Similarities of Open Data and Open Source: Impacts on Business

Juho Lindman

Hanken School of Economics, Information Systems Science, Helsinki, Finland, juho.lindman@hanken.fi

Abstract

What are the similarities of open data and open source software when building a business? Despite their differences as phenomena (one is about applications and one is about data), the two also have many similarities. Both for example share the idea that the transparency of the artifact enables contribution. Many developers of open data have experience with open source development. But do the companies that build their offerings on open data and open source have similarities, and if so, what are the similarities? Drawing on fieldwork and interviews with software entrepreneurs and managers, this paper investigates these questions through an empirical focus on openness in business and clarifies the links between commercial organizations engaged with open source and open data. The article reports similarities on how the managers use the terms *open data* and *open source* to describe their business dynamic. These similarities are of importance to those who are interested in developing services that rely on open source or open data or who are interested in community management and legal and business issues or policy.

Keywords: Open data, Open source, Software business, Business model, Developer communities

1 Introduction

Sometimes open data and open source are seen as the opposite of commercial service provisions because they both include the word *open.* However, for open source, the openness of the software artifact mainly limits the traditional subscription sales revenue. Other revenue sources such as offering services, dual-licensing, and other well-documented strategies offer revenue potential for open source companies [8], [18].

For open data, there are two ways commercial companies can turn released data sets into (societal) public value: the release of datasets can 1) increase the transparency of governments and other institutions, fight corruption, and provide new, more participative, services and 2) create economic value, growth, jobs, and thus tax revenue by designing new services that governments do not offer and selling them [20], [29]. This paper focuses on the latter (economic value), and thus the main focus is not on individual developers, citizens, or developer communities, but rather on organizations driven by a profit motive. Measuring the impact of open policies in different levels of analyses [31] or that of open data in specific industries [4] is not always straightforward. Instead of the quantitative measurement of this impact, we focus on the micro-level and on the managers and entrepreneurs who build their offering on open source and open data.

We assume that different meanings exist for the terms *open source* and *open data* among the people concerned (for a more thorough discussion on the definitions of the terms *open source* and *open data*, see section 2). Open source diffusion in the past has held this kind of multifaceted appeal [26]. Open source was marketed as a solution that would solve many different problems inherent in organizational software production. When systems were then implemented, it was often the case that, despite their merits, open source had trouble meeting these expectations [26].

Any new invention goes through transformations as it is accepted into local organizational use [24]. Previous examples of such inventions include, for example, intranet and case tools [24]. First evangelists, academics, industry papers, tradeshows, and consultants push the term forward in society, then the popular press becomes interested in the term, and ultimately experts in organizations figure out what adopting the latest invention would mean to them and what the organizational impacts would be.

This article focuses on the final part of the process: when organizations start to use these inventions. The goal is to study the local meanings of open source and open data (as narrated by managers and entrepreneurs) in companies and compare them. A series of interviews were carried out with both open source and open data entrepreneurs to better understand the marketplace and the business models of these organizations. The research question addressed in this paper is: *How the managers and entrepreneurs perceive the transparency of open source and open data for their own businesses?* We build on this by comparing observed empirical similarities and differences in their business environment.

The results of this study show that open data proponents struggle with similar issues related to the terms and impacts of implementing open data as happened with open source in the past. For example, the business logics of both open source and open data require external and often voluntary contributions to the development process. Issues related to attracting these contributions and managing them are similar in the two business environments.

This information helps to show the similarities in the business logics and also has implications for innovation policy. For example open data can avoid some of the pitfalls open source proponents have been accused of, such as overselling the benefits, cost-savings and voluntary contributions attainable to projects.

In addition to practical relevance to managers and entrepreneurs, we also hope to build links between mature open source research and the emerging field of open data research. We believe that cross-fertilizing these two traditions might create interesting new avenues. Both, for example, share similar views on the transparency of the development process engaging new developers, contributions and thus result in higher quality services.

In what follows we will first provide definitions of open source and open data. Then we review earlier literature on open source and open data business to position our research in relation to earlier work. After that we move to the empirical part of the paper: we describe the paper's methodology and the case studies. In the findings section we compare open source and open data businesses. The discussion section is organized around issues related to community, licenses and links between the open data and open source. This section is followed by research implications to practitioners, companies and innovation policy and managerial implications. The final section contains conclusion as well as a discussion on the limitations of the paper.

2   Background

In this chapter we discuss definitions of open data and open source in detail. A short summary of earlier research follows on open source business and open source related issues in organizations. Then we proceed to discuss open data and review the characteristics of a business that sells services based on open data stacks.

2.1   The Definition of Open Source and Open Data

Computer science relies on the separation of data and applications [16]. Following this distinction, open source concerns the software license of applications, and open data refers to the access and reuse of data. Open source means a piece of software licensed under an OSI-approved license produced by virtually distributed communities that in most cases follow open source governance, practices, and tools and view themselves as open source communities.

Open data on the other hand means data that is *openly* available (on the Internet). It is based on *the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control (*Wikipedia*)*. According to another commonly used definition, *content or data is open if anyone is free to use, reuse, and redistribute it—subject only, at most, to the requirement to attribute and/or share alike* (Opendefinition).

Concepts related to open data are Public Service Information (PSI) and Open Government Data (OGD). Both of these terms concern only data that is released by governments. Open data can on the other hand be released by private actors as well. Furthermore, open data is released in a manner that not only aims to make reuse possible but welcomes the reuse of the data. This openness of data then enables developers to build services on the released datasets. Openness in this context can have three different elements: technical openness (interfaces and standards), legal openness (concerning copyright, open licenses, and other issues), and commercial openness (that commercial usage is possible) [17].

2.2   Open Source Business

Voluntary collective action systems often include semipublic goods [22]. These have the benefits of being 1) non-excludable to the network partners and 2) jointly supplied, assuming that partners' uses are non-competing [28]**.** These public goods can be for example open source or open data, which enables collective action and may create network effects.

Open source companies cannot sell software subscriptions because they are disseminating their software free of charge on the Internet. In a similar way, open data companies cannot sell data access if they are publishing the same data free on the Internet. From a strategic management perspective, mixing open and proprietary product strategies offers potential to many companies that sell software or consulting services [9]. Another way to enjoy the benefits of openness is to enhance internal processes by mimicking open source production [10], for example using Corporate Source [5] and Inner Source [27]. Sometimes open sourcing can also offer possibilities to externalize development tasks to an external workforce [1].

Open source in organizations is still an under-researched phenomenon when compared to, for example, developer motivation and community-driven projects [10]. Open source software literature has identified several different ways to engage in open source [11] in addition to its use [28]**.** These include, for example, [28] using open source CASE (Computer-Aided Software Engineering) tools within organizations, integrating open source into software systems, participating in open source development, or providing the company's products as open source.

Organizational open source implementation is often constrained by different internal and external organizational pressures. Open source is leveraged using either open source business models [12], [20] or management strategies [23]. From a buyer perspective, open source can limit a situation called "vendor lock-in" [22]. To simplify, this is a situation in which earlier procurement decisions make the costs of switching to another system so high that it is virtually impossible to change the vendor. The result of such a lock-in is a loss of control over the organization's own infrastructure to the IT-vendor. One of the key issues of progressive industrial policy from a citizen's oversight perspective would be to limit vendor locks to proprietary IT systems and vendors, because open source solutions would be available on the Internet.

2.3   Open Data Business

Open data provides economic growth through services based on open data sets [13]. The process of data transformation and business has been theorized in different ways [2]-[3], [32]. One way is to focus on the general models to create a tenable offering on open data [19]: 1) freemium, 2) cross subsidy, and 3) network effects. In the freemium model, users are given certain data for free but are then charged for value-added service. The cross-

subsidy model refers to price discrimination to certain groups to create services or gain a wider user base. Network effects means collaborating in a way that reduces costs or increases the service reach for some of the parties involved.

Another way to classify open data access is to divide services according to price mechanisms for open data [6], [7]: 1) premium, 2) freemium, 3) free. Premium access has subscription fee, freemium access is limited in features, time, or size and free access comes with either advertising or some method of cross-subsidizing.

Latif et al. [15] offered a model to describe the roles of entities in open data business: 1) raw data provider, 2) linked data developer, 3) data application provider, and 4) end user.

In more detail:

1. Raw Data Provider (or Data Provider) provides the data;

2. Linked Data Provider (or Data Service Provider) converts the raw data into linked data (in machine-readable format);

3. Data Application Provider (or Application Developer) has the expertise to develop applications, visualizations, and mash-ups on top of data and linked open data.

4. End Users are people who consume the data.

The business model concept can be used as a unit of analysis, and it uses a holistic approach to describe how companies carry out their business. Business models are centered around the activities of the companies, and they help explain how (economic) value is created and captured. However, business model research has not agreed on a single definition of the term "business model" [30].

We have elsewhere developed a conceptualization building on Latif [15] and Rajala's [21] business model classification, and this conceptualization focuses on the different business models of the actors [25]. To simplify, value capture (of the small open-data companies) may follow three different paths as summarized below (Table 1).

Table 1: Service offering of open data companies [25]

Revenue sources	Service offering
1) Consultancy	Pre-publication analysis, project management, legal and privacy issues, user training, etc.
2) Conversion	Scraping, cleaning, adding semantic richness, combining with other data-sets, etc.
3) Application development	Designing the service, maintenance, etc.

The different stages of this data chain offer a variety of activities that may generate and capture economic value. Organizations may opt to use one or more of them. In addition, third-party organizations can build their offerings by implementing these processes in the data chain.

Earlier research provides background and positions the paper. The main revenue sources and other key characteristics in building service business are described. Now we turn from theory to practice. Informed by earlier research we report our methodology and the findings of the interviews that were conducted on managers of open source and open data companies.

3 Methodology

We conducted a round of interpretative semi-structured interviews [14]. The data source was small open data companies. We focused on the meanings and benefits of open data. Thus, the perspective of the interview was on how to use and develop services based on open data in commercial organizations. There are several reasons why we opted to focus on small companies. First of all we wanted to get a holistic entrepreneurial view on building and selling services. Secondly we we have earlier worked with small open source companies. Thus we wanted to extend earlier work to open data companies. Third, we wanted organizations that build their main service offering on open data and there are not very many large companies that operate in this manner.

As the field of open data is strongly developing, we chose the respondents using our existing connections, and we also asked those interviewed to recommend other people we should interview. All the respondents are from the same country (Finland), and their profiles are listed below (Table 2). We chose one respondent per organization. The research took place in the context of a research project on open data in organizations and in service development. This interview data was then compared to data collected earlier on open source entrepreneurs. The organizations

were chosen in a way that limits the overlap of open source and open data – we chose organizations that had their main service offering either in open source or in open data.The focus of the analyses was on the similarities with open data and open source in a business setting (Table 2).

Table 2: Informants of the interviews

Open data companies	Respondent	Revenue source	Customers	
1	Small (5 persons) web technology and application development company	Project manager / Consultant	Application development	Public and private organizations
2	Small (5 persons) web technology and application development company (same as above)	Chief Executive Officer (CEO) / Consultant / Developer	Application development	Public and private organizations
3	Small (10 persons) software development company	Project manager / Developer	Conversion, Application development	Mainly government data
4	Small (2 persons) consultancy and software development company	CEO / Consultant	Consultancy	Any open data releaser
Open source companies				
1	Small (3 persons) company developing collaborative learning tools	CEO	Application development	Mainly public organizations
2	Small (3 persons) company developing collaborative learning tools (same as above)	Developer	Application development	Mainly public organizations
3	Small (1 person) company developing relational database tools	Entrepreneur	Application development	Public or private organizations
4	Small (10 persons) company developing web services	Developer	Application development	Public or private organizations

The interview guide included questions about the definition of the phenomenon (open source or open data) in business contexts. Following Rajala [21], the questions explored the different elements of the business model framework i.e. is offering, revenue model, relationships and resources. The respondents described the phenomena and we then grouped similarities in business environments into the categories of competition, customers, revenue and community. In addition, we focused on how the respondents described *openness*. The different elements of the business models were also analyzed and summarized in a table. Then the interviews were transcribed and emailed back to the respondents for comments. All the respondents corroborated that they had been quoted correctly. We also made efforts to publish the results back to the respondents to get their input and to create a discussion as to whether the respondents agreed with the results.

Our analyses focused on how the respondents talked about the terms "open data" and "open source" [24] as well as their business benefits. First, we tabled all the instances in which respondents mentioned *open source* and *open data*. In the second stage, we highlighted business uses. The third part of the analyses was the comparison of open source and open data companies. The results are reported in what follows.

4   Findings

In this chapter we discuss the findings of our empirical work in detal. First we report similarities in business environment, then the meaning of the term "open" in open source and open data and finally similarities in business models. Then we proceed to discuss these similarities and differences as well as their impact.

4.1   Similarities in Business Environment

The empirical research and analyses of the interviews showed several similarities in the business environments of the open data and open source companies as perceived by their managers. The similarities that emerged from the interviews were grouped together. The similarities formed the categories of 1) competition, 2) customers, 3) revenue sources, and 4) community following the work of Rajala [21]. Table 3 summarizes the similarities below.

Table 3: Similarities between open data and open source software business

	Open data business	Open source business	Similarity
Competition	Market is divided between small and large software companies	Market is divided between small, medium, and large software companies	Some larger companies are familiar both with open data and open data markets. They may use or develop the same software. Most of their large competitors are the same.
Customers	Emphasis on public organizations (for example, cultural institutions, municipalities). There is potential in the media industry	Emphasis on public organizations (for example, schools) as well as private companies	Public sector as a large customer
Revenue	Consultancy, conversions, application development, maintenance	Consultancy, application development, maintenance	Develop services on top of public goods
Community	Often enjoy popularity and community support	Often enjoy popularity and community support	Developer communities have links with developer communities (they have contributed to open data and open source projects).

The competition environment for both types of companies in Finland is very similar. Both open data and open source software companies share many of the same large software vendor competitors, especially in the public sector market. In addition to the international and large players, there is a considerable number of small entrepreneurial software companies. Both open data and open source companies operate in markets characterized by high market polarization between small and large companies.

Public organizations are among the largest customers for both open source and open data organizations. The customer base for both open data and open source companies is similar. However, open data companies are almost exclusively focusing on public organizations and cultural institutions while open source software companies also serve private sector customers to a large extent. Many of the developers have participated in the same programming courses or know each other some other way.

As their main product (software artifact or access to an open database access) cannot be sold directly, both rely on indirect revenue sources. The revenue sources for both types of companies are based on consultancy and building services on top of public goods. The open data revenue base was still developing, but it is growing at least according to respondent estimates. Companies also relied heavily upon community goodwill. Many open source tools are extensively used, and demands were made to keep the development processes as open as possible for both open data and open source software applications.

The philosophy of openness was discussed in the interviews, and both open source and open data companies were interested in building synergies between developer communities.

4.2 What is Open: Open Source and Open Data?

This refers to how open open data is related to the degree of data openness (open to whom) and technical format (raw data, machine-readable data, and human-readable data). Some respondents regarded opening up datasets inside organizations as an open data approach while other respondents saw data as open only when it was released over the internet (as indicated by our earlier definitions).

The benefits of data openness were of course related to the scope of openness. If the data was opened inside an organization, the benefits were limited inside that organization. Examples of such benefits include improved international communication or organizational performance. If data was opened over the Internet to anyone interested, respondents saw benefits on a national scale—for example, increased transparency or boosted economic development.

Different actors can thus open their data on different scopes depending on the goals they set for the release of the data publication. One option is to pursue open models while others see more scaled-down approaches to be more beneficial. There are also hybrid approaches to the question of openness, but ultimately this choice depended on organizational goals for publication.

4.3 Similarities in the Business Models

The companies had some differences in their business models. Table 4 summarizes some of the similarities in the business models of these companies. The business model elements are [21] revenue, offering, resources, and relationships.

Table 4: Similarities in business models [21]

	Open data business	Open source business	Similarity
Revenue	Public funding, end-user (mobile) subscription, service sales	Dual-licensing, service sales	No revenue from the sales of the open artifact (software product or access to the open database)
Offering	Service, consultancy	Service, consultancy, software product	Main offering related to service built on a public good. Another offering in most cases on consultancy and sales of expertise.
Resources	Application development, different skills related to data, sales, management	Application development, sales, management	Strong emphasis on application development
Relationships	Data publisher, development community	User community, development community, business ecosystem, and open source community	Strong ties to the development community

The main similarity in the revenue model was, unsurprisingly, that the companies could not gain subscription revenue from their main service (open source application or open database). Instead they relied on a number of different revenue sources such as public support, app sales, or, in the case of open source, dual-licensing.

A similar constraint was on offering, in which the offering was built on their main service: instead companies sold added-value services or consultancy. In the open data cases these services were related to how to open the dataset and in the open source cases they were related to their application.

All the companies were driven by application development, so their resources were in application development and maintenance, both for open data and open source. In the relationships category small companies had good relationships with the developer communities related to open source and open data. In addition, open data companies had good relations to the actor who was the original data publisher.

5 Discussion

This chapter contains discussion regarding the research results. First we discuss issues related to development community, then legal differences and similarities between open source and open data and finally discuss briefly the links that currently exist between open data and open source.

5.1 Development Community

The openness of the source code or data is not often the main concern for the user of a given service. Instead, usability and functionalities play a larger role. The expectation is that the developer has handled the issue of openness already during design. Often it does not really matter to most end users how the software is produced or what the origin of the data is, as long as they work as expected.

The situation is very different for developers and organizations that want to build their businesses on open source applications or open data stacks. They need to be convinced about the maturity, availability, long-term sustainability, maintenance, legal issues, and so on related to the open data or application. Often this certitude is linked to the developer communities in question and therefore requires organizations to build links with the developers they depend on. This is probably one of the reasons many small organizations had close links to the development community.

A specific issue related to open data was the origin of the data: if it was collected and released by a certain party, the organization that built services relying on those released datasets needed to be sure that the original publisher would continue to release up-to-date data stacks in a standard format (and normally without fee); otherwise the provision of the service would be discontinued, and the service providers would likely lose their development input and investment.

Open source companies depended heavily on their development communities, even though almost all of the code contributions were made by the company's own workers. Organizations had developed different ways to manage their relationship with the community—for example, concerning communication, the handling of contributions, bug reports, etc. Open data companies in contrast were just developing similar kinds of mechanisms. Building a relationship with the data publisher was also a key issue: many data publishers were making efforts to enlist community support and build feedback loops so that they could gain a better understanding of their downward data stream.

Different governmental actors have also been developing different kinds of industrial innovation contests to create small service companies and build new services (for example, in the areas of journalism, healthcare, energy, traffic, etc). These contests are seen as a good way to establish more actors in the field as well as to incentivize innovation. However, the long-term sustainability of these ventures was often considered problematic. New institutional arrangements were created to support these organizations after their seed-money had run out.

5.2 Legal Issues

Although most of the discussion related to legal issues falls outside the scope of this study, it is necessary to discuss the impacts of legal issues briefly. Legal issues in general were of concern to many of the respondents. The open source situation was a bit clearer because of the wide use of and interest in open source licenses that govern what developers and users can and cannot do with the software. Attached to the source code, the internationally accepted open source licenses handle issues related to copyright and derivative works. Licenses ensure that the derivative works of the source branch (fork) stay open.

The situation for open data was not yet so mature. National legislation was more diverse in different countries—a consideration for those services that were expecting an international user base. Data publishers also often had different conditions for the release and maintenance of data. The legal relations were in many cases far from clear and also interwoven with the question of who should fund the openly available data maintenance and service development. The respondents agreed that when states provided material to increase transparency and citizen oversight the usage of governmental data was quite straightforward. However, when governments produced datasets that private actors used free of charge to create services and build their businesses, the situation was seen as much more volatile.

The accuracy of data is a key question. Risks related to publication of datasets that posed threats to individual privacy as well as national security were also mentioned in the interviews.

The open data marketplace can be threatened if some providers gain similar lock-ins, as was the case with open source. An example would be a service that combines data from open and proprietary data sources and thus by the control of the private data source is able to build switching costs for the buyer. In this situation, organizations risk finding themselves losing control over the development of services to their vendors. If this vendor-lock situation occurs for public organizations, it could result in a shift away from citizen control over the infrastructure required to carry out public service.

5.3   Links between Open Data and Open Source

Many respondents viewed that open data and open source had several things in common. Respondents who were versed in technical details of course identified the core differences between developing open data and open applications. On a process-level, however, the respondents had similar views on the transparency of the process and that the end-results should likely be open in most of the cases. Many respondents also had doubts about how effective the current way to approach proprietary licenses and database rights were for new service development.

The links between open source and open data has been discussed widely in the open data community. The discussion concerns whether services built on open data should be licensed under open source licenses, what the relation is between open source licenses and the question of license and copyright regarding data. Some have argued that in order for open data services to be credible, there should be an option to have a look at the source code of the application that handles, controls, maintains, or visualizes the released datasets. Others argue that open source licenses would limit the commercial potential of open data too much, and thus services that are based on proprietary products can and should be allowed to benefit from open data stacks. This discussion is still ongoing.

6   Research Implications

This chapter describes implications of our research. The chapter is divided into three different parts. First we describe the practical impact of our findings. Then we detail what are the implications of this research to companies operating in the field. *Finally we give some recommendations and link our research to ongoing policy discussion.*

6.1   Impacts for Practitioners

The main research impact is that despite its characteristics, open source research and practice can be used to understand open data. This is especially true for issues related to community management and developer motivation that have been extensively studied related to open source. It is also worth stating the obvious: knowing open source development very well does not yet guarantee the necessary technical skills to be able to provide open data services or vice versa.

Open source literature has also delved quite deeply into the question of how to provide services when traditional software sales are not an option. The processes related to the production of the open service (open application or open dataset) are different. However, both share the ideology of the transparency of the development process. In addition, if we take the end-user perspective of the service, the process that creates the provided service is not the main concern. Instead, issues such as trustworthiness of the data, usability of the application, and the possible price of using the service are likely to be first on a customer's mind.

6.2   Impacts for Companies

At the level of a company, the similarities offer possibilities for imitation as well as learning opportunities to avoid pitfalls. One of the most interesting opportunities is related to what are called hybrid-models in open source research. These models pertain to the ways in which companies can *close* or constrain a part of their offering to extract profit. One example model is that of dual-licensing, in which the company has two versions of the open source software: the open one and the closed one. Anyone can download the open one from the Internet or participate in its development, but the closed version is only available as traditional proprietary software.

There are several good reasons why companies want to buy a product even if it would be freely available. There reasons might include for example better support service, questions related to liability, or even the need to appear to be a good pro-open source company that takes care of its obligations. Similar models might be possible for open data.

Cross-fertilization of ideas related to open data and open source is another interesting avenue to explore further. One quite obvious business model is to build proprietary software on top of open data stacks. If this is the business model pursued, releasing the software as open source will not gain traction with the company. The *open ideology* many of the respondents shared was quite interesting. We are however a little hesitant to discuss the merits and impacts of that ideology based on our respondents alone. All of the respondents had personally invested either time or money in setting up their companies, so the open development spirit was not seen as something that is opposed to profit motives that the companies had.

6.3   Impacts for Open Data Policy

Because of their similarity, open data proponents are able to draw useful lessons from open source regarding licensing issues, evasion of lock-ins, and the need to push public policies. In some sectors of the industry, interventions to public policy and procurement seem vital.

Open data enthusiasts might be able to benefit from open source experiences in several ways. In the past, open source was sometimes pushed without carefully assessing its impacts (especially in the public sector and consultancy). Sometimes it was unclear what open source could and could not deliver in the short term. This in turn led to a situation in which open source benefits were oversold and resulted in disappointment when the expectations for the new technology were not met. Now, open data proponents in organizations are facing similar challenges and could learn valuable lessons from open source.

There is also a wider discussion concerning the production of public value in society. In this paper we focused on economic value creation and capture. However, the production of public value also requires citizen oversight of the government industrial policy and commercial organizations. Open data can increase societal transparency in different ways and thus also lead to other good outcomes (for example less corruption). However, we omit these discussions from this paper due to its limited scope.

7   Managerial Implications

In what follows (Table 5), we have summarized the managerial implications of the article. They are listed in the form of guidelines that are backed by both earlier research and the empirical analysis of the respondents. First, we have listed an issue already addressed by open source companies, and then we have formulated a guideline for open data companies. The elements are competition, customers, revenue, and community, as in Table 3.

Table 5: Manager take-aways for open data business

	Open source experience	**Open data take-away**
Competition	The license fee is only a part of the total costs for the customer. Even free service does not guarantee long-term success against companies that sell their product.	Price competition might favor an open artifact, but the company still needs to focus on marketing, prospecting, and customer relations.
	Legal issues related to copyright and licensing were the focus and were also resolved early on.	Work on data licensing is still ongoing. Databases often fall outside copyright protection, but there are other legal issues to take into account: privacy, database rights, national security, etc.
Customers	The customers main concern in most cases was not the openness of the software product, but other issues (usability, price, availability of service, etc.).	The openness of the offering in itself is often not enough to sell the service. Other aspects of the service need to also be in place.
Revenue	Issues related to the business model and especially revenue source were critical for long-term business.	There are a limited number of tenable business models for companies whose offering is based on open stacks. The hybrid models might be worth pursuing.
Community	Development communities ensured long-term staying power	Community dynamics for open data may be different from open source. The data publisher's strong role especially is a potential cause for friction.
	Managing the community required know-how.	New capabilities are needed for the management of the open data development community
	More eyeballs made the bugs shallow.	External developers might be able to spot errors and problems in the data and service, but this requires feedback channels.
	Scratching an itch—attracting voluntary contribution required perks. Offering financial resources was one way, but maybe not the best one.	Developers and customers are not lured in by openness alone. If there is an expectation of outside development help, perks, and processes are needed to attract contributions.

The main managerial implications are related to building tenable business propositions in a situation characterized by offerings based on public goods. The price might be an issue that favors open data companies, but it is hardly the only factor that needs to be taken into account when developing and marketing a service. Other issues include usability, availability, maintenance, reputation, and issues such as marketing of the service. Legal issues constrain business opportunities but might also provide means to extract revenue.

The most prominent lessons might be in the area of developer motivation and *managing* a development community. New resources are needed, but new contributors must also know the ethics and expectations for good behavior in a virtual development community. External contribution to the production of service is the ultimate target for open development, but this requires processes and feedback channels to work as expected.

8   Conclusion and Limitations

There are several reasons why open data research has not studied and learned lessons from open source research, but one of the main issues may be related to open source research focusing on community-driven development and individual developer motivation. From these perspectives, many research findings are not applicable to software entrepreneurs who want to make sense of their business environment.

In this article we hope to have shown how drawing on both research traditions may be useful to better understand open source and open data. Furthermore, combining the data and application development fields offers interesting avenues for future research. We have only scratched the surface with our small round of exploratory interviews, which mostly serves as a demonstration of the dynamics of the field.

The aim of this paper was to look for some similarities between open data and open source in the context of small software companies. The marketplaces they operate in as well as the offerings of these organizations were found to be similar. Both ecosystems are populated by small and networked software companies that build services on top of public goods. We have also listed the similarities in business models as well as discussed community management, legal issues, and open data policy. We have also provided a list of guidelines for open data managers who want to benefit from earlier open source research and practical experiences.

We have excluded different dataset-specific legal concerns from this paper. We agree that the different application areas and industries offer very different business environments that may require a more thorough review than was possible in the scope of this paper. The legal and policy environment concerning open data is currently changing in Finland as well as in other national contexts. We welcome research efforts that would compare differences in legislation as well as other parts of the business environment related to open data. We conclude that open data research can draw valuable lessons from open source research. These lessons can help practitioners and managers as well as the companies to build tenable businesses as indicated.

References

[1]   P. Ågerfalk and B. Fitzgerald, Outsourcing to an unknown workforce: Exploring opensourcing as a global sourcing strategy, MIS Quarterly, vol. 32, no. 2, pp. 385-409, 2009.
[2]   T. Berners-Lee. (2006, July) Linked data design issues. W3. [Online]. Available: http://www.w3.org/Design Issues/LinkedData.html
[3]   C. Bonina, The value of open data: definitions, challenges and opportunities, RCUK Digital Economy, NEMODE, London, Final report, 2013.
[4]   M. T. Borzacchiello and M. Craglia, The impact on innovation of open access to spatial environmental information: a research strategy, International Journal Technology Management, vol. 60, no. 1/2, pp. 114-129, 2012.
[5]   J. Dinkelacker, P. Garg, R. Miller, and D. Nelson, Progressive open source, in Proceedings of ICSE 2002, Orlando, FL, US, 2002, pp. 174-184.
[6]   E. Ferro and M. Osella, Business models PSI Re-Use: A multidimensional framework, presented in Using Open data: Policy Modeling, Citizen Empowerment, Data Journalism Workshop, European Comission, Brussels, June 19-20, 2012.
[7]   E. Ferro and M. Osella, Eight business model archetypes for PSI Re-Use, presented in Open Data on the Web workshop, Google Campus, London, April 23-24, 2013.
[8]   B. Fitzgerald, The transformation of open source software, MIS Quarterly, vol. 30, no. 3, pp. 587-598, 2006.
[9]   A. Fosfuri, M. Giarratana and A. Luzzi, The penguin has entered the building: The commercialization of open source software products, Organization Science, vol. 19, no. 2, pp. 292-305, 2008.
[10] V. Gurbani, A. Garvert and J. Herbsleb, Managing a corporate open source asset, Communications of the ACM, vol. 53, no. 2, pp. 155-159, 2010.
[11] Ø. Hauge, C. Ayala and R. Conradi, Adoption of open source software in software-intensive industry – a systematic literature review, Information and Software Technology vol. 52, no. 11, pp. 1133-1154, 2010.
[12] F. Hecker, Setting Up Shop: The business of open-source software, IEEE Software, vol. 16, no. 1, pp. 45-51, 1999.
[13] N. Huijboom and T. Van den Broek, Open data: an international comparison of strategies, European Journal of ePractice, vol. 12, pp. 4-16, 2011.
[14] H. Klein and M. Myers, A set of principles for conducting and evaluating interpretative field studies in information systems, MIS Quarterly, vol. 23, no. 1, pp. 67-94, 1999.

[15] A. Latif, A. U. Saeed, P. Hoefler, A. Stocker, and C. Wagner, The linked data value chain: A light weight model for business engineers, in Proceedings of I-SEMANTICS '09 International Conference on Semantic Systems, Graz, Austria, 2009, pp. 568-575.

[16] J. Lindman and L. Nyman, The businesses of open data and open source: Some key similarities and differences, Technology Innovation Management Review, vol. 1, pp. 12-17, 2014.

[17] J. Lindman, M. Rossi and V.K. Tuunainen, Open data services: Research agenda, in Proceedings of HICSS, 46th Hawaii International Conference on System Sciences, Hawaii, 2013, pp.1239-1246.

[18] J. Lindman and R. Rajala, How open source has changed the software industry: Perspectives from open source entrepreneurs, Technology Innovation Management Review, vol. 1, pp. 5-11, 2012.

[19] Open Data Institute. (2013) How to make a business case for open date. Open Data Institute. [Online]. Available: http://www.theodi.org/guide/how-make-business-case-open-data

[20] A. Osterwalder, Y. Pigneur and C. L. Tucci, Clarifying business models: Origins, present, and future of the concept, Communications of the Association for Information Systems, vol. 16, pp. 1-25, 2005.

[21] R. Rajala, Determinants of business model performance in software firms, Ph.D. dissertation, Aalto University School of Economics, Helsinki, Finland, 2010.

[22] C. Shapiro and H. R. Varian, Information Rules: A Strategic Guide to the Network Economy, Harvard Business School Press, Boston, MA, 1998.

[23] M. Shaikt and T. Cornford, Letting go of control to embrace open source: Implications for company and community, in Proceedings of HICSS, Kauai, Hawaii, 2010, pp. 1-10.

[24] B. Swanson and N. Ramiller, The Organizing vision in information systems innovation, Organization Science, vol. 8, no. 5, pp. 458-474, 1997.

[25] Y. Tammisto and J. Lindman, Open data business models, in Proceedings of IRIS2011, 16-19.8., Turku, Finland, 2011, pp. 762-778.

[26] L. Udéhn, Twenty-five years with the logic of collective action, Acta Sociologica, vol. 36, no. 3, pp. 239-261, 1993.

[27] F. van der Linden, B. Lundell and P. Marttiin, Commodification of Industrial Software – A case for open source, IEEE Software, vol. 6, no. 4, pp. 77-83, 2009.

[28] K. Ven, J. Verelst and H. Mannaert, Should you adopt open source software?, IEEE Software, vol. 25. no. 3, pp. 54-59, 2008.

[29] G. Vickery. (2011) Review of recent studies on PSI Re-Use and related market developments. Umic. [Online]. Available: http://www.umic.pt/images/stories/publicacoes6/psi_final_version_formatted-1.pdf

[30] C. Zott, R. Amit and L. Massa, The business model: Recent developments and future research, Journal of Management, vol. 37, no 4, pp. 1019-1042, 2011.

[31] A Zuiderwijk and M. Janssen, A policies, their implementation and impact: A framework for comparison, Government Information Quarterly, vol. 31, no 1, pp. 17-29, 2014.

[32] A Zuiderwijk, N Helbig, J.R. Gil-García and M Janssen, Guest editors' introduction, innovation through open data: A review of the state-of-the-art and an emerging research agenda, Journal of Theoretical and Applied Electronic Commerce Research, no. 9, vol. 2, pp. I-XIII, 2014.