

Innovation in Times of Big Data and AI: Introducing the Data-Driven Innovation (DDI) Framework



Sonja Zillner

Abstract To support the process of identifying and scoping data-driven innovation, we are introducing the *data-driven innovation (DDI) framework*, which provides guidance in the continuous analysis of factors influencing the demand and supply sides of a data-driven innovation. The DDI framework describes all relevant aspects of any generic data-driven innovation and is backed by empirical data and scientific research encompassing a state-of-the-art analysis, an ontology describing the central dimensions of data-driven innovation, as well as a quantitative and representative research study covering more than 90 data-driven innovations. This chapter builds upon a short analysis of the nature of data-driven innovation and provides insights into how to best screen it. It details the four phases of the empirical DDI research study and discusses central findings related to trends, frequencies and distributions along the main dimensions of the DDI framework that could be derived by percentage-frequency analysis.

Keywords Data-driven innovation · Business models · Data ecosystems · Value proposition · Collaboration · Platform economy

1 Introduction

To support the process of identifying and scoping data-driven innovation by reflecting the dynamics of supply and demand trends, we are introducing the *data-driven innovation (DDI) framework*, which provides guidance in the continuous analysis of factors influencing the demand and supply sides. The framework systematically addresses the challenges of identifying and exploring data-driven innovations. It guides start-ups, entrepreneurs and established companies alike in scoping

S. Zillner (✉)
Siemens AG, Munich, Germany
e-mail: sonja.zillner@siemens.com

promising data business opportunities by analysing the dynamics of both supply and demand.

The DDI framework is based on a conceptual model represented as ontology. The DDI ontology describes all relevant aspects of any generic data-driven business. On the supply side the focus is on the development of new offerings. For a clearly defined *value proposition*, this includes identifying and accessing required *data sources*, as well as the analysis of *underlying technologies*. On the demand side the focus is on understanding the dynamics of the addressed markets and associated ecosystems. This includes the development of a *revenue strategy*, a way forward to harness *network effects* as well as an understanding of the *type of business*. As data-driven innovations are never created in isolation, identifying potential *partners* and a *viable ecosystem* helps to align supply and demand in order to achieve a competitive advantage.

The DDI ontology and framework were developed and tested in the context of the Horizon 2020 BDVe project¹ and are backed by empirical data and scientific research encompassing a quantitative and representative research study covering more than 90 data-driven business opportunities. The objective of the empirical research study was to systematically analyse and compare successfully implemented data-driven business innovations.

By relying on the DDI ontology and framework, we now have a method in place that we can share with members of the big data value ecosystem to explore data-driven business opportunities. The DDI ontology and framework are complemented by a comprehensive set of methods and guiding questions that are used for industrial trainings and university lectures. The derived characteristics and patterns of successful data-driven innovation help entrepreneurs, innovators and managers to scope their data-driven business opportunities in such a way that industrial investment decisions will become more successful and sustainable.

In what follows, Sect. 2 aims to define the notion of data-driven innovation. Section 3 details the four phases of the empirical research study establishing the foundation for developing the DDI framework. Section 4 summarises the main findings of the empirical DDI research study and Section 5 concludes the chapter.

2 Data-Driven Innovation

Finding a way to identify and scope data-driven innovation requires an understanding of the business opportunities in general as well as of the characteristics of data-driven innovation and an appropriate way forward to scope them. The following section briefly describes the overall layout, characteristics and specific challenges for data-driven innovations.

¹<https://www.big-data-value.eu/>

2.1 *What Are Business Opportunities?*

The term business opportunities is a broad concept that is used to describe the chance to address a particular market need through the creative combination of resources that allow the delivery of advanced value propositions (Ardichvili et al. 2003).

From this definition, we can derive that promising business opportunities are based on a smooth balancing of two perspectives, i.e. the mainly technical capabilities on the supply side with the market dynamics and user requests, motives and interests on the demand side.

This argument is supported by a study by Timmons and Spinelli (2007) showing that most successful entrepreneurs and investors continuously observe the demand side very carefully in order to understand what customers and marketplaces want and never lose track of it. The insights gained about the demand side is used to guide the scoping of offerings by combining innovative technology components with reusable and available assets in a way that fosters competitiveness.

We observe several economic properties that play a crucial role when developing of data-driven business opportunities. For instance, when re-using a data source as input for producing of data-driven offering, it will never lose its initial value. However, the value of the data is not given per se but depends on availability of complementary assets that allow to extract the relevant information from the raw data.

The mentioned economic properties of data are impacting the dynamics of the market. In particular, due to network effects and the increasing flexibility of how offerings are scoped and priced for the different customer segments, the success of data-driven innovation requires continuous alignment between the needs on the demand side and the capabilities on the supply side.

2.2 *Characteristics of Data-Driven Innovation*

Data-driven innovation refers to the use of data and analytics to improve and foster new products and processes, new organisational processes, and new markets and business models (OECD 2015). We observe several economic properties that play a crucial role when developing of data-driven business opportunities. For instance, when re-using a data source as input for producing of data-driven offering, it will never lose its initial value. However, the value of the data is not given per se but depends on availability of complementary assets that allow to extract the relevant information from the raw data.

The mentioned economic properties of data are impacting the dynamics of the market. In particular, due to network effects and the increasing flexibility of how offerings are scoped and priced for the different customer segments, the success of data-driven innovation requires continuous alignment between the needs on the demand side and the capabilities on the supply side.

2.3 *How to Screen Data-Driven Innovation?*

The data economy is perceived as highly dynamic market: This is supported by the rapid growth of the European data markets, recent technical breakthroughs and the continuous growth of data assets.

Same, same but different: It is expected that the development of data-driven offerings will speed up as the existing data technologies along the data value chain are getting reused, combined and aligned with each other. For instance, systems such as Watson that required development over several years with the involvement of a large team will in the future become available to ordinary software engineers.

This in consequence leads to situations where entrepreneurs aiming to bring new offerings to the market need to continuously scan market offerings in order to identify promising available technology components – such as specific algorithms, knowledge models or hardware assets – that can be reused to speed up the development time of their innovation. At the same time, they need to constantly investigate their own unique selling point and the competitive advantage of their offerings in a highly dynamic environment. In such settings, innovations are no longer implemented by one organisation alone but rather a population of organisations and entrepreneurs that copy from each other as much as possible to ensure that technological assets can be reused and combined.

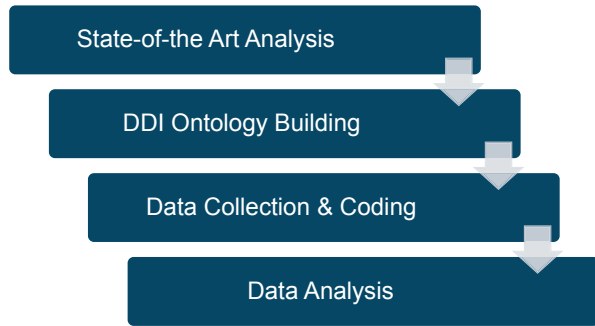
Of course, it is still necessary to put in enough effort to ensure that they make a difference in the market with a unique offering. This can be compared to a swarm of birds flying in the same direction with each bird continuously observing where the others are flying to have enough distance to avoid collision, but at the same time to be close enough to benefit from the wind shadow (Baecker 2007). In this way entrepreneurs need to continuously reassess what is part of their core offering and in which areas they are partnering with others in order to stay competitive in a fast-moving market.

The matching of supply and demand is a key success criterion for data market growth: The high-growth scenario² in the comprehensive European data market study (IDC & OpenEvidence 2017) is based on supply-demand dynamics that shift from technology push to demand pull. In other words, any means that provides guidance in match-making between market needs on the demand side and technical capabilities on the supply side helps to stimulate the adoption of data-driven innovation and in consequence the growth of the European data market. This can become possible through a fully developed ecosystem that is generating positive feedback loops between data/technology companies and users.

Accordingly, data-driven business opportunities that are described with a clear scope of offering per market segment (*supply side*) and reflect the ecosystem

²Which estimated 4% of GDP growth between 2016 and 2020.

Fig. 1 The four phases of the DDI research study



dynamics and benefits of network effects (*demand side*) are more likely to find a promising market fit. Given the dynamics of the growing data economy, the relation between the scope of offering on the supply side and the type of attributed value (e.g. price) on the demand side requires continuous reassessment. In consequence this leads to a *co-evolution between the supply side (e.g. the offering) and the demand side (e.g. adjacent ecosystems)* for each data-driven business opportunity.

To summarise, data-driven business opportunities should be described with a clear scope of offering per market segment (*supply side*) and reflect the ecosystem dynamics and benefits of network effects (*demand side*).

3 The “Making-of” the DDI Framework

This section describes the set-up of the DDI framework. The ontology and framework were developed in four phases (see Fig. 1).

By first *reviewing the literature* on existing proven methods and the theoretical concepts for scoping data-driven innovation/business opportunities, we could identify the relevant aspects of the data-driven innovation. The learnings from the literature review guided us in developing a *conceptual model* in the form of ontology describing the central aspects of supply and demand in data-driven ecosystems. Based on the conceptual model, data from a representative sample of data-driven start-ups could be collected and coded. Subsequently, the data was *analysed*, and best-practice insights and *patterns* identified.

3.1 *State-of-the-Art Analysis*

So as not to reinvent the wheel, we aimed to reuse and combine existing business modelling methodologies whenever possible – and to complement them with a *meta-analysis of demand- and supply-side trends* in order to guide the process of identifying data-driven offerings.

In our state-of-the-art analysis, we investigated to which extent existing frameworks, research results and methodologies can be used to describe the supply and demand sides of data-driven innovation. The DDI approach builds upon popular existing business modelling methodologies and related research, such as Osterwalder and Pigneur (2010), Nooren et al. (2014), Gassmann et al. (2014), Hartmann et al. (2014), Attenberger (2016) and Johnson et al. (2008).

We could reuse valuable content from the OECD (2015) to scope the actors in data ecosystems and learn about the characteristics and nature of data-driven innovation in general. From Adner (2006) we use findings about the handling of risks involved either when working with partners to develop innovations or when engaging with partners required to adopt the innovation. In our work we relied on findings about emerging disruptive business and market patterns (Hagel et al. 2015), as well as insights about the different strategic roles in the governance of ecosystems (Iansiti and Levien 2004). In addition, we used important concepts and findings from research about emerging platform businesses, such as Parker et al. (2016) and Choudary (2015).

The data and technologies along the *data value chain* are the central aspect of the supply side of data-driven business opportunities. To explore the data value chain, we relied on a simplified version of the DAMIAN methodology that we developed and prototyped in particular for the scoping of data-driven scenarios. This approach could be complemented with our findings in Cavanillas et al. (2016) and with methodologies for exploring the value proposition (Osterwalder et al. 2014) and co-innovation partners (Adner 2006).

3.2 DDI Ontology Building

Based on the above-mentioned literature review, the dimension of data-driven innovation could be identified. This leads to an initial version of a conceptual model as an ontology, covering relevant dimensions and concepts to describe data-driven innovations in a comprehensive manner. The objective of the DDI ontology is to cover all relevant aspects of data-driven innovations and establish the basis for analysing these aspects in an effective way. Recognising the findings of IDC and OpenEvidece (2017), the dimensions/concepts of the DDI ontology have been divided into two areas: the *supply side* and the *demand side*. Figure 2 gives an overview of all dimensions of the DDI ontology.

On the *supply side* the focus is on the development of new offerings. For a clearly defined value proposition, this includes the identification of and access to required *data sources* and the analysis of *underlying technologies*, as well as of all the



Fig. 2 Overview of all DDI dimensions on the supply and demand sides

partners that are required for the development and implementation of the data-driven innovation.

On the *demand side* the focus is on the dynamics of the addressed markets. The analysis includes the development of a *revenue strategy*, a way forward to harness *network effects* as well as an understanding of the *type of business*. As data-driven innovations are often built into established value chains, the partners in the *ecosystem* are analysed to understand under which conditions value chain partners are willing to adopt the innovation and thus will facilitate market access.

The initial version of the DDI ontology was continuously updated by incorporating lessons learned and insights gained by running DDI university lectures, seminars and workshops, as well as by performing a coding test run on a smaller set of 20 start-ups. For further details related to the different versions of the DDI ontology as well as the description of the final version of the DDI ontology, we refer to the following technical reports: Zillner et al. (2018), Zillner (2019) and Zillner and Marangoni (2020).

3.3 Data Collection and Coding

Based on three selection criteria, a representative sample set of data-driven innovation could be collected. In accordance with the dimensions described in the DDI ontology, the initial sample set of data-driven start-ups was enriched by findings from manual research (data coding).

3.3.1 Selection Criteria

To identify a representative data set, the following three selection criteria have been identified:

Focus on start-ups: Being well aware that data-driven innovations are developed in all types of organisation, i.e. in large, medium and small enterprises, we decided simply due to two practical reasons to focus in our study on data-driven start-ups only. First, as larger corporates and SMEs barely share information about their business or innovation designs and decisions, no public information was available. Second, as innovation activities in large corporates and SMEs are often influenced by existing infrastructures, legacy systems, prior systems and the existing customer base, organisational implications, such as changes in the sales channels, customer bases, migration issues, pricing models and processes, and customer expectations, need to be incorporated into the analysis. Those interdependencies with existing operations make it difficult to analyse data-driven innovation in isolation or to derive generic patterns.

Success criteria: To identify successful data-driven start-ups, we needed to define a *measurement for success*. We decided to choose start-ups with funding between

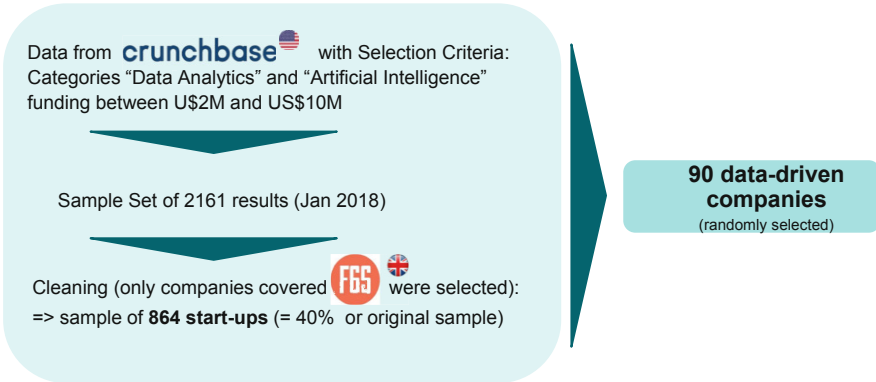


Fig. 3 Overview of the generation of the start-up data set

US\$2 M and US\$10 M³ to cover the ones that had already convinced some ventures to invest in them, meaning that they would already have their product validated, but still are a “younger” start-up.

Technology focus: To identify data-driven start-ups, keywords/selection criteria such as *data analytics* and *artificial intelligence* seemed to be promising.

3.3.2 Sample Data Generation

To ensure high data quality, we decided to cross the data from two start-up databases. The initial database was Crunchbase,⁴ an American-based platform for finding business information about private and public companies, and this served as the primary source for generating our sample data set. The second data source was F6S,⁵ the largest platform for founders based in Europe.

The start-up data was extracted on 16 January 2018 from Crunchbase using the aforementioned filters:

- Categories “Data Analytics” and “Artificial Intelligence”⁶
- Funding between US\$2 M and US\$10 M

³The decision criteria for the values (between two and ten million dollars) were made in the light of venture capital theory. Although there is no consensus regarding the exact amount of money that determines each stage, we decided to follow the criteria used by Crunchbase: Angel is the first round, normally financed with less than US\$10,000. The following stage is Seed, ranging from US \$10,000 to US\$2 M. Then there are the venture rounds that could have many series (A–Z), with A and B series normally valued between US\$1 M and US\$20 M.

⁴<https://www.crunchbase.com/>

⁵<https://www.f6s.com/>

⁶Crunchbase is using 46 categories to classify all of its companies.

Based on these filters, we could extract a sample set of 2161 data-driven companies.

From this larger sample set, we extracted a statistically valid sample set of 90 start-ups with entries in both databases. Figure 3 provides an overview of how the initial data set of start-ups was generated.

3.3.3 Coding of Data

The start-up data was coded in accordance with the categories of the data-driven innovation framework. For each start-up, relevant background information was manually searched and investigated to identify relevant statement(s) related to certain categories of the DDI framework.

To ensure reliability, the different categories of the DDI model were defined before the coding exercise started. To avoid coding errors, a test run of the coding exercise based on a manually selected sample of 20 start-ups was performed. After coding of this initial set of start-ups by two independent coders, all categories or concepts with a high percentage of disagreement in coding were discussed in detail and then *redefined or removed*.

The start-ups from the sample set were coded by three independent coders. For each start-up the three coders manually annotated a binary feature vector covering all DDI dimensions and concepts. In case a specific feature was present, it was annotated with “1”; in case it was not present, it was annotated with “0”; and in case no information could be found, it was indicated with “2”.⁷ This was done by searching the Internet for relevant statements indicating a specific feature of the DDI ontology.

For each start-up at least three websites (Crunchbase, F6S and company website) were consulted. Very often additional webpages, e.g. linked press releases, were analysed, and complementary Internet searches were conducted to ensure that all categories and concepts were addressed.

After having performed the manual annotations, the coders met online to compare coding results and to discuss and resolve disagreements. The result of the coding process was 90 binary feature vectors representing the presence or absence of each DDI category or concept for each start-up.

⁷Although the feature vector can be annotated with three values (0,1, 2), we still treat it as a binary feature vector, as the third value category “2” was only introduced for practical reasons, to indicate that for a specific feature the accomplished search did not reveal any related information. This helped us to monitor the progress of the coding exercise as well as to remove start-ups from the analysis.

3.4 *Data Analysis*

Based on the three previous phases, it was possible to generate a sample data set that had 90 variables (dimensions and categories of the DDI ontology) and 90 observations (start-ups) that were marked either by the presence of the variable (1) or by the absence of it (0). For example, one of the variables described whether a start-up was doing business in the B2B domain. For start-ups for which this was true, we marked a (1), and for start-ups that did not target B2B, we marked a (0). In the percentage-frequency analysis, we then counted how many start-ups were marked with (1) and divided this by the total number of observations for that variable. Using the same example, we could observe that 88 start-ups out of 90 were marked with (1), which means that 98% of companies target B2B customers.

The first method employed to assess which variables could shape data-driven business innovation was a percentage-frequency analysis. The goal of using this method was to understand how frequently a variable was observed in our data.

4 Findings of the Empirical DDI Research Study

To derive meaningful insights into trends, frequencies and distributions, a *classical statistical data analysis* was used. Based on a percentage-frequency analysis, many insightful findings along the main dimensions of the DDI framework could be identified. In the following subsections, we will summarise all findings derived from the percentage-frequency analysis. We will represent these findings by first discussing some generic findings before discussing the findings in relation to the dimension the DDI framework.

4.1 *General Findings*

It was important for us to find out whether the distinction between B2B and B2C has an influence on the design of data-driven innovation. In addition, we wanted to better understand the possible impact of the (non-)sector focus of data-driven innovations.

Target Customer: The majority of data-driven start-ups (78%) are addressing B2B markets. Only 2 out of 90 start-ups in our sample focused solely on end-customer markets. Start-ups addressing end-user needs prefer already established channels to deliver their offering to the users. They tend to rely on partnerships with established business partners to bring their offering to users. A second, quite frequent, strategy used by 19% of start-ups is positioning data-driven

solutions as multi-sided market offering combining complementary offerings to align private and business needs.

Seventy-five per cent of our start-up sample have developed a clear *sector focus*. Companies with clear *sector focus* have a concrete customer segment in mind for whom a concrete value proposition is delivered. Those companies have a concrete customer segment(s) in mind for which a concrete value proposition is delivered.

For example, *CloudMedx*⁸ Inc. designs artificial intelligence-driven software for medical analytics. Clinical partners at all levels can derive meaningful and real-time insights from their data and intervene at critical junctures of patient care. Its underlying clinical AI computing platform uses healthcare-specific NLP and machine learning to generate real-time clinical insights at all points of care to improve patient outcomes. By relying on evidence-based algorithms and deep learning, a wide variety of structured and unstructured data being stored in clinical workflows can be understood and used for decision making.

In comparison, we also found start-ups that focus on technology with cross-domain impact. In general, their solution will be used by other intra- or entrepreneurs to build data-driven solutions for end users.

For instance, the start-up *DGraph Labs*⁹ is offering an open-source distributed graph database. The company is planning to release an enterprise version that is closed source, as well as a hosted version (as it is easier to run hosted services for customers than trying to help them debug every issue on their own). Customers are using the service to build their own sector-specific applications.

In summary, sector-specific data-driven offerings are much more frequent than technology-driven sector-agnostic solutions. This is due to the very different pre-processing challenges of data sources in the various sectors, as well as the higher possibilities of identifying target groups in concrete sector settings. Most sector-agnostic offerings are intermediate functionalities addressing developers to build customised solutions.

4.2 Value Proposition

To analyse the value proposition in the context of data-driven businesses, our main focus is on the different ways data is used to generate value. *Data value* refers to the insights that can be generated out of data and how this can be used in a particular user or business context. In accordance with its value and complexity, we distinguish four different types of analytics that are used for generating different types of insights, i.e. descriptive analytics explain what happened, diagnostic analytics highlight why something happens, predictive analytics forecast what will happen in the future, and prescriptive analytics identify optimal actions and strategies (Zillner 2019).

⁸<http://www.cloudmedxhealth.com/>

⁹<https://dgraph.io/>

Two out of every three start-ups rely on *data analytics* in general for generating insights. Among the start-ups using data analytics, 83% rely on descriptive analytics in their offering (i.e. every second start-up).

For instance, the start-up *Apptopia*¹⁰ is using *descriptive analytics* to provide app analytics, data mining and business intelligence services. They collect, measure, analyse and provide user engagement statistics for mobile apps and visualise the aggregated data in classical dashboards. The unique selling point of their offering is the high number of data points they are able to integrate and visualise, i.e. they state that they rely on “more different data points than nearly any other app data provider in the world”. The insights, which can be generated by descriptive data in this large data set, are of interest to the worldwide mobile app developer community as they allow them to compare their own app performance with competing or related apps. Whenever app developers are engaging with the *Apptopia* platform to benchmark their own apps, additional valuable data sets can be generated. By offering free-of-charge descriptive analytics-based dashboards, *Apptopia* are able to attract a large number of developers to use their platform, which again allows them to produce high-value data sets that can be sold to business customers.

Four out of ten start-ups in our sample set relied on *predictive analytics* to generate value for their users.

For instance, the start-up *Visiblee*¹¹ collects IP addresses and cookies of all website visitors and uses these to predict the identity of unknown visitors in real time. By relying on these real-time predictions, the company is able to increase the leads¹² threefold.

Compared to *descriptive and predictive analytics*, we can observe that diagnostics and prescriptive analytics are used less frequently. Only every fifth data-driven start-up is offering solution for automating manual tasks or activities, and match-making is observed in only 16% of cases.

To implement data-driven offerings, in general, several algorithms and approaches are combined. This is also true for the four different types of data analytics discussed earlier. In our sample, 4 out of 10 start-ups use more than 2 different types of data analytics, and 19% of start-ups rely even on 3 or more types of analytics to generate value.

For instance, *Eliq*¹³ provides a comprehensive platform for the intelligent energy monitoring of utilities. The AI-powered app offers a wide range of insights:

- By relying on *descriptive analytics*, *Eliq* shows periodic energy consumption patterns that can be drilled down into different time frames, i.e. yearly, monthly, hourly, etc.).
- By relying on *diagnostics analytics*, *Eliq* helps users to identify potential “energy leaks” or potential sources of energy theft.
- By integrating external data sources, such as extreme weather change forecast, *Eliq* can inform users that their energy consumption is likely to change significantly (*predictive analytics*). Utilities benefit from such information as they can customise marketing communication accordingly.
- By relying on *prescriptive analytics*, the *Eliq* platform can not only inform users about increased energy consumption but also recommend strategies to overcome such high

¹⁰<https://apptopia.com/>

¹¹<https://www.visiblee.io/en/home/>

¹²In a sales context leads refer to contacts with potential customers.

¹³<https://eliq.io/>

consumption scenarios, e.g. by upgrading or replacing devices with higher efficiencies. This allows utilities to establish a personalised and targeted user engagement.

Eliq is an example of a start-up that establishes a unique value proposition and competitive edge by offering a wide range of analytical services. We want to highlight that this is not a frequent pattern. The majority of start-ups (62%) is focusing on only one analytical offering.

4.3 Data

Data is *the* key resource for realising data-driven innovation. In general, we observe that the used data sources greatly influence the efforts in data pre-processing as well as the scope of the data-driven offering. In case a data-driven innovation is based on image data, we can conclude that an image segmentation algorithm needs to be in place. In accordance with how specific or domain specific the underlying image data set is, a new pre-processing image algorithm needs to be developed. Or in the case of personal data and of industrial or operational data, GDPR-compliant services and data privacy methods need to be in place, respectively.

For that reason, we recommend exploring the data assets early when scoping one's data-driven innovation. Data exploration will help to understand:

- Whether the envisioned value proposition can be realised. Very often, we face the situation that the data quality is not good enough to generate the needed insights.
- How much effort is needed to create data of high quality. Often the raw data is not yet the data quality needed. The good news is that there exist many approaches to increase the quality of data for this scoped purpose. However, the expected return always needs to be aligned with the efforts needed. Other projects in the Big Data Value Public-Private Partnership (BDV PPP) have reported similar experiences (Metzger et al. 2020).

In the following, we will give an overview of which data types and sources are used and how frequently in data-driven innovations.

A wide range of different types of data sources exist that are relevant for developing data-driven innovation. Although only 19% of start-ups were addressing B2C markets, personal data was still the most frequently (67%) used in the analysed data-driven offerings. This is a very impressive number given the fact that only a very low number of companies in our sample (19%) were addressing business-to-consumer markets. In consequence this also implies that a high percentage of start-ups addressing business customers in Europe¹⁴ need to handle the constraints of the General Data Protection Regulation (GDPR).

¹⁴Our sample set is not restricted to European start-ups only, as we wanted to make sure that our analysis covers worldwide excellence. As we do not have precise numbers for European data companies, the sentence is formulated with some ambiguity.

For example, *Oncora Medical*¹⁵ is using personal data to fight cancer. The US-based company collects data on cancer patients including information related to treatments and clinical outcomes through an intuitive software used by doctors. Their objective is to deliver predictions that can help design better radiation treatments for patients, as well as enabling precision medicine in radiation oncology. The data collected is personal data and is thus sensitive and has higher standards of protection.

Industrial data, i.e. any data assets that are produced or used in industrial areas, is a second type of data which has high data protection requirements. In comparison to personal data, industrial data is used only half as often. Organisations seem to be reluctant (in particular if they do not see the immediate value) to share their industrial and operational data with third parties, such as start-ups, because they are afraid to reveal relevant business secrets.

One successful example, *PlutoShift*,¹⁶ offers a platform that is helping industrial customers to improve their operational efficiency by identifying inefficient patterns of energy usage by analysing customer data stored in the cloud and operational sensor data. With energy being a high-cost driver, *PlutoShift* can help industrial customers to reduce resource consumption and operating costs.

The second most popular types of data source are time-series and temporal data. Fifty-six per cent of start-ups in our sample rely on these types of data to generate value. The high frequency might be due to the popularity of using behavioural data that is tracked within each user interaction on the web and mobile devices and is thus very likely to cover time-series data. Another very frequently used data source is geo-spatial data (46%), and the usage of Internet of Things (IoT) data is seen in 30% of our sample.

4.4 Technology

The BDV Strategic Research and Innovation Agenda (SRIA) (Zillner et al. 2017) describes five technical priorities identified by the BDVA ecosystem and experts as strategic technical objectives. In our study, we were interested in which of these technical areas were most frequently covered when realising data-driven innovation.

Among the five technology areas listed in the BDV SRIA, *data analytics* is used most frequently. Eighty-two per cent of our start-up samples relied on some type of data analytics to implement data-driven value proposition. The usage of technologies in the *data management* area is seen in 41% of cases and is very much in line with offerings addressing the challenges of processing unstructured data sources. Solutions for *data protection* are the least frequently addressed research challenge with 13%. When looking at to which extent BDV SRIA technologies are used in

¹⁵<https://oncoramedical.com/>

¹⁶plutoshift.com; previously called Pluto AI

combination, we observed that more than half of the start-ups, precisely 59%, combine two or more technologies.

*Uplevel Security*¹⁷ is one example that combines *data management* with *data protection*. They redefine security automation by using graph theory for real-time alert correlation. Their product creates a dynamic security graph (data management) for an organisation based on incoming alerts, prior incident investigations and current threat intelligence (data protection). *Uplevel Security* then transforms the ingested data into subgraphs that continuously inform the main security graph. By automatically surfacing relationships, investigations no longer occur in isolation but begin with context.

Less frequently observed, 22% of the companies combine more than three technologies.

One example of this is the medical company *CloudMedx*,¹⁸ which started with the aim to make healthcare affordable, accessible and standardised for all patients and doctors. The company uses NLP and proprietary clinical contextual ontologies (data management) and deep learning (data analytics) to extract key clinical concepts from electronic health records, which serve as insights for physicians and care teams with the goal to improve clinical operations, documentation and patient care. In addition, *CloudMedx* is presenting the results to dedicated teams through a user-friendly platform that allows for interactive predictive and prescriptive analytics to assess current metrics and build a path forward with informed decisions.

4.5 Network Strategies

For digital and data-driven innovations, network effects are important phenomena to reflect. In our study, 57% of start-ups rely on network effects. A network effect occurs when a product or a service becomes more valuable to its users as more people use it (Shapiro and Varian 1999). Network effects are also known as demand-side economies of scale and predominately exist in areas where networks are of importance, such as online social networks or online dating sites. A social network or dating site is more appealing to its user when it is able to continuously attract and add more and more users. In consequence, harnessing network effects requires developing a broader network of users in order for the network or site to differentiate itself from its competitors. For that reason, the critical mass of users and timing are key success factors in a network economy.

Due to the high impact of the network effects, competitors starting from “ground zero” with no users in their network will face difficulties in entering the market success fully. In this context we are using the expression “network effect” to

¹⁷<https://www.uplevelsecurity.com/>

¹⁸<https://www.cloudmedxhealth.com/>

highlight the positive feedback (positive network externality¹⁹), i.e. the phenomena that already existing strengths or weaknesses are reinforced, might lead to extreme outcomes. In the most extreme case, positive feedback can lead to a winner-takes-all market (e.g. Google).

Network effects impact the underlying economics and operation of data-driven innovation. Instead of creating products that are early on the market and different from other offerings, the focus here is on scaling and scoping the demand perspective. Understanding network effects and their underlying market dynamics is crucial to successfully positioning data-driven products, services and businesses in the market. In doing so, data-driven innovation can harness network effects on *three different levels*.

First, data-driven businesses are relying on **network effects at data level**, if they are able to improve their offerings by the sheer amount of data they hold available. In our sample this was the case in 49% of start-ups.

For instance, the already mentioned company *Apptopia*²⁰ uses big data technology to collect, measure, analyse and provide user engagement statistics for mobile apps. The more app providers produce data being connected to the platform, the more valuable the service becomes. In order to gain more real-time data, they attract app developers to connect to their platform by providing free data analytics products. With this free-of-charge value proposition, developers benefit in registering their mobile apps on the platform while giving the platform the permission to analyse user engagement data of the mobile app. Professional and expensive subscription fee models for business customers, including Google, Pinterest, Facebook, NBCUniversal, Deloitte and others, benefiting from real-time engagement insights of mobile apps, complement the revenue strategy of this offering.

In this context, multi-sided business models are the usual way forward. Typically, a multi-sided business model brings together two or more distinct but interdependent groups of customers. Value is only created if all groups are attracted and addressed simultaneously. The intermediary, in our example the company *Apptopia*, generates value by facilitating interactions between the different customer groups, whereas the value increases when more users are attracted. The more app developers register on the platform, the more accurate the statistics become. With an increasing number of business customers, *Apptopia* then creates the required resources to invest in advanced functionalities for app developers.

Second, when businesses are providing a technical foundation for others to build upon, we can observe **network effects at infrastructure level**. In our sample these have been 12% of start-ups. Based on a layer of common components, third-party players are invited to develop and produce an increasing number of data-driven offerings.

This set-up is also known as product platforms (Hagel et al. 2015). A prominent example is the Android platform – it provides the technical foundation for others to

¹⁹For completeness we want also to mention the phenomena of negative network externalities which occur when more users make a product less valuable (e.g. traffic congestion). Negative network effects are also referred to as “congestion”.

²⁰<https://apptopia.com/>

build apps. This includes any type of tool and service that enables the plug-and-play building of data-driven offerings, e.g. (open) standards, de facto standards, APIs and standardised data models. The more functionalities are available that help others to build and position innovative offerings better, faster, etc., the more attractive the offering itself becomes. The infrastructure layer has little value per se unless other users and partners create value on top of it.

An example of this dynamic is the agricultural-robotics technology company *Skyx*.²¹ This company is offering neither hardware nor agriculture end-customer applications, but a software that enables a modular swarm of autonomous drones for spraying. By providing a technology to plan and control the mission of drones in real time as well as to auto-pilot the entire fleet/swarm, it addresses the need for agri-spraying application developer applicators in building their solutions at a higher quality and at less cost by relying on a standardised approach. In addition, as the software is compatible with any commercially available hardware, the cost of connecting the wide range of drones can be significantly reduced. Thus, *Skyx* provides tools and connectors for agri-spraying application developers to build their own solutions. The more drone hardware can be connected, and the more spraying functionalities can be provided, the more attractive the overall offering for applicators.

Third, in cases where the number of marketplace participants is the key source of value, data-driven offerings can harness **network effects at marketplace level**. Offerings that are able to connect participants in their specific roles, such as buyer and seller, and consumer and producer, allow two participants to easily interact with each other.

The low number of network effects at marketplace level in our study (10%) indicates the difficulties and challenges in building them. The challenges are less at the technical level and more at the level of building critical size and balanced user communities. Several strategies to attract users from the different communities have been implemented by start-ups.

4.6 Revenue Strategy

We have been interested in the question of how data-driven businesses are making money. Is this different from traditional businesses? And can we identify some dominant revenue models?

Our first finding is that it was often difficult to find information about the type of revenue models used. Especially in cases when start-ups have been focusing on emerging technical advances, such as drones or autonomous driving, information about revenue models was – understandably – not available.

As emerging technology businesses are often seen as a risky investment or bet on the future in a market not yet established, the absence of revenue-related information is not surprising. This was the case for 10% of the companies analysed: We couldn't find or extract any information about the revenue model.

²¹<https://www.skyx.solutions/>

Our study confirmed the findings of Attenberger (2016) that revenue models have not changed through the usage of data technologies per se. The major difference to traditional businesses is that data-driven innovations rely on different types and combinations of revenue streams that are continuously changing over time in order to address the specific user needs of each customer segment. On the one hand, we observe *new forms of value propositions*, ranging from service offerings, to the bundling and unbundling of offerings, to intermediate offerings, to product differentiations through versioning, that allow the specific user needs to be addressed.

On the other hand, the majority of data-driven innovations have – in comparison to traditional businesses – a different cost structure. With data and data offerings being cheap to reproduce and deliver, the typical cost structure of data-driven innovations relies on fixed costs for the development of the offerings but low variable cost. This kind of cost structure leads to substantial economies of scale as with more offerings sold, the average costs of development decrease dramatically. In addition, as the reproduction and distribution costs are often marginal, the danger of price dumping and surplus of offerings in the competitive market is a frequent phenomenon. For instance, Aitken and Gauntlett (2013) counted more than 40,000 health apps in the app store being offered for free or for a very low price.

With this new cost structure for most data-driven innovations, organisations have a new flexibility to *adjust the equation between value proposition and price* in accordance with the user needs of various customer segments. In this context, companies elaborate the specific price level the targeted user group is willing to pay. The main objective for aligning the product version with the pricing version for each customer segment is to attract more users and interactions, as well as to grow the community.

The most frequently used revenue model in our study was the subscription model. We observed in this context a strong correlation between the spread and high adoption of software as a service (SaaS) approach, which brings a lot of flexibility when used for deploying data-driven innovations. The second most frequently used revenue model is the selling of services in which the person's time is paid for. These revenue models are very often used for open software offerings as well as when offerings are not standardised or off-the-shelf. Advertisement as a revenue model is rarely observed. In our sample, only 2% of start-ups are applying it. Although this might seem surprising, it merely reflects the high percentage of B2B models.

4.7 Type of Business

Data-driven innovations can disrupt existing value chains. However, at the same time, we observe a large number of “low hanging fruits”, i.e. business opportunities in the scope of established processes (intern) or value chains (cross-organisational).

To classify data-driven business opportunities we will introduce four strategies with a significant impact on markets and associated value chains:

- (a) Providing new value to customer with established market position
- (b) Developing a new marketplace/ecosystem
- (c) Leverage an existing ecosystem by scoping a niche offering
- (d) Building technology assets that ensure a future competitive advantage

The following remarks describe the four strategies in detail and illustrate them with an example from our sample of start-ups.

In general, this classification is based on approaches available for the classification of traditional business opportunities. One important work in this context is Ardichvili et al. (2003), who classified business opportunities into two dimensions: *value creation capability* and *value sought*. Although both dimensions have at first glance a good mapping to the DDI supply and demand side, they did not reflect the changing nature of underlying business ecosystems. As already discussed at the beginning of this chapter, data-driven innovations are rarely developed alone but rely on the collaboration between many partners in the value chain.

When positioning data-driven offerings in the market, it is also necessary to reflect the associated business strategy and innovation ecosystem.

Data-driven services are often associated with the strategy of “**Finding a new business partner**”. This strategy tries to focus on one single customer (segment) and his or her business processes. Based on a detailed understanding of his/her business processes (including the pain points, happiness points and unaddressed user needs), new values/services for specific user needs are built. As the service is heavily focusing on this one specific partner, the overall market and business ecosystem is only observed in an indirect manner. In our study, the *data-driven service business* was the most frequently observed approach (with 78%) to position offerings in the market.

For instance, the company *Arable* provides an agricultural solution based on in-field measurements as a software-as-a-service (SaaS)-based service offering. To enable growth, advisors and businesses are invited to play a proactive role in ensuring high quality and longevity of their agricultural operations. As a consequence, the company can derive real-time, actionable monitoring and predictions related to weather risk and crop health by means of a tiered SaaS offering with different levels of services combined with IoT businesses. The tier I service includes reporting, integrating and visualisation, whereas the tier II services include predictions and advanced analytics.

Compared to data-driven services, the second type of business strategy – **developing a data-driven marketplace** – is significantly more complex as a new marketplace/ecosystem needs to be built up. Only 16% of companies in our sample relied on this approach. Market participants on the supply as well as on the demand side need to be attracted. In addition, it is necessary to ensure that a critical number of participants are providing their assets and at the same time a critical number of participants are requesting them.

The growth of the marketplace needs to be balanced on both sides – the supply and demand sides – in order to retain its attractiveness. It seems that organisations have been developing very different strategies to attract the different participant groups, e.g. by providing necessary IT services and analytics services, and offering services for free.

One example of this strategy is *Zizoo*,²² a Vienna-based company that established a global boat rental platform. *Zizoo* is building a global digital booking platform and website connecting suppliers (charter companies) to travellers worldwide, similar to “[Booking.com](https://www.booking.com) for Boats”. When the building of this marketplace started, the founders of the company were entering a market (the boat rental market) which was 10 years behind any other travel sector. As the majority of boat charter companies had not yet been digitalised, they needed to put a lot of effort into attracting the supply side to join their emerging marketplace. For instance, they offered charter companies a powerful inventory management tool and business intelligence for free. As they are making boat holidays affordable and accessible to everyone (bookings start at €20 a day), they were also able to attract the demand side.

Another strategy is to identify an existing healthy ecosystem that is already in place which gives the opportunity to position one’s own offering as a niche application. The so-called niche player leverages *an existing ecosystem* by scoping a niche offering in accordance with the defined constraints of the dominant or key player of the ecosystem. Typical examples of such strategies are the thousands of apps offered in the iOS or Android ecosystems for mobiles. In our sample we could observe this in 12% of cases.

One good example of this strategy is *AIMS Innovation*.²³ This start-up develops AI and machine learning technologies to give the world’s largest companies deep insights into and control of their most business-critical processes – such as safely distributing electricity, shipping thousands of daily orders to ecommerce customers or delivering the results of medical tests to doctors quickly and reliably. They are positioning their offering in the Microsoft ecosystem. According to their website, they offer the only artificial intelligence solution in IT operations covering all core Microsoft enterprise technologies.

The last type of business category is the **emerging technology business** that anticipates a future ecosystem or market. In our study this was seen in 9% of the sample. As the market is not yet settled and the technology is often in a very early stage, it is scoped as investment in the future. Thus, revenue strategies cannot be implemented. The main focus of emerging technology businesses is building capabilities/assets ensuring a future competitive advantage.

For instance, the company *Carfit*²⁴ is working on creating the most comprehensive library of car vibrations. They collect and generate systematically data related to noise, vibration or harshness. An enhanced data analytics algorithm is in place to incorporate automotive domain expertise. The company is aiming at a car vibration tracking device that can help to lower car maintenance costs and increase the efficiency and transparency of the car’s operations. But the self-diagnostic and predictive maintenance platform only brings real value to end users when vehicles are moving autonomously. Thus, the company is addressing a future market (as today drivers are in general good at detecting abnormal noises in their car). However, when cars are moving autonomously the need for remote monitoring will become critical.

²²<https://www.zizoo.com/>

²³<https://www.aims.ai/>

²⁴<https://car.fit/>

5 Conclusion

The data-driven innovation (DDI) framework addresses the challenges of identifying and exploring data-driven innovation in an efficient manner. It guides entrepreneurs systematically in scoping promising data-driven business opportunities by reflecting the dynamics of supply and demand through investigating the co-evolution and interactions between the scope of the offering (supply) and the context of the market (demand). The DDI framework consists of eight dimensions that are divided into a supply side (value proposition, data, technology and partners) and a demand side (ecosystem, network strategy, revenue strategy and type of business).

The DDI framework was developed and tested in the context of the BDVe project and is backed by empirical data and scientific research encompassing a quantitative and representative study of more than 90 data-driven business opportunities.

The data-driven innovation framework offers a proven method for all members of the BDV ecosystem to provide guidance in exploring and scoping data-driven business opportunities. The comprehensive content can be used for industrial workshops and educational set-ups.

References

- Adner, R. (2006). Match Your Innovation Strategy to Your Innovation Ecosystem Match Your Innovation Strategy to Your Innovation Ecosystem. *Harvard Business Review*, 84(4), 98–107. https://doi.org/10.1007/978-1-4614-3858-8_100487
- Ardichvili, A., Cardozo, R., Ray, S., & S., R. (2003). A theory of entrepreneurial opportunity identification and development. *Journal of Business Venturing*, 18(1), 105–123. [https://doi.org/10.1016/S0883-9026\(01\)00068-4](https://doi.org/10.1016/S0883-9026(01)00068-4)
- Aitken, M., & Gauntlett, C. (2013). *Patient apps for improved healthcare: From novelty to mainstream*. IMS Institute for Healthcare Informatics.
- Attenberger, Y. (2016). *Business Models and Big Data*. Master Thesis at the Munich School of Management, LMU Munich.
- Baecker, D. (2007). *Studien zur nächsten Gesellschaft* 1. Aufl., Frankfurt am Main: Suhrkamp.
- Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*. <https://doi.org/10.1007/978-3-319-21569-3>
- Choudary, S. P. (2015). *Platform Scale—How a new breed of startups is building large empires with minimal investment*. Singapore: Platform Thinking Labs.
- Gassmann, O., Frankenberger, K., & Csik, M. (2014). Revolutionizing the business model—St. Gallen business model navigator. *Management of the Fuzzy Front End of Innovation*, 18(3), 89–97. https://doi.org/10.1007/978-3-319-01056-4_7
- Hagel, J., Brown, J. S., Wooll, M., & De-Maar, A. (2015). *Patterns of disruption—Anticipating disruptive strategies in a world of unicorns, black swans and exponentials* (p. 28). Texas: Deloitte University Press.
- Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. *Cambridge Service Alliance Whitepaper*, 1–29. <https://doi.org/10.1016/j.im.2014.08.008>
- Iansiti, M., & Levien, R. (2004). *The keystone advantage*. Boston: Harvard Business School Press.

- IDC, & OpenEvidece. (2017). *The European data market study: Final report*. 275. Retrieved from <http://datalandscape.eu/study-reports/european-data-market-study-final-report>
- Johnson, M. W., Christensen, C. M., & Kagermann, H. (2008). Reinventing your business model. *Harvard Business Review*, 86(12). <https://doi.org/10.1111/j.0955-6419.2005.00347.x>
- Metzger, A., Franke, J., & Jansen, T. (2020). Ensemble deep learning for proactive terminal process management at duisport. In Brockke, Mendling, & Rosemann (Eds.), *Business process management cases*. (Vol. 2) Springer.
- Nooren, P., Koers, W., Bangma, M., Berkers, F., & Boertjes, E. (2014). Media-internet-telecom value web introducing the damian method for systematic analysis of the interdependencies between media-internet-telecom value web. *TNO report*.
- OECD. (2015). Data-driven innovation: Big data for growth and well-being. In *OECD digital economy papers* (p. 456). Paris: OECD. <https://doi.org/10.1787/97892264229358-en>
- Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. Toronto: Self Published.
- Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). Value proposition design. In *Strategyzer series*. <https://doi.org/10.1017/CBO9781107415324.004>
- Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). Platform revolution: how networked markets are transforming the economy--and how to make them work for you. *W.W. Norton & Company*, 256. <https://doi.org/0393249131>
- Shapiro, C., & Varian, H. R. (1999). *Information rules: A strategic guide to the network economy*. Harvard Business School Press. <https://doi.org/10.1145/776985.776997>
- Timmons, J. A., & Spinelli, S. (2007). New venture creation: entrepreneurship for the 21st Century. In *Business*, vol. 2009. Retrieved from <http://www.amazon.com/dp/0071276327>
- Zillner, S. (2019). *D2.6. Annual report in opportunities, BDVe deliverable 2.6*.
- Zillner, S., & Marangoni, I. (2020). *D2.7 Annual report on opportunities; BDVe deliverable 2.7*.
- Zillner, S. et al. (eds) (2017). *Big data value strategic research and innovation agenda*.
- Zillner, S., Timan, T., & Kotterink, B. (2018). *D2.5 Annual report in opportunities*.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

