Technovation xxx (xxxx) xxx

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Open data: Lost opportunity or unrealized potential?

Serdar Temiz [a], Marcus Holgersson [b,*], Joakim Björkdahl [b], Martin W. Wallin [b,c]

[a] Department of Civil and Industrial Engineering, Uppsala University, 751 04, Uppsala, Sweden
[b] Department of Technology Management and Economics, Chalmers University of Technology, 412 96, Gothenburg, Sweden
[c] Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Big data
Business model
Digitalization
Open data
Open innovation
Value capture
Value creation

ABSTRACT

The promise of open data is grand, but the results are often meager. To resolve this conundrum and make headway in the adoption of effective open data practices, we take a step back and investigate the underlying reasons for investing in open data. Based on survey results, interviews, and complementary evidence from secondary sources, we explore the motives and beliefs about open data investment expressed by open data experts in both public and private organizations. To our surprise, in both public and private organizations we find that open data investments are driven more by legitimacy-seeking than a quest to realize the value creation potential of open data. The results are worrisome, as such motives and beliefs do not necessarily lead to investment in the complementary assets needed to realize the potential associated with open data—instead, open data risks becoming a lost opportunity. Clearly, it's time to move beyond the open data hype and get down to business. Our paper provides insights for practice and calls on future research to unpack antecedents and mechanisms for value creation, and to identify appropriate complementary investments in open data, for example in terms of technologies, tools, and systems.

1. The rise of open data and challenges ahead

Open data is currently gaining momentum in both policy and business communities (Gurstein, 2011; Kiron, 2017; Mention, 2019). Inspired by the practices of open innovation (Chesbrough, 2003; Chesbrough et al., 2006) and open source (Stallman, 1999; von Krogh et al., 2012), open data rests on the idea that data should be generated and shared freely (Janssen et al., 2012; Murray-Rust, 2008; Perkmann and Schildt, 2015). But just as open source software initially struggled to identify ways for all stakeholders to benefit, so are open data proponents struggling today (cf., Dahlander and Magnusson, 2005, 2008; Janssen et al., 2012). In this paper, we put forward a simple yet powerful argument: There is great potential in open data, but current motives and beliefs about open data do not necessarily lead to the investments needed to realize this potential. Instead, open data risks becoming a lost opportunity.

Proponents of open data have often argued from two different positions: either that open data has a societal value in terms of improved transparency, accountability, and democracy (Gurin et al., 2019), or that open data has an economic value, waiting for someone to utilize it (Gurin et al., 2019; Manyika et al., 2013; Huyer and van Knippenberg, 2020). Much research on open data has focused on governments and the use of publicly owned and generated data, and was mainly concerned with societal gains (e.g., Bonina and Eaton, 2020; Janssen et al., 2012; Kassen, 2013; Zuiderwijk and Janssen, 2014). However, literature on the economic potential of open data is gaining momentum (Bonina, 2013; Gurin et al., 2019; Manyika et al., 2013; Huyer and van Knippenberg, 2020). For example, NASA's Landsat mapping program has demonstrated that when satellite data were first released, it led to an increase in private gold discoveries (Nagaraj, forthcoming). The literature also includes analyses of different business models that are based on open data (e.g., Janssen and Zuiderwijk, 2014; Magalhaes and Roseira, 2020). However, a more granular analysis of why organizations (private and public) invest in open data is still lacking.

The aim of this paper is twofold. First, to explore the motives and beliefs underpinning the investment in open data. Second, to provide insights to narrow the gap between the potential for open data and what has actually been realized. We do so through a combination of a novel survey of open data initiatives, in-depth interviews with industry experts, and complementary evidence from secondary sources. Some

* Corresponding author.
E-mail addresses: serdar.temiz@angstrom.uu.se (S. Temiz), marcus.holgersson@chalmers.se (M. Holgersson), joakim.bjorkdahl@chalmers.se (J. Björkdahl), martin.wallin@chalmers.se (M.W. Wallin).

https://doi.org/10.1016/j.technovation.2022.102535
Received 30 November 2020; Received in revised form 29 March 2022; Accepted 12 April 2022

results surprised us, as they contradicted commonly held beliefs about open data. For instance, many organizations invest in open data for reasons unrelated to business needs and, as a result, miss out on experiencing the true benefits of open data.

2. The value of open data

The concept of data has shifted from being something incomprehensible, best handled by the IT department, to the pinnacle of boardroom strategizing and government policy-making. But the use of data to improve business is not a new phenomenon. For example, loyalty cards have long provided retailers and airlines with data that allowed for targeted marketing (Berman, 2006), and with the advent of social media and artificial intelligence, both the need for and potential of data have accelerated (Davenport et al., 2020). Data points are rarely of immense value on their own, but as soon as they are combined—for example, by linking customer characteristics to buying behavior, or process data to performance data—the value can quickly multiply (Bauer and Kaltenböck, 2011; Trantopoulos et al., 2017; Cappa et al., 2021; Dahlander et al., 2021). For example, Facebook and Google are able to target campaigns with unprecedented precision, motivating customers to pay good money for their services. Indeed, in the past few years, the business community has gradually realized that data is a truly valuable resource (Alharthi et al., 2017). In 2017, *The Economist* published an influential issue on "The world's most valuable resource: Data and the new rules of competition." No longer is data only a question for hyped internet companies and Silicon Valley startups, but an accepted source of competitive advantage for any type of business. Fast forward to the issue of open data in the present day: when viewed through the lens of traditional strategic analysis, data is a valuable asset that should be heavily guarded and protected, not freely shared, and thus in complete opposition to the ethos of open data. Nevertheless, there has been an increasing interest in open data and the potential of opening up certain data to outsiders, as well as utilizing available open data provided by others (Janssen et al., 2012; Huyer and van Knippenberg, 2020).

The potential of open data relies on the core characteristics of data: data is cumulative and combinatorial, and the value derived from it is often exponentially related to the size of the dataset. By pooling open data, organizations can create value at a level no single organization could accomplish alone. Consider the words of Håkan Samuelsson, the CEO of Volvo Cars, discussing automotive safety data at a European Commission conference in 2017[1]:

> "*We think this type of anonymized data sharing should be done for free, for the greater good, and to the wider benefit of society. It saves lives, time and taxpayer money […] I call on other car makers and governments to work with us on realizing this type of data sharing as widely as possible.*"

The philosophy behind open data takes many of its cues from the open source software movement that has long pushed for computer code to be *libre* or free. The key similarity between open source software code and open data is that code and data can be used, reused, studied, and modified without permission (Murray-Rust, 2008; Feller and Fitzgerald, 2002).

The concept of open data was first applied in scientific communities to ensure free access to academic data published in digital repositories (Murray-Rust, 2008). It was later adopted by proponents who wanted government data to be freely available (Janssen et al., 2012). Consider, for example, the Structural Genomics Consortium that brought together leading universities such as the University of Oxford, Karolinska Institutet, and the University of Toronto, and major pharmaceutical companies such as GlaxoSmithKline, Novartis, and Merck. The consortium pooled resources to determine the three-dimensional shape of proteins and then shared their results as open data, freely available without restriction (Perkmann and Schildt, 2015). More recently, universities, research institutes, research journals, and companies pledged to share data that was useful for finding a vaccine for COVID-19.[2] The Organization for Economic Co-operation and Development (OECD) sent a powerful message about the need for sharing open data across public and private sectors and throughout civil society[3]:

> "*To strengthen the contribution of open science to the COVID-19 response, policymakers need to ensure adequate data governance models, interoperable standards, sustainable data sharing agreements involving public sector, private sector and civil society, incentives for researchers, sustainable infrastructures, human and institutional capabilities and mechanisms for access to data across borders.*"

Indeed, grand challenges often require organizations, public as well as private, to open up for collaboration across actors (McGahan et al., 2021). Sharing data openly can spark solutions to societal challenges that, at first glance, are unrelated to the data. Consider for example the large negative environmental impact of online shopping of clothes and fashion. For some types of goods as much as 80% of orders are returned, often because consumers struggle to select the right size when shopping online. While many online retailers deal with the problem by pruning supply chain inefficiencies, Norwegian startup Zizr attacked the problem's root causes by leveraging open data and artificial intelligence. The company used purchase history across individuals, resellers, and brands to create digital twins for consumers' body parts. The technology significantly improves the likelihood of ordering the right size, and only the right size, when customers shop for well-fitting clothes. Despite the simplicity and convenience of the technology, its success is dependent on several actors along the value chain opening up and sharing some of their data, sometimes even to competitors.

These examples paint a bright future for open data—clearly it has great potential. But to make headway in terms of adoption of open data practices, we need to take a step back and explore the underlying reasons for investing in open data. While a large amount of research is beginning to materialize around what organizations can do *when they have open data*, very little is known about *why companies invest in open data* in the first place. This is an important distinction, because realizing value from open data requires alignment within organizational functions and between several actors. Therefore, the first step is to explore the motives and beliefs for investing in open data.

3. About the research

To empirically investigate why organizations invest in open data, we developed a questionnaire that explores motives and beliefs about open data. By reviewing extant literature on open data—focusing on operational performance and productivity (Hung et al., 2006; Yang and Wu, 2016), effort (Venkatesh et al., 2003), reputation and legitimacy (Chwelos et al., 2001; Yang and Wu, 2016), and risk (Yang and Wu, 2016)—we identified 21 such motives and beliefs that could potentially impact the decision to invest in open data. A draft of the questionnaire was discussed with five experts (both researchers and open data practitioners) to review the wording and length, and to ensure consistency and correct interpretation. Open data initiatives can be both inbound

[1] "Volvo Cars joins groundbreaking, pan-European safety data sharing pilot project", https://www.media.volvocars.com/global/en-gb/media/pressreleases/253968/volvo-cars-joins-groundbreaking-pan-european-safety-data-sharing-pilot-project [accessed on September 5, 2020].

[2] The COVID-19 pandemic led to several initiatives to open up both data and proprietary technologies for the use of others in order to find a cure for the disease quickly and to mitigate its negative impact on society (e.g., Contreras et al., 2020).

[3] "Why open science is critical to combatting COVID-19", http://www.oecd.org/coronavirus/policy-responses/why-open-science-is-critical-to-combatting-covid-19-cd6ab2f9/[accessed on October 19, 2020].

(accessing external sources of open data) and outbound (making one's own data accessible to outsiders), but they often include elements of both in a coupled mode, which is why respondents were instructed to view open data initiatives broadly and to include both modes of open data. Respondents provided input regarding the degree to which they agreed (from 1, strongly disagree, to 5, strongly agree) with statements linked to whether their organization invests resources in open data projects (see Table 1).

The questionnaire was distributed in Sweden—the Swedish open data ecosystem is vibrant and developing fast, with government, public, and private actors actively working with open data projects. Respondents were identified among Swedish data professionals who had previously participated in open data events and/or whose organizations had received government funding for open data projects. We also relied on snowballing to identify additional respondents having prior experience of open data. From an initial sample of 304, useful responses were collected from 96 respondents from private IT and energy firms (34), non-governmental organizations (NGOs) (8), universities and research institutes (17), and other public organizations (37). For descriptive statistics of the sample, see Table 2. The data were collected from mid-November 2017 to early January 2018.

Due to the richness of the data and the range of the variables, and due to a large number of significant correlations between the variables, we conducted a principal component analysis to explore the underlying factors in the dataset. The properties of our data allowed us to use this analytical method. For example, the Kaiser-Meyer-Olkin measure was 0.779 and Bartlett's test of sphericity was significant at the 0.1% level (cf., Cerny and Kaiser, 1977; Dziuban and Shirkey, 1974). Five factors were extracted using varimax with Kaiser normalization (see next section). In the final part of our statistical analysis, we used linear regression analysis to explore how these five factors and control variables that included size, age, and sector were related to open data investment.

To provide a better understanding of the open data phenomenon, we complemented the survey data and statistical analysis with insights from 16 in-depth interviews with open data project managers representing small (4) and large (3) private firms, NGOs (1), public universities and research institutes (4), and other public organizations (4). These were identified through their involvement in a national program for open data support and knowledge exchange. In the next section, we will present the results of the statistical analysis, supported by data from the interviews.

4. Open data—are organizations investing for the right reasons?

From the principal component analysis of the survey data, we identified five factors that explain most of the variation in the dataset (see Table 1). These factors represent motives and beliefs about open data: 1) *ease and need of use*, 2) *business potential*, 3) *capabilities and openness for open data,* 4) *legitimacy*, and 5) *business and legal risks*.

In the second stage of our analysis, we conducted a linear regression analysis to assess the extent to which these factors are linked to the degree organizations invest resources in open data projects (Wang and Lo, 2016), see Tables 2 and 3. In Table 3, Model 1 is the baseline model, with controls for organizational attributes. The main independent variables were added separately in Models 2–6. We then combined all independent variables and controls in Model 7. After controlling for the organization's size, age, and sector, we found that only *ease and need of use*, *capabilities and openness for open data*, and *legitimacy* could positively and significantly predict the degree to which an organization invests resources in open data projects. Surprisingly, *business potential* and *business and legal risks* have no significant effects.

The insignificant factors—relating on the one hand to the performance, productivity, effectiveness, and mission of the organization (Factor 2: business potential) and on the other hand to the concern of legal implications and loss of valuable assets (Factor 5: business and legal risks)—would typically be assumed to be relevant for different types of organizations, but possibly less so among public organizations where data could more likely be opened on purposes to improve transparency and legitimacy alone. Therefore, in Models 8–10 we continued with additional analyses aiming to test the robustness of our results within different types of organizations in our sample. In Model 8 we conducted the regression on only private/non-governmental organizations and in Model 9 we conducted it on only public/governmental organizations. Our results hold also in the specific subsets, with small differences in significance levels and effect sizes. In Model 10, we further investigated possible differences between private and public organizations by adding interaction effects between the type of organization and the different factors. [4] All in all, the small differences in effect sizes and significance levels between the main factors in Model 7 and Model 10, respectively, the lack of significance of interactions reported in Model 10, and the small and insignificant increase in the fit of Model 10 (R^2 = 0,548) as compared to Model 7 (R^2 = 0,531) again supports that our results hold for both private and public organizations.

The finding that business potential and risk are not of greater importance, stands in sharp contrast to common views expressed by business leaders as well as to prior research. Rather, it seems that organizations are investing in open data for legitimacy reasons—because similar peer organizations are investing in open data or because such investments are regarded as positive by the media and the public. This was echoed in our interviews. For example, one project manager reported that "the project of [open data] was more like a social responsibility project." Another project manager told us that when engaged in open data, they could "brag about it, that they are more open." Other than legitimacy, capabilities matter. Organizations that invest in open data do not start from scratch—they often possess some kind of relevant experience. This experience is reflected in the importance of higher-order capabilities to work with open data (captured by the factor *capabilities and openness for open data*) as well as in lower-order capabilities (captured by the factor *ease and need of use*) such as project-level tools and processes.

Our results demonstrate that organizations invest in open data for reasons other than realizing the business potential of open data. Yet, in our interviews we could observe some frustration. A case in point is the project manager who told us that "eventually we need to have returns […] we need to have some sort of leverage to justify that opening data is good." Another project manager said that "I think that there will come a time when this is not the new thing anymore and it will be questioned. Why should we do open data? It doesn't really pay off. People don't use it. We are publishing data that no one is using. Should we really do that?"

Overall, the survey results and the interview data suggest that organizations might not yet be taking full advantage of the potential of open data. In the next section we will delve deeper into these findings and discuss insights for research and theory.

5. Insights for research and theory

How can these puzzling results provide insights for research and theory on open data? While previous research has pointed at the potential for open data to create significant value (Gurin et al., 2019; Manyika et al., 2013; Huyer and van Knippenberg, 2020) we provide new insights for open data research and theory by connecting the dots from motives and beliefs about open data to the investments made in open data. A core insight is that motives and beliefs need to align with investments to create value. Specifically, to realize the immense potential of open data substantial complementary investments in technologies and capabilities may be needed, but such investments are at

[4] We also tested robustness by including interactions with the private firm dummy (instead of private/non-governmental organizations more broadly), with very similar results as Model 10. This analysis is not reported in Table 3.

Table 1
Component matrix from principal component analysis.

	Components					Rotated components				
	1	2	3	4	5	1	2	3	4	5
Implementing open data projects can increase the performance of our organization's operation.		0,776					0,862			
Implementing open data projects can raise the productivity of our organization's operation.		0,760					0,878			
Implementing open data projects can enhance the effectiveness of organization operations.		0,646					0,781			
It is easy for our organization to implement open data projects.	0,694					0,711				
The process of how to deliver open data project is easy for our organization.	0,712					0,665				
It is easy for our organization to use related tools and platforms to implement open data projects.	0,580					0,737				
Our organization has sufficient incentives to participate in open data projects	0,693					0,601				
Our organization can obtain rewards through open data project implementation.						0,528				
Our organization can receive positive reputation with open data projects.										
Our organization is concerned with data quality/data content								0,517		
Our organization is concerned with inability to follow the legal framework with open data projects			0,538	0,595						0,824
Our organization has potential concerns such as losing valuable assets with open data (projects).			0,526							0,763
Open data projects fit with our organization's mission.	0,509						0,592			
Our organization is open to new ways of delivering services, such as doing open data projects or letting third parties use our data	0,522							0,712		
Our organization encourages working with open data projects	0,766							0,722		
Our organization has knowledge on how to work with open data.	0,668							0,656		
Policymakers and high-level authorities expect our organization to work with open data (projects).	0,528							0,504		
Similar organizations have engaged in the development of open data projects.			-,700						0,850	
Media and the public encourage our organization to participate in open data projects.	0,505								0,607	
Work processes in our organization require the use of open data/open data tools/services	0,717					0,583				
Functional areas in our organization require the use of open data (tools, projects, services)	0,636					0,747				

Notes: Coefficients with less than a 0.5 factor loading are not shown. Extraction method: principal component analysis; rotation method: varimax with Kaiser normalization. Rotation converged in nine iterations.

Table 2
Descriptive statistics.

	N	Minimum	Maximum	Mean	Std. Deviation
Size (employees)	96	1,00	55 000	4 808	11 847
Size (log employees)	96	,00	4,74	2,32	1,51
Age	96	,00	600	74	107
Age (log)	96	,48	2,78	1,55	,53
Private IT/Energy firms (sector dummy)	96	,00	1,00	,35	,48
University and research institute (sector dummy)	96	,00	1,00	,18	,38
NGO (sector dummy)	96	,00	1,00	,08	,27
Factor 1: Ease and need of use	96	−2,27	2,22	,00	1,00
Factor 2: Business potential	96	−3,19	1,67	,00	1,00
Factor 3: Capabilities and openness for open data	96	−3,19	2,42	,00	1,00
Factor 4: Legitimacy	96	−3,32	1,98	,00	1,00
Factor 5: Business and legal risks	96	−2,91	2,53	,00	1,00

risk when the underlying reason is clouded or insufficient.

Many organizations—public organizations as well as private firms—engage in open data without a clear plan for how open data can contribute to some form of value creation. Both public organizations and private firms invest in open data for reasons not primarily associated with the "business potential" (i.e., improving productivity or operational performance), but because of more indirect needs, such as seeking legitimacy from the public and other relevant peers. Similarly, both public organizations and private firms pursue open data not only because they "should," but also because they "can," specifically in terms of having the resources and capabilities to engage in open data—regardless of the ultimate goal. Future research needs to be mindful of differences in motives and beliefs about open data, where some, for example, are clearly linked to achieving a specific financial goal (e.g., open data could help the organization to cut costs or to grow sales) whereas others are not (e.g., complying with expectations from relevant peer groups). This organizational level insight corroborate and extend previous research conducted on a national level that has shown that policymakers adopt open data policies more due to inspiration from other countries and owing to an ambition to maintain an image of being an "advanced information society" (Huijboom & Van den Broek, 2011). While it is not surprising that public organizations invest in open data projects because they have a legal mandate to do so, it is worrisome if they invest in open data without plans for how open data can be turned into productive use. Equally if not more worrisome is that many private firms are similarly confused about why they invest in open data.

At this point we might ask ourselves: does it really matter *why* organizations engage in open data, as long as they do it? Our answer is a sounding yes. It matters for public organizations and for private firms, regardless of their different mandates to engage in open data. Earlier research has shown that organizations tend to be overly naïve in believing that open data will automatically lead to value creation (Janssen et al., 2012). The problem is that for open data to create value, and for organizations and individuals to capture a share of that value, several complementary investments and capabilities are needed. For example, the data must not only be open, but also useful, useable, cleaned, and technically and legally accessible (Zuiderwijk et al., 2014; Gurin, 2019), and it must be matched by investments in information, metadata, software, quality management, and social tools that can cultivate the ecosystem around the open data (Zuiderwijk et al., 2014; Bonina and Eaton, 2020), in addition to data analytics capabilities (El-Darwiche et al., 2014). Down the road, for open data to create value someone will need to be able to use it—open data is not productive in itself—and for this to happen complementary investments and

Table 3
Regression results.

	Dependent variable: Our organization invests resources in open data projects									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	4,754***	4,624***	4,758***	4,632***	4,339***	4,753***	4,117***	2,586***	3,592***	4,004***
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Size (log employees)	−0,190	−0,185	−0,194	−0,099	−0,148	−0,189	−0,054	−0,347*	0,234	−0,022
	(0,150)	(0,134)	(0,147)	(0,404)	(0,230)	(0,153)	(0,585)	(0,014)	(0,212)	(0,838)
Age (log)	−0,508+	−0,390	−0,502	−0,572*	−0,429	−0,508+	−0,380+	0,913*	−0,623*	−0,382
	(0,095)	(0,173)	(0,103)	(0,036)	(0,131)	(0,098)	(0,096)		(0,025)	(0,102)
Private IT/energy firms (sector dummy)	−0,433	−0,677	−0,438	−0,391	−0,080	−0,432	−0,313			−0,222
	(0,327)	(0,107)	(0,325)	(0,322)	(0,849)	(0,334)	(0,360)			(0,537)
University and research institutes (sector dummy)	−0,084	−0,001	−0,086	−0,121	0,104	−0,084	0,139		0,194	0,198
	(0,796)	(0,998)	(0,791)	(0,675)	(0,732)	−0,797	(0,568)		(0,425)	(0,432)
NGO (sector dummy)	−1,001	−0,928	−1,017+	−0,979+	−0,573	−1,000	−0,511	−0,173		−0,439
	(0,101)	(0,104)	(0,099)	(0,072)	(0,320)	(0,103)	(0,268)	(0,586)		(0,368)
Factor 1: Ease and need of use		0,432***					0,437***	0,436***	0,453**	0,482**
		(0,000)					(0,000)	(0,000)	(0,002)	(0,001)
Factor 2: Business potential			0,027				0,004	0,024	−0,116	−0,070
			(0,819)				(0,964)	(0,848)	(0,333)	(0,568)
Factor 3: Capabilities and openness for open data				0,505***			0,514***	0,589***	0,475***	0,436***
				(0,000)			(0,000)	(0,000)	(0,000)	(0,000)
Factor 4: Legitimacy					0,418***		0,391***	0,365**	0,436***	0,452***
					(0,000)		(0,000)	(0,006)	(0,001)	(0,000)
Factor 5: Business and legal risks						0,002	−0,013	0,042	−0,237	−0,167
						(0,986)	(0,881)	(0,699)	(0,101)	(0,244)
Interaction: Factor 1 x Priv. and NGOs										−0,059
										(0,756)
Interaction: Factor 2 x Priv. and NGOs										0,056
										(0,761)
Interaction: Factor 3 x Priv. and NGOs										0,195
										(0,299)
Interaction: Factor 4 x Priv. and NGOs										−0,052
										(0,777)
Interaction: Factor 5 x Priv. and NGOs										0,243
										(0,200)
Observations	96	96	96	96	96	96	96	42	54	96
R^2	0,087	0,210	0,087	0,281	0,216	0,087	0,531	0,656	0,563	0,548
Adjusted R^2	0,036	0,156	0,026	0,233	0,163	0,025	0,476	0,572	0,485	0,463

Notes: Regression coefficients (unstandardized β), with significance in parentheses. +,*,**,*** significant at the 10, 5, 1, and 0.1 percent levels, respectively. The base category for sector dummy variables is Public administration in models (1)–(7) and (9)–(10), and Private IT/energy firms in model (8). Models (1)–(7) and (10) include all organizations, while model (8) focuses on private/non-governmental organizations and model (9) focuses on public/governmental organizations.

capabilities are often needed.

These are important issues, but a more fundamental question follows: if data is opened without a clear intention to create value, will these necessary investments in complementary resources and capabilities really be made? We argue that when there is a significant disconnect between the true value of open data and the reasons for investing in open data, organizations will most likely not make the necessary investments in complementary resources and capabilities that would allow for value creation and capture on a large scale. An important insight for future research is to explore how organizations can be prompted to make such necessary complementary investments.

If open data initiatives are primarily related to factors such as legitimacy-seeking, in part analogous to “green washing” (Delmas and Burbano, 2011), there is limited incentive to really work toward a productive use of the data. Indeed, while gaining legitimacy may be a necessary condition, it is far from sufficient. And if data is primarily opened for the sake of being open, and not for any productive use, we are unlikely to see the expected potential come to life. Going forward, research on open data needs to put more emphasis on ‘outcomes’, specifically in terms of how open data initiatives can deliver value to various constituents. Next steps could include research to help decision-makers evaluate the decision to open up data, and how value creation and value capture potential connects to strategic decisions and investments. At the same time, research will need to identify and disentangle other motives and beliefs that can prompt initial as well as sustaining investments in open data. Here we have identified legitimacy-seeking as such a motive not directly linked to value creation but clearly affecting investments in open data. Future research could expand in this direction and explore other motives already observed in open source software development, e.g., enjoyment and obligation-based motives (Osterloh and Rota, 2007), and social and technological motivation (Bonaccorsi et al., 2006). Based on our insights here, we provide several managerial takeaways below.

6. Insights for practice

The concept of openness has inherently positive connotations that resonate well with many. This may explain our empirical results, signifying that investments in open data are driven more by availability, slack resources, and legitimacy, than by actual business potential. However, to succeed with open data, organizations need to consider how open data can empower their organizational goals and enhance their potential to create value for their stakeholders. This perspective connects to prior work (e.g., Magalhaes and Roseira, 2020), and our empirical investigation highlights several important issues that decision-makers need to consider.

First, decision-makers must *identify and decide what investments in complementary assets they need to make* in order to realize the potential from outbound open data (making their own data accessible and useable to others) as well as inbound open data (accessing and using externally shared data), while at the same time considering who currently controls such assets. On the one hand, the idea of open data builds on the logic of

distributed complementary assets that do not need to be concentrated in one organization. On the other hand, the distribution of complementary assets directly impacts where value will be created and captured (Teece, 1986, 2018). Sometimes these assets may be controlled by the "wrong" entity, and sometimes the complementary investments needed are too hefty for an open data strategy to make sense. But in any case, complementary investments are typically needed for organizations to be able to create and/or capture value from open data (e.g., Bonina and Eaton, 2020). For example, open data must be matched with technology. This was the case in the investigation of the Panama Papers scandal. The International Consortium of Investigative Journalists (ICIJ) organized hundreds of journalists to work on analyzing the leaked data. The big breakthrough did not come until the data were complemented with the new graph database technology developed by the Swedish startup firm (now with a unicorn valuation) Neo4j, as described by ICIJ's data editor, Mar Cabra: "My journalists were amazed. We felt like we had superpowers, because the reaction was, 'Oh my God, I did not see these connections before by looking through the documents, I'm finding more stories.' To them, this was magic."[5] This underscores an important insight for practice: making data openly available does not necessarily create a lot of value if not matched with complementary technologies, systems, and tools for users. Clearly, this holds for public and private organizations, and both types of organizations may need to broaden their scope of open data investments to reach their goals—from assembling and publishing data to providing tools and tutorials so users can make productive use of the data shared. For example, Microsoft made investments in tools, frameworks, and templates to enable data sharing in their open data initiative,[6] and the European Patent Office invested in expert support and online communities to enable better use of the data in their PATSTAT project.

Second, our results highlight the need for decision-makers to *align incentives and policies*—both within and between organizations—if open data is truly to create value beyond signaling and legitimacy. Within organizations, incentives across units or employees are often misaligned. As data is often spread across units within an organization, technically and contractually (Björkdahl, 2020), it is difficult to use or share the data effectively (DalleMule and Davenport, 2017; Gandhi et al., 2018). This difficulty is reinforced by data quality issues, as expressed by one project manager trying to convince other parts of his organization to release some of its data openly: "I have heard several times that, 'OK, we have this data but we are not sure that it is 100 percent correct, so we don't want to publish it'." Between organizations, incentives also need to be aligned. In some cases, a data provider may need to adapt its open dataset to build a prosperous ecosystem of external actors (Adner, 2012; Granstrand and Holgersson, 2020; Zuiderwijk et al., 2014). To give outsiders sufficient incentive to use a dataset, "you are forced to structure and clean up your data," as expressed by one of our interviewees. Another interviewee told us that "you need to do some refinements [to the open data] in order to satisfy the [external] developers." In other cases, organizations may need to manipulate part of its dataset to get internal units with sensitive data on board with an open data initiative. All in all, there need to be policies in place for open data governance (Bonina and Eaton, 2020; Zuiderwijk et al., 2014). This is sometimes more challenging than accessing the right technology, as one interviewee told us: "technology can be fixed. The real challenges are much more like culture and trust and legal issues and whatever else that might be." The alignment of incentives and policies is not only a question of data quality and organizational and contractual design, but it also relates directly to our previous point about the need for complementary investments. Recent research shows that when data analytics capabilities are asymmetrically distributed across the involved organizations, interaction risks becoming competitive rather than collaborative (Cepa, 2021).

Third, decision-makers must *consider what type of open data business model to use*. For public organizations the open data business model is mainly focused on value creation, often prompted by a legal mandate for open data. For private firms, the situation is often more complex as their open data business models need not only to create value but also to capture a portion of that value to establish competitive advantage. Indeed, open data is not a single strategy, but relates to a plethora of different business models (that may differ between organizations of similar type). A decision to engage in open data driven by legitimacy motives and slack resources might not consider the full complexity of open data business models, which is also why such an open data initiative is probably not going to reach its full potential. A decision driven by value creation potential, on the other hand, will force organizations towards a deeper and more nuanced analysis of *how* value can be created (and captured). Hence, the decision to engage in open data is a complex one, tightly integrated with the organization's business model. Simply releasing data into the open, without a plan for how to create (and capture) value is likely to produce meager results. Other authors have already made excellent progress in exploring different open data business models (e.g., Hartmann et al., 2016; Janssen and Zuiderwijk, 2014; Magalhaes and Roseira, 2020; Urbinati et al., 2019). A good starting point for decision-makers—when evaluating ways to create and capture value from open data—is to take cues from these extant accounts.

7. Conclusion

Policy reports and research articles have long argued for the large economic potential of open data (e.g., Huyer and van Knippenberg, 2020; Manyika et al., 2013) but open data initiatives have thus far failed to fully deliver on this promise (e.g., Gurin et al., 2019; Huyer and van Knippenberg, 2020). This unsatisfactory situation prompted us to empirically investigate how investments in open data initiatives are driven by underlying motives and beliefs about open data. We found that investment is largely driven by legitimacy-seeking and the availability of slack resources rather than by a quest to realize a business potential from open data. Further analysis revealed that complementary investments in technologies, tools, and systems are needed to realize the value from open data initiatives. However, when investments are mainly driven by legitimacy and slack, organizations often fail to make the complementary investments needed to create value for users of open data. This is not the first time a new managerial, organizational, or technological concept has been initially hyped and used without a clear connection to value creation (e.g., Ehigie and McAndrew, 2005), and it partly explains the slow realization of value from open data. Our message is, first and foremost, that we need to move beyond the hype and get down to business—literally—and secondly, only when matched with complementary investments and assets, aligned incentives, and the right business model will the true value of open data be realized. We call on future research to more granularly identify which complementary investments are needed to achieve specific open data goals and exactly how organizations can be induced to make the appropriate complementary investments. Overall, our research suggests a stronger focus on value creation, both in research and practice.

This study does not come without limitations. For example, it is based on results from a particular national context (Sweden) in a time when open data quickly gained popularity but was still a relatively new phenomenon. Legitimacy, a central concept in our study, is likely both context-specific and temporal, and given our results here it is of importance for future empirical studies to consider the contextual setting. Another limitation is that we did not discriminate for inbound

[5] "Neo4j Enables Pulitzer Prize-Winning Investigation into Global Tax Evasion", https://neo4j.com/case-studies/the-international-consortium-of-investigative-journalists-icij/[accessed September 8, 2020].

[6] "Closing the data divide: the need for open data", https://blogs.microsoft.com/on-the-issues/2020/04/21/open-data-campaign-divide/[accessed November 15, 2020].

and outbound activities in the survey as we were more interested in finding a general pattern behind the motives for investing in open data. More research is needed to unpack important mechanisms for how value can be created from open data both in relation to organizations that are opening up data and for organizations that use open data. In particular, we see a greater need for research to focus on in-depth case and process studies that help to identify inner and outer contextual factors that influence the creation of value from open data in public organizations and the establishment of competitive advantage based on open data in private firms.

References

Adner, R., 2012. The Wide Lens: A New Strategy for Innovation. Penguin, London.
Alharthi, A., Krotov, V., Bowman, M., 2017. Addressing barriers to big data. Bus. Horiz. 60 (3), 285–292.
Bauer, F., Kaltenböck, M., 2011. Linked Open Data: the Essentials. Edition mono/monochrom, Vienna.
Berman, B., 2006. Developing an effective customer loyalty program. Calif. Manag. Rev. 49 (1), 123–148.
Björkdahl, J., 2020. Strategies for digitalization in manufacturing firms. Calif. Manag. Rev. 62 (4), 17–36.
Bonaccorsi, A., Giannangeli, S., Rossi, C., 2006. Entry strategies under competing standards: hybrid business models in the open source software industry. Manag. Sci. 52 (7), 1085–1098.
Bonina, C., 2013. New Business Models and the Value of Open Data: Definitions, Challenges and Opportunities. Department of Management, London School of Economics and Political Science. Retrieved June 22, 2020. Retrieved from. http://www.nemode.ac.uk.
Bonina, C., Eaton, B., 2020. Cultivating open government data platform ecosystems through governance: lessons from buenos aires, Mexico city and montevideo. Govern. Inf. Q. 37 (3), 101479.
Cappa, F., Oriani, R., Peruffo, E., McCarthy, I., 2021. Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety and veracity on firm performance. J. Prod. Innovat. Manag. 38 (1), 49–67.
Cepa, K., 2021. Understanding interorganizational big data technologies: how technology adoption motivations and technology design shape collaborative dynamics. J. Manag. Stud. 58 (7), 1761–1799.
Cerny, B.A., Kaiser, H.F., 1977. A study of a measure of sampling adequacy for factor-analytic correlation matrices. Multivariate Behav. Res. 12 (1), 43–47.
Chesbrough, H.W., 2003. Open Innovation: the New Imperative for Creating and Profiting from Technology. Harvard Business School Press, Boston.
Chesbrough, H., Vanhaverbeke, W., West, J. (Eds.), 2006. Open Innovation: Researching a New Paradigm. Oxford University Press, Oxford.
Chwelos, P., Benbasat, I., Dexter, A.S., 2001. Research report: empirical test of an EDI adoption model. Inf. Syst. Res. 12 (3), 304–321.
Contreras, J.L., Eisen, M., Ganz, A., Lemley, M., Molloy, J., Peters, D.M., Tietze, F., 2020. Pledging intellectual property for COVID-19. Nat. Biotechnol. 38, 1146–1149.
Dahlander, L., Magnusson, M.G., 2005. Relationships between open source software companies and communities: observations from nordic firms. Res. Pol. 34 (4), 481–493.
Dahlander, L., Magnusson, M.G., 2008. How do firms make use of open source communities? Long. Range Plan. 41 (6), 629–649.
Dahlander, L., Gann, D.M., Wallin, M.W., 2021. How open is innovation? A retrospective and ideas forward. Res. Pol. 50 (4), 104218.
DalleMule, L., Davenport, T.H., 2017. What's your Data Strategy. Harv. Bus. Rev. 95 (3), 112–125.
Davenport, T., Guha, A., Grewal, D., Bressgott, T., 2020. How artificial intelligence will change the future of marketing. J. Acad. Market. Sci. 48 (1), 24–42.
Delmas, M.A., Burbano, V.C., 2011. The drivers of greenwashing. Calif. Manag. Rev. 54 (1), 64–87.
Dziuban, C.D., Shirkey, E.C., 1974. When is a correlation matrix appropriate for factor Analysis? Some decision rules. Psychol. Bull. 81 (6), 358.
Ehigie, B.O., McAndrew, E.B., 2005. Innovation, diffusion and adoption of total quality management (TQM). Manag. Decis. 43 (6), 925–940.
El-Darwiche, B., Koch, V., Meer, D., Schehadi, R., Tohme, W., 2014. Big data maturity: an action plan for policymakers and executives. In: Bilbao-Osorio, B., Dutta, S., Lanvin, B. (Eds.), The Global Information Technology Report. World Economic Forum, Geneva, pp. 43–53.
Feller, J., Fitzgerald, B., 2002. Understanding Open Source Software Development. Addison-Wesley, London.
Gandhi, S., Thota, B., Kuchembuck, R., Swartz, J., 2018. Demystifying Data Monetization. *MIT Sloan Management Review*. In: November 27. Retrieved June 15, 2020, Retrieved from. https://sloanreview.mit.edu.
Granstrand, O., Holgersson, M., 2020. Innovation ecosystems: a conceptual review and a new definition. Technovation 90–91, 1–12.
Gurin, J., Bonina, C., Verhulst, S., 2019. Open data stakeholders – private sector. In: Davies, T., Walker, S., Rubinstein, M., Perini, F. (Eds.), The State of Open Data: Histories and Horizons. African Minds and IDRC, Cape Town and Ottawa, pp. 418–429.
Gurstein, M.B., 2011. Open data: empowering the empowered or effective data use for everyone? Clin. Hemorheol. and Microcirc. 16 (2) https://doi.org/10.5210/fm.v16i2.3316.
Hartmann, P.M., Zaki, M., Feldmann, N., Neely, A., 2016. Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. Int. J. Oper. Prod. Manag. 36 (10), 1382–1406.
Huijboom, N., Van den Broek, T., 2011. Open data: an international comparison of strategies. European J. ePractice 12 (1), 4–16.
Hung, S.Y., Chang, C.M., Yu, T.J., 2006. Determinants of user acceptance of the E-government services: the case of online Tax filing and payment system. Govern. Inf. Q. 23 (1), 97–122.
Huyer, E., van Knippenberg, L., 2020. The Economic Impact of Open Data: Opportunities for Value Creation in Europe. European Commission.
Janssen, M., Zuiderwijk, A., 2014. Infomediary business models for connecting open data providers and users. Soc. Sci. Comput. Rev. 32 (5), 694–711.
Janssen, M., Charalabidis, Y., Zuiderwijk, A., 2012. Benefits, adoption barriers and myths of open data and open government. Inf. Syst. Manag. 29 (4), 258–268.
Kassen, M., 2013. A promising phenomenon of open data: a case study of the chicago open data project. Govern. Inf. Q. 30 (4), 508–513.
Kiron, D., 2017. Lessons from becoming a data-driven organization. MIT Sloan Manag. Rev. 58 (2).
Magalhaes, G., Roseira, C., 2020. Open government data and the private sector: an empirical view on business models and value creation. Govern. Inf. Q. 37 (3), 101248.
McGahan, A.M., Bogers, M.L.A.M., Chesbrough, H., Holgersson, M., 2021. Tackling societal challenges with open innovation. Calif. Manag. Rev. 63 (2), 49–61.
Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., Almasi Doshi, E., 2013. Open data: unlocking innovation and performance with liquid information. McKinsey & Company. Retrieved November 15, 2020. Retrieved from. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information#.
Mention, A.L., 2019. Digital Innovation: Harnessing the Value of Open Data. World Scientific Publishing Co. Pte. Ltd, London.
Murray-Rust, P., 2008. Open data in science. Ser. Rev. 34 (1), 52–64.
Nagaraj, A. (Forthcoming). The private impact of public data: Landsat satellite maps increased gold discoveries and encouraged entry. Manag. Sci..
Osterloh, M., Rota, S., 2007. Open source software development—just another case of collective invention? Res. Pol. 36 (2), 157–171.
Perkmann, M., Schildt, H., 2015. Open data partnerships between firms and universities: the role of boundary organizations. Res. Pol. 44 (5), 1133–1143.
Stallman, R., 1999. The gnu operating system and the free software movement. In: DiBona, C., Ockman, S., Stone, M. (Eds.), *Open Sources: Voices from the Open Source Revolution*, 53-70. O'Reilly Media, Sebastopol, CA.
Teece, D.J., 1986. Profiting from technological innovation: implications for integration, collaboration, licensing and public policy. Res. Pol. 15 (6), 285–305.
Teece, D.J., 2018. Profiting from innovation in the digital economy: standards, complementary assets, and business models in the wireless world. Res. Pol. 47 (8), 1367–1387.
Trantopoulos, K., von Krogh, G., Wallin, M.W., Woerter, M., 2017. External knowledge and information technology: implications for process innovation performance. MIS Q. 41 (1), 287–300.
Urbinati, A., Bogers, M., Chiesa, V., Frattini, F., 2019. Creating and capturing value from big data: a multiple-case study analysis of provider companies. Technovation 84–85, 21–36.
Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: toward a unified view. MIS Q. 27 (3), 425–478.
von Krogh, G., Haefliger, S., Spaeth, S., Wallin, M.W., 2012. Carrots and rainbows: motivation and social practice in open source software development. MIS Q. 36 (2), 649–676.
Wang, H.J., Lo, J., 2016. Adoption of open government data among government agencies. Govern. Inf. Q. 33 (1), 80–88.
Yang, T.M., Wu, Y.J., 2016. Examining the socio-technical determinants influencing government agencies' open data publication: a study in Taiwan. Govern. Inf. Q. 33 (3), 378–392.
Zuiderwijk, A., Janssen, M., 2014. Open data policies, their implementation and impact: a framework for comparison. Govern. Inf. Q. 31 (1), 17–29.
Zuiderwijk, A., Janssen, M., Davis, C., 2014. Innovation with open data: essential elements of open data ecosystems. Information Policy 19 (1–2), 17–33.