To Share, Curate or Sell: Three Pathways of Using Data in Open Innovation

ABSTRACT

The open data market size is estimated at €184 billion and is forecast to reach between €199.51 and €334.21 billion in 2025. Although the data volumes are increasing, the literature is still far from understanding the role of data in open innovation. This study asks: how can companies use data in open innovation? We interviewed 25 professionals in business transformation, data science, and domain experts in animal health to answer this question. First, we provide evidence of the role of data in open innovation for extracting value. Second, we theorize how data in inbound and outbound open innovation can be shared, curated, or sold. The study is limited by the selection of a highly regulated animal health industry and welcomes further research to confirm or extend its findings.

Keywords: Open innovation; Business model innovation; Qualitative Methods

INTRODUCTION

In the current environment of high volumes of data and increasing technology use, firms are changing their work practices and looking for alternative ways of gaining a competitive advantage. As part of these changes, large, established organizations are considering how they can capitalize and create strategic value by leveraging data. As a result, they turn to external data sources from their suppliers and customers to sustain new services and product offerings (Smith et al., 2016). However, with studies showing that open data could unlock more than $3 trillion in additional value worldwide across different application domains (Manyika et al., 2013), the literature is still far from understanding how data contributes to value realization.

Recent research has considered the benefits of using data for supporting artificial intelligence. For instance, among many ventures in the oncology domain, Lauer et al. (2021) contend that Cambridge Cancer Genomics is exemplary in its use of open data to train machine learning models. As a result, it develops precision oncology solutions that can detect the relapse of patients sooner than the norm, predict responses to cancer therapy, and reduce ineffective treatment protocols. Another example is ChatGPT (Generative Pre-trained Transformer). This AI tool uses learning on vast amounts of data from the internet before 2021 and provides responses that may sound human-like[1]. Considering examples, it is important to systematize how data realize value across application domains. Particularly, we focus on animal healthcare, where vital data is locked between animal owners, animal clinics and veterinarians due to privacy restrictions. Our study aims to understand how using data in open innovation can improve value realization. By so doing, we will provide incentives for data owners to unlock and share data for open innovation, therefore increasing the chances of developing predictive models in various sectors.

[1] https://help.openai.com/en/articles/6783457-chatgpt-faq (accessed 7.01.2023)

Open innovation (OI) has captured the attention and efforts of scholars and industry practitioners, who noticed a shift from closed to open models (Chesbrough, 2003). OI rests on the principle that firms leverage internal and external ideas and paths to market to innovate, defining new organizational architectures and systems (Bogers et al., 2018; Smart et al., 2019). Chesbrough and Bogers (2014, p.17) define OI as "a distributed innovation process based on purposefully managed knowledge flows across organizational boundaries." In essence, this involves harnessing inflows and outflows of knowledge to drive innovation within the firm. Whilst open innovation has attracted significant research attention (Michelino et al., 2015; Bagherzadeh et al., 2020; Nguyen, Huang & Tian, 2021; Bagherzadeh et al., 2022; Barczak, Hopp, Kaminski, Piller & Pruschak, 2022), there is limited current literature that provides empirical and academic evidence and suggestions for the best practice ways to deliver data-driven open innovation (Lichtenthaler, 2008). A few studies focused on aspects such as open innovation governance modes (e.g., Bagherzadeh et al., 2022) and open innovation platform design (Osorno & Medrano, 2022) or the effect of data to stimulate innovation (van Veenstra and van den Broek, 2013; Janssen et al.,2012) yet we lack a holistic understanding of how data impacts open innovation. Particularly, studies such as Bogers et al. (2018); Eckartz, van den Broek and Ooms (2016); Monino (2021) and Zuiderwijk et al. (2015); have studied the use of technologies for data generation, but no study places clear evidence of the data-driven value. There is less research specifically focused on how data in open innovation can create strategic organizational value (Beck et al., 2020; Bogers et al., 2018; Eckartz, van den Broek, Ooms, 2016; Fritsch, Titze, & Piontek, 2020; Monino, 2021). Given the focus on explaining how to extract strategic value for organizations, the research question is: *how can companies use data in open innovation?*

Firstly, we provide evidence of the role of data in extracting value for organizations (Smith et al., 2016), based on interviews with specialist professionals working with data and managing organizational data. Secondly, we theorize how data influences open innovation by bringing data sharing, data curation, or data selling. Further, we have constructed three purposeful pathways for using data in open innovation: (1) data disseminator – for research purposes; (2) data curator – for data aggregation and servitization purposes; or (3) data marketplace – for data selling purposes. By doing so, we challenge the existing requirements of open innovation organizational systems (Bogers et al., 2018; Naqshbandi, 2018).

In the sections that follow, we review the literature on open innovation to show the state-of-the-art positions and debates currently considered important. We then present the review of open data, business models, and business model innovation and its key theories and frameworks, which inform current debates on data-driven open innovation. This review positions the current study and has led to our research question. The methodology explains the data collection strategy for the 25 qualitative interviews and how it was collated and analyzed using thematic analysis (King and Horrocks, 2010). The thematic framework is presented in the methodology section. The findings are based on the thematic framework, and the key explanation of the results is presented. The discussion is then presented, three pathways of using data in open innovation. The paper concludes with a focus on generalizing the theoretical implications of the findings to organizations in further domains. Finally, the practical implications are presented, and the paper concludes with limitations and suggestions for future studies.

LITERATURE REVIEW AND THEORETICAL FOUNDATIONS

The rise of open data and open innovation

Open data has been described as the "digital fuel of the 21st century" (Kundra, 2012) as an enabler of new economic activity and innovations (Davies and Perini, 2016). Indeed, open data can be leveraged to realize value and drive economic growth (Smith et al., 2016). Referring to "a system of informative and freely accessible databases that public administrations make generally available online in order to develop an informative network between institutions, enterprises and citizens" (Maretti et al., 2021, p.1134), open data currently supports value creation across all sectors of the economy. "Open" refers to how "liquid" and transferable data are (Manyika et al., 2013, p.4).

While open data has traditionally been associated with the open government movement, it has quickly been adopted by the industry. However, while enterprises generate substantial amounts of data as part of their operations, their true value-generation potential is yet to be realized (Fritsch, Titze, & Piontek, 2020; Smith et al., 2016; Bonina et al., 2021). Open data is closely related to open innovation (OI) which has captured the attention and efforts of scholars and industry practitioners. Introduced by Chesbrough (2003), who noticed a shift from closed to open innovation models, OI rests on the principle that firms leverage both internal and external ideas and paths to market to innovate, defining new organizational architectures and systems (Bogers et al., 2018; Smart et al., 2019), along research, development and commercialization phases of the famous funnel. In essence, OI involves harnessing inflows and outflows of knowledge, including data, to drive innovation within the firm (Chesbrough and Bogers, 2014, p.17). Moreover, OI has become even more important in the context of the opportunities brought by the wider digital transformation of business and society, which enables value to be created in new ways and promises to radically transform all industries and sectors (Chesborough, 2003; Huizingh, 2011).

For example, a key aspect of OI is placing the user at the centre of innovation processes within an ecosystem of people, organizations, and sectors that supports value co-creation (Bogers et al., 2018). Importantly, a key consideration in the context of OI is the business model, which captures and defines the (new) requirements of OI organizational systems and architectures

(Bogers et al., 2018; Barczak, Hopp, Kaminski, Piller & Pruschak, 2022). Thus, OI and the leveraging of open data cannot happen without consideration of business models, specifically how these may need to change and transform to meet the demands of OI.

Open data, Business Models, and Business Model Innovation

Bogers et al. (2018, p.10) emphasize that OI requires "business models—the logic of creating and capturing value—that dynamically transcend organizational boundaries within that innovation ecosystem." It contrasts with large organizations' traditionally closed business models, whereby value is largely created from internal knowledge sources. As OI opens business models to external knowledge flows and inputs, this may involve significant reconfiguring ways of doing business. Therefore, business model innovation is critical in examining how organizations can unlock value through OI, especially in an increasingly digitalized economy and society.

In broad terms, the business model can be understood as a framework that explains how companies "do business" or, as Osterwalder et al. (2005, p.4) put it, "the blueprint of how a company does business". Zott and Amit (2013, p.404) define the business model as a "system of interdependent activities that are performed by the firm and by its partners and the mechanisms that link these activities to each other." In other words, the business model embeds the logic of value creation in an organization. Magretta (2002, p.4) explains that "a good business model answers Peter Drucker's age-old questions: Who is the customer? Moreover, what does the customer value? How do we make money in this business? Finally, what underlying economic logic explains how we can deliver value to customers at an appropriate cost?"

Furthermore, BMs comprise a series of front-end and back-end components that together provide the building blocks of the business (Günzel and Holm, 2013). These include value propositions, product/service offerings, customers, key partners, value creation mechanisms, and value appropriation mechanisms (Fjeldstad and Snow, 2018). While business model definitions vary, they all point to three key dimensions: value creation, value delivery, and value capture (Arend, 2013; Berends et al., 2016).

Moreover, the business model lends itself to innovation. Casadesus-Masanell and Zhu (2013, p.464) define business model innovation as "the search for new logics of the firm and new ways to create and capture value for its stakeholders; it focuses primarily on finding new ways to generate revenues and define value propositions for customers, suppliers, and partners." Importantly, firms that develop new business models enjoy substantial returns (Euchner, 2016), with studies showing a positive impact on firm financial performance (e.g., Pohle and Chapman, 2006). Business model innovation has been described as "the cornerstone of long-term performance" (George and Bock, 2010).

However, when it comes to OI, many companies "find it hard to find a good fit" (Van der Meer, 2007), where "fit" relates to the difficulty of identifying appropriate business models to leverage the opportunities of OI. Indeed, Euchner (2016, p.10) highlights that "new business models conflict with well-entrenched practices, requiring the entire organization to move beyond its comfort zone". However, scholars have shown that large innovative organizations tend to display closed behaviour toward open innovation when the potential benefits of OI become clear (Fritsch, Titze, & Piontek, 2020; Van der Meer, 2007).

Importantly, Zott and Amit (2007, p.181) note that the business model "elucidates how an organization is linked to external stakeholders and how it engages in economic exchanges with them to create value for all exchange partners". It relates closely to the concept of OI. As a more hidden "locus of innovation" (Zott and Amit, 2007, p.183), the business model can drive innovation beyond firm boundaries and is thus a suitable conceptual vehicle for exploring opportunities for data-driven value realization. Täuscher (2018) suggests that future research can further leverage qualitative research methods to investigate configurations of business models, competitive strategy, and industry characteristics needed for superior firm performance. Given the opportunities presented by the unlocked data and open innovation, the key question is: *how can companies use data in open innovation?*

To answer this question, the paper develops three pathways of using data in open innovation. The next section presents the methodology and empirical focus of the paper, picking up on how value is created, delivered, and captured.

METHODOLOGY

The data collection for this research is based on semi-structured interviews, which are expected to shed light on the role of data in open innovation. Qualitative data were collected through in-depth interviews performed between May and July 2021. A total of 25 professionals working in animal healthcare and data science were interviewed about data-driven innovation and the value of data for open innovation, Table 1. We focused on the animal healthcare sector specifically for several reasons; firstly, this is a £7.8 billion industry in the UK where much data is generated and accessible, meaning that there are significant opportunities for data-driven innovation. However, this data is often left underexploited by companies in the industry (Chui et al., 2014). Secondly, these data are less regulated than human healthcare; therefore, the ongoing analytical works on animal health and welfare can inform similar human projects in the future. We present an overview of our respondents in Table 1.

Insert Table 1 here

Data collection was performed following a rigorous procedure. In the first step, six interviews used open-ended questions to explore the value of data, namely how data sharing could be used for value creation, delivery, and capture. Next, we targeted the most experienced stakeholders in business transformation, data science, and entrepreneurship to distil how companies leverage data for open innovation and gain a first insight into types of data-enabled value realization. Thus, we identified three pathways of using data in open innovation, constructed following the earlier questions as dimensions, illustrated later in the paper in Figures 2 and 3. Later, these were discussed with other interviewees to enhance the description of each type. In this step, we also validated the previous findings and stopped when we reached saturation, as the improvements brought by additional cases (interviews) were marginal (Eisenhardt, 1989). Additionally, rigorous procedures were adopted to increase the equivalence, reduce the bias, and increase the credibility and the dependability of the results, which can be considered alternative terms to reliability and validity as used in quantitative studies (Guba & Lincoln, 1989; Sinkovics et al. 2008).

Thematic analysis

We analyzed the data thematically, a method that seeks to find themes of interest in the dataset (Braun & Clarke, 2006; King& Horrocks, 2010). It allowed us to reassess the pre-defined initial ideas through "inductive contact with the data" (Fereday & Muir-Cochrane, 2006, p.88). This approach is suggested for investigating "commonalities between [a] set of companies rather than analysis of individual cases in greater depth" (King & Horrocks, 2010, p.159). We followed the abductive procedure of Dubois and Gadde (2002), which involves simultaneously evolving the theoretical, empirical fieldwork, and case analysis. Figure 1 illustrates the coding scheme. Figure 1 illustrates the coding scheme.

Insert Figure 1 about here

In the process, we agreed on including new sub-categories as they appeared in the data (King& Horrocks, 2010). When the analysis no longer identified new codes, we confirmed theoretical saturation and, therefore, the final coding scheme's structure (Bryman, 2008; Silverman & Marvasti, 2008). Inter-coder reliability was ensured by following an iterative process for analyzing content using documented peer-review and reflection of our coding (Creswell & Miller, 2010), contributing to the reliability of the study (Miles & Huberman, 1994). Any differences were resolved through peer review and discussion (Kassarjian, 1977). External validity was ensured by reviewing previous literature on data and business model innovation to remove observer bias (Eisenhardt & Graebner, 2007). After the thematic analysis was completed, we presented three pathways of using data in open innovation to a large pharma company that participated in our study to get feedback and enhance its validity (Crabtree & Miller, 1999). Furthermore, we ensured the respondents' full anonymity and confidentiality and addressed the potential of key informant bias.

RESULTS

Creating value via data

In response to the disparate and siloed data in many industries, it is critical to regulate data sharing to combine the available data and fill the gaps in this area. For example, IP15 stated that in animal healthcare, "much data does not talk to each other". IP18 confirmed that in the pharmaceutical industry, "data [also] operates in silos" and therefore requires aggregation around the topic of interest. Firstly, 23 experts mentioned the need for clear data governance to resolve the initial data-sharing concerns during this aggregation. For instance, IP8 highlighted that: "Big pharma has a challenge of privacy and feasibility. Where is my data going? How could I check this?". Additionally, IP5 stressed there should be a standard Non-Disclosure Agreement (NDA) where "third-party sharing must be allowed". The expert added that "asking for data does not work", and there should be incentives for data holders to realize value by sharing their data. For instance, large institutions could sign "explicit and multiple 1-1 arms-length data partnerships" to supply their analytical partners with sustainability. Furthermore, it should enable "data re-use from commercial projects" as customers of these companies would leave data, as suggested by IP6. Smaller companies, such as SMEs, and individuals have an unclear purpose for sharing due to the B2B inequality problem; those who contribute data are not those who get value. SMEs will contribute their data once there is a clear concept of value, for instance, "matchmaking with the right people to turn data into insights" (IP13), or to easily access actionable insights, with access to dive into how the insight was generated, as suggested by IP14. As IP18 emphasized, key questions need answering: "What is my own gain? Will it improve my productivity? A quick provision of insights could incentivize many smaller companies, as they have neither statistics departments nor significant analytical skills (i.e., no personnel to run analysis) on board. IP8 highlighted:

> Data sharing can make a difference for many companies which have data but do not understand probability theory.

Furthermore, 12 experts agreed that data is often unstructured, unformatted, and requires additional work. Therefore, *data preparation* is needed to add value**.** The experts suggested that data anonymization, cleaning, and standardization activities must be carried out before any analysis occurs. Business stakeholders, such as IP5, suggest anonymizing the data to avoid the problem of data threshold, namely "identifiable data of individuals (e.g., companies) whilst moving from the top to bottom". Tech entrepreneurs like IP11 highlighted the need to remove NULL values and corrupted entries. Responding to that, IP5 suggested that a way to improve the quality of datasets could be as simple as "making [anonymized] data available for research teams". Other experts stressed that dataset standardization is needed. For example, IP7 stated that "data should be validated, aligned to the model".

Most experts mentioned that the data is a handful for extending the existing thematic catalogues with accessible datasets, for example, via public data catalogues, data marketplaces or social media. IP7 claims this would provide a foundation for powerful predictive and prescriptive models. Another expert suggested, "put[ting] together pieces of data" (IP1) from the existing wellness devices, such as fitness monitors, into advanced deep learning models "to uncover important links between behaviour, activity, and health", which is applicable for animals as well as for human beings. Further open data, such as open weather forecasts (e.g., weather.com), air, soil and water quality could provide a fruitful opportunity for allergy predictor diagnostics. IP13 commented that the regular actualization of datasets must be assured via data ownership.

All experts for this study agreed that a clear pathway is *data-driven services.* These should enable evidence-based precision for customer decision-making, actionable insights with ease of access, and "give a certain outcome, such as risk-based analysis" (IP18). It would allow them to "understand the sudden changes of behaviour" and "to provide an early intervention to improve health" (IP1). IP7 claims that these services should deliver insights into powerful predictive models, for instance, to prevent epidemic outbreaks in organic fish farming.

Taken together, unlocking data and sharing data is a win-win situation for all stakeholders in the animal healthcare industry, creating value for humans, animals, doctors, pharmaceutical companies, and society and helping to address sustainable development goals: good health and wellbeing (3), life below water (14), and life on land (15). Furthermore, IP1 discussed that data contributors would receive personalized health management, namely "personalized care driven by objective, evidence-based, precision-medicine", whereas data-driven companies will gain "confidence, credibility and reliability" during treatments.

Delivering value through data

Whilst discussing value delivery from data, we found few changes compared to the conventional value delivery in information technology. The experts mentioned the following customer interfaces: (i) dashboards, (ii) mobile applications, and (iii) digital platforms.

Firstly, the experts suggested using *dashboards* for data visualization and comparing datasets, for instance, object monitoring and regression analysis results. Furthermore, the experts suggested that such dashboards could separate "sustainability-oriented" research from "business-oriented" work. For business-oriented purposes specifically, most experts advised using apps as the most suitable interface. For instance, IP15 commented that "people will download the app" in their industry, while IP7 agreed that "an app as a customer interface is preferred". For such customers, speed and readiness for analysis are important. For instance, IP8 commented that "getting faster is the priority", and IP2 agreed that "readiness for short-term contracting" should be provided.

Finally, a *digital platform* was suggested as a preferable interface for the data selling/ analytics marketplace. By using data platforms, smaller companies could market their datasets, larger firms may buy data for their specific needs, and data analysis freelancers could provide on-demand analytical services. As mentioned by IP14, the platform should enable "ease of access" for buyers and sellers in this market and provide information about the actors and resources.

Capturing value via data

Finally, the interviews considered ways of capturing value from data. Most experts suggested subscription to data-driven services, arguing that the "subscription rate turns an important cycle of data into better insights", said a business transformation partner in a large pharmaceutical company (IP13). IP5 agreed that subscription is the best capture mechanism for commercial users, while users in academia could attract research projects to cover the costs. Four experts have named research grants as another way to cover the costs of data cleaning, standardization, and analysis via the scientific community[2]. For instance, IP3 said that "self-sustaining" work, which is necessary for platform operation, could be funded by a research grant. In addition, universities can participate in data-sharing projects by (1) supplying data from experimental work and a "pipeline of talent" to work on these datasets; (2) increasing credibility for the data-sharing projects; and (3) providing infrastructure to store data. It enables the reduction of operational costs and the capture of value indirectly. Finally, the experts also provided examples of secondary activities that could help capture the value from data. For example, hackathons could be organized to get co-funding from industrial partners. These events focus data analysis efforts on a particular industrial challenge, which requires the collaborative work of analytical teams. During hackathons, the ideas are tested, realized into data-driven prototypes, and validated with the potential customer. Another expert suggested that digital platforms can help to "engage audiences" around data sets, which can be used to "sell ads to different user groups" (IP6) as a part of marketing campaigns of larger companies.

In summary, the respondents' answers visualized the process of data-driven value creation, delivery, and capture. Firstly, data governance enables a legal framework, sharing partnerships, and incentives for data owners. Secondly, data is anonymized, cleaned, and standardized, ready for application. Thirdly, the dataset is accumulated or enhanced using data, and data could be specialized for selling to a particular customer group. Further, the dataset is exploited by sharing, curating, or selling (see Figure 3 in the next section). Finally, the value could be delivered using statistical dashboards, mobile apps, or digital platforms and captured by subscription, research, or commercial or university partnerships.

Pathways to using data in open innovation

[2] see above section on data-driven value creation

Our analysis suggests that data does not create value directly but increases the volume of datasets available in open innovation research and development phases. In these data-intensive conditions, firms can undertake open innovation with new possibilities to create value. This value requires infrastructure for customer delivery and contracts to capture revenues. Thus, data improve the commercialization phase of open innovation. Considering the open innovation process, we position data as underpinning the commercialization phase of the open innovation funnel to enable inbound and outbound open innovation, Figure 2.

Insert Figure 2 about here

Based on the findings, we have constructed three pathways of using data in open innovation: (1) data disseminator, (2) data curator, or (3) data marketplace.

Data disseminator for outbound open innovation

Data disseminator considers the research perspective of data-driven value realization. A data disseminator "requires data volumes to be able to address notable questions" (IP13). IP2 clarified that this type requires "attracting strategic partners, e.g., universities, corporate companies for long-term research projects up to 5 years [for data sharing] and seeking funding through grants". These stakeholders accumulate datasets around the topic of interest and engage communities to investigate the data. Data can trigger the start of an analytical ecosystem by providing incentives to the community for data analysis and app development. As such, the issue of data attraction has been widely discussed, with IP15 commenting:

> The data is as good as you put it […] if you constrain it to one company, you will never get robust data […] It should be mutually beneficial to share data

Some mentioned that significant data could be collected about the area of interest. For example, communities can launch surveys with dichotomous questions, analyze answers, and sell results to large companies' marketing departments for targeting ads: "Once we have many data, we could do a lot with it … then we can target big companies… and run ads" (IP17). However, others advised that some data-sharing regulation is needed to advance research. For instance, "an agreement [through] KPI how much people share the data … how many datasets we have in the public domain…" (IP19). IP9 added that data disseminators "encourage users to report what they found and used, bring people to do little jobs, and they compete", and IP13 commented that data disseminators "should [make the data sharing process] more attractive for people looking at datasets than Kaggle". However, it does not exclude commercialization, as there is in-direct funding support via grants, sponsorship payments, and barter-style services. IP12 agreed that data disseminators enable "easier access to funding, but for the short term", with IP4 adding that universities play an important role here, and, therefore, "university researchers use the tool without paying, commercial users pay subscription ". IP5 added that university research "could be aimed at data harmonization [to data model]", while IP17 emphasized that data disseminators are good from a scientific point of view as they are much connected with sustainability. Finally, IP13 warned that for a commercial company, "the value is unclear … it has a very academic focus".

Data curator for inbound open innovation

Data curation considers the business perspective of data-driven value realization using a closed data platform, which follows commercial goals by enhancing private datasets and developing new services using complementary data. IP8 visualized the data-driven value chain from data search to cleaning, augmenting with open data, sorting, analyzing, and reporting. IP7 conceptualized the outcome as a "preventive analytic lab for delivering digital reports for industries." For this purpose, the hub aggregates data by signing "explicit data partnerships with institutions to capitalize on data", where profit, not the data per se, is a driver, said IP5:

"Social good is not enough, profit must be a driver". IP1 added that data curation could enable data scientists and app developers to develop "a high-value service". As digital analytics departments of large organizations are still a bit internally focused, they are missing the infrastructure required to attract data and try to look externally (IP13). The resulting data-driven products (e.g., based on diagnostics) could improve the brand and enable more transactional data to the ecosystem, thus opening a reinforcing loop of further ecosystem growth; IP21 commented about data curators:

> The data owner benefits from the analysis […] once the issues with the legal framework [are resolved]. (IP9)

Partnerships with close companies could give [the data owner] the complementary data (IP21). The expert added that, in contrast to the data disseminator, the data curator would "not work with competitors, [as] much legal tidying is needed to make it work", and an important concern is to get enough projects so that data curation can be economically viable. To do so, curators can "sell anonymized data products, requesting further transactional data [from customers] using the developed data model" or "take responsibility for data preparation" (IP5).

Data marketplace for outbound open innovation

Data selling considers the business perspective of data value realization on an open digital platform, which follows commercial goals by interconnecting data contributors, data scientists, and data collaborators for selling/purchasing data and data-driven insights. The main goal of a data marketplace is to lower barriers to data access by providing market information and enabling standard procedures for contracting. One expert stated that "standard procedures for contracts attract buyers and sellers [enable participation of] freelancers and crowdsourcing for ad-hoc projects up to 3 months" (IP2). IP8 added:

> Too much data – we could get lost in the project. Layering different data helps to build more success; therefore, there is a need for on-demand data-driven services, like a report

For example, a large organization could be "a private partner, sponsor or customer" for a digital platform as "a secondary marketplace [where users can do] own data upload" (IP12). IP19 introduced the problem of the perceived data value estimate:

> Data is not valued, but when a nominal fee is assigned, people want to sell it […] it is better to create a market for it and see if people are ready to pay [by using] concrete examples, like mosquito nets[3]

Finally, once the marketplace is created, the data marketplace is expected to "*lower barrier access to data"* (IP5). Table 2 describes these three types.

Insert Table 2 about here

DISCUSSION

Contribution to theory

Our study provides two contributions to the theory of open innovation: (1) First, we provide evidence of the role of data in open innovation for extracting value; (2) Second, we theorize how data in inbound and outbound open innovation can be shared, curated, or sold.

Three pathways for the longer sustainable growth of large organizations by using data enable different kinds of substantial returns (Euchner, 2016). For example, in sustainability research and in-direct value capture, such as branding (*Data disseminator*), high-value services proposition and partnership gains (*Data curator*) and financial performance (*Data*

[3] Mosquito nets suddenly became popular in Africa when the price to the product was assigned

marketplace, (see also Pohle and Chapman, 2006). By doing so, we contribute to the current literature on business models by finding a hidden "locus of innovation" (Zott and Amit, 2007). Specifically, using these models for open innovation can drive data-driven service development beyond firm boundaries and realize additional value. By so doing, this paper highlights that companies that generate and own substantial amounts of data as part of their operations (Barczack, 2022; Smith et al., 2016) can leverage the underutilized data they own or have access to for open innovation. Furthermore, by doing so, we challenge the existing requirements of open innovation organizational systems (Bogers et al., 2018; Naqshbandi, 2018; Fritsch et al., 2020; Smart et al., 2019).

The question of which model is most appropriate depends on the industry. Firstly, the volume of data and its standardization indicates if a data marketplace model can be used; unstructured data that is not cleaned or not applicable for automated analysis implies the need for a crowdsourcing process, such as the one described in data disseminator. Once firms employ data modellers and analysts to clean/standardize datasets, it enables companies to follow the data curator path, which implies open innovation and commercialization of datasets. Potential spin-offs from this activity can further improve the general business model of large established organizations. Further, once a large organization develops automated data preparation tools, such as machine learning for data cleaning, the range of appropriate business models increases so that large organizations can as data intermediaries in their industry and leverage selling transactions.

The paper provides evidence of how organizations can foster a culture of innovation by adopting one of the data pathways[4] (e.g., Chesbrough and Rosenbloom, 2002). It was a long-term concern for many companies who found it hard to identify appropriate business models

[4] Data availability, preparedness, and ability to derive value following business goals are important.

to leverage the opportunities of OI (Van der Meer, 2007). The proposed data-driven models should help large established organizations to avoid disruption from start-ups (Euchner, 2016), as many business models of organizations became obsolete in the past. Managers can choose from these pathways of using data in open innovation to extract more value from their datasets, add data-driven components to their business models, and contribute to sustainability research.

They provide evidence that any large organization can start with building a dedicated web portal for data analytics, attract earlier customers, and move to a digital platform strategy. For example, the business models could work within a business unit, which is mostly data-driven in large established organizations, attracting corporate partnerships for data sharing and earlier customers for commercialization.

CONCLUSION

Although the data volumes are increasing, the literature is still far from understanding the role of data in open innovation. This study asks: how can companies use data in open innovation? First, we provide evidence of the role of data in open innovation for extracting value. Second, we theorize how data influences open innovation by sharing, curating, or selling data. While the paper contributes to the open innovation literature, it is also important to acknowledge the limitations. First, more data must be collected from other industries to generalize and confirm the findings. Second, it is still uncertain whether the pathways for using data in open innovation derived from animal healthcare are representative for other industries. Third, the data-driven business model innovation process should be investigated in other industries to gain insights into how others are leveraging open data for innovation and whether and how industry dynamics may be shaping opportunities for data-driven business model innovation. Finally, future research should investigate how data-driven innovation addresses business performance and its impacts on society from a time perspective.

REFERENCES

Arend, R. J. 2013. The business model: Present and future—beyond a skeumorph. ***Strategic Organization***, 11(4): 390-402.

Bagherzadeh, M., Markovic, S., Cheng, J., & Vanhaverbeke, W. 2019. How does outside-in open innovation influence innovation performance? Analyzing the mediating roles of knowledge sharing and innovation strategy. ***IEEE Transactions on Engineering Management***, 67(3): 740-753.

Bagherzadeh, M., Gurca, A., & Brunswicker, S. 2019. Problem types and open innovation governance modes: A project-level empirical exploration. IEEE Transactions on Engineering Management, 69(2): 287-301.

Barczak, G., Hopp, C., Kaminski, J., Piller, F & Pruschak, G. 2022 How open is innovation research? – An empirical analysis of data sharing among innovation scholars, ***Industry and Innovation***, 29(2):186-218.

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T.M., Conradsen, M.L., Di Marco, D., Distel, A.P., Dobusch, L., Dörler, D., Effert, A. and Fecher, B., 2020. The Open Innovation in Science research field: a collaborative conceptualization approach. ***Industry and Innovation***, 29(2):1-50**.**

Berends, H., Smits, A., Reymen, I. and Podoynitsyna, K. 2016. Learning while (re)configuring: Business model innovation processes in established firms. ***Strategic Organization,*** 14(3): 181–219.

Bogers, M., Chesbrough, H. and Moedas, C. 2018. Open Innovation: Research, Practices, and Policies. ***California Management Review***, 60(2): 5–16.

Chesbrough, H. and Bogers, M. 2014. Explicating Open Innovation: Clarifying an Emerging Paradigm for Understanding Innovation, in ***New Frontiers in Open Innovation***, H.

Chesbrough, W. Vanhaverbeke, and J. West (eds). Oxford: Oxford University Press, 3–28.

Chesbrough, H. and Rosenbloom, R.S. 2002. The role of the business model in capturing value from innovation: Evidence from Xero Corporation's technology spin-off companies. ***Industrial and Corporate Change***, 11(3): 529–555.

Chesbrough, H.W. 2003. ***Open innovation: The new imperative for creating and profiting from technology***. Boston, MA: Harvard Business Press.

Chui, M., Manyika, J., & Kuiken, S. V. 2014. What executives should know about open data. ***McKinsey Quarterly***, January.

Davies, T. and Perini, F. 2016. Researching the emerging impacts of open data: revisiting the ODDC conceptual framework. ***The Journal of Community Informatics***, 12(2): 148–178.

Eisenhardt, K. M. 1989. Building theories from case study research. ***Academy of management review***, 14(4): 532-550.

Eckartz S., van den Broek T., Ooms M. 2016 Open Data Innovation Capabilities: Towards a Framework of How to Innovate with Open Data. In: Scholl H. et al. (eds) Electronic Government. ***EGOV 2016. Lecture Notes in Computer Science***, vol 9820. Springer, Cham.

Euchner, J. 2016. Business Model Innovation. ***Research-Technology Management***, 59(3): 10–11.

Fritsch, M., Titze, M., & Piontek, M. 2020 Identifying cooperation for innovation—a comparison of data sources, ***Industry and Innovation***, 27(6): 630-659

Fjeldstad, Ø.D. and Snow, C.C. 2018. Business models and organisation design. ***Long Range Planning***, 51(1): 32–39.

George, G., and Bock, A.J. 2010. The business model in practice and its implications for entrepreneurship research. ***Entrepreneurship Theory & Practice***, 35(1): 83–111.

Guba, E. G., & Lincoln, Y. S. 1989. ***Fourth generation evaluation***. Newbury Park: Sage Publications.

Günzel, F. and Holm, A.B. 2013. One size does not fit all — Understanding the front-end and back-end of business model innovation. ***International Journal of Innovation Management***, 17(1), 1340002.

Huizingh, EKRE. 2011. Open innovation: State of the art and future perspectives, ***Technovation***, 31 (1): 2-9

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. 2012. Benefits, adoption barriers and myths of open data and open government. ***Information Systems Management***, 29(4): 258–268.

Lauer, D.K.B., Smith, A., Blomberg, D.N. et al. 2021 Open data: A driving force for innovation in the life sciences. ***F1000Research***, 10(828).

Lichtenthaler, U. 2008. Open innovation in practice: an analysis of strategic approaches to technology transactions. ***IEEE transactions on engineering management***, *55*(1): 148-157.

Lustgarten, J. L., Zehnder, A., Shipman, W., Gancher, E., & Webb, T. L. 2020. Veterinary informatics: forging the future between veterinary medicine, human medicine, and One Health initiatives—a joint paper by the Association for Veterinary Informatics (AVI) and the CTSA One Health Alliance (COHA). ***JAMIA Open***, 3(2): 306-317.

Magretta, J. 2002. Why business models matter. ***Harvard Business Review***. Available at: https://hbr.org/2002/05/why-business-models-matter (accessed 1 August 2021).

Manyika, J, Chiu, M., Farell, D., Van Kuiken, S., Groves, P. and Doshi E.A. 2013. Open data: Unlocking innovation and performance with liquid information. ***McKinsey***.

Maretti, M., Russo, V. and del Gobbo, E. 2021. Open data governance: civic hacking movement, topics, and opinions in digital space. ***Quality and Quantity***, 55(3): 1133–1154.

Michelino, F., Lamberti, E., Cammarano, A., & Caputo, M. 2015. Open innovation in the pharmaceutical industry: An empirical analysis on context features, internal R&D, and financial performances. ***IEEE Transactions on Engineering Management***, 62(3): 421-435.

Monino, JL. 2021 Data Value, Big Data Analytics, and Decision-Making. ***J Knowl Econ*** 12(1): 256–267.

Moustakas, A. 2017. Spatio-temporal data mining in ecological and veterinary epidemiology. ***Stochastic Environmental Research and Risk Assessment***, 31(4): 829-834.

Naqshbandi, M. M. 2018. Organizational Characteristics and Engagement in Open Innovation: Is There a Link? ***Global Business Review***, 19(3): 1–20.

Nguyen, T. P. T., Huang, F., & Tian, X. 2021. When and how does Open Innovation enhance innovation output? A Meta-Analysis. ***IEEE Transactions on Engineering Management***.

Osorno, R., & Medrano, N. 2020. Open innovation platforms: A conceptual design framework. ***IEEE Transactions on Engineering Management***.

Osterwalder, A., Pigneur, Y. and Tucci, C. 2005. Clarifying business models: origins, present, and future of the concept. ***Communications of the Association for Information Systems***, 16: 1–25.

Pohle, G. and Chapman, M. 2006. IBM's global CEO report 2006: business model innovation matters. ***Strategy& Leadership***, 34(5): 34–40.

Sinkovics, R. R., Penz, E., & Ghauri, P. N. 2008. Enhancing the trustworthiness of qualitative research in international business. ***Management International Review***, 48(6): 689-714.

Smart, P., Holmes, S., Lettice, F., Pitts, F.H., Zwiegelaar, J.B., Schwartz, G. and Evans, S., 2019. Open Science and Open Innovation in a socio-political context: knowledge production for societal impact in an age of post-truth populism. ***R&D Management***, 49(3): 279-297.

Täuscher, K. 2018. Using qualitative comparative analysis and system dynamics for theory-driven business model research. ***Strategic Organization***, 16(4): 470-481.

Van der Meer, H. 2007. Open Innovation – The Dutch Treat: Challenges in Thinking in Business Models. ***Creativity and Innovation Management***, 16(2): 192–202.

van Veenstra, A.F., van den Broek, T.A. 2013. Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organization. In: Wimmer, M.A., Janssen, M., Scholl, H.J. (eds) ***Electronic Government. EGOV 2013. Lecture Notes in Computer Science***, vol 8074. Springer, Berlin, Heidelberg.

Voss, C., Tsikriktsis, N., & Frohlich, M. 2002. Case research in operations management. ***International journal of Operations & Production Management***, 22(2): 195-219.

Weerakkody, V., Irani, Z., Kapoor, K. et al. 2017 Open data and its usability: an empirical view from the Citizen's perspective. ***Inf Syst Front*** 19: 285–300.

Yin, R. K. (2009). ***Case study research: Design and methods*** (4th Ed.), Thousand Oaks, CA: Sage.

Zuiderwijk, A., Janssen, M., & Dwivedi, Y. K. 2015. Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. ***Government Information Quarterly***, 32(4): 429-440.

FIGURES AND TABLES

Figure 1: Representation of the coding scheme for business model dimensions, after King and Horrocks (2010)

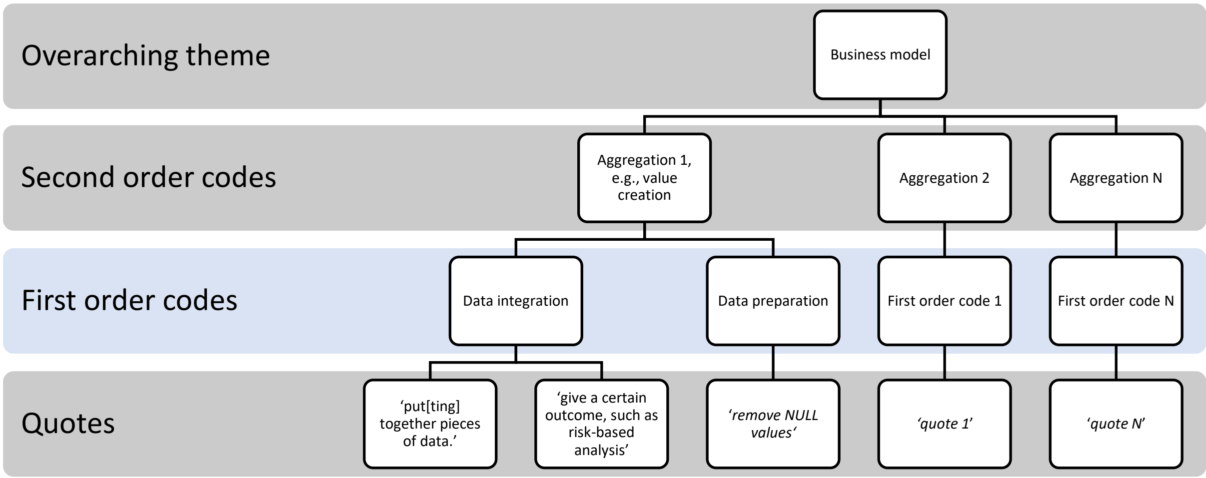
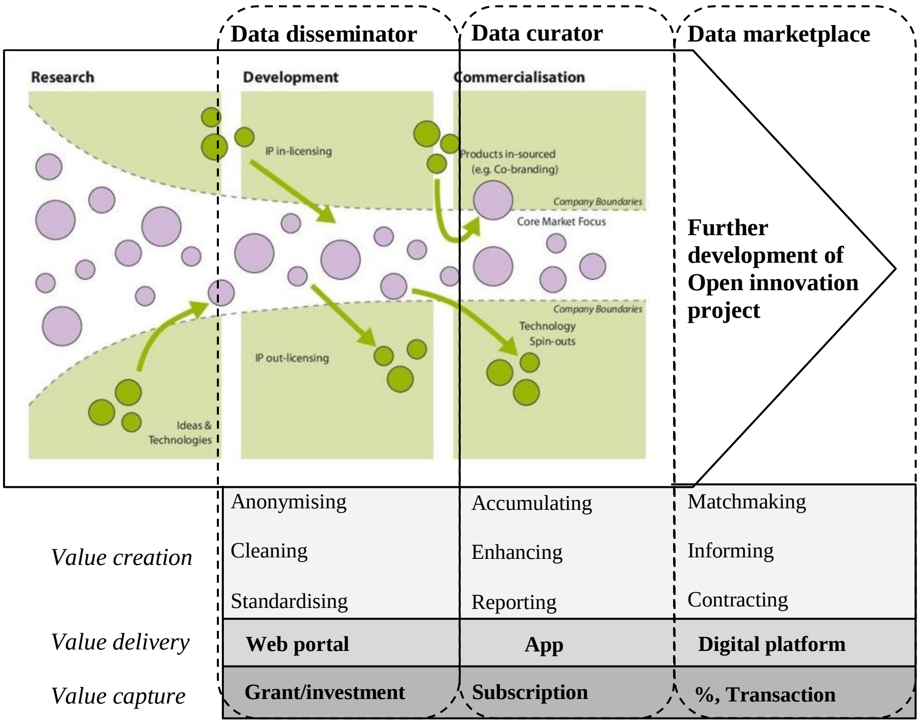


Figure 2: Three pathways for using data in open innovation, a modified picture of open innovation funnel[5]



[5] https://www.rndtoday.co.uk/open-innovation/open-innovation/ accessed 9.01.2023

Table 1: Overview of respondents

IP	Company	Interviewee Position	Interviewee Role	Experience, years
IP1	Large pet-care company	Director of digital health innovation	Data collaborator	30+
IP2	The medium-sized software development company	Innovation & business development director	Domain expert	30+
IP3	UK-based University	University senior relationship manager	Domain expert	15+
IP4	Global animal health company	Strategy lead in companion animal health	Domain expert	15+
IP5	Global animal health company	Head of digital, data and analytics	Domain expert	15+
IP6	Data-driven start-up	Tech entrepreneur	Data collaborator	15+
IP7	Global animal health company	Director strategic development	Data collaborator	15+
IP8	Strategic consulting firm	Managing partner in outcomes research[6]	Domain expert	20+
IP9	UK-based data innovation unit	Data scientist	Data scientist	15+
IP10	Global animal health company	Partner in outcomes research	Domain expert	15+
IP11	UK-based data innovation unit	Data scientist/ App developer	Data scientist	10+
IP12	UK-based data innovation unit	Strategy lead	Domain expert	15+
IP13	Global animal health company	Business transformation partner	Domain expert	20+
IP14	Global animal health company	Director of outcomes research	Domain expert	30+

[6] a data-driven method of determining the result of healthcare interventions for a patient or a system.

IP15	Global animal health company	Global medical director in livestock diagnostics	Domain expert	15+
IP16	UK-based University	Research assistant in data-driven virology	Data scientist	3+
IP17	Global animal health company	Director of Business Operations and Strategy	Domain expert	20+
IP18	UK-based data innovation unit	Expert in tropical animal health and data-driven innovation	Domain expert	15+
IP19	UK-based University	Associate dean of research and innovation	Domain expert	20+
IP20	Global animal health company	Pig and poultry business unit director	Domain expert	15+
IP21	Global animal health company	Companion animal veterinary lead	Domain expert	20+
IP22	Strategic consulting firm	Consultant in data outcomes research	Domain expert	15+
IP23	Strategic consulting firm	Managing partner in data outcomes research	Domain expert	20+
IP24	Global animal health company	Senior director of a large animal health company	Data collaborator	20+
IP25	Global animal health company	Customer relations director of a large animal health company	Data collaborator	15+

Table 2: Three pathways of using data in open innovation

	Data disseminator	Data curator	Data marketplace
Focus	data sharing	data curating	data selling
Data collection	open data	private and open data	private data
Data preparation	n/a	data anonymization, cleaning, standardization, open data augmentation	legal framework
Data governance	general terms and conditions	individual contract arrangements	standard contract arrangement
Value creation	analytics via crowdsourcing	data set integration & analysis	data seller & data buyer matchmaking
Value delivery	web portal, research reports	mobile app	digital platform
Value capture	research grant, ads	% from each project	% from each transaction
Impact	public good	private and public good	private good
Open innovation	Outbound	Inbound	Outbound

APPENDIX A: The list of open-ended questions

1. What is the value of data in your industry?
2. What are the main barriers to data sharing?
3. Why is the data not accessible?
4. What incentives can motivate data owners to share data?
5. How to enable data preparation?
6. What is the role of universities in the data-driven economy?
7. How could this value be delivered?
8. How to commercialize data?
9. What is the appropriate contracting for data-driven services?