Chapter 7
Open Data Value and Business Models

"Open data is the new oil of the digital economy."

7.1 Introduction

The chapter focuses on innovation processes aspiring to generate value through a purposeful and effective exploitation of data released in an open format. On the one hand, such processes represent a great opportunity for private and public organizations while, on the other, they pose a number of challenges having to do with creating the technical, legal and procedural preconditions as well as identifying appropriate business models that may guarantee the long term financial viability of such activities. As a matter of fact, while information sharing is widely recognized as a value multiplier, the release of information in an open data format through creative common licenses generates information-based common goods characterized by nonrivalry and nonexcludability in fruition. An aspect posing significant challenges for the pursuit of sustainable competitive advantages.

The objective of this chapter is to shed light on some of the challenges highlighted above, with particular reference to the business models that may be adopted for igniting data-driven value generation activities. More specifically, the chapter will start by providing some background on a few key concepts having to do with the notion of value, the economics of information and business models. Subsequently, an overview of the most prominent studies on business models for open data will be presented. Finally, the main exploitation opportunities and some real-life cases will be discussed to exemplify a number of good practices of open data valorization in both the private and the public sector.

Y. Charalabidis et al., *The World of Open Data*, Public Administration and Information Technology 28, https://doi.org/10.1007/978-3-319-90850-2_7

7.2 Key Concepts

The discussion conducted in the following sections will address the value of open data and the different exploitation avenues that may be pursued from both a public and private perspective. The brief review presented in this section will thus glimpse at three concepts that are at the heart of open data exploitation processes: the notion of value, the cost structure of information and the concept of business model. The aim of this section is thus to create a clear and shared understanding to be used as a starting point for further discussion.

7.2.1 Value

As Adam Smith (1776) reminds us, when talking from an economist's perspective "the word value has two different meanings, and sometimes expresses the utility of some particular object, and sometimes the power of purchasing other goods which the possession of that object conveys. The one may be called 'value in use'; the other, 'value in exchange'. The things which have the greatest value in use have frequently little or no value in exchange; on the contrary, those which have the greatest value in exchange have frequently little or no value in use".

When taking a philosophical stance, traditional axiology shows how it is possible to distinguish between intrinsic value and instrumental value. In other words: if something is good only because it is related to something else, then its value is instrumental to the achievement of a given objective. To exemplify, money is supposed to be good, but not intrinsically good: it is supposed to be good because it leads to other good things such as the possibility to buy food and water (Schroeder, 2008).

In addition, the so called point of view theory (Schroeder, 2008) clarifies the difference between what is good simpliciter from what is good for a specific stakeholder: the former defines what has value from a more generic point regardless of the circumstances, while the latter is perspective-dependent.

Finally, the perception of value is strictly correlated with the needs of a society. In this respect, it is useful to mention that individual as well as collective needs may be hierarchically organized in order to provide a priority ranking. The work conducted at the beginning of the last century by the American psychologist Abraham Maslow represents a cornerstone in this field (Maslow, 1943). His celebrated hierarchy of needs identifies five categories of needs having to do with physiology, security, belonging, esteem and self-actualization. In a resource constrained situation, such classification represents a useful tool in identifying and prioritizing the long term strategic priorities that should be targeted in order to create value for the society. A value that – as Savitz (2006) reminds us – unfolds along a number of dimensions touching upon financial, social, and environmental aspects.

7.2.2 Public Value

Moving on to the concept of public value, it may be described as the analogue of the desire to maximize shareholder value in the private sector: in fact, according to Kelly, Mulgan, and Muers (2002), all governments should want to maximize "public value added", i.e., the benefits of government action when weighed against the costs (including the opportunity costs of the resources involved). In addition, the notion of public value spawned the development of performance measurement/management frameworks, attracting the attention of practitioners and management enthusiasts.

Taking this stance, Kelly et al. (2002) discuss public value as an analytic framework for public sector reform where public value becomes "the value created by government through services, laws, regulations and other actions" thereby creating a "rough yardstick against which to gauge the performance of policies and public institutions". Cole and Parston (2006) crafted the Accenture Public Service Value Model's methodology for measuring how well an organization achieves outcomes and cost-effectiveness over a period of years and, adopting a sectorial perspective, Cresswell, Burke, and Pardo (2006) outlined a public value framework for the return on investment (ROI) analysis of government IT estate. Despite some difficulties in operationalizing the concept through wide-ranging measurement systems, the notion of public value may offer a promising way of measuring government performance and guiding policy decisions.

7.2.3 Business Model

The notion of value is at the heart of business models. They have been integral to trading and economic behaviour since pre-classic times (Teece, 2010) nevertheless, the business model concept became prominent with the advent of the Internet in the 1990s and it has been gathering momentum since then. As it often happens in the academic field, no consensus has been reached on a common definition for such concept. The literature, in fact, refers to a business model as a statement (Stewart & Zhao, 2000), a description (Applegate, 2000; Weill & Vitale, 2001), a representation (Morris, Schindehutte, & Allen, 2005; Shafer, Smith, & Linder, 2005), an architecture (Dubosson-Torbay, Osterwalder, & Pigneur, 2002), a conceptual tool (Osterwalder, 2004; Teece, 2010) a structural template (Amit & Zott, 2002), a method (Afuah & Tucci, 2002), a framework (Afuah, 2004), a pattern (Brousseau & Penard, 2006) and as a set (Seelos & Mair, 2007).

For the purpose of the present discussion, the notion of business model will be intended as a representation of the value architecture through which a given enterprise generates, delivers and appropriates value (Osterwalder & Pigneur, 2010). Business models thus provide an enterprise centric view and are tightly connected with the notion of value. Specifically, the key challenge that we will be discussing in this chapter is the identification of the value architectures (business models) that may be put in place for the generation of both public and private value.

In order to properly design financially sustainable and strategically cunning business models it is important to have a deep understanding of the economics of information and its impact on exploitation strategies. In this respect one of the most authoritative contribution on the topic present in the literature has been offered by Carl Shapiro and Hal Varian in "Information rules: a strategic guide to the network economy" (1999). In their bestselling book, the two authors provide a clear and detailed account of the cost structure of information in terms of production, reproduction and distribution.

According to Shapiro and Varian (1999), one of the most fundamental features of information goods is that their cost of production is dominated by the "first-copy costs". Once the first copy of a DVD or an MP4 file has been generated, the cost of producing additional units is very low. In addition, the cost of distributing information is falling, causing first-copy costs to comprise an even greater fraction of total costs to get an information good in the hands of the final consumer. In the language of economics, the fixed costs of production are large, but the variable costs of reproduction are small. This cost structure leads to substantial economies of scale: the more you produce, the lower your average cost of production. But there's more to it than just economies of scale: the fixed costs and the variable costs of producing information each have a special structure. The dominant component of the fixed costs of producing information are sunk costs, costs that are not recoverable if production is halted. If you invest in a new office building and you decide you do not need it, you can recover part of your costs by selling the building. But if your film or your song flops, there isn't much of a resale market for its script or score. Sunk costs generally have to be paid up front, be/one commencing production. In addition to the first-copy sunk costs, marketing and promotion costs loom large for most information goods. The variable costs of information production also have an unusual structure: the cost of producing an additional copy typically does not increase, even if a great many copies are made. Unlike Airbus, Google does not face appreciable and lasting capacity constraints. Normally there are no natural limits to the production of additional copies of information: if you can serve one customer you can serve a million customers at roughly the same unit cost. The low variable cost of information goods offers great marketing opportunities. Just as sellers of new brands of toothpaste distribute free samples via direct mail campaigns, sellers of information goods can distribute free samples via the Internet. The toothpaste vendor may pay a dollar or two per consumer in production, packaging, and distribution to promote its product; but the information vendor pays essentially nothing to distribute an additional free copy.

The first-copy costs common to information goods are "merely" the extreme version of what we see in other industries where scale economies are powerful, which includes many high technology industries like chip fabrication. To summarise the brief overview on the economics of information, we may say that:

- Information is costly to produce but cheap to reproduce.
- Once the first copy of an information good has been produced, most costs are sunk and cannot be recovered.
- Multiple copies can be produced at roughly constant per-unit costs.
- There are no natural capacity limits for additional copies.

The cost structure of information goods is a key aspect to keep in mind when designing economically sustainable (and profitable) products or services leveraging open data as a strategic resource.

7.3 Open Data Value Chain and Business Models

The process that leads from the generation of a data asset to its consumption is far from being linear and subject to diverse interpretations. Many studies have embarked in providing a high-level representation of such process (Capgemini, 2015; DG Connect, 2013; Ferro & Osella, 2011; Pira International, 2010). The various attempts provided representations at different levels of granularity and units of analysis. For the purposes of this discussion a revisited version of the value chain proposed by Ferro and Osella (2011) will be used in order to include information generated both by public and for-profit actors as well as to clearly distinguish three aspects: (1) activities conducted, (2) relevant actors and (3) outputs generated in each step of the value chain.

As it may be noticed from Fig. 7.1, the main added-value activities conducted along the chain are: data generation, dissemination, retrieval, storage, categorization, exposure, re-use and consumption; while the outputs of the different steps are: raw data, refined data, and "fit-for-purpose" products and services; finally, 11 of archetypical actors (four public and six for-profit) operate along the value chain.

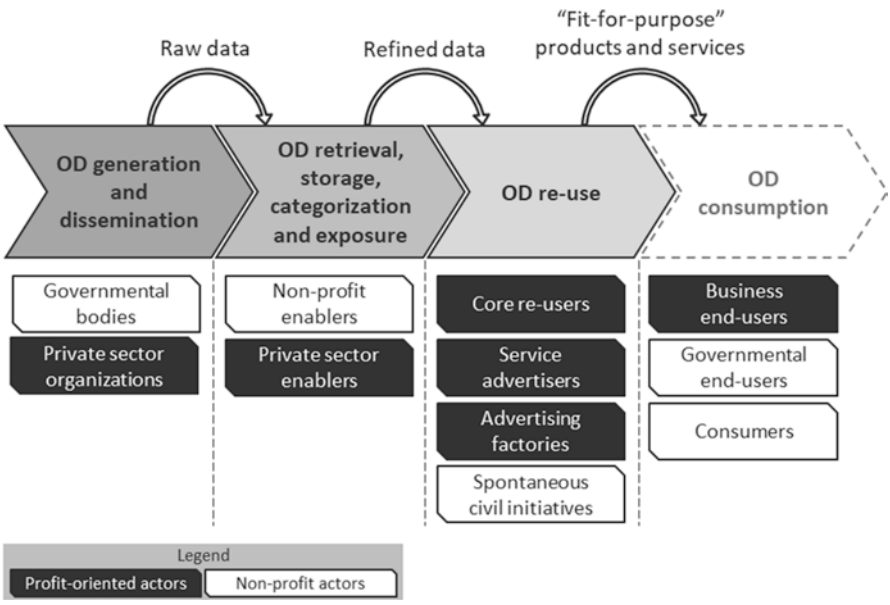

Fig. 7.1 Open data value chain. (Elaborated from: Ferro and Osella (2011))

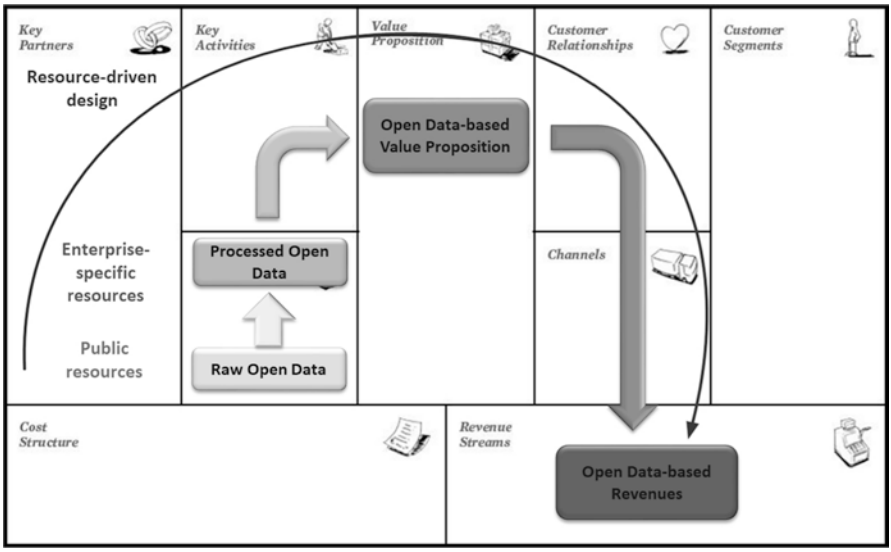

Fig. 7.2 A resource-driven design. (Source: Ferro and Osella (2011))

The discussion about which business models may be adopted in the exploitation of open data mainly applies to private for-profit organizations as they are the actors more challenged by finding a financial sustainability in leveraging a public good. It is important to underline that such discussion does not merely offer a representation of the activities conducted or the position covered in the value chain. As a matter of fact, to provide actionable insights to a would-be open data entrepreneur it is essential to depict the value architecture through which an organization creates, delivers and appropriates value. For this reason, the business model canvas methodology devised by Osterwalder & Pigneur, 2010 represents a useful and comprehensive tool (Fig. 7.2).

As highlighted in Ferro and Osella (2012), in the case of open data reuse the epicenter of the business model lies in a resource (i.e., one or many data sets) which is accessible by everyone when released in accordance with the open data paradigm (i.e., without technical, legal and price barriers). Subsequently, such a raw resource is elaborated in order to become an enterprise-specific asset that distinguishes the respective owner from the rest of the world. Such processed data is an ingredient of the value proposition that the enterprise offers to the market. In other words, elaborated data is "packaged" and embedded in the bundle of products and services which is supposed to create value for at least one customer segment. In return to such a value, customers generate revenues for the enterprise through alternative forms of payment. The discussion about business models employable in the exploitation of open data will focus on for-profit actors operating in the second and the third step of the value chain. More specifically, on two archetypal actors directly facing the end consumer (core-users and service advertisers) and two operating behind the front lines (enablers and advertising factories). For each archetype one or

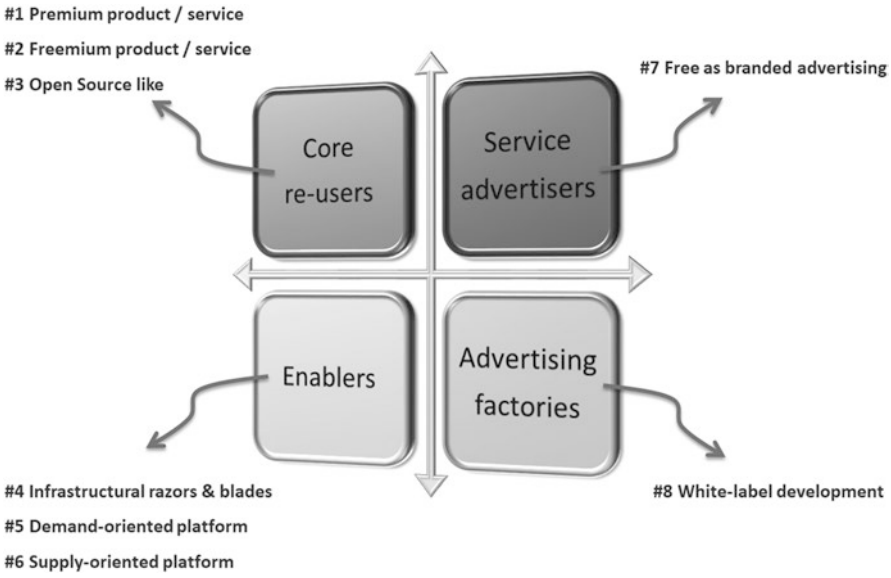

Fig. 7.3 Archetypal actors & business models. (Source: Ferro and Osella (2011))

more potential business model was identified and briefly described in natural language. A more formal representation of such business models may be found in Ferro and Osella (2013) (Fig. 7.3).

#1 Premium Product/Service While implementing this business model, a core re-user offers to end-users a product or a service presumably characterized by high intrinsic value in exchange for a payment that could occur à la carte or in the guise of a recurring fee: while the former implies the payment of an amount of money for each unit of product purchased (pay-per-use), the latter has an "all-inclusive" nature since it grants for a given timeframe the access to certain features in accordance with contractual terms. In this business model, probably associated to the "mainstream" model by the majority of analysts, the high intrinsic value, coupled with the price mechanism, calls for B2B customers often called "high-end market" (De Vries et al., 2011) and for long or medium terms relationships going beyond single transactions.

#2 Freemium Product/Service Core re-users resorting to this business model offer to end-users a product or a service in accordance with freemium price logic: one of the offerings is free-of-charge and entails only basic features, while customers willing to take advantage of refined features or add-ons are charged. In the open data realm, the implementation of this business model has its roots in limitations deliberately imposed by the core re-user in terms of data access: as a result, ad-hoc payments may be required to enjoy advanced features, to have recourse to additional formats or, sometimes, to weed out advertising. In contrast with the previous model, here the prominent target market is the consumer one often called "low-end market"

(ibid) with which the firm establishes medium or short terms relationships that usually do not involve the customization. Target customers are generally reached via the Web or via the mobile channel, which are promising to "hit" a considerable number of installed bases.

#3 Open Source Like This very peculiar business model takes place on top of products, services, or simple unpackaged data that are provided for free and in an open format. In terms of economics, a cross-subsidization (Anderson, 2009) occurs in the enterprise under examination since the costs incurred for free offering of data are covered by revenues stemming from supplementary business lines that are still open-data-based: in fact, trickles of revenue for the core re-users may stem only from added-value services or from license variations (dual licensing). The resemblance with Open Source software is given by the fact that in this circumstance data is provided in a totally open format that allows free elaboration, usage and redistribution without any technical barrier.

#4 Infrastructural Razor & Blades Entering in the realm of enablers, this business model is chosen by enterprises acting as intermediaries that facilitate the access to open data resources by profit-oriented developers or scientists not driven by commercial intent. As it happens in the well-known model "razor & blades", the value proposition hinges on an attractive, inexpensive or free initial offer ("razor") that encourages continuing future purchases of follow-up items or services ("blades") that are usually consumables characterized by inelastic demand curve and high margins. Applying this model in the open data environment, datasets are stored for free on cloud computing platforms being accessible by everyone via APIs ("razor") while re-users are charged only for the computing power that they employ on-demand in as-a-service mode ("blades"). This business model exhibits another case of cross-subsidization whereby profits accrued from the provision of on-demand computing capacity cover costs attributable to the storage and maintenance of data. Finally, it goes without saying that application of this model is limited to contexts and domains in which the computational costs are significant.

#5 Demand-Oriented Platform Following this business model, the enabler acting as intermediary provides developers with easier access to open data resources that are stored on proprietary servers having high reliability. Once collected, open datasets are subsequently catalogued using metadata, harmonized in terms of formats and exposed through APIs, making it easier to dynamically retrieve data in a meaningful way. As a result, a wide range of critical issues pertaining to original raw data are made irrelevant due to the usage of platforms capable to convert datasets in data streams, contributing significantly to the "commoditization" and "democratization" of data. In addition, developers may reap the benefits given by the "one stop shopping" nature of such platforms: they may resort to one supplier and access a variety of information resources through standardized APIs – even beyond the borders of the open data – without having to worry about interfaces connecting to each original source. This "procurement" approach is crucial to minimize search costs and, by

consequence, transaction costs. In terms of pricing, as a good that was born free and open (such as Open Government Data) cannot be charged in absence of added value on top of it, enablers adopting this business model earn revenues in exchange for advanced services and refined datasets or data flows. To sum up, re-users are charged according to a freemium pricing model that sets the boundary between free and premium in light of feature limitations.

#6 Supply-Oriented Platform To conclude with enablers, this business model entails the presence of an intermediary business actor having again an infrastructural role. However, on the contrary of the previous case, according to this logic open data holders are charged in lieu of developers. In fact, the enabler, following the golden rules of two-sided market (Eisenmann, Parker, & Van Alstyne, 2006) fixes the price according to the degree of positive externality that each side is able to exert on the other one. Consequently, this approach is beneficial for both sides of the resulting arena: from developers' perspective, their barriers are wiped out (i.e., they can retrieve data without incurring cost) while, from the governmental angle, open data holders become platform owners taking advantage of some handy features such as cloud storage, rapid upload of brand-new datasets by public employees, standardization of formats, tagging with metadata and, above all, automated external exposure of data via APIs and GUI. Public agencies that adhere to such programs in order to dip their toes into the water of Open Data establish long term relationships with providers and are required to pay a periodic fee that depends on the degree of sophistication characterizing the solutions purchased and on some technical parameters.

#7 Free as Branded Advertising Service advertising is an emerging form of communication aimed at encouraging or persuading an audience towards a brand or a company. Conversely to the more famous "display advertising", where commercial messages are simply visualized, in service advertising the advertiser strives to conquer the customer by providing him or her with services of general usefulness. That said, in the open data realm, services offered in this way do not generate any direct revenue, but they are supposed to bring positive return in a broad sense, driving economic results on other business lines – unrelated to open data – that represent the enterprise's core business. The rationale fueling this "enlightened" business model is twofold. Firstly, it may be based on a powerful advertising boost that leads the company to consider the cost as a promotional investment in the marketing mix. Secondly, it seems to be very convenient in presence of zero marginal costs (Anderson, 2009), a situation that occurs when the costs of distribution and usage are not significant.

#8 White-Label Development Last but not least, if service advertisers do not have in-house sufficient competencies required to develop their business endeavors, they can knock the door of advertising factories. Such firms, in fact, come into play as outsourcers carrying out duties that otherwise would be handled by service advertisers. Hence, the development of PSI-based solutions is particularly compelling for

companies willing to use open data as "attraction tool" but not equipped with competencies required to do so (e.g., data retrieval, software development, service maintenance, marketing promotion). In order to let the service advertiser's brand stand out, solutions are developed in a white-label manner, i.e., shadowing the outsourcer's brand and giving full visibility to the sole service advertiser's brand. Taking into account the "one stop shopping supply" and the business-criticality of the solutions in terms of corporate image, the resulting one-to-one relationship between provider and customer is tailor-made and "cemented". Concerning financials, advertising factories collect lump-sum payments or recurring fees in exchange for turn-key solutions so developed, depending on whether the crafted solution takes the form of product or service: whilst in the former case service advertisers perceive the cost as CAPEX, in the latter one the respective cost assumes an OPEX nature.

To provide and clear and explicit link among: archetypal actors, business models and real life business ventures, some examples are provided in Table 7.1.

Although the table does not have any expectation of statistical representativeness or exhaustiveness, it is possible to note a concentration trend around few positions in the value chain. More specifically, the lack of market maturity seems to have led the majority of companies to either lean towards enabling open data fruition for third parties by helping public agencies to expose data sets in a machine-readable format or towards leveraging open data as a marketing attraction tool through the provision of branded value-added services free of charge.

The business models presented above are stemming from the results of the exploratory study conducted by Ferro and Osella (2013). Other attempts to shed light on the topic have been conducted by scholars and professionals around world with different slants and foci. To exemplify, Shuhaka and Tauberer (2012) looked into business models for the reuse of legislative data and identified a six business models mostly overlapping with those identified by Ferro & Osella (pay services (or premium), freemium, advertising, startup, crowdfunding, nonprofit, government).

Table 7.1 Examples of for-profit open data ventures

Archetypal actors	Business model	Companies
Core re-user	Preemium	HospitalRegisters
Core re-user	Freemium	Voglio il Ruolo
Core re-user	Open source-like	OpenCorporates, OpenPolis
Enablers	Infrastructural razor & blades	Public Data Sets on Amazon Web Service
Enablers	Demand-oriented platforms	Data Markets, Infochimps, Factual, SpazioDati
Enablers	Supply-oriented Platforms	Socrata, OpenDataSoft, Microsoft OGI
Service advertizers	Free as branded advertizing	IBM City Forward, IBM Many Eyes, Google Public Data Explorer
Advertizing factories	While-label development	Datamarketing

Source: Ferro and Osella (2011)

The work conducted by Suhaka and Tauberer looked at both for profit and nonprofit venture and took into consideration provisional business models as in the case of "startup" (a company operating on venture capitalists' funds). Another effort worth mentioning is that of Jennifer Tennison (2012) focusing on a number of pricing logics for open data that take inspiration from the open source world. More specifically, she identified the eight logics briefly explained below:

Cost Avoidance: may help organisations avoid the costs of Freedom of Information (FOI) requests. This applies only to data that is likely to be requested or has a very low publishing cost. Organisations that have a high FOI spend with lots of successful requests may find that they can lower that FOI spend by proactively releasing data (and making it easy to find).

Sponsorship: the reverse of cost avoidance is finding sponsors for open data publication. If there are people who strongly believe that a particular dataset should be open and available to all, they may be prepared to sponsor its publication (which is not the same as licensing it; the consequence is that the data is open for all, not just for those who pay). How to persuade others to sponsor opening up data? Perhaps, if it is the type of dataset that is hard to close up again after it has been made open, they might gamble that it would lower their long-term costs. Perhaps they sell analysis or visualisation products that they know those who use the data will find useful, and so getting the data available widely will aid their business.

Freemium: the freemium model has been used with some success for web-based services; it might also work for open data. Under this model, an organisation would publish open data in a basic form – perhaps with some limitations on formats and throttling of API calls – and offer advanced access to those who are willing to pay. There are many ways in which open data can be made more useful than static publication of spreadsheets or a basic API; under a freemium model some of these enhancements would only be offered to those who pay for it:

- availability of different machine-readable formats
- unconstrained numbers of API calls
- more sophisticated querying
- access to data dumps rather than through an API (or vice versa)
- provision of feeds of changes to the data
- enhancement of the data with additional information
- early access to data
- provision of data on DVDs or hard disks rather than over the net

Dual Licensing: data publishers could provide data under an open license for certain purposes, and under a closed license for others. This technique has worked for some open source products. The "certain purposes' might not be simply 'non-commercial": publishers could still encourage start-up use of the data by charging based on the size or revenue of the organisation. Or the license could state that the data can be used in products but cannot be used in further "added value" data feeds without being licensed (this is roughly equivalent to dual-licensing with a share-alike license).

Support and Services: offering support and services is a business model which seems to work well for companies built around open source. In the open data world, data publishers could offer paid packages with:

- guarantees on data availability
- prioritisation on bug fixes (both in data and its provision) for paying customers
- timely help for customers using the data
- services around data visualisation, analysis and mashing with other data

These kinds of services still tend to be coupled with licenses in the data world, whereas in open source they have been successfully disentangled.

Charging for Changes: in some cases, individuals or organisations are obliged to provide information to public bodies (and they have a statutory duty to collect it), so that it is available within government and more generally in society. These public bodies can (and sometimes do) charge the providers of that information "administration costs". Examples of this are Companies House information, the Gazettes, Land Registrations, VAT Registrations and so on. In these cases, those who supply the information to the register are bound to by law, so it would be possible to charge them whatever it took to support providing the data as open data. Indeed, supplying the data as open data is likely to increase its usage (both within government and more widely), and therefore the political pressure to retain the registry and thereby maintain its longevity.

Increasing Quality through Participation: the model used by legislation.gov.uk is based on increasing the quality of the data that we have to publish – bringing the statute book up to date – by enlisting the help of other parties who would benefit from having an up-to-date open statute book. Because otherwise this information is very costly to get hold of, there are any number of potential contributors, including publishers, lawyers, academics, and government itself. This model doesn't entirely cover the costs of opening up data: contributors are not generally paying money to be involved but donating effort to maintaining the published data. Thus, this business model does not completely cover costs, but it is a very useful one for organisations that have an obligation to publish information but lack the resources to do it well.

Supporting Primary Business: the final business model may be used when releasing open data naturally supports the primary business goal of the organisation. The best example of this is around the Barclays Cycle Hire in London, where releasing open data about the bikes drives the development of Apps that make it easier for potential customers to use the scheme, thus bringing in revenue to the core business. Another example is the recent release of data about Manchester City football players which, they hope, will lead people to create better ways of measuring player performance, which they will then be able to take advantage of.

A further, and final, perspective is offered by Janssen and Zuiderwijk (2014) who conducted a study on the business models for infomediaries, i.e.: organizations positioning themselves between open data producers and users. The authors identified six business models (single-purpose apps, interactive apps, information aggre-

gators, comparison models, open data repositories, and service platforms) some of which describing the purpose of the tool developed and some others describing the activities conducted by the organizations building the tool.

As it may be noticed from the overview provided above, the topic of business models for open data exploitation still requires time and efforts to reach a maturity stage. As the availability and the quality of open data increase, it could be worth conducting a new wave of studies that go beyond mapping and formalizing business models by looking at their performance and long-term sustainability from a financial, legal and operational point of view.

In the following sections the discussion will shift from an enterprise centric view to a macro level perspective highlighting market and governance aspects that need to be addressed for the creation of a vibrant open data socioeconomic system.

7.4 Open Data Exploitation in the Private Sector

In order to understand what business opportunity may reside in the exploitation of open data for Europe-based enterprises, it is important to provide an overview of the estimated current market size, the expected growth trends as well as of the breakdown by sector and member state.

Figure 7.4 provides a quantification of the European market size together with some projections to 2020 along three main dimensions. The total market value was estimated to be close to 300 billion euros in 2016 and expected to more than double by 2020. The foreseen increase in value is reflected in a nonlinear fashion in the amount of companies that will integrate data in their core business as well as in workers whose main duties will revolve around data treatment or exploitation.

From a geographical point of view, the European data market follows a negative exponential distribution with a concentration of over 60% of the value in 4 of the 28 member states. Figure 7.5 shows the distribution among the different member states both in absolute value and in terms of percentage over the total.

Moving now from geography to industrial sectors it is interesting to note that the manufacturing and the financial sectors seem to offer the most significant opportunities. This may justify the significant attention that is currently being paid to topics such as industry 4.0 and fintech (Fig. 7.6).

With respect to the effort still necessary in terms of data liberation as a prerequisite for an effective and fruitful data exploitation, Fig. 7.7 shows that a long way is still lying ahead. A yearly survey conducted by the Open Data Barometer shows how at a global level only 10% of data is currently released in an open data format, thus significantly limiting the potential for reuse and exploitation. Having said that, of course, not all data should be released in an open format, especially those containing personal or sensitive information.

From a public-sector information standpoint, it is important to understand what could be the commercial appeal of the different datasets for private sector organizations in order to: prioritize investments in data liberation, allocate the resources

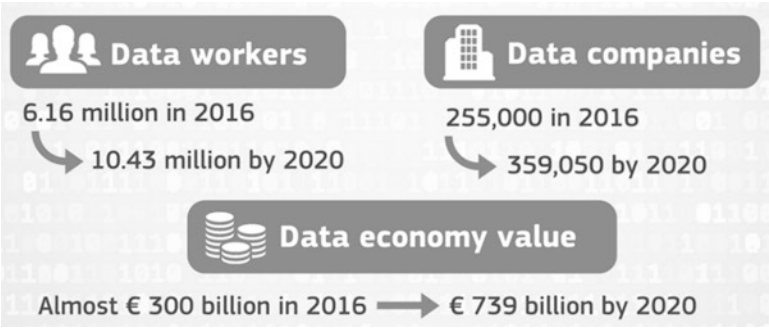

Fig. 7.4 EU data market overview. (Source: IDC (2017))

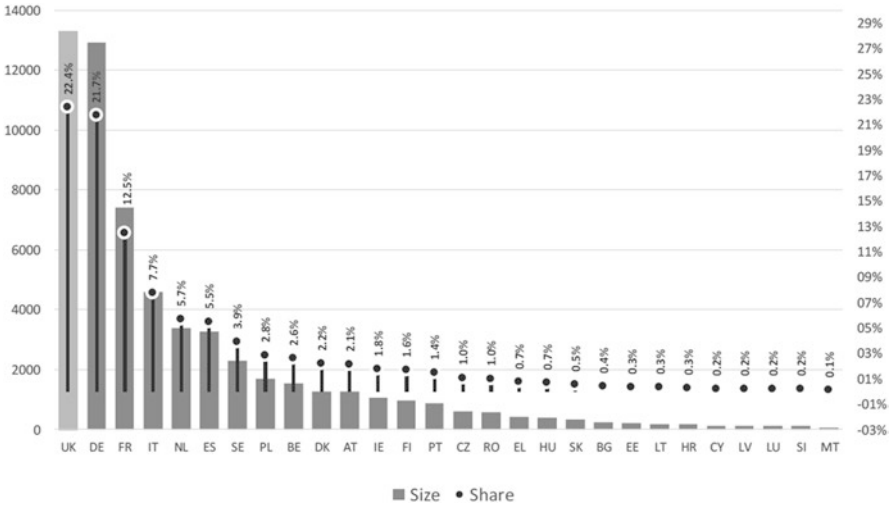

Fig. 7.5 Data Market Value (€M) & Share (%) by MS. (Source: IDC (2017))

necessary to guarantee the required levels of data quality and, finally, define a fair pricing model that may lead to a long-term sustainability of the process of data provision.

In this respect a study, conducted by Capgemini (2015) looked at the commercial reuse of open data sets. This study shows the different types of data generated by the public sector during its daily operations by appeal in terms of commercial reuse for profit-oriented business (see Fig. 7.8). Aside from noting that geographical together with meteorological and economic information seat of the podium of the classification, it is important to notice that not all data carry the same appeal and, as a consequence, should be exploited at the same time. This is to say that some data set are more readily reusable by the business ecosystem, while other types of datasets (e.g. cultural content) may require a longer lead time to find a viable exploitation avenue.

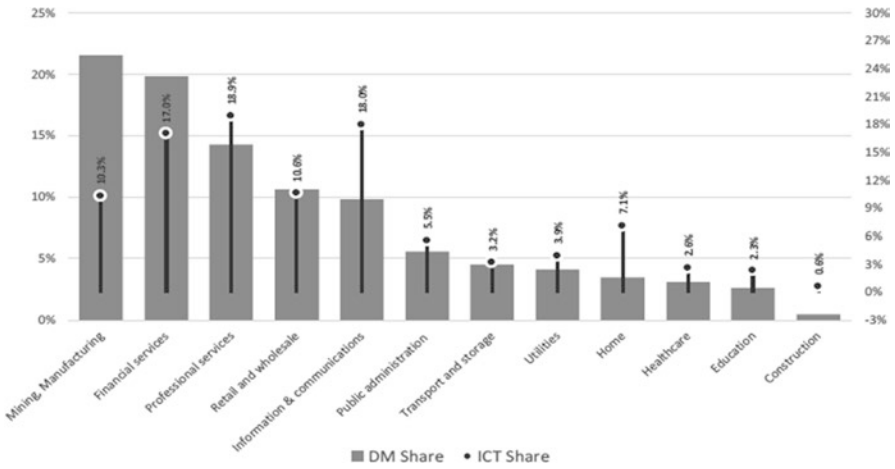

Fig. 7.6 Market size and ICT spending per sector. (Source: IDC (2017))

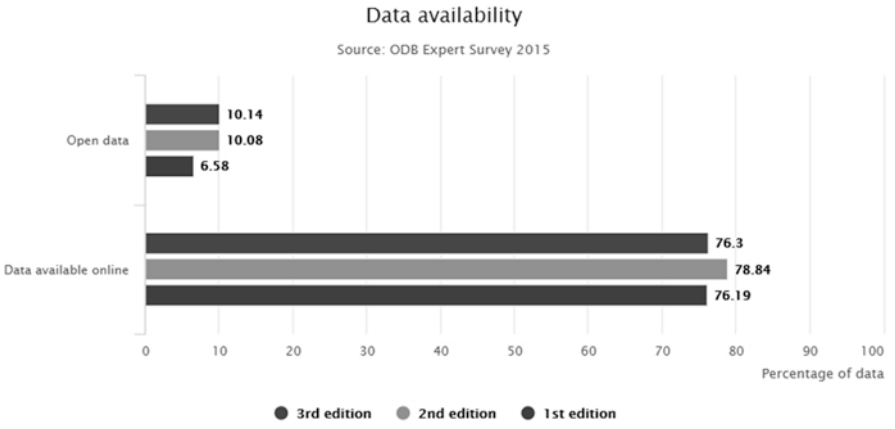

Fig. 7.7 Evolution of the availability of online data and open data. (Source: ODB (2016))

Finally, a strategic aspect to consider in the exploitation of open data as a key ingredient of a company product or service offering, is the identification of possible sources of competitive advantage necessary to consistently generate a performance that is superior to that of the other actors operating in the same competitive arena. When released in a fully open and reusable format, information may duly be considered a public good characterized by non-rivalry and non-excludability in consumption. As a consequence, the access to this type of resource may not be considered in itself a source of competitive advantage. Figure 7.9 shows how the focus in effort allocation shifts as a function of the degree of openness of the data sets exploited. In a situation in which legal, technological and price barriers are present, the company willing to exploit a given data set is required to spend significant resources in the process of data acquisition (especially for what concerns tech-

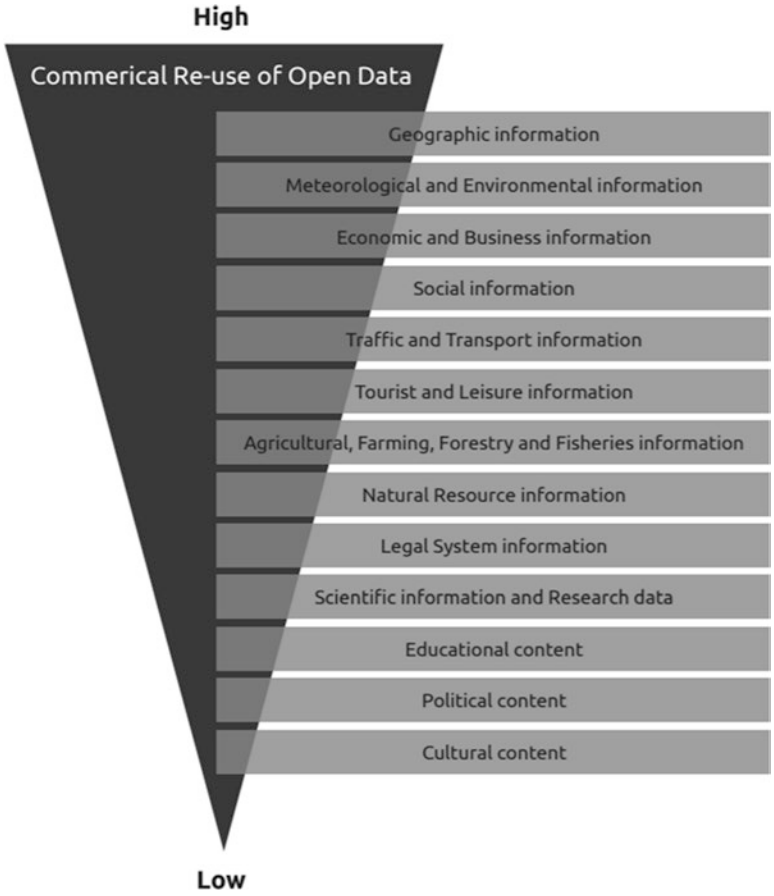

Fig. 7.8 Commercial reuse of open data. (Source: Capgemini (2015))

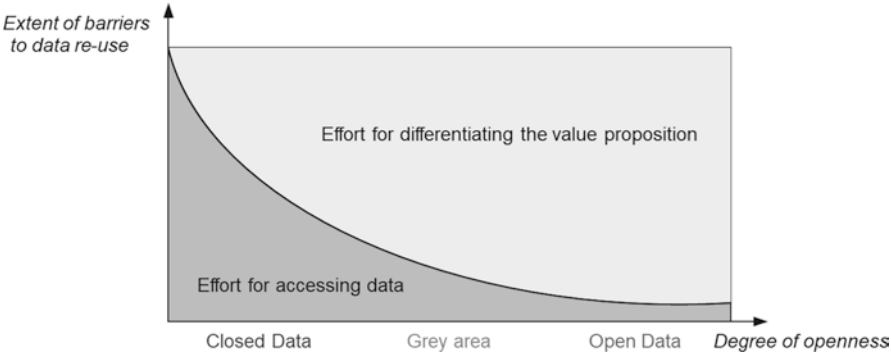

Fig. 7.9 Effort allocation as a function of data openness. (Source: Ferro and Osella (2011))

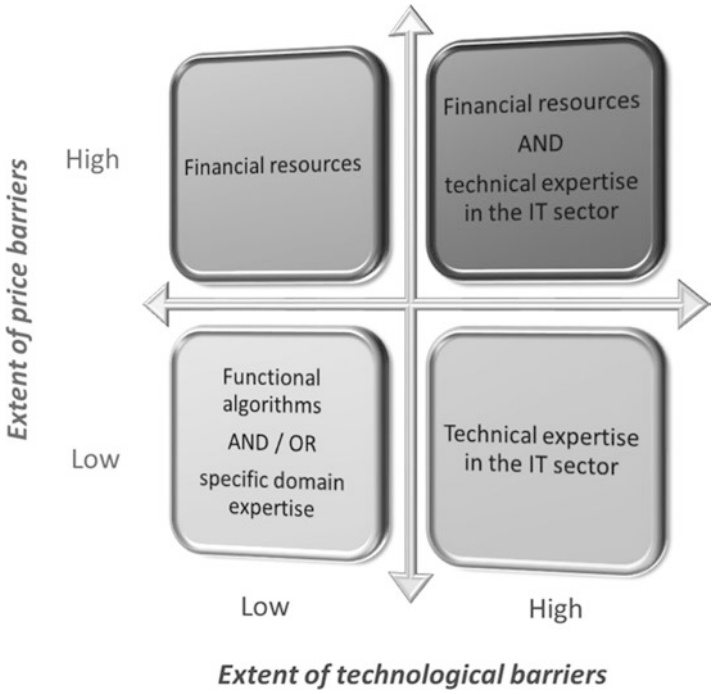

Fig. 7.10 Barriers and sources of competitive advantage. (Source: Ferro and Osella (2011))

nological and price barriers, as legal barrier may not be overcome). As the barriers to data re-use diminish, the focus of the company efforts moves from the process of data acquisition to the differentiation of its value proposition with respect to the competitors who, due to lower barriers to entry, increase in terms of numerosity.

The matrix depicted in Fig. 7.10 further clarifies the potential sources of competitive advantage that a company may exploit based on the presence and extent of price and technological barriers. When price barriers are significant and technological obstacles are negligible the availability of financial resources become the primary competitive edge discriminating between who can afford to access the information asset and who cannot. When, instead, technological barriers dominate over price barriers, technological skills become a must have to excel in the process of data acquisition, harmonization and integration. In contexts in which both type of barriers are present, the presence of both ingent financial resources and robust technological competences is required. Finally, when both price and technological barriers are not present or negligible, it is interesting to note that the sources of competitive advantage are no longer connected to the process of data acquisition, but rather they are related to functional algorithms for the treatment of data as well as to the presence of domain-specific expertise. While the former play a horizontal role and allow to differentiate the application logic of the service provided, the latter allow to contextualize the offering within a given vertical market.

In the final part of this section a use case will be presented and discussed in order to allow the reader to contextualize the knowledge and concepts presented in the previous sections into a practical and real-life example. More specifically, we will draw from and elaborate on the Open Corporates case study conducted by Becky Hogge (2016).

In 2010 the World Bank published a report showing that of 213 grand corruption investigations across 80 countries, 150 involved corporate vehicles that shielded the true beneficiaries of financial transactions. In these 150 cases, the total proceeds of corruption amounted to approximately $56.4 billion (Van de Does de Willebois, Halter, Harrison, Park, & Sharman, 2011). Open Corporates is the largest open database of companies in the world. It launched at the end of 2010 covering 3.8 million UK past and present companies. As founder told the Open Data Institute in 2012: "we take messy data from government websites, company registers, official filings and data released under the Freedom of Information Act, clean it up and using clever code make it available to people". The launch of Open Corporates predates the decision by Companies House to release all the data it holds as open data. But Companies House has made more basic datasets available for several years, and it was this data, combined with other government data sources (for example government spending data and Health and Safety notices) that fuelled Open Corporates in the beginning. Taking the same mixed input approach, Open Corporates has now expanded its coverage to over 105 jurisdictions and 85 million companies.

The added value that Open Corporates brings is the very detailed knowledge of how their database works. In addition, Open Corporates did "data-based advocacy", UK Department for Business were consulting on whether directors' and shareholders' full dates of birth should be published on the register, Open Corporates was able to demonstrate through real data that were dates of birth to be partially redacted, investigators would be unable to identify individual directors and shareholders robustly in cases numbering in the tens of thousands. OpenCorporates was also instrumental in pushing NGOs to demand the registry be made publicly available.

Open Corporates represents a very interesting case study in our discussion for a number of reasons: firstly, the business model they are implementing falls under the "open source-like" category identified by Ferro and Osella (2013) according to which the costs incurred for free offering of data are covered by revenues stemming from supplementary business lines that are still open-data-based. In this respect, considering that the whole Open Corporates database is freely available online and covered by an open license, the source of competitive advantage that the company may leverage to maintain its economic sustainability comes from a deep and detailed knowledge of the data base as well as of the domain. The second aspect of interest has to do with the fact that Open Corporates, not only acts as a open data advocate in the country in which they operate, but it helps breaking the silos present among public agencies working in countries both within and outside the European Union. Finally, Open Corporates may represent the dawn of a new paradigm in the pricing of data assets. More specifically, data released with an open license requiring any user to release derivative products in the same manner, may create the space for a new pricing logic that could require third parties to pay to maintain closed information assets

generated by combining both closed and open data sources. This represents an inversion with respect to traditional pricing logics aimed at opening the access to information assets that could build on the diffusion of "open-by-default" as a mainstream approach as well as the diffusion of distributed ledger technologies like blockchain as an instrument to further promote transparency in the treatment of data.

7.5 Open Data Exploitation in the Public Sector

Shifting now the perspective from private sector actors to public agencies, this section intends to provide two contributions. The first has to do with the creation of a fully engaged and sustainable supply side, the second has to do with the investigation of the benefits that the public sector may enjoy as a savvier re-user of open data.

Despite the efforts put in place by an international and a highly motivated community of open data advocates operating from both within and outside the public sector, the "open-by-default" approach to date is still struggling to become a widespread practice and to generate the expected impact on the European socio-economic system. For this reason, there is an urgent need to take a new perspective on the topic in order to put cities, companies and citizens in the position to benefit from the significant, yet untapped, value residing in public sector's data vaults. More specifically, it is important to acknowledge the self-interested nature of human behavior by focusing on the benefits that public administrators may gain as stewards of government data vaults while viewing current drivers as significant, yet second order, positive externalities. Drawing on the principle that a thriving open data ecosystem requires the attainment of sustainability from the demand as well as from the supply side, the perspective proposed endorses governments' ROI as yardstick for gauging the ultimate feasibility of open data programs.

As a result, a new open data paradigm entails a radical shift in the way civil servants look at open data. This wave of change may be summarized as follows:

- From legal obligation to operational necessity
- From outward orientation to inward orientation
- From cost to opportunity
- From clerical function to strategic function
- From requiring a leap of faith to generating evidence-based impact

At an operational level, the implementation of such paradigm requires to rid of the "data liberation" approach in favor of an "open-by-design" principle allowing data to be born open through a revision of their generation process. This would represent a valuable tool in facing the challenges posed by a steadily growing pressure on public budgets. In addition it could contribute to make a further step towards the obtainment of an outcome-based government whose actions demonstrate a clear link with the results generated (i.e., outcomes) in terms of value that, in turn, could be internalized by the governments (e.g., efficiency, effectiveness) without overlooking the quest for the creation of value for society at large ("public value"). The adoption

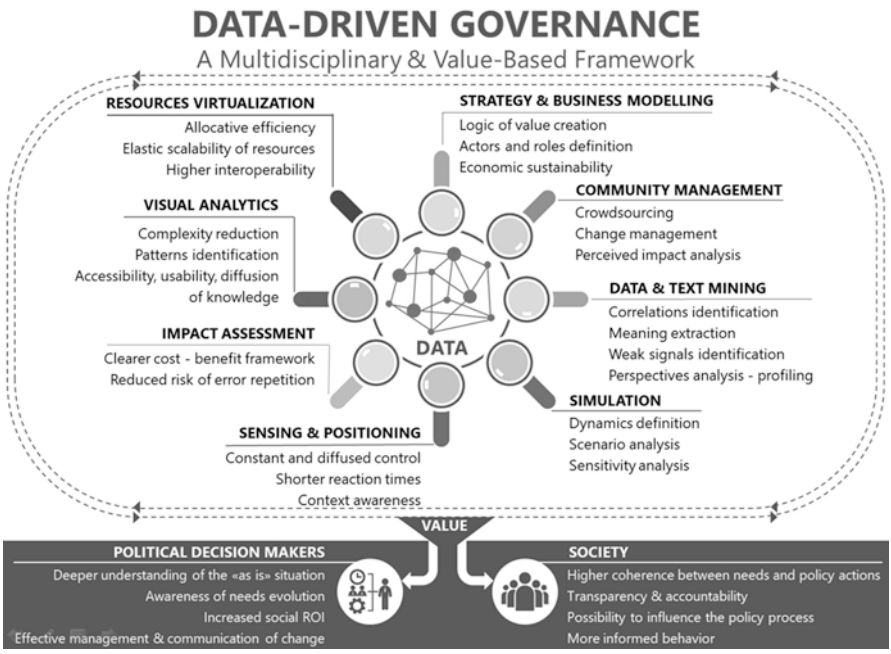

Fig. 7.11 Data-driven governance

of such an approach could represent a foundational step in the path leading to a data-driven governance paradigm briefly outlined in Fig. 7.11.

Placing data at the center of the governance process and combining it with a plurality of skills drawn from multiple knowledge domains represent the key ingredients for significantly improve the opportunities for value creation of a public decision maker. As a matter of fact, a data-driven multidisciplinary and value-oriented modus operandi may greatly benefit both decision makers and society at large. The former may gain a deeper understanding of the "as is" situation over which a given policy should be implemented to obtain a desired outcome, increase her awareness of evolution of needs to address, manage and communicate change more effectively and ultimately, increase the social ROI of any public investment. The latter, instead, may enjoy a higher level of alignment between perceived needs and policy responses, be more informed and incentivized to engage in the public debate thanks to higher levels of transparency and accountability. The creation of such virtuous cycle is believed to lead to a more effective and efficient allocation of taxpayers' money representing a key goal in times of shrinking public budgets.

To exemplify the benefits that the implementation of this approach may bring in terms of generation of value for society, a brief description of a use case conducted by OECD (2016) on the city of San Francisco is reported below. In the city of San Francisco, the heads of the foster care, juvenile probation and mental health departments, crafted an agreement with the city's attorney to permit the limited exchange of case information among agencies. The sharing enabled a new level of care for

children interacting with any of these agencies. Case coordination improved, invisible populations emerged (overlapping clientele). This was made possible by the fact that the new integrated data system recognizes and focuses on the families that are most vulnerable, most troubled and most in need. Prior to data integration and data analysis the agencies had not realised that only 2000 users of services were using half of the resources of the department, and most of these families lived within walking distance.

As a follow up, the Human Service Agency concentrated delivery of services in specific neighbourhoods and co-located services at community centres, and this improved efficiency. Results included savings and better service delivery. Analysis of open linked data enabled a better assessment of needs of high risk youngsters diverting them from negative future events, the understanding of where youth were falling through, identification of what services were needed to intervene earlier and prevent negative outcomes. Initially supported by a low-tech system the solution was transferred to a more sophisticated platform to enable the three agencies to better understand the overlaps among their users. The crossover users of multiple systems were at higher risk of committing a crime (51% of San Franciscans involved in multiple systems were convicted of a serious crime, 1/3 had been served by the three agencies and 88% of these youths committed a crime 90 days after having become a crossover user – a critical window of opportunity for the case worker to intervene). A report produced highlighted a specific need: a web-based integrated case management system to make this connection in real time.

As services started being delivered by non-institutional care providers, the awareness grew of the need to balance the right of excellent care with the right to privacy protection. Hence, the need to carefully avoid sharing unneeded information. What made it so difficult where legal related matters. The preliminary good results convinced the district attorney's office that the integrated database could support better prevention services and gave the authorisation through a new statute that justifies the sharing of records on youth at particularly elevated risk levels. The school district decided to join to target students with high probability of dropping out to structure early intervention. Multi perspective on client's risk and identify protective factors. This can help agencies to determine which programmes are more effective, who needs to be targeted (most vulnerable, in trouble and in need) and how to coordinate the responsibilities. The San Francisco case study represents an excellent example of how a smarter exploitation of data by public agencies may lead to significant increases in performance.

7.6 Conclusions

The re-use of open data is believed to contribute to the world improvement for its potential to empower citizens, businesses, change how government performs, and improve the delivery of public services (Zeleti, Ojo, & Curry, 2014). The aim of the present chapter was to go beyond the glorification of the opportunities lying behind

open data exploitation by exploring potential strategic viable choices from both a private and a public-sector perspective. Despite still being a phenomenon in its initial stages, the literature studying applicable business models to open data ventures offers some preliminary guidelines about possible strategic avenues that may be pursued in the design and implementation of potentially successful businesses leveraging open data. A portfolio of business models has been compiled as a toolkit from which would-be entrepreneurs or managers operating in established organizations may draw inspiration in the process of giving light to new companies or business lines. A reflection was also offered on the potential sources of competitive advantage may leverage in crafting their competitive strategy. As the barriers to data access decrease, it is possible to note a shift in the sources of competitive advantage for an organization. More specifically, the availability of financial resources and technical skills to be leveraged in the process of data acquisition becomes less relevant, while the presence of sophisticated functional algorithms and domain specific knowledge gains importance in the process of data elaboration and value extraction.

Shifting to a government perspective, a new approach to open data conceptualization and management in the public sector was proposed as a key complementary activity for the creation of flourishing open data ecosystem in which government agencies in addition to becoming reliable and efficient providers of quality data sets, become their first beneficiaries thus enabling a process of data-driven governance with significant positive spillovers for both policy makers and society at large.

Finally, to conclude the chapter, five synoptic principles are suggested to guide both public and private sector actors in a more purposeful valorisation of data assets. The principles are briefly described below:

- Size is not synonymous of value. That is to say, the assessment of data value should be based on a plurality of criteria: relevance for decision making, quality, and availability over time to name a few.
- Data science skills and the development of an evidence-based culture represents a key complementary ingredient to technological investments.
- Openness is a key driver of value multiplication. In other words, data should be released in formats maximizing the opportunities for the generation of economies of scope.
- Move beyond retrofitting. Rather than liberating data ex-post, the processes of data generation have to be open by design in order to minimize the cost of making them available to relevant stakeholders.
- Shared and clear values. The exploitation of data should be driven by shared values clearly identifying priorities in terms of advancing the environmental, social and economic conditions of the city.

The adoption of the above principles in the application of a long-term approach to data generation, exploitation and management may represents the necessary foundations to turn open data exploitation from a niche activity to a mainstream phenomenon as well as to make sure that the innovations contribute to the generated a positive impact on society in the quest towards the construction of a more sustainable and equitable world.