# Chapter 4
# Data Products

**Jürg Meierhofer, Thilo Stadelmann, and Mark Cieliebak**

**Abstract** Data science is becoming an established scientific discipline and has delivered numerous useful results so far. We are at the point in time where we begin to understand what results and insights data science can deliver; at the same time, however, it is not yet clear how to systematically deliver these results for the end user. In other words: how do we design data products in a process that has relevant guaranteed benefit for the user? Additionally, once we have a data product, we need a way to provide economic value for the product owner. That is, we need to design data-centric business models as well.

In this chapter, we propose to view the all-encompassing process of turning data insights into data products as a specific interpretation of service design. This provides the data scientist with a rich conceptual framework to carve the value out of the data in a customer-centric way and plan the next steps of his endeavor: to design a great data product.

## 1 Introduction

Analytics[1] provides methodologies and tools to generate insights from data. Such insights may be *predictive*, for example, a traffic forecast, a recommendation for a product or a partner, or a list of customers who are likely to react positively to a marketing campaign (Siegel 2013). Insights may also be *descriptive*, that is, providing us with a better understanding of a current or past situation, for example, our company's performance right now or during the previous month. Insights will probably in any case be *actionable*, for example, by enabling a smart controller to

[1]We use the word "analytics" throughout this chapter to refer to those methods and tools from data science that pertain directly to analyzing, mining, or modeling the data: Statistical methods, machine learning algorithms, the application of data management tools, etc.

J. Meierhofer (✉) · T. Stadelmann · M. Cieliebak
ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland
e-mail: juerg.meierhofer@zhaw.ch



M. Braschler et al. (eds.), *Applied Data Science*,
https://doi.org/10.1007/978-3-030-11821-1_4

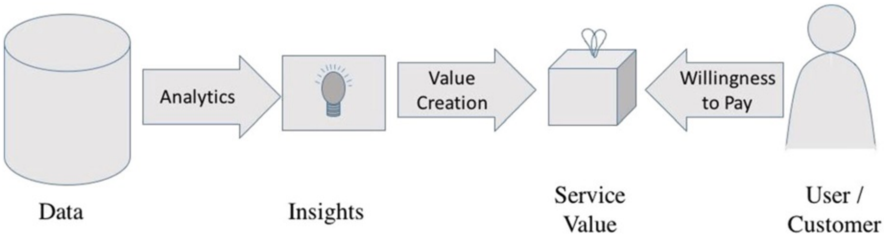


**Fig. 4.1** The value chain of a data product

drive a car, operate a building control system, or regulate electricity production according to market demands. In an extension of the purely exploratory paradigm of data mining, a data scientist purposefully plans to build such data insights that benefit the user (Veeramachaneni 2016). This automatically moves the *result* to the center of the analytics process.

But do these kinds of insights already make up a data product? To find the answer, we go back to the definition of a product (Kottler 2003): a product provides a set of benefits for which the customer has a willingness to return a value, typically in the form of money. Thus, insights generated from data can be considered a data product if there are "users" willing to give back value for these insights. The user may be an external customer (e.g., a "consumer") or a user in an organization, for example, inside the company. The value given back may be in the form of a financial payment, but not necessarily (there are other dimensions of value like emotional or social value (Jagdish et al. 1991), or the collected data, e.g., health data from wearables, search patterns, etc.). This is illustrated in the complete value chain of a data product (see Fig. 4.1).

In other words: in order to have a data product, we need to design insights generating relevant benefits for which users pay. Service science provides us with concepts to solve this problem: according to Lusch and Vargo (2014), a service is defined as the application of competences (knowledge and skills) for the benefit of another entity. With respect to data products, the application of competences refers to the competence of applying data science—the "*unique blend of skills from analytics, engineering & communication aiming at generating value from the data itself*" (Stadelmann et al. 2013).

Therefore, a data product is defined as the application of data science competences to provide benefit to another entity. This makes perfect sense if we substitute "data science" for its original definition cited above, thus resulting in:

> A data product is defined as the application of a unique blend of skills from analytics, engineering & communication aiming at generating value from the data itself to provide benefit to another entity.

Data products are a subset of services (every data product meets the definition of a service, but not every service is a data product). Therefore, the concepts and methods of service science and service design can be applied to systematically design data products. This rounds off earlier work of defining a data product as the result of value-driven analysis that generates added value out the analyses of the underlying data (Loukides 2010). There is a vast field of application examples available for added value generated by analysis. Siegel (2013), for instance, provides an extensive list of 182 examples grouped in the 9 categories: (1) family and personal life; (2) marketing, advertising, and the web; (3) financial risk and insurance; (4) healthcare; (5) law enforcement and fraud detection; (6) fault detection, safety, and logistical efficiency; (7) government, politics, nonprofits, and education; (8) human language understanding, thought, and psychology; (9) workforce: staff and employees. There are also other literature sources providing similar application examples with different groupings, for example, Marr (2016).

In the next section, we provide a very short introduction to general service design before explaining the specific characteristics of applying it to the design of data products. We then identify the gap between current service design and the development of data products, and subsequently propose a framework specific for data product design. We conclude by a discussion of the essential building block of each data product—the data itself, and how to potentially augment it—and a review of the current state of the field, including an outlook to future work.

## 2 Service Design

Service design starts from the user perspective, which means understanding the tasks and challenges the user faces in his context. Customer insight research methods such as depth interviews, participant observation or shadowing, service safari, focus groups, cultural probes. etc. (Polaine et al. 2013), serve to understand the user in his context. The value proposition design framework (Osterwalder et al. 2014) describes a practical template to map the customer jobs, pains, and gains, which together constitute the so-called customer profile (see right hand side of Fig. 4.2). The customer jobs are challenges and tasks that the user needs to tackle and solve. The pains are factors that annoy the user during his job, and the gains provide the benefits that the customer aims at. For the design of the data product, features fitting

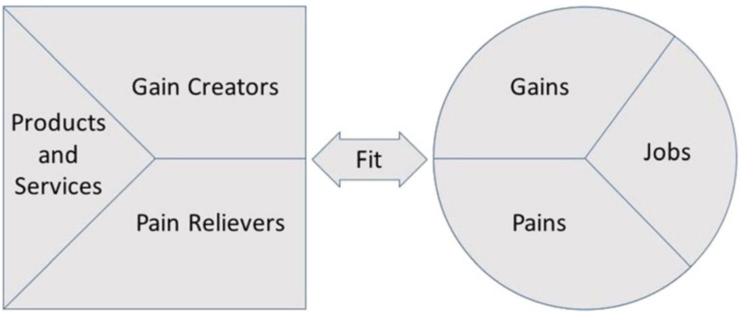


**Fig. 4.2** Fit of value proposition (left) with customer needs (right) (Osterwalder et al. 2014)

with the customer jobs, pains, and gains need to be designed (right hand side of Fig. 4.2). In this context, it is very important to note that service design systematically considers also non-functional customer jobs, for example, emotional or social jobs (Osterwalder et al. 2014; Smith and Colgate 2007).

Additionally, we can apply many more of the useful tools for service design, like customer journey mapping, emotion curve, service blueprinting, service ecosystem design, etc. (Polaine et al. 2013; Stickdorn and Schneider 2010).

A word about terminology: the term "service design for customers" often evokes the connotation of consumer services. However, according to the concept of service-dominant logic, the so-called customer generically may be any person getting benefits of the service (Lusch and Vargo 2014). The human being may well be a consumer, but also an employee getting support for doing his job, a citizen getting support for his everyday life, or also an individual representing a societal stake.

## 3 The Gap Toward Data Product Design

Keeping in mind our definition of data products (the application of data science competences to provide benefit to another entity), the service design approach discussed so far clearly satisfies the second part of that definition, that is, providing benefit to another entity. However, there is still a gap w.r.t. to the application of data science competencies: service design per se does not systematically consider using analytics competences to bring forth benefits for the customer. In cases where the respective data is available, leveraging analytics capabilities in service design (i.e., doing data product design) generally yields more value to the customer and in return more revenue to the provider.

Two scenarios are conceivable—enhancing existing or creating completely new services:

1. First, we may assume that an existing product or service is effective in meeting the customer needs but could do this *more efficiently* if insights from data were used. For example, assume a service giving advice to customers when to replace existing factory equipment (machines). Leveraging data about the status of the old machines (i.e., condition monitoring) as well as forecasted production volumes, market evolution, etc., the service can become much more efficient and more effective. In this scenario, an existing solution becomes more efficient and is provided with higher quality.
2. Second, by leveraging data science, we can find *completely different and new products* which are much more effective in meeting the customer needs. Although new data products do not create new customer needs[2] (the fundamental

[2]There is often the belief that technology can create new customer needs, which is only true at a superficial level. If we dig deeper in the hierarchy of customer needs, which we do in service design, we find underlying needs which are given by the customers' tasks.

underlying motivations and needs of customers have been there before, often not at the conscious level), the new data products may provide completely new and previously inconceivable ways to satisfy those needs. For example, we may develop a configurable music player that continually evaluates data about the context and situation of the user via a connection to his smartphone and adapts the playlist to meet the circumstances, smoothly adapting to events like new music releases or sensed moods and environmental conditions.

Designing the resulting data products requires methodologies that go beyond those covered by the service design literature. Meierhofer and Meier (2017) propose an approach to data product design which we are going to discuss in the following section.

## 4 Bridging the Gap (Then and Now)

From the previous discussion we see that service design provides us with a framework to systematically design products that generate relevant benefits for the customer. These benefits could be quantitatively or qualitatively higher if the potential of data was leveraged.

However, data scientists made the experience in recent years that insights generated by sophisticated analytics algorithms are often not properly adopted or undervalued by the users: the insights may be considered technology driven, not relevant for the user, or simply not trusted by experts (Finlay 2014; Veeramachaneni 2016). Hence, there is a gap between analytics results and value creation. This gap needs to be bridged in order to exploit the potential of data products (see Fig. 4.3).

Of course, many excellent data products available today show that this gap can be bridged: the examples of Siegel (2013) in nine different industries have already been mentioned. Such cases, in which insights from data are developed into data products that fit with the customer needs, might be successful because of the situative combination of good ideas: interdisciplinary teams formed by so-called "T-shaped-people"[3] (Stickdorn and Schneider 2010) (i.e., by the ideal profile of a data scientist) may be sufficiently creative to exploit the potential of analytics while deriving a value proposition that is consequently driven by the customer needs. However, a more systematic methodology for the development process is desirable.

First approaches for systematic data product design have been presented in the literature after Loukides (2011) pointed out that "...the products aren't about the data; they're about enabling their users to do whatever they want, which most often has little to do with data." Howard et al. (2012) then suggested the so-called drivetrain approach that we will briefly review below. Recently, Scherer et al.

[3]The horizontal part of the T-shape refers to the broad skills in a large field like data science, with additional depth in a specific sub-field, e.g., service design or analytics (the vertical part).

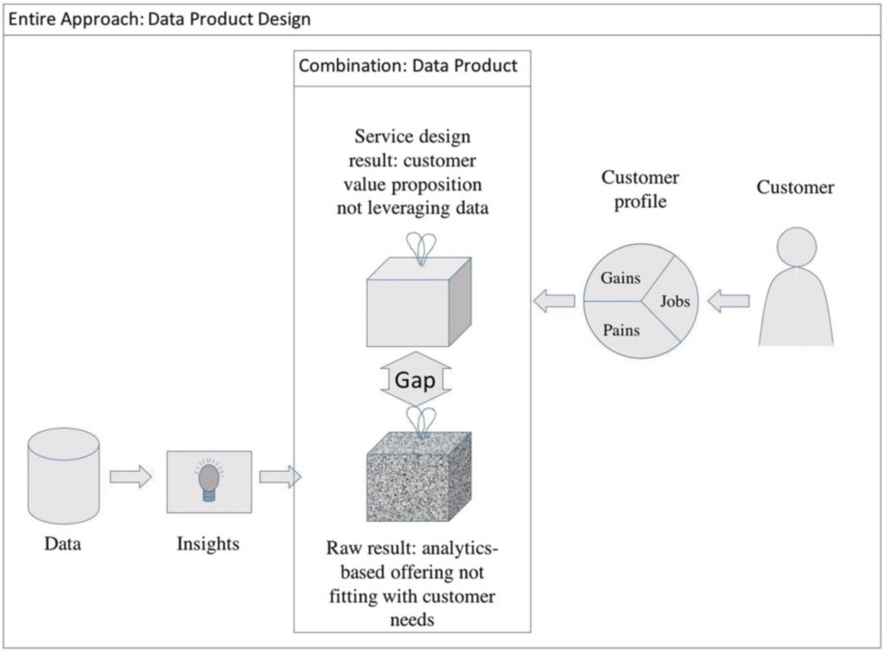


**Fig. 4.3** Data products bridging the gap between analytics and service design

(2016) presented another approach on how to use data analytics for identifying user patterns.

The 4-stage drivetrain process starts with the definition of the goal: Which of the user's needs shall be addressed next? Let us assume for a moment the example of "web search"—finding information on the web, based on keyword queries. The second step is then the identification of the levers that the data scientist can set to reach this goal. This may, for example, be a new analytics idea, as has been the case with the "PageRank" algorithm within Google's then new answer to the web search example created above: it is based on the idea that the number of incoming and outgoing links to web pages (so-called off-page data) contain information about its relevance with respect to a query. The third step consists of collecting the necessary data sources to enable setting the identified levers of the previous step. In the Google example, this data was collected by the company in their search index. The data may thus already be available internally. However, the combination of internal and external data has great potential for (and often holds the key to) the realization of new analytics ideas. For this reason, the question of how to design good data products is closely linked with knowledge of the international data market as well as of the open data movement and respective options: publicly available datasets may at least augment one's internal data, as the next section will show. The fourth step finally involves building analytical models, as the options of which modeling technique to apply are to a large extent predetermined by the previous three steps.

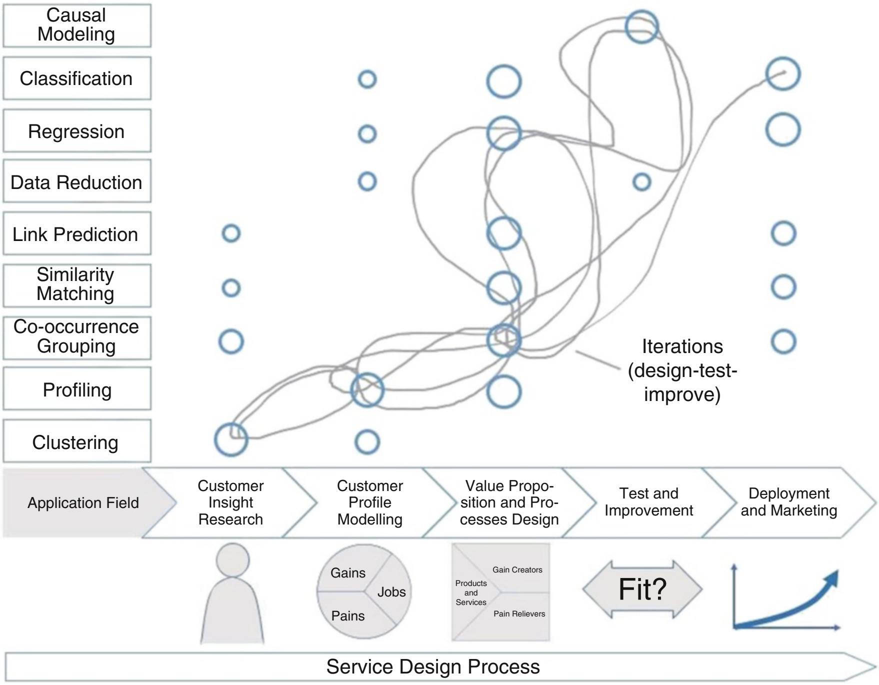


**Fig. 4.4** Methodological bridge between the service design process and data analytics tools

The drivetrain approach is the distillate of the lessons learned of hundreds of publicly held data science competitions. While capturing indispensable knowledge, it is still quite abstract, being more descriptive than prescribing next actions: it serves well as a model to conceptualize a successful data product design project in retrospect but is hard to use as a model to decide on the next concrete step.

To overcome this weakness and provide an all-encompassing approach for data product design, we propose to cover all phases of the service design process and additionally exploit the full spectrum of data analytics methods and tools as much as possible, in a way that allows for planning ahead. We use the framework shown in Fig. 4.4.:

- The horizontal axis depicts the stages of a typical service design process: for a given application field (e.g., "customer searches and purchases a new product of our portfolio") we start the process with collecting data about potential users or customers ("customer insight research"), then build the customer profile (jobs, pains, gains), followed by the phases for designing the value proposition and the service processes. In the next phase, we test the fit between value proposition and the customer profile and improve our solution in several iterations (indicated by the squiggles in the figure). In the last step, we bring our new data product to the market (deployment and marketing).

- The vertical axis in Fig. 4.4 shows a structure of several data analytics methods w.r.t. their potential to provide benefits for data products (raw results according to Fig. 4.3). The terms on the vertical axis (from “clustering” to “causal modeling”) stand for fundamental data analytics methods (or tools, use cases) according to Provost and Fawcett (2013).
- The dots in the matrix framework of Fig. 4.4 indicate in which stage of the data product design process which data analytics tool can be typically applied. Larger dots qualitatively indicate a stronger value contribution in the corresponding combination.

As practical service design cases often do not follow the service design process in a linear way from left to right in Fig. 4.4, we exemplarily discuss the matrix in Fig. 4.4 by a case study according to Meierhofer (2017). This example is in the application field of customer service representatives (agents) in a company providing consumer services. The goal is to provide employee support services to the agents in order to make inconvenient tasks easier for the employees, to reduce sources of errors, and to increase efficiency. Such tasks may be, for example, to detect the relevant contact reason of the customer in real-time (e.g., while having the customer on the phone) and find related contacts of the same customer in the past. For instance, a customer may have contacted the firm for various matters several times before and this time has a complaint for one specific topic, which makes it difficult for the agent to dig the details relevant for this complaint out of the contact history in a short time. Or the customer may call because of a specific question concerning his customized product instantiation or his customized contract. It is likely that the company has the answer ready in its data base in the form of a documented solution of a similar problem in the past. An individual agent can impossibly keep all these cases in mind or find them while talking to a customer.

Data and technical tools, e.g., records of past customer interactions as well as algorithms for speech-to-text and natural language processing, are assumed to help in this process and provide benefits to the agents. Finding the relevant nuggets in the bulk of past customer contacts, which often are documented as unstructured data, can be heavily supported by such analytics tools. Hence, this case study starts from the perspective of data and technology according to Meierhofer and Meier (2017) instead of a precise understanding of the user’s jobs, pains, and gains. It can be considered a technology push approach.

In traditional service engineering procedures, the project would deploy as follows:

- An interdisciplinary project team is set up consisting of (a) analytics specialists, (b) IT specialists in the company-specific CRM system, and complemented by (c) business process specialist of the customer service department.
- In a requirements engineering process, the required features for the agent support tool are elaborated and then stripped down to a feasible set in the framework of the project constraints (cost, time, quality).
- The tool is implemented, technically tested, and deployed to the users. This last step includes training as well as change management aspects in order to convince

the agents of the benefits of the new tool. This development and deployment phase would typically span over several months and result in high resources costs.

Unfortunately, this procedure often turns out not to be effective in the sense that the tool delivered after the long development period does not solve relevant jobs, pains, or gains of the users. As a consequence, the users consider the tool irrelevant and are not ready to invest the energy to get sufficiently familiar with it in order to leverage at least some benefit. This is the point where cultural change management comes in to get the agents to use the new tool, which is often not successful. At the end, the project may be considered a disappointment.

To circumvent this problem, best-practice approaches have come up in the recent years tackling the problem from a design perspective in combination with agile methodologies. The challenge to support the agents in their daily work would consequently start by understanding and modelling the agents jobs, pains, and gains. Next, a value proposition would be developed which helps the agents to do their job, overcome the pains, and increase their gains. However, this procedure would typically miss out the potential of the new possibilities in analytics, which may be assumed in the fields of mining data (e.g., past customer interaction records) or process automation (e.g., speech recognition). As a consequence, for the case study described above, an agent support tool may be built which turns out to be useful for the agents, for example, by providing search tools for similar problems, but could possibly provide much more benefit by systematically applying analytics.

Now, applying the new data product design scheme shown in Fig. 4.4, we proceed as follows:

- To start, remembering that we have a technology-driven case, we elaborate a map of the data-driven assets available which we assume to provide benefits for the given problem statement. In this case, this is:
  - Generating a layout of insights that can be gained from past customer interactions. The data of closed customer contacts, which is stored in records in the CRM tool, is mined and interpreted by data scientists in co-creation with process experts of the customer service department.
  - Exploring the possibilities of natural language processing and speech-to-text conversion in the context of the agents' work with a CRM system (e.g., the environment of the use case, the languages applied, the acoustical environment, the real-time requirements, etc.).

  This collection of the data-based value contributions as a starting position corresponds to tackling the problem from the left-hand side in Fig. 4.1 and to establishing the vertical axis in Fig. 4.4.
- In the next step, we develop the horizontal axis of Fig. 4.4 and proceed with understanding the agents' jobs, pains and gains. To do so, we research insights about their jobs, pains, and gains by shadowing a qualitative sample of agents in their daily job (i.e., accompanying the agents as an observer). A practical tool to do this can be found in the "a day in the life of" concept: accompanying a person during a typical day and observing what she does, where she struggles or needs

too much time or energy for doing a job, and where she gets the desired output in a satisfactory manner.

This qualitative customer insight research step is complemented by a quantitative analysis of process data found in the agent workflow tool. The process steps found completed in past customer interactions are stored with their timestamp as well as the type of process step and free text remarks entered by the agent. This analysis backs up the qualitative insights about the jobs, pains, and gains found so far, and eventually verifies or falsifies the hypotheses.

- This collection of agent data also enables the potential segmentation of the agents into different profiles (so-called "personas" in the service design terminology) by clustering approaches. Based on this, different profiles of agents can be described. If the analysis yields different relevant profiles with clear differences in the pains and gains (the jobs are assumed to be the same in the same job context), the service for the agents needs to be developed with different flavors depending on the profile.
- Next, we tackle the task of developing the actual service for the agents, which means developing the value proposition (left-hand side of Fig. 4.2). In this step, we now make use of the collection of the data-based value contributions that we prepared at the start of our technology-driven approach. We confront the elaborated agents' jobs, pains, and gains with those value contributions differentiated according to the customer profile. This step yields the following outcomes:

  – There are jobs, pains, or gains to which we can responded by the given data-based value contributions. For example, finding similar cases in the past may be supported by similarity matching of the current case description with past descriptions by means of Information Retrieval methods.
  – There are jobs, pains, or gains for which we do not have a data-based value contribution. This situation takes us to making additional data sources accessible or to solving the problems by non-data-based means. For instance, it would be very helpful for the agents to get an indication of the customers' current emotional tension and the evolution of this in the past. We may not have sufficient data of the past cases to detect this reliably and may suggest a conversational script for the agent to find this out while talking to the customer.
  – There are data-based value contributions for which we do not have a corresponding job, pain, or gain (yet). In this case, we may find that the particular data has no value for our problem. Or, alternatively, we may find a way to utilize the data for solving the problem in a new way which was not seen before. Example: for a given customer enquiry, we may have data indicating that other users already had the same problem before, but the solution could not be standardized enough to generate a script for the agents for solving future problems. However, we can leverage this information to create a user support community and defer users whose problems have sufficient similarity to this community for peer-to-peer problem solving.

- The new service for supporting the agents (i.e., the value proposition) designed in this way is developed in several prototyping steps and tested with a sample of agents. These tests reveal technical bugs, but much more important, make transparent whether our hypothesis on the agents' jobs, pains, and gains as well as the corresponding value proposition are validated or falsified. If falsified, we introduce an additional iteration and adapt our service to test it again until we find a sufficient fit of our solution with the problem.
- Finally, when we deploy the new tool to the entire group of our customer service representatives, we measure how the tool is used by collecting data from the process workflows and the CRM data records. We detect where the solutions can be improved and enter the continuous improvement process.

## 5 The Essential Building Block of a Data Product

We finally turn our attention to the essential building block that distinguishes a data product from universal services: the supporting data and its analysis. Here, we focus on the data sources, since methods and technologies for data analytics are covered in detail in several other chapters of this book.

A data product can only be as good as its supporting data. While this statement might sound trivial at first sight, it has enormous impact on the design of a data product: if the underlying data is unreliable, all efforts to get high-quality analytics results and creating value to the customer must fail. Here, "unreliable" includes various types of issues, for example, incomplete or faulty data entries, unavailability, legal issues, etc. Hence, careful selection of appropriate data sources is an important step in the development of a data product.

Many data products are based on internal data. This data is proprietary to the data service provider, who often has full authority over its content, data format, access rules, licenses etc., which makes it comparably[4] easy and straightforward to incorporate it in a data product. However, there are still some reasons why internal data might not be used for a data product:

1. It is *personal* data, that is, "all information relating to an identified or identifiable person" (FACH 1992); this could be, for instance, customer profiles, phone call history or transcripts, customer feedback, etc. All personal data is subject to privacy regulations, which vary significantly from country to country. For instance, in the USA any data that might be traced back to an identifiable person is considered private and, thus, protected. When Netflix, a video-on-demand service, released a dataset of movie ratings of its users, researchers were able to

[4]Numerous hardships are attached to the process of extracting, transforming, and loading (ETL) even internal data into a form that is amenable for analytics. The topic of (automatic) data integration and corresponding engineering efforts toward a data warehouse is huge. For the sake of this chapter, however, we will assume the respective organization has already taken care of it.

identify individual persons within this supposedly anonymized data, thus, forcing Netflix to withdraw the dataset (Singel 2009).

2. The data is *confidential*, for example, emails, internal documents, meeting minutes, etc., and an automated data product might unwittingly reveal such information to the customers.
3. The data was *intended for a purpose* different from the data product: for instance, the "Principle of Earmarking" in German and European data protection regulations explicitly prohibits usage of personal data for any other than the intended purpose without consent.

As a way out, data products may augment internal data with additional external sources to provide maximum benefit to the user. There exist literally hundreds of thousands of external datasets that are available for download[5] (static) or via an "Application Programming Interface (API)" (dynamic). Thus, the question often is not *if* a useful dataset exists, but *where* to find it in the vast expanse of accessible data sources. To this end, *data marketplaces*, such as datahub, Amazon AWS Datasets, or Microsoft Azure Marketplace, come into play, which are useful in three major ways: they are a central point of discoverability and comparison for data, along with indicators of quality and scope; they handle the cleaning and formatting of the data, so that it is ready for use (this step, also known as data wrangling or data munging, can take up to 80% in a data science project (Lohr 2014); and they offer an economic model for broad access to data that would otherwise prove difficult to either publish or consume.

On the other hand, there exists a vast amount of *open data*, which is ever-increasing since more and more governments, research institutions, and NGOs are adapting open data strategies. These data include, for instance, human genome sequences, historic weather data, or voting results of the Swiss National Council. Data collections such as data.gov (USA), open-data.europa.eu (European Union), or data.gov.uk (United Kingdom) contain thousands of public datasets (see Chap. 14 on the usage of open data). While most of these datasets are stand-alone, *Linked Open Data* (LOD) provides methods to interlink entities within open datasets. Linked data, which goes back to Tim Berners-Lee (2006), uses a Uniform Resource Identifier (URI) for each entity, and RDF triples to describe relations between these entities. This allows machines to traverse the resulting graph, which contains nowadays billions of entities, and collect required information automatically.

Once the underlying data of the data product is clear, it can be collected, pre-processed, combined, and analyzed to provide the desired service to the customer. Since most data products rely on data that changes over time, it is important to track the data sources closely, because API's can be updated, data formats may change, or entire data sources may vanish completely. Only then it can be ensured that the data product works reliable and to the benefits of the customer.

[5]See, for example, http://cooldatasets.com/

# 6 Discussion and Conclusions

We have reviewed the state of the art in data product design and concluded that up to now no systematic approach has been presented that allows for planning the next steps in designing a product based on data insights specifically to the needs of a certain customer. We suggested to extend the methodology found in the discipline of service science by concrete ideas on how and where to invoke certain analytics methods and tools. We argued that using the methodology, and hence vocabulary, of service-dominant logic and service design gets data scientists a long way toward such a development processing on the broad range of possible data science use cases, not just in typical "service business" settings. In Fig. 4.4, we presented a concise but all-encompassing framework of how to develop data products from a user-centric point of view, including suggestions of typically helpful analytics methods and tools per design phase. Finally, we gave pointers to potential external data sources to enhance the essential building block of each data product—its underlying data collection.

We see data product design as a discipline that is still in its infancy.[6] Its core and borders are still very much under development:

- While one of the first university-level courses on the topic mentions to ". . .focus on the statistical fundamentals of creating a data product that can be used to tell a story about data to a mass audience" and then focuses on technical details in building web applications (Caffo 2015), others are based on a curriculum that focuses on service design, leaving analytics aspects to other modules (Stockinger et al. 2016).
- While the drivetrain approach has been too abstract to guide new design endeavors, our approach is conceptually nearer to certain kinds of applications and thus may in practice be more difficult to apply to a problem of, say, the internal control of a machine (where no user is directly involved) than in marketing (although it really is generally applicable).

We thus see our presented approach as a contribution to an ongoing discussion: all data scientists need, besides deep analytics know-how, the business-related skills to not just design a successful algorithm, but to think through a whole product. This is the all-encompassing process we have sketched above. For the engineering-heavy data scientist, who daily mangles data and thinks in terms of algorithms, this may seem far away: she is more involved in CRISP-DM-like processes (Shearer 2000) to assemble the smaller parts of the final solution. But these smaller parts are then treated as black boxes within the all-encompassing data product design process as outlined above.

In this sense, the data product design approach presented here is not the process to create each data insight (smaller part). It is the packaging of one or many of these into "publishing" form through optimization, wrapping, and finally marketing.

[6]Borrowing a phrase from Michael M. Brodie that he frequently relates to data science as a whole.

Future investigation has to answer the question of how to bring both processes into one conceptual framework: the "internal" CRISP-DM-like data insight creations, and the "external" data product design wrapper.

## References

Berners-Lee, T. (2006). Linked data. *Blog post*. https://www.w3.org/DesignIssues/LinkedData.html

Caffo, B. (2015). *Developing data products*. MOOC at Johns Hopkins University, Coursera. https://www.coursetalk.com/providers/coursera/courses/developing-data-products

Federal Assembly of the Swiss Confederation. (1992). *Federal act of 19 June 1992 on data protection (FADP)*, paragraph 3. https://www.admin.ch/opc/en/classified-compilation/19920153/index.html

Finlay, S. (2014). *Predictive analytics, data mining and big data: Myths, misconceptions and methods*. Basingstoke: Palgrave Macmillan.

Howard, J., Zwemer, M., & Loukides, M. (2012). *Designing great data products – the drivetrain approach: A four-step process for building data products*. Retrieved March 28, 2012, from https://www.oreilly.com/ideas/drivetrain-approach-data-products

Jagdish, N., Sheth, B., & Newman, I. (1991). Why we buy what we buy: A theory of consumption values. *Journal of Business Research, 22*(2), 159–170.

Kottler, P. (2003). *Marketing management*. Upper Saddle River, NJ: Prentice Hall.

Lohr, S. (2014). For big-data scientists, 'janitor work' is key hurdle to insights. *The New York Times*, blog post. https://mobile.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html

Loukides, M. (2010). *What is data science?*, blog post. https://www.oreilly.com/ideas/what-is-data-science

Loukides, M. (2011). *The evolution of data products*. O'Reilly Media. ISBN 978-1-449-31651-8.

Lusch, R. F., & Vargo, S. L. (2014). *Service-dominant logic*. Cambridge: Cambridge University Press.

Marr, B. (2016). *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*. Chichester: Wiley. ISBN 978-1-119-23138-7 (hbk).

Meierhofer, J. (2017). Service value creation using data science. In E. Gumesson, C. Mele, & F. Polese (Eds.), *Service dominant logic, network and systems theory and service science: Integrating three perspectives for a new service agenda*. Rome: Youcanprint Self-Publishing.

Meierhofer, J., & Meier, K. (2017). From data science to value creation. In *Proceedings of International Conference on Exploring Service Science* 2017.

Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014, November). *Value proposition design*. Hoboken, NJ: Wiley.

Polaine, A., Løvlie, L., & Reason, B. (2013). *Service design – From insight to implementation*. Brooklyn, NY: Rosenfeld Media.

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

Scherer, J. O., Kloeckner, A. P., Duarte Ribeiro, J. L., Pezzotta, G., & Pirola, F. (2016). Product-service system (PSS) design: Using design thinking and business analytics to improve PSS design. *Procedia CIRP, 47*, 341–346.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *J Data Warehousing, 5*, 13–22.

Siegel, E. (2013). *Predictive analytics – The power to predict who will click, buy, lie, or die*. Hoboken, NJ: Wiley. ISBN 978-1-118-35685-2.

Singel, R. (2009). Netflix spilled your brokeback mountain secret, lawsuit claims. *Wired*, blog post. https://www.wired.com/2009/12/netflix-privacy-lawsuit

Smith, J. B., & Colgate, M. (2007). Customer value creation: A practical framework. *Journal of Marketing Theory and Practice, 15*(1), 7–23.

Stadelmann, T., Stockinger, K., Braschler, M., Cieliebak, M., Baudinot, G., Dürr, O., & Ruckstuhl, A. (2013). *Applied data science in Europe – Challenges for academia in keeping up with a highly demanded topic*. European computer science summit ECSS 2013. Amsterdam: Informatics Europe.

Stickdorn, M., & Schneider, J. (2010). *This is service design thinking*. Amsterdam: BIS.

Stockinger, K., Stadelmann, T., & Ruckstuhl, R. (2016). Data Scientist als Beruf. *Big Data*. Edition HMD, Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-11589-0_4.

Veeramachaneni, K. (2016, December 7). *Why you're not getting value from your data science*. Harvard Business School Publishing Corporation. https://hbr.org/2016/12/why-youre-not-getting-value-from-your-data-science