Business models for open government data

Gustavo Magalhaes  
UT Austin-Portugal Program  
The University of Texas at Austin  
Austin, Texas 78705, USA  
gustavo.magalhaes@fe.up.pt

Catarina Roseira  
Faculdade de Economia  
Universidade do Porto  
4200-464 Porto, Portugal  
croseira@fep.up.pt

Laura Manley  
New York University  
295 Lafayette Street  
New York, NY 10012, USA  
laura.manley@nyu.edu

ABSTRACT
The commercial re-use of open government data is broadly expected to generate economic value. However, the practice and study of this trend is still in its infancy. In particular, the issue of value creation in the commercial re-use open government data remains largely unknown. This study aims to further understand how open government data is used to develop commercial products and services. Grounded in the comprehensive data obtained from a sample of 500 U.S. firms that use open government data as part of their business model, we propose a taxonomy that encompasses three business model archetypes (enablers, facilitators, and integrators). Furthermore, we discuss the value proposition of each business model archetype, and subsequently present a framework that describes the value created in the context of the open government data ecosystem. Our framework can be used by both scholars and practitioners in the field of open government data to effectively frame the debate of the value created by the commercial re-use of open government data. Simultaneously, our work can be of benefit to entrepreneurs as it provides a systematic overview, as well as practical insights, of the growing use of open government data in the private sector.

Categories and Subject Descriptors
H.3.5 [Information storage and retrieval]: Information Services – Commercial services; H.4.2 [Information Systems Applications]: Type of systems – e-government applications; J.1 [Computer Applications]: Administrative Data Processing – Government

General Terms
Management, Economics, Standardization, Theory

Keywords
Open government data; business model; intermediaries; services; value creation;

1. INTRODUCTION
In the last few years, governments of various countries have encouraged current businesses and entrepreneurs to utilize public sector data to create economic and social value. Consequently, public sector data is viewed today not only as an essential resource for increased transparency [20, 40] and public accountability [46, 50], but also as a valuable tool for innovation [7, 23]. The re-use of public data in innovative ways is expected to create jobs and generate wealth [41, 55]. In light of this, governments around the world have rapidly set forward measures and policies that promote both the supply and use of public sector data [14, 31]. This new wave of public sector data release and promotion, frequently referred to as open government data, entails the participation of all levels of government, as well as the private sector and civil society, i.e. open government data ecosystem [27, 47]. Open government data is concurrent with other open movements such as open source, open access. Much in the same way it has evolved from its foundational principles of a free, equitable, and cooperative model of access and redistribution of information to something of a "political economy" [57]. However, despite the emerging public attention on this topic, the practice and study of open government data is still in its infancy. In particular, regarding the commercial re-use of open government data, which emerges as a rather promising application of public data, very few studies have been published. This article aims to further understand how open government data is used to develop commercial products and services. By bearing the similarity of studying the issue of business models in an emerging context, we base our approach on Timmers' groundwork on business models for electronic markets [54]. As such, we address the following question: what types of business models are present among firms relying on open government data as a resource for their business? To our knowledge, this study is the first to provide a systematic analysis of the value created by businesses that use open government data to generate products and services.

We first provide an overview of the main concepts and challenges regarding the commercial re-use of open government data and position this debate in the realm of the open government data ecosystem. Subsequently, we use the research method of content analysis to propose a taxonomy of business model archetypes using open government data to develop products and services. Additionally, we propose a framework that describes the value proposition of these services for the supply and the user side of the data.

2. THEORETICAL BACKGROUND
Data is the new buzzword. Today, companies across various industries are probing disruptive data-driven applications to improve competitiveness. However, data by itself has no value [13, 16]. The extraordinary value-adding potential of data lies in the ability to extract meaningful and actionable information from it [6]. In the case of open government data (hereafter OGD), data is not open to the public in any meaningful sense [11, 21]. Consequently, the onus is on the users to extract information from the available public data. This creates a problem since the average citizen is generally unable to explore the array of datasets available due to the lack of necessary computational and statistical skills [25, 37]. In other words, the increasing amount and diversity of data available limits the exploration and interpretation of data to those with the time, resources and expertise to comprehend it [10, 32, 61]. Due to these barriers in the adoption of OGD, the role of intermediate and innovative services becomes therefore a particularly important factor of OGD initiatives [27, 48].
2.1 The re-use of Open Government Data

The re-use of OGD can serve many purposes. For example, the exploration, analysis, and presentation of public data by civic-minded citizens (e.g. activists, citizen-journalists) is a critical element regarding the empowerment of citizens to scrutinize government and hold authorities accountable [12, 49]. Also, the re-use of OGD in scientific research, through open access to public datasets (e.g. genome database or census data), is acknowledged as highly beneficial to society. Additionally, OGD is used by the private sector for the development of for-profit products and services. The commercial re-use of OGD is expected to generate an economic impact in society by boosting innovation, and creating jobs and wealth [36]. Regarding this debate, Yu and Robinson [59] make a very useful distinction between “Political Accountability” and “Technological Innovation”, and point out that over the last few years the movement of OGD “has shifted toward open technology” (p. 178). However, this recent shift has generated real concerns among civic activists that the current top-down strategies enacted by governments to promote the so called ‘open data economy’ are based on prevailing capitalist interests that enable the private sector to get access to expensively produced data for no cost [5, 33, 51]. Considering the noticeable growth of this phenomenon, it seems relevant therefore to further understand how OGD is used to develop commercial products and services, and what is the value created by these firms. The business model concept, as a unit of analysis that seeks to describe the architecture of value creation [2, 44] seems to present itself as an excellent starting point for this study.

2.2 Open Government Data business models

Data has become a fundamental aspect in modern business management and strategy. However, extant literature highlights a general gap concerning how data actually creates value for companies [28]. In terms of OGD, few authors have described or analyzed the issue of business models. Two noteworthy analyses were produced by research firms McKinsey, in the US [36], and Deloitte, in UK [26]. These reports seem to substantiate the growing interest in the innovative potential of OGD. Deloitte’s research work on open data, which was conducted in collaboration with the Open Data Institute, identifies five business model archetypes in the open data “marketplace” – suppliers, aggregators, developers, enrichers, and enablers. In a similar fashion, Gurin [24] distinguishes business using OGD in two different categories: Better Business Through Open Data refers to business models from sectors such as healthcare, energy, education, and finance, that seamlessly use OGD to improve service; Open Data Pure Plays are disruptive business models that simply would not exist without OGD. However, these analyses are based on anecdotal evidence, thus suggesting the need for a more rigorous scientific study of the phenomenon. Alternatively, academic research on this topic is still very scarce. Ferro & Osella [17, 18] for example, proposed eight archetypal business models for the re-use of public data based on an exploratory study consisting of thirteen case studies of firms having business lines devoted to public data re-use in Italy. However, the work by Ferro & Osella offers a rather financially-grounded analysis of the OGD business model archetypes. We argue that their focus on the value capture element of the business model provides a limited view of the value created by these firms in the context of the open government ecosystem. Foulonneau, Martin, & Turki [19] study the development of services based on OGD by monitoring the actual re-use of the data. To do so, they present the thematic distribution of apps based on datasets published in open data portals. Although their work provides an informative view on the re-use of public datasets, it arguably falls short of describing the value created by the commercial re-use of OGD since it is confined to the apps listed in open data portals. As observed, the current research on this topic presents several limitations. Consequently, a systematic analysis of the value created by the commercial re-use of OGD based on robust empirical data regarding the manifold use of OGD by the private sector is needed.

3. METHODOLOGY

Our research is based on a data sample of 500 firms that re-use OGD to develop commercial products and services as its main source of data. This data was primarily collected during the months of September 2013 to April 2014 as part of an ongoing project, the Open Data 500, conducted by The GovLab at New York University. The objectives of the study were to (1) provide a basis for assessing the economic value of open government data, (2) encourage the development of new open data companies, and (3) foster a dialogue between government and business on how government data can be made more useful. The list of 500 companies was compiled through outreach campaigns, advice from experts and professional organizations, and additional research from publicly available sources1. The companies were identified based on the following criteria: be U.S.-based; earn revenue from its products and services; use open government data as a key resource for its business. A survey was sent by email to all companies to collect 1) information about the company such as description, number of employees, industry, type, etc.; and 2) information about their use of OGD such as the source and the specific dataset used. Of the total of 500 companies, 240 responded to the survey. For the remaining 260 companies, the information was collected online. Other descriptive data, such as internal notes and other online references (e.g. press releases, interviews) that were amassed during the sampling process, was also considered in order to enrich the sampling data and achieve more robust results [53].

To analyze our data we employ the method of content analysis in order to systematically classify the data into a taxonomy. A taxonomy can be described as a classification schema to find patterns or common characteristics from the empirical data [53]. This is justified by the need to organize the large quantity of text data for analysis into manageable groups with similar characteristics. Content analysis enables researchers to scientifically gain replicable and valid inferences from text concerning the phenomenon under investigation [35] by reducing the complexity of dealing with many instances [45]. In content

---

1 The original data file, as well as detailed information on how the list was compiled, is available for download at http://www.opendata500.com
analysis, the research question generally takes a central part in being able to generate results of scientific value [39, 52], which allows the researcher to choose the most adequate approach to tackle the problem and phenomenon under study [1, 56]. Considering the scarcity of theory in the field of OGD, it seems suitable to avoid preconceived categories and allow the categories to flow from the data [30, 34].

Our content analysis approach consisted firstly in analyzing the data thematically in grounded theory approach [22] due to the missing theoretical foundations on the characteristics of the phenomenon [53]. The coding was done by assigning 'keywords' or 'indicators' for each case (firm) using a matrix. Subsequently, in order to reduce the data and synthesize concepts, we initiated an unsupervised conceptual clustering [38]. Our approach is based on Al-Debei and Avison [1]. Therefore, we used an evaluation function to discover clusters. Companies were clustered into the same class only if they satisfied the following three conditions:

1. They are thematically similar to each other, that is, they communicate similar mission statements or business ideas about the use of data, in particular open government data

2. They have contextual relationships within the open government data ecosystem that complement each other, thus they become more useful if clustered.

3. The clustered definitions as a whole articulate a unique business model archetype.

The content analysis described here was performed by iteratively comparing within-group similarities [15] to achieve a manageable and generalizable set of conceptual clusters that share common characteristics. Consequently, this research method produced a taxonomy which encompasses three OGD business model archetypes: enablers (62 companies), facilitators (161 companies), and integrators (277 companies). In the following section we describe each of the OGD business model archetypes identified in the data analysis.

4. OPEN GOVERNMENT DATA BUSINESS MODEL ARCHETYPES

4.1 Enablers

These are businesses that provide customers with technologies such as apps or software programs mainly built on, or for the use of, OGD. On the supply-side, enablers can be understood as serving the backbone of the open government ecosystem by providing governments and public agencies with state of the art technologies for the collection, management, and disclosure of public data. For example, services such as hosting, cloud computing, or data management software, may turn out to be essential in offering all levels of government the capacity to achieve higher performance in response to the data-intensive challenges of the next few years. For example, companies such as Capticity and Xcential help governments unlock and transform static documents, including handwriting on paper, into actionable data. On the other end of the process, companies such as Socrata, Granicus, and Junar provide cloud-based technologies so that government agencies can streamline data efficiently and affordably, and offer citizens the ability to view, download, analyze, and share the data.

On the user-side, enablers develop products or services that aggregate data from different sources, including public sources, or creatively combine OGD with other types of data such as user data (e.g. location, medical records), crowdsourced data, or third-party data. Companies such as (Leg)Cyte (Congressional Research Service data), Aunt Bertha (charity program data), Embark (transportation data), FutureAdvisor (financial data), Geolytics (census data), Intermap Technologies (space data) are examples of technologies that collect, analyze, and present customized information to users in a simplified and valuable form.

4.2 Facilitators

This business model archetype is an emergent, yet important intermediary in the open government ecosystem. Facilitators support or accelerate the access and exchange of data between the supply side (government) and the user-side (developers, e.g. civic hackers, entrepreneurs) by simplifying and promoting the access to OGD. This may imply repackaging or reformatting data through technologies such as web applications, databases and APIs so that both developers and end-users can navigate, explore, and subsequently gain insights from the data. Quandl for example helps users access and explore complex finance, economics, society, health, energy, demography data. PolicyMap is a national online mapping tool offering over 20,000 indicators, many of which are from the U.S. Census Bureau, the Bureau of Labor Statistics, and the Consumer Financial Protection Bureau. Another example is Enigma.io, which is a platform that centralizes, mines and relates big public data about companies, people and locations, currently offering one of the largest and broadest repositories of public data.

Additionally, facilitators assist other private sector organizations and government agencies in the process of submitting mandatory information in compliance with public regulations. For example, Level One Technologies co-created with the Wyoming Department of Education a web based application that allows state agencies to manage and report their use of federal stimulus funds, e.g. ARRA reports. Another noteworthy example of facilitators is the emerging group of services that harness the power of the new government mandated data standard XBRL. Companies such as IPHX, Ez-XBRL, and Calbench help their clients (mostly firms in the financial sector) access, process, and submit data efficiently from the SEC’s corporate financial data repository.

4.3 Integrators

This business model archetype refers to firms that integrate OGD into their existing business models in order to further improve their existing offer. An integrator can be any business that makes use of open government data by combining it with internal data or other types of proprietary data in order to augment its business capabilities. These firms operate in various sectors (e.g. biotechnology, investment banking, real estate, insurance), and can potentially extract value from OGD at any stage of the value chain (e.g. logistics, R&D, marketing, and sales). There are many examples of firms in this category, ranging from large firms such as LexisNexis and Experian (which use records from thousands of sources including public, private, regulated, and derived data) to Van Lines moving companies that use census zipcode data to provide prospective customers with price quotes.

Broadly, the re-use of OGD in this case seems to generate a competitive advantage. However, it can also be argued that the
impact created by the re-use of OGD by integrators falls somewhat beyond the scope of the open government ecosystem since OGD is not considered a core resource in their business model.

5. ANALYZING VALUE CREATION

5.1 Value proposition

In order to further understand the value created by the commercial re-use of OGD, we first consider the concept of value proposition. Various authors have provided the definition of value proposition [4], however, although its definition is not consensual, it is commonly posited that the value proposition is a central building block of the business model concept [44]. In our work, we adopt the view of [3], who define value proposition as a statement of benefits that are delivered by the firm to its external constituencies. In light of the analysis presented in the previous section, we briefly describe in Table 1 the value proposition of each business model archetype for both the supply-side and the user-side of OGD.

<table>
<thead>
<tr>
<th>Archetype</th>
<th>Supply-side</th>
<th>User-side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enablers</td>
<td>Efficiency</td>
<td>Personalization</td>
</tr>
<tr>
<td>Facilitators</td>
<td>Standardization</td>
<td>Empowerment</td>
</tr>
</tbody>
</table>

We posit that enablers provide efficiency to the supply-side of OGD by offering high-performance technology for better data management at all levels of government. The contribution made by these businesses is very significant as the increase of the capacity of government agencies to collect, maintain, and publish the data, is essential to the development of the open government ecosystem. Additionally, enablers provide end-users with personalization through tailored solutions that combine OGD from multiple public sources or with personal data. These services offer customers valuable tools for better decision making in a wide range of areas, such as health care, civic and political interests, home utility bills, etc.

Furthermore, we argue that facilitators provide the user-side with empowerment through services that improve access to and exploration of OGD. For example, these services repot data that is relevant only to a specific target group of citizens or professionals, or produce modern interactive user-friendly web platforms for developers and end-users to navigate and explore the data.

Furthermore, facilitators contribute to the standardization of data being submitted to government. This is a crucial aspect since greater standardization and reliability in the process of submitting mandatory data to government, either from the private sector or the public sector itself, can reduce fraud and eliminate waste.

Finally, we consider that due to the diversity of businesses that compose the integrators business model archetype, it becomes unfeasible to characterize a common value proposition for this group.

5.2 Value created in the context of the OGD ecosystem

The re-use of OGD is part of a wider trend, open government, which consists mainly in achieving greater transparency, participation, and collaboration at all levels of government [42]. Harrison et al. [27] propose the “open government ecosystem” metaphor to represent “government in relations of interdependence with innovators on the one hand and citizens on the other hand, nested within the larger environment of the economy, legal system, and policy expertise” (p.908). Consequently, when discussing the commercial re-use of OGD it becomes beneficial to employ a more network-centric approach. The business model concept broadly described by Zott, Amit, & Massa [60] as “a new unit of analysis that is distinct from the product, firm, industry, or network; it is centered on a focal firm, but its boundaries are wider than those of the firm” (p. 2) is therefore an appropriate tool for assessing to what degree a firm’s business architecture is designed to collaborate with (external) stakeholders within the ecosystem. The concept of business model openness, as discussed in open innovation [8, 9, 58] and business model literature [29, 43, 58] seems to provide a useful dimension for this particular aspect. Despite much discussion regarding a consensual definition of the concept of open business model, the term open “is generally seen as referring to a firm's boundaries and its collaboration with the outside world across these boundaries – be it with other firms, communities, or customers” [58]. This concept becomes particularly relevant when studying a business model or industry if one wishes to describe the surroundings of the focal firm into which it is embedded, i.e. ecosystem [58]. In our research, we adopt the concept of business model openness described by Holm et al. [29] as the degree to which the focal firm depends on one or a few partners to perform its business. In the context of the open government data ecosystem, this relates to how much a firm relies on its value network, or in other words, its ecosystem, to provide an offering. The concept of business model openness forms the first dimension of our framework.

In the second dimension of our framework, we introduce the concept of open government data centrality as the degree to which OGD is a central resource of the business model. We posit that open government data centrality is a practical way of representing to what degree OGD contributes to the firm’s sustained value creation and capture.

Thus, we conclude our analysis by mapping qualitatively the previously identified business model archetypes (facilitators, enablers, and integrators) in a two-dimensional framework comprising the dimensions of business model openness and open government data centrality (see Figure 1).

Our analysis shows that facilitators have both a) high business model openness, since they do generally collaborate with the ecosystem, namely their customers, to provide a joint offering, and b) high open government data centrality because their business model is based on OGD as the main resource for value creation and capture. Google Public Data Explorer is an example of a service that is completely based on OGD (high OGD centrality) and co-created value with its users by providing the platform in which users explore the data to generate insights (high business model openness).
Conversely, integrators imply a low degree on both dimensions since, in their case, OGD is commonly used for in-firm activities such as R&D, and is not a central resource in their business model. For example, JP Morgan Chase combines vast amounts of credit card information and other transactional data about U.S. consumers with publicly available economic statistics to develop proprietary insights into consumer trends (low OGD centrality and low business model openness).

Finally, enablers are positioned between low and high in both dimensions since most applications and software developed by these firms only partially involve collaborating with the customer (often in the form of data captured from the user). SimpleTuition helps students plan for college costs and college-related expenses. Their service makes use of the Integrated Postsecondary Education Data System (IPEDS) of the Department of Education in combination with personal data supplied by the user (medium OGD centrality and high business model openness).

6. CONCLUSION
The re-use of open government data is said to contribute to greater transparency, increased public sector efficiency, and open innovation. On the supply-side government officials look to reap the benefits of external innovation. Simultaneously from the user-side, businesses are finding more and ways to re-use and add value to this data. Grounded in the rich data obtained from a sample of 500 firms, we propose a taxonomy comprising three different business model archetypes based on how open government data is used. We also present a framework that describes each business model archetype in relation to its openness and open government data centrality. Our framework can be used by government officials and civic-minded individuals to effectively frame the debate of the value created by the commercial re-use of open government data. Simultaneously, our work can benefit entrepreneurs as it provides a systematic overview, as well as practical insights, of the growing use of open government data in the private sector. New avenues for research may consider empirical data from other contexts (e.g. international data, sector data) or explore the point of view of the end-user.

7. BIBLIOGRAPHY
Proactive Disclosure and Open Data in the United States and Other Countries.


[55] Vries, M. de 2012. *Re-use of public sector information – Catalogue and highlights of studies, cases and key figures on economic effects of changing policies.*


