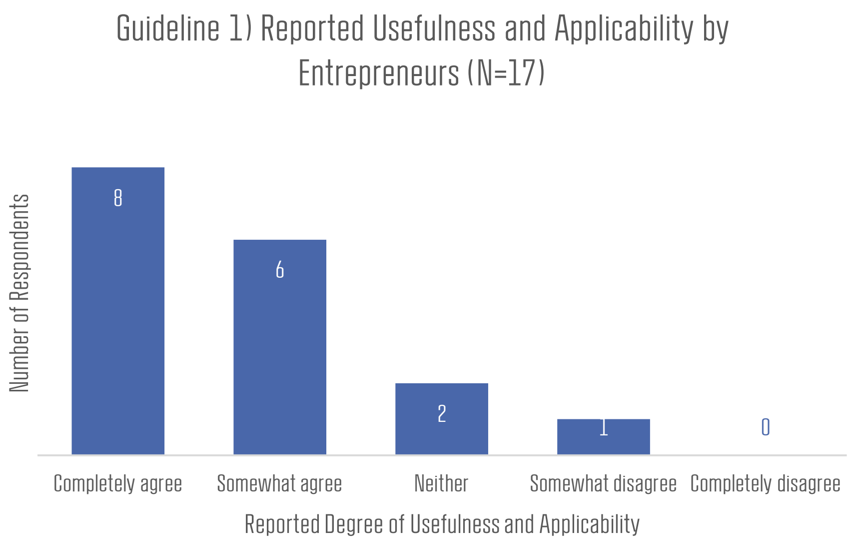

Chart 4 - Guideline 1, Reported Degree of Usefulness and Applicability

Overall, the guideline has been received positively by both Verbossen and the entrepreneurs who completed the survey, as shown in chart 4.

7.2.2 Evaluation of Guideline 2

The second guideline is concerned with how to set up an analytics framework to track and store identified metrics and behavioral user activity at a granular and individualized level.

During the proof of concept phase, I found the approach of starting with relatively unsophisticated techniques to derive learnings beneficial in terms of speed. Initial focus on simple, quick wins simplified the coding – and drifting was avoided. However, the proposed data collection and processing model for websites was found to increase complexity due to the suggestion of using JavaScript for all tracking. Regular PHP and MySQL was sufficient to track the identified metrics.

Surveyed entrepreneurs (14 of 17 respondents) found this guideline insufficiently cover the topic. Similarly, Verbossen commented that "i*f you don't have an analytical background it's some very difficult stuff […] It makes sense to me, talking about star-schema, but maybe for entrepreneurs, it gets scary immediately*". (Verbossen, Appendix 11.3) During the proof of concept development, I too experienced the guidelines to provide insufficient guidance due to a relatively high level of programming and analytical proficiency needed to correctly, or usefully apply the guideline.

The table below presents the respondents (N = 17) reported degree of increased knowledge by the four subjects covered by guideline 2.
Yes = the guideline increased the respondent's knowledge.
No = the guideline did not increase the respondent's knowledge

	Yes	**To some degree**	**Neither**	**To a lesser degree**	**No**
Useful and applicable	4	10	2	0	1
Learned more about setting up granular and individualized tracking	5	8	2	0	2
Benefits and trade-offs related to In-House and Vendor solutions	2	11	3	0	1
How to start data collection process	1	6	4	4	2

Table 2 - Reported increase of knowledge by Respondents

Overall, by examining the respondents' answers, the knowledge on the subject has generally increased. However, more in-depth coverage of complex concepts, such as star-schema is perhaps needed.

The guideline has focused greatly on benefits and trade-offs of developing a proprietary analytics platform or choosing vendor solutions. Verbossen has noted this as a good aspect of the guideline, however, guideline 2 is in favor of an in-house approach, whereas Verbossen strongly emphasizes that startups are better of going with a vendor solution – at least data differentiation has greatly increased (that is data from multiple sources).

"You also raised the good question of doing it in-house or use a vendor. I think that also very much depends on the difficulty and differentiation of their data. If you are solely a website e-commerce platform, I think 90% can go for a standard vendor approach […] I believe you should stay away from building it yourself as long as you can." (Verbossen, Appendix 11.3).

7.2.3 Evaluation of Guideline 3

The last guideline is about making data easily accessible and sharable. Emphasis is placed on considerations when making e.g. dashboards to facilitate faster learnings.

During the proof of concept, I found this guideline particularly useful due to making metrics readily available. Addressing the considerations did provide further insights into how to build the—simple—analytics platform. However, it does require a relatively high level of programming proficiency.

Verbossen (Appendix 11.3) noted this as the clearest guideline and the one that entrepreneurs might find the most applicable because their wish to see results. He also suggested to add setting up notifications to metrics on the dashboard as a part of the data presentation, so the entrepreneur would get a "heads up" when e.g. a goal was plus or minus five percent. This is opposed to having to read reports of for instance 20 metrics when only a few are interesting in that moment.

Secondly, Verbossen proposed instead of entrepreneurs developing their own proprietary platform, they should *"look for tools that are widely spread, so that you have a community behind it. For instance Tableau, you have a big community, so if you have a question you can look online in the community"* (Appendix 11.3).

Verbossen (Appendix 11.3) suggests improvements that relate to addressing the question "should I focus on my core business or should I do the analytics part by myself as well?". There is a great deal of overhead related to building a proprietary platform – which I found as well. Overall, it is a time-consuming process because of the need to build the platform, incorporate metrics, write database commands, ensure high quality results are passed back, and defining the presentation of results.

15 of 17 surveyed entrepreneurs found this guideline to increase knowledge on benefits of making data easily accessible and sharable and 13 of 17 found it to be useful and applicable to their startup. However, only 5 of 17 found the guideline sufficiently covers the topic.

8 Discussion

This section provides a discussion of my empirical findings and process, which includes reflections, possible artifact improvements, the contribution of knowledge to the wider body of research on lean startup methodology, and lastly, future research.

8.1 Possible Artifact Improvements

Areas of improvements have been pointed out during the expert interview and identified when applying the guidelines to a theoretical case. This led me to further consult the literature on business analytics to substantiate and strengthen the guidelines' rigor and applicability.

Guideline 1 provides a limited approach to the process of identifying metrics as I found during proof of concept, as well as noted by Verbossen (Appendix 11.3). Verbossen's notion of using KPI trees with respect to a startup's goals and strategy may prove as a useful addition due to a more comprehensive approach that includes important metrics that otherwise would be overlooked. Metrics inform e.g. data infrastructure, but without a strategic context, organizations cannot decide which data to focus on (Acito & Khatri, 2014).

The guideline's structured way of understanding the business and its relevant metrics may pose both as its weakness and strength. Strength in terms of offering a structured approach – however, it diminishes the value of ingenuity and creativity. Structure can constrain innovation and ability to see problems from different perspectives. Ingenuity and creativity plays a large role in understanding business (Provost & Fawcett, 2013) and fosters innovative approach to derive learnings (Stubbs, 2013).

Furthermore, Verbossen (Appendix 11.3) suggested finding new metrics by data exploration. Through data analysis, a startup might spot patterns and correlations that impact future performance – such findings can be standardized, i.e. instead of having to conduct manual data analysis, you automate the approach by which you found the metric (for instance by saving an advanced query or model). Such metrics are nearly impossible to articulate up front as they are hidden in the startup's data.

Guideline 2 suggests a way to setup appropriate data collection and processing, encouraging the development of a proprietary analytics platform tailored to the individual business.

Both Stubbs (2013) and Verbossen (Appendix 11.3) advocates for holding back with developing custom analytics platforms until data differentiation has reached a point whereby vendor solutions no longer provide enough value. However, Stubbs has also pointed out that using-in built, already configured analytics hinder differentiation. Business analytics is said to have the ability to create competitive advantage, but when everyone does the same thing (e.g. use Google Analytics), differentiation is impossible. In the context of lean startups, a compromise can be made by creating custom dashboards using vendor solutions. In this way, even when using tools everyone else have access to, it can provide business tailored insights to drive learnings and further product enhancements. Verbossen suggests using 3rd party solutions with a broad community. In terms of guideline improvements, it may be feasible to incorporate this aspect of customizing existing solutions.

In efforts to improve guideline 3, Verbossen (Appendix 11.3) suggests setting up notifications that alerts entrepreneurs when they are drifting off target or has exceeded it. Automating the monitoring process adds value to the startup by freeing up time that would otherwise have been spent checking reports manually (Stubbs, 2013). Sensitivity thresholds for when alerts are triggered must be balanced, otherwise you might run the risk of having alerts being set off so frequently that they start to appear “whiny” (Croll & Yoskovitz, 2013), and if too insensitive, the alert is being triggered after preventative action could have been taken.

8.1.1 Separation of Concern

As of now, each guideline covers a wide area and are more or less dependent on each other. For instance, guideline 1 covers getting to understand the business problem, how to identify metrics, and what comprises a useful metric. Guideline 2 is even more ambitious and not only covers more subjects, but also slightly more complex ones that require technical proficiency. Feedback from surveyed entrepreneurs reveals that topics covered by the guidelines were insufficiently covered.

The concept *separation of concern* is used within programming literature: a design principle that aims at separating code into distinct sections. Another benefit of this approach is modularity. The concept may have proved useful when creating the artifact by ensuring only one subject or topic was covered within each guideline. This allows for thorough and in-depth artifact that leads to higher performing guidelines. Additionally, using this design principle when designing the artifact would likely lead to specialized and modular guidelines.

8.2 Knowledge Contribution: to Practice and Research

First, business analytics are difficult. What is even more difficult is to take relatively complex concepts and boil them down, repurpose them to fit the context of lean startups, and serve them to an audience that might not be proficient in the "art" of analytics or programming. I truly experienced Stubbs (2013, p. 11) notion of difficulty in explaining business analytics in any other way than "it creates better outcomes".

This thesis has been conducted using design science research, which is fundamentally a problem-solving paradigm that aims at creating artifacts to solve problems or provide better solutions. An artifact denotes a construct that is created by people as opposed to occurring naturally. The artifact must be designed and evaluated (Hevner & Chatterjee, 2010).

This thesis claims current literature on Lean Startup Methodology provides insufficient guidance on how to approach data collection, storage, and use to generate actionable insights and learnings. There is a need for a more developed approach to deriving learnings during the build-measure-learn iterations. Hence, an artifact that outlines a strategy for generating learnings has been developed, explicated, and evaluated.

In efforts aimed at understanding how my research contributes to the knowledge creation within the body of research on Lean Startup Methodology, I have adopted Gregor's (2006) taxonomy on types of theories in Information Systems and Gregor & Hevner's (2013) knowledge contribution framework. The taxonomy and DSR knowledge contribution framework provide a shared vocabulary necessary to classify and explicate what encompasses contribution to knowledge and its position in design science research. Importance is placed on the balance of scholarly contributions in the form of design, generalization, and theorization, and practical contribution in the form of usefulness to the community of practice (Beck, Weber & Gregory, 2012).

Using Gregor's taxonomy, my research is classified as theory for design and action, which is characterized by articulating a way *how to do* something. Theory for design and action includes principles of form, function, and justificatory theory. Contributions toward creating data strategy guidelines for deriving learnings have been made by incremental artifact construction and considered to be positioned partially in both the *improvement* and *exaptation* quadrant of the knowledge contribution framework (discussed later). Knowledge contribution for this theory category is proposedly evaluated by addressing the following criteria: utility to a community of users and novelty of the artifact. Contribution regarding the artifact can also include the evaluation of completeness, simplicity, ease of use, and quality of results (Gregor, 2006).

First, when discussing implications for the community of practice, it is fitting to briefly outline the objective of the guidelines with respect to practice: I strive to develop a greater understanding of how to implement a data strategy in startups to derive learnings, and ultimately, better startups and products. According to a press release from the European Commission (2013), investing in entrepreneurship is one of the activities that bring the highest return on investment, however, 50% of startups fail within the first five years. Furthermore, claims have been made that 98% of new products fail (Bosch et al., 2013). Educating entrepreneurs how to, in practice, validate their *leaps of faith* assumptions—i.e. how they think value and growth is generated—the rate of failure in respect to either product launch or business in general may drop, which is beneficial to the general society in terms of workplaces, know-how, and offerings. Contributions toward increasing entrepreneurs' knowledge on the matter have been made by first developing a deep understanding of the problem environment by reviewing literature and justificatory theories from lean startup methodology, agile programming principles, and dispersed literature on business analytics. Based on this, I argue the designed artifact has high relevance for the community of practice that its intended to help. With the knowledge on how to measure, validate, or reject initial assumptions, entrepreneurs will be more likely to create startups and products that succeed.

The initially designed artifact is merely a nascent design theory with an ambitious goal. Efforts to improve its design and applicability has been made, but there is still room for plenty more improvements. I recognize the limitations of the proposed artifact, which are reflected upon in the subsequent section. Based on evidence from an evaluation of the artifact—in terms of usefulness, applicability, and subject depth—by self-selected members of the entrepreneurial community, I argue there is a convincing degree of utility to the intended community.

Novelty of the artifact is worth considering when evaluating contribution to knowledge. An in-depth analysis of existing literature and theories must be undertaken to identify a research gap. Startups following the Lean startup methodology is a relatively new branch of startups. The methodology was popularized by Ries (2011) with his seminal work "The Lean Startup" and has since received much attention throughout the entrepreneurial and business management community. Due to LSM's novelty, there is an obvious lack of research in this domain. The challenge of deriving learnings has been addressed by few authors (Croll & Yoskovitz, 2013; Ries, 2011), however, no IT artifact has been designed to provide a concrete solution, as far as I could find during my initial theory discovery. I argue filling this research gap is highly relevant to LSM as learning through collecting, storing, and analyzing data is at the heart of LSM. Novelty of the designed artifact emerges from its attempt to offer a concrete approach to the challenge.

Lastly, completeness of the artifact is a key evaluand in knowledge contribution. First, the artifact designed during this thesis is considered a first attempt and has been designed as an incremental construction that needs to undergo further development to achieve its goal. Second, due to every startup's situation is unique, the artifact only provides a generic approach. These two factors do limit the overall completeness of the artifact. The scope of the thesis was offer an artifact that address known problems by providing a new solution based on existing literature and theories.

8.3 Future Research

As working with this thesis, the artifact design, and its evaluation progressed, an array of interesting, possible future research approaches came into light: 1) Additional research can be undertaken to incorporate the areas of improvements that have been identified. 2) Implementation of the guidelines into an actual startup to monitor the artifact's performance and the startup's subsequent performance. 3) Longitudinal study of the same, or even 5) study the implementation from a programming perspective.

8.4 Reflections and Limitations of the Artifact

The initial survey guided my choice of focusing on proprietary analytics platform because of the interesting discovery that only 25% of startups had developed a custom analytics solution. An argument can be made that my focus was slightly misplaced as I paid little attention to the relevance—in terms of what entrepreneurs perceive to be relevant—when selecting which problems to solve and solutions to provide. The low number of entrepreneurs which develop their own proprietary platform may be attributed to the majority perceiving such undertakings involve excessive and unnecessary overhead. Despite my interest in and focus on lean startups with the implied lean thinking, I have approach my artifact creation in a linear, plan-and-execute manner. An artifact of higher relevance would likely have been designed if formative evaluation episodes had been conducted when research was first commenced to ensure relevance as well as problem maturity.

Furthermore, the initial survey revealed entrepreneurs do employ methods to collect data, but may lack the analytical skills to derive meaningful learnings from their data. A more productive guideline, from the viewpoint of entrepreneurs, might have been to layout the process of analyzing data – as Verbossen (Appendix 11.3) also noted when evaluating the second guideline.

The result of exploring data strategy from the perspective of developing proprietary analytics platforms led the artifact to target, or to be considered useful, only to a subsection of entrepreneurs with lean startups—those with a relatively high degree of programming and analytical proficiency. The guidelines may not prove to "work" under the conditions of low programming proficiency.

Additionally, emphasis was placed greatly on flexibility in guideline two and three. Following the guidelines may lead to poorly constructed databases and increased complexity by creating ad-hoc tables and dimensions due to the need of supporting a high level of flexibility. As ad hoc tables are created and vast amounts of dimensions are tracked complexity increases. This potentially results in a lack of resources later because of time spend maintaining an ever-increasing complex system.

9 Conclusion

This research was motivated by the fact that lean startups are inherently experimental and data driven in terms of seeking validated learnings through the Build-Measure-learn cycle. Deriving learnings is a key activity in the learn startup methodology (LSM), however, this thesis claims current literature on LSM provide insufficient guidance on how to approach data collection, storage, and use to generate actionable insights and learnings. Hence an attempt to bridge this research gap was made by developing, explicating, and evaluating an artifact that outlines a strategy for generating learnings.

The process of designing the artifact started by consulting justificatory theories and literature by authoritative authors from the lean startup methodology, agile programming, and business intelligence. The designed artifact features three guidelines on how to implement a data strategy (collection, storage, and use of data):

1. Understand the Business Problem & Relevant Metrics
2. Appropriate Data Collection & Processing
3. Make Data Easily Accessible and Sharable

The three guidelines were evaluated during multiple summative and formative evaluation episodes that include survey questionnaire to assess the artifact’s performance by the entrepreneurial community, an expert interview to mainly identify areas of improvements, and a proof-of-concept to reveal its feasibility and whereby practical insights and pitfalls were documented.

The first guideline aimed at laying out a structured approach to identifying critical metrics by investigating the startup's business model and its components. From the evaluation episodes, I found the approach somewhat unilateral and simplified which was likely to overlook important metrics that was not necessarily derived from the business model, but the startup's overall goal and strategy. On a positive note, it was also found to enforce a disciplined approach ensuring the relevance and had clear go/no-go decisions associated with outcomes. The second guideline was concerned with setting up appropriate data collection and processing techniques. The proof of concept found that quick wins were facilitated by starting out with relatively unsophisticated techniques. Nonetheless, it was also found that the guideline requires a high level of analytical and programming proficiency which may limit its applicability in some situations. The last guideline was about making data easily accessible and sharable. Emphasis was placed on considerations when making e.g. dashboards to facilitate faster learnings. The guideline was considered to be particularly useful due to making metrics' performance readily available. Additional improvements were suggested, such as setting up notifications (also referred to as alerts) which may reduce the time required to overlook reports and provide the ability of *alerting* the entrepreneur when the startup is starting to drift away from targets.

Contributions to knowledge were derived from the artifact's attempt to bridge the research gap by offering a concrete approach to the challenge of collecting, storing, and using data in lean startups to derive learnings. However, only limited empirical data was collected due to time constraints, and therefore the artifact cannot be considered as a full theory for design and action, as it needs further testing in the actual environment. Despite the artifact was mere a first attempt, it provided some interesting paths for future research on how to create better, and more informed startups.

10 References

Acito, F., & Khatri, V. (2014). Business analytics: Why now and what next? *Business Horizons*, *57*(5), 565-570. doi:10.1016/j.bushor.2014.06.001

Beck, K., Beedle, M., Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., … Grenning, J. (2001). Manifesto for Agile Software Development. Retrieved March 21, 2017, from https://agilemanifesto.org/iso/en/principles.html

Beck, R., Weber, S., & Gregory, R. W. (2012). Theory-generating design science research. *Information Systems Frontiers*, *15*(4), 637-651. doi:10.1007/s10796-012-9342-4

Berry, T. (2012, July 10). Lean Startups Need Business Plans, Too. Retrieved from https://www.entrepreneur.com/article/223878

Blank, S. (2010, March 4). Perfection By Subtraction – The Minimum Feature Set. Retrieved from https://steveblank.com/2010/03/04/perfection-by-subtraction-the-minimum-feature-set/

Blank, S. (2013, May). Why the Lean Start-Up Changes Everything. *Harvard Business Review*. Retrieved from https://hbr.org/2013/05/why-the-lean-start-up-changes-everything

Bosch, J., Holmström Olsson, H., Björk, J., & Ljungblad, J. (2013). The Early Stage Software Startup Development Model: A Framework for Operationalizing Lean Principles in Software Startups. *Lean Enterprise Software and Systems*, 1-15. doi:10.1007/978-3-642-44930-7_1

Bosch, J., & Olsson, H. H. (2016). Data-driven continuous evolution of smart systems. In *SEAMS '16 Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems* (pp. 28-34). Retrieved from http://dl.acm.org/citation.cfm?doid=2897053.2897066

Clifton, C., & Thuraisingham, B. (2001). Emerging standards for data mining. *Computer Standards & Interfaces*, *23*(3), 187-193. doi:10.1016/s0920-5489(01)00072-1

Cohn, M. (2014, July 22). Thinking Critically About Scrum. Retrieved from http://www.mountaingoatsoftware.com/blog/my-primary-criticism-of-scrum

Cooper, R. G. (1990). Stage-gate systems: A new tool for managing new products. *Business Horizons*, *33*(3), 44-54. doi:10.1016/0007-6813(90)90040-i

Cooper, R. G., & Sommer, A. F. (2016). Agile-Stage-Gate: New idea-to-launch method for manufactured new products is faster, more responsive. *Industrial Marketing Management*, *59*, 167-180. doi:10.1016/j.indmarman.2016.10.006

Croll, A., & Yoskovitz, B. (2013). *Lean analytics: Use data to build a better startup faster*. Sebastopol, CA: O'Reilly.

European Commission. (2013). *Press release - Entrepreneurship as a main driver for economic growth*. Retrieved from European Commission website: http://europa.eu/rapid/press-release_MEMO-13-5_en.htm

Fawcett, T., & Provost, F. (2013). *Data science for business*. Sebastopol, CA: O'Reilly.

Firstround.com. (n.d.). Your Data Is Your Lifeblood — Set up the Analytics It Deserves. Retrieved April 5, 2017, from http://firstround.com/review/your-data-is-your-lifeblood-set-up-the-analytics-it-deserves/

Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. *interactions*, *19*(3), 50-59. Retrieved from http://delivery.acm.org.esc-web.lib.cbs.dk/10.1145/2170000/2168943/p50-fisher.pdf

Fitzgerald, B., Conboy, K., Power, K., Valerdi, R., Morgan, L., & Stol, J. (Eds.). (2013). *Lean Enterprise Software and Systems: 4th International Conference, LESS 2013, Proceedings*. Galway, Ireland.

Gage, D. (2012, September 20). The Venture Capital Secret: 3 Out of 4 Start-Ups Fail - WSJ. Retrieved from https://www.wsj.com/articles/SB10000872396390443720204578004980476429190

George, J. A., & Rodger, J. A. (2010). *Smart data: Enterprise performance optimization strategy*. Hoboken, NJ: Wiley.

George, M. O. (2013). *The lean six sigma guide to doing more with less: Cut costs, reduce waste, and lower your overhead*. Hoboken, NJ: Wiley.

Goldkuhl, G. (2011). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, *21*(2), 135-146. doi:10.1057/ejis.2011.54

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611-642. Retrieved from http://www.jstor.org/stable/25148742

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337-355.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, *53*(6), 59. doi:10.1145/1743546.1743567

Hevner, A., & Chatterjee, S. (2010). *Design research in information systems: Theory and practice*. doi:10.1007/978-1-4419-5653-8

Hines, P., Holweg, M., & Rich, N. (2004). Learning to evolve. *International Journal of Operations & Production Management*, *24*(10), 994-1011. doi:10.1108/01443570410558049

Hodeghatta, U. R., & Nayak, U. (2017). *Business analytics using R - a practical approach*. doi:10.1007/978-1-4842-2514-1

Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Indianapolis, IN: Wiley.

Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professiona*. Hoboken, NJ: John Wiley & Sons.

Ladd, T. (2016, March 7). The Limits of the Lean Startup Method. *Harvard Business Review*. Retrieved from https://hbr.org/2016/03/the-limits-of-the-lean-startup-method

Minelli, M., Dhiraj, A., & Chambers, M. (2013). *Big data, big analytics: Emerging business intelligence and analytic trends for today's businesses* [Paper]. Retrieved from http://ebookcentral.proquest.com.esc-web.lib.cbs.dk/lib/kbhnhh-ebooks/reader.action?docID=818100

Nayak, A., Poriya, A., & Poojary, D. (2013). Type of NOSQL Databases and its Comparison with Relational Databases. *International Journal of Applied Information Systems*, *5*(4), 16-19.

Nerur, S., Mahapatra, R., & Mangalaraj, G. (2005). Challenges of migrating to agile methodologies. *Communications of the ACM*, *48*(5), 72-78. doi:10.1145/1060710.1060712

Ojala, A. (2016). Business models and opportunity creation: How IT entrepreneurs create and develop business models under uncertainty. *Information Systems Journal*, *26*(5), 451-476. doi:10.1111/isj.12078

Olsson, H. H., & Bosch, J. (2015). Towards Continuous Customer Validation: A Conceptual Model for Combining Qualitative Customer Feedback with Quantitative Customer Observation. *Lecture Notes in Business Information Processing*, 154-166. doi:10.1007/978-3-319-19593-3_13

Oxley, S. (2014, August 20). Building a Custom Tracking and Analytics Platform: Information is Power. Retrieved from https://engineering.groupon.com/2014/misc/building-a-custom-tracking-and-analytics-platform-information-is-power/

Parikh, R. (2014, May 25). The Single Most Useful Analytics Strategy for Early-Stage Startups - Heap Blog. Retrieved from https://blog.heapanalytics.com/the-single-most-useful-analytics-strategy-for-early-stage-startups/

Parker, G. G., Van Alstyne, A. M., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. New York, NY: W. W. Norton & Company.

Pasanen, J. (2014, September 25). Jussi Pasanen on Twitter: "Minimum Viable Product". Retrieved March 17, 2017, from https://twitter.com/jopas/status/515301088660959233

Periscope Data. (n.d.). Building Your Own User Analytics System In SQL. Retrieved April 5, 2017, from https://www.periscopedata.com/blog/how-to-build-your-own-user-analytics-in-sql.html

Rich, S. R., & Gumpert, D. E. (1985, May). How to Write a Winning Business Plan. *Harvard Business Review*.

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. New York, NY: Crown Business.

Sarasvathy, S. D. (2001). Causation and Effectuation: Toward a Theoretical Shift from Economic Inevitability to Entrepreneurial Contingency. *The Academy of Management Review*, *26*(2), 243-263. doi:10.2307/259121

Saunders, M. N., Lewis, P., & Thornhill, A. (2012). *Research methods for business students*(6th ed.). Harlow, England: Pearson Education.

Schlossnagle, G. (2004). *Advanced PHP programming: A practical guide to developing large-scale Web sites and applications with PHP 5*. Indianapolis, IN: Sams Pub.

Stubbs, E. (2013). *Delivering Business Analytics: Practical Guidelines for Best Practice*. Hoboken, NJ: Wiley.

Toren, M. (2015, June 5). How to Craft a Business Plan That Will Turn Investors' Heads. Retrieved from https://www.entrepreneur.com/article/223814

Towill, D., & Christopher, M. (2002). The Supply Chain Strategy Conundrum: To be Lean Or Agile or To be Lean And Agile? *International Journal of Logistics Research and Applications*, *5*(3), 299-309. doi:10.1080/1367556021000026736

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77-89. doi:10.1057/ejis.2014.36

11 Appendix

11.1 Initial Survey Design

A PDF presenting the initial survey can be found here:

https://www.dropbox.com/s/pxv7v0fpriqrv11/initial_survey.pdf?dl=0

11.2 Evaluation Survey Questionnaire Design

A PDF presenting the evaluation survey can be found here:

https://www.dropbox.com/s/lukv00z8xjdysz1/second_survey.pdf?dl=0

11.3 Expert Interview Transcript & Audio file

The audio file can be accessed using the link below:

https://www.dropbox.com/s/19yl3un7eyi2y8w/Joeri%20Interview.m4a?dl=0

Interview conducted with Joeri Verbossen, Business Intelligence manager in Plato group. Conducted 5th may 2017 using Skype video call.

Verbossen had received a document containing the type of questions before it was carried out, and instructed to read, evaluate, and find areas of improvement for the proposed guidelines.

Transcript	**Speaker**

So first, the interview is about evaluating the guidelines, but mostly about areas of improvements for the guidelines. Because I also have entrepreneurs who have filled in a survey questionnaire about how well they think the guidelines perform... like if they can use them or not. So this is mainly to find areas of improvement.	NM
Okay, shall we walk through it and make remarks where I think... or where I have some questions or stuff that can be added.	J
First of, lets start with you and your position and what you do at IGO-POST and CLIPPER	NM
Yes, I work as a business intelligence manager for the PLATO group, that's IGO and CLIPPER, I have been working here for 6 years now, I started as marketing data analyst. Since 3 years I'm responsible for business intelligence in general. And what we try to do for all the departments in the company, we try to convert the data we have into actionable information insights and at the end into action. We do that for the commercial departments, sales, marketing, and purchase, but as well for the regional departments. We try to cover the total scale of the analytics.	J
Okay, so you should be well equipped to evaluate the guidelines I suppose.	NM
Yea, I think so, at least from an analytical perspective, otherwise I'm at the wrong position.	J
We start with guideline 1, 2, and 3. We start with one of them and evaluate that with the four questions and if you have anything else you might wanna add. The interview is mostly structured, but we can talk if there is something special you want to talk about, about the guidelines. But otherwise we just stick to the four questions.	NM
Yea, but I think we will cover some areas, so that will be good. First guidelines is to understand the business problem... and metrics. What I was wondering, the focus is business problems, but is it also business goals. That's what we try to do as well. Of course we also have problems, but, do we achieve our goals. We use analytics to solve problems but also just see if we reach our goals. And	

why are or aren't we. But that's a small twist. You approach it from a critical problem, but you can also approach it from business goals. You can look at it both ways. But that was my question...	
Yeah, it was actually both business problem and goals. But I guess I kinda left out the part about goals. Good point.	NM
What I find very strong, is the example "If a metric does not promote action, it's probably not that important". That's a thing we are implementing right now as well. What we found out is that when start measuring things, to begin you think everything is interesting, but then it gets too broad, so, indeed think about each measurement and what are you going to do if there's this or this. But that is technically what you say. What we also do is to try to define the most important metrics. That's why I'm mentioning goals. How strongly are they correlating with out goals. The influence on goals. Most of the time the goals are about turnover or profit, or maybe cost prediction. Those goals can help you prioritising what you are going to measure. We call that KPI trees. So, again, from out business goals, the strategic goals, just think about which things are influencing for instance turnover - this is a strategic goal. What is influencing turnover... such as leads, lead conversion. So in that way, you can easily define what are the most important metrics and also decide if they are actionable or not. That is how we define relevant metrics. You can take that stuff further even, but we are not there yet. For instance, if we look at turnover, we can identify possibly 15 metrics that are a big on that. But with correlation techniques, you can also theoretically prove which is the most important metric. So for instance, if best predictor is [inaudible] for your turnover next week or two weeks, if it has more predictive value then that is the most important KPI. Theoretical, maybe for a startup, it's difficult because you don't have much data. But that is at least the way we think. What has the most influence on our strategic goals, that is a way to define your measurement framework.	J
Very good point	NM

What I liked very much is your rule of thumb for what a metric should be, so: comparable, easily understandable, promote action, be a ratio or rate. We also use "influenceable"... But maybe for a startup with less people, it's maybe not that important, but you have to show the metrics to the people that can influence those metrics. What we did in the past was to show too many metrics to people who cannot influence those metrics. To a salesperson, we can show the financial results, but for a salesperson at a lower level, it better to show them how many open leads they have, because that's the thing they can work on. That's their circle of influence. But for a startup, that is small, that is maybe more difficult... But, in relation to that, we always, something I learned as well from another guy, is to strive to have a balance with metrics in quality, quantity, and timeliness. So what we did in the past, for instance, for predictions, we just measured how many items are printed. But if you only think of metric in terms of quantity, then you are not having a good touch on the quality. So yeah, now we have productivity and number of complaints. Because there can be a correlation if the only focus is on quantity... we will reach a quantity goal, but complaints go up. So that's also a thing... To balance your metrics. There is also a big difference between lacking and leading information. For instance, your financial results, very important, but it's really lacking, it's a result of what we did last month maybe. So, try to balance lacking and leading. For instance, open quotations, that's a leading KPI, because that we can still influence. So also there, try to have a balance of the different kind of metrics.	J
Yeah, very good point.	NM
It's all theoretical ideas that I gained and I try to keep them in mind when I build a report and stuff like that.	J
Do you see any difficulties entrepreneurs might have, when they apply these guidelines. I know you are not in the entrepreneur business, but...	NM
Maybe, what they can be missing is, where do I start. Because you approach it critical problems. For entrepreneurs, is it clear enough. So, how do I set up	J

a measurement framework. But... I think it should work. If you don't have an analytical background it still very difficult.	
I am actually doing a proof of concept, documentation, where I am coding a web analytics platform using PHP, MySQL and even R... I am looking at the guidelines one by one, and that is also what I see, that, if you don't have a technical or analytical background, you might run into some problems.	NM
Maybe what could help... we use an easy approach, take it from a business strategy, and from there, look at influencers of your strategic goals. Maybe that can help setting up your measurement framework, keeping also in mind all your tips and tricks. I think it's good, but still it's difficult. I think, from there, also the problems and goals should be the starting point for creating a measurement framework. But it really also depends on the entrepreneur the you deal with. Some have analytical skills and some don't have them. What I wanna add is, Google Analytics, for a standard startup, 70% of the most important KPI will be in there already. Those were my small remarks for this topic. It will maybe be too many tips if I add all I gave you as well.	J
Yeah, my problem when formulating the guidelines was also the space that I had. Because if I need them to evaluated I can't make guidelines of 30 pages, because then no one will read them. So I really had to keep it tight. Which was a bit of a problem. But that is what you are here for, to evaluate if I did a good enough job.	NM
I understand your problem the you have to keep it tight. But... it looks good, maybe some additions maybe.	J
Yeah, some of the points you made were really good. I will try to implement them. For the next guideline, appropriate data collection and processing, if it covers its subject sufficiently.	NM
Yeah, for me it's clear and good. Also here, if you don't have an analytical background it's some very difficult stuff. Also, you said, you approach it form mainly website data. You also raised the good question of doing it in-house or use a vendor. I think that also very much depends on the difficulty and	J

differentiation of their data. If you are solely a website e-commerce platform, I think 90% can go for a standard vendor approach... For instance a startup, most of the time, it will work fine. But for us, we have multiple brands, we have an online and offline approach, we have commercial departments, we have so many sources of data we cannot go with standard vendor solutions. We have to build a datawarehouse with star-schema. That's what we do. I think it really depends on the differentiation of data, whether if you can go for a standard solution or if you can go for a custom. But I think most of the time, for a startup, at least at the beginning, it won't be that difficult until...	
It's maybe a bit too complex for the usual entrepreneur?	NM
I think the average entrepreneur cannot cope with this. It's hard. For me it's still difficult to be honest. It keeps getting more difficult because of the more data we have the more... It's a difficult part. What I also stated is, what we see more and more, if you are looking at software solutions, it's becoming more and more important how they integrate with other platforms. Because, if they integrate in an easy way, that can keep you from building it yourself. I believe you should stay away from building it yourself as long as you can. Because, you need employees, it takes up a lot of money. But, this was clear, but difficult for entrepreneurs. I think if you are a fully online startup, you should go with a standard solution, otherwise, it will become your main problem of your company. I'm not saying it's not worth that, but for startups...	J
Yeah, startups are all about speed, agility and flexibility... Any... besides the things you have mentioned, any direct improvements that you see could be made to maybe... the data collection and processing model, or, just anything?	NM
Hmm. You approach it from a website perspective, but that's also what you say, but maybe what you could add, in the more broader sense, and whether it's website data or ERP, it's all about you have your raw data, and then you have... what is called an ETL process. You extract it from the source, you have to transform it, you have to load it into the data warehouse or BI tool, but that's maybe... It could be a small thing that is easy to understand: this is how	J

data works. You have to extract it and transform it, and then at the end into an information platform. That maybe gives an easy view of how it works.	
Funny enough, that was actually my first approach to this guideline. That was the first model I made. That was, determine your primary and secondary sources of data, put them into some database or where you can perform analytics... and then present the data... but then I skipped it because I was more into this granular and individualised level of tracking. But yeah, maybe just explaining how data works is probably more useful to entrepreneurs.	NM
Yeah, I can read it.. It makes sense to me, talking about star-schema, but maybe for entrepreneurs, it gets scary immediately. But maybe it can help getting a grip of at least the flow of how to turn data into information. To me it's clear, to my BI department it would make sense, but for anyone else it might get messy.	J
Alright, moving on to the third guideline... so, same process, does it cover its subject sufficiently. This one is all about making data easily accessible and sharable with the team, but yea... does it cover the subject sufficiently and any problems the entrepreneurs might have?	NM
I think this is maybe the most clear one. Also to entrepreneurs, because they also want to see results in an easy way so that makes sense what.... What I find very strong about the newer BI tools is that they also have alert functions. You can... maybe this is too in detail... Instead of making available standard reports to people... People having to look at 20 metrics each time, and each week 2-3 stand out. You have the articles you setup in such a way, for instance, turnover goal, I only want to get a notification if it's + or - 5% of what my goal was. In that way you can really set up a measurement framework, and immediately push you to think when this metric is standing out and if it needs attention. So if my conversion rate is 5% up or down, I get a notification. I think that is really a strong thing about BI tools. It can make the work for everyone on the floor much easier. And that way people don't have to go over reports each week or month. You get a notification if something you think is important	J

stands out. Maybe this can be used in the data presentation. I think, also for startups... There are some really good platforms you can start with for a low budget. You have two kind of tools. The ones where you have to buy for 20-30.000 euros and you have all the access you want. And then you have the one that are building up - the cost of the tool is going up with the number of users, for startups that... for instance Tableau, you can just start with 1000 euros... If you look at IBM then you have to pay 50.000 euros immediately. Maybe a good idea for startups is to look for tools that are widely spread, so that you have a community behind it. For instance Tableau, you have a big community, so if you have a question you can look online in the community instead of having to get a consultant for 1000 euros per day. But that's small things. But I think it's quite clear.	
My guidelines have mainly been about creating your own platform. That was a result of my data collection from entrepreneurs. If you go to the third page, the context page, you see only 25% have developed their own analytics platform... That was kinda my focus that I want to help them be able to build this. But the thing is, I didn't ask them if they are interested in having their own analytics platform because they are using 3rd party platforms. Even though I have mainly focused on their own analytics platform, I still think they can use some of the ideas from my guidelines. That was just a small note for the third guideline.	NM
I think that's an important choice for an entrepreneur "should I focus on my core business or should I do the analytics part by myself as well?" That's the choice they have to make.	J
Yeah, because it does take away their focus right... They do have to focus on a totally different aspect...	NM
But it's also have a lot to do with the complexity of your business, and the complexity of your data. Whether a standard solution can be suitable or not. The more difficult it gets, the more custom made things you need. It also really is depending on the startup.	J

Okay, the last one... Analytics Process Framework... Just a note, because I'm not sure if I made this clear enough in my guideliens at least... But do you know the agile programming framework Scrum?	NM
I know some things... We are starting with it in IGO-POST as well.. I know the backlog, sprints, reviews...	J
And do you know CRISP-DM... It's a... data mining...	NM
Yeah, for more advanced...	J
Yeah, exactly. So, I tried to combine these two concepts into the thing you see in the middle.. That was just some extra info for you...	NM
I think that's... again you approach this from a business problem, but how do you see... Of course we always have problems and projects, and I totally agree to approach it this way. But how do you approach you standard report structure, do you also think it can be done this way?	J
I think it maybe can, because in the backlog, you put in all the different metrics you would use in the report and the move on to collect... retrieve the data... either set up tracking for the metrics or retrieve the data you already have... and I guess explore data could be made into compile report instead.	NM
Yeah okay, so you have a question, you go to investigate and after that you decide to implement in a structural way... and maybe decide if an interesting metric should be standardised. It can be the outcome of the process. We are actually starting with this next monday, where we have a project around this. We are going to do this with the scrum method as well. Where we find like 15 business questions we think we can solve with advanced analytics like R - can we predict when someone will do his next order, is a question. We are going to approach it with sprints, so it's more or less like this. So, I cannot disagree with this.	J
One thing that... I don't know which topic, but, I think it's very important with A/B testing. I don't know if that's too detailed, but it could be in your measurement part. I think that is also a thing that is understandable and gives	J

a lot of good information and can improve your business very fast. I think it fits in.	
Good point.	NM
For A/B testing there are a lot of standard offerings available. It's not that rocket science. [...] It's a very powerful technique.	J
Yeah, well... I think that is pretty much all... By the way, are there any areas of improvement for the analytics process framework.	NM
Maybe what can be... you have your backlog, with a lot of points. What we are doing right now, is, we have a very long list... We have a system where we look at the possible impact and the workload of the business question. We are then going to prioritise the questions we have. For instance, for advanced analytics, we have 15 IDs. For each ID how do we think it's going to impact our strategic goal, so, turnover, profits, efficiency... and how complex is it terms of costs, time, and risk of failure. We are going to score them based on these six items. So that's how I choose which topic to touch first. And again, we go back to the relation to our strategic targets. It can help you prioritise which metrics and questions to answer. So, look at impact and complexity for prioritising actions.	J
That's actually a really good point.	NM
Yeah, it can help you to focus on the right things. If you help a long list... If we involve people from the business in this part, you really get them on track and onboard on the project you are going to start. You give the people the feeling that they are joining what you are doing.	J
I think it's a good document and I could bring something with my remarks.	J

11.4 Designed Artifact

The designed artifact can be accessed at http://nickmillard.com/dsg-v1.pdf