

Article

The Unexploitable Smartness of Open Data

Grazia Concilio * and Francesco Molinari

DAStU, Politecnico di Milano, Via Bonardi 3, 20133 Milano, Italy; mail@francescomolinari.it

* Correspondence: grazia.concilio@polimi.it; Tel.: +39-02-23995473

Abstract: The paper identifies a contradiction between data openness and economic value, possibly hiding a ‘market failure’ requiring a more active intervention from the public hand. Though the sheer quantity of data available for free usage is steadily increasing worldwide, its average quality usually stays well below the minimum threshold required for value creation. In contrast, there is now growing evidence that the use of data has enormous potential for the economy and society, including research and the progress of science. Unfortunately, useful datasets are usually locked in and when actually made accessible, suffer from the same limitations mentioned before. Maybe the time is ripe to undervalue the generalized disclosure of government data in favor of an appropriately incentivized and targeted creation of actionable bases of new IT applications. We present four cases touching upon the issues and potentials of service design, urban innovation, and data-related policies. We identify two possible ways of tackling the highlighted market failure: direct subsidies to government bodies or agencies engaged in disclosing their own datasets and keeping them clean and accessible over time or new regulations that establish more productive data ecosystems, rewarding knowledge creation rather than mere data ownership.

Keywords: open data; data smartness; open government; data policy; data governance



Citation: Concilio, G.; Molinari, F. The Unexploitable Smartness of Open Data. *Sustainability* **2021**, *13*, 8239. <https://doi.org/10.3390/su13158239>

Academic Editors: Michele Grimaldi and Martin De Jong

Received: 18 April 2021
Accepted: 19 July 2021
Published: 23 July 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Started in the late 2000s, the so-called open data revolution has never fulfilled its ambition [1] to create a diffused supply of data-based IT public and private services, as most observers and critics are now willing to recognize [2–4]. Though the sheer quantity of data available for free usage is steadily increasing all around the world, its average quality (however measured) usually stays well below the minimum threshold compatible with value creation. By data quality we mean an acceptable level of compliance with the FAIR principles (findability, accessibility, interoperability, and reusability) [5] in order for data to be exploitable. The most diffused and shared issues and criticisms are well summarized in the following citation from the 2017 edition of the open data barometer: “Government data is usually incomplete, out of date, of low quality, and fragmented. In most cases, open data catalogues or portals are manually fed as the result of informal data management approaches. Procedures, timelines, and responsibilities are frequently unclear among government institutions tasked with this work. This makes the overall open data management and publication approach weak and prone to multiple errors” ([2], p. 14).

In contrast, there is now robust evidence that data-based IT services have enormous potential for the economy and society, including research and the progress of science. The size of the European data market for display and programmatic advertising was estimated at USD 3.2 billion in 2013 and growing at a double-digit rate, to reach USD 4.2 billion in 2019 [6]. Three times bigger is the dimension of the (western) European big data and analytics market, including artificial intelligence and machine learning technologies, which hit USD 14.1 billion in 2017 and is expected to grow with a five-year CAGR of 9.2% [7]. An order of magnitude higher is the global size of the European data market, also including system infrastructure software, IT hardware, and services, which was estimated at around

EUR 54.4 billion in 2015 [8] and projected to reach EUR 106 million by 2020, or 4% of the EU GDP [9].

Unfortunately, the current data market seems to show that the value of data essentially lies in its being ‘locked’: sometimes patented, like the human genome sequence, quite often anonymized, also for GDPR (General Data Protection Regulation)-related reasons, but always restricted in access, circulation, and reuse by unauthorized (i.e., non-paying) third parties. We therefore face a paradox: while everybody would subscribe to the argument that data are getting more and more prominence as an engine of growth and jobs—up to the point of speaking about the emergence of a “data driven economy” as the next step beyond the more familiar “knowledge based economy” [10]—the prevailing value generation mechanism is the exact opposite of what would be expected from a proper knowledge revolution: data seem to be less valued when they are shared and public than if private and restricted.

Take research as an example. Compared with 2006, the share of papers published in the American Economic Review (AER) that obtained an exemption from the AER’s data availability policy grew by more than 500% in 2014, showing a clear change of mindset towards the eternal dilemma between protection of discovery and replicability of results [11]. The growing (and well deserved) attention of global publishers to the never reproached enough practice of self-plagiarism by scientific writers has certainly been doing justice of some abuses and misconducts, but is also sending the contradictory message that only after a research work has been definitely elaborated and finalized by its author can it be properly shared and thus contribute to furthering the scientific or technological frontier in a given domain.

The public sector makes no exception to this rule: according to the 2017 Open Data Barometer, 9 out of 10 government datasets are not open. “We assessed 1725 datasets from 15 different sectors across 115 countries [and] found that only 7% of the data is fully open, only 1 of every 2 datasets is machine readable and only 1 in 4 datasets has an open license” ([2], p. 12). Anyone, including the authors of this paper, can easily think of examples of public sector organizations refusing to disclose certain datasets they own, based on the allegation that this would dissipate, rather than increase, their embedded value [12].

In our opinion, this paradox configures itself as a failure of one of the main dimensions of ‘data smartness’ (meaning the capacity of gathered and processed data to create useful insights for new smart solutions), particularly in, but not limited to, the experience of our cities. To summarize a number of related arguments set forth in research and practice, it is the meaning—or semantics—associated to existing datasets, not their sheer availability, that turns them into ‘the new oil’, i.e., an exploitable source of diffused economic and societal value. As the analogy with oil implies, there may be some wells that are still too deep to dig at a sustainable cost, or so expensive to keep in operation that their upkeep may not be considered a priority. Likewise, adding more and more data to the public domain is not necessarily the best way to stimulate the market to generate new data-based applications, as the costs of doing so may prove unsustainable compared with effective returns.

This brings us back to our initial argument, that openness and exploitability do not seem to go hand in hand. In fact, not only has public policy globally stayed below expectations in achieving the goal of ‘servitizing’ open data, i.e., promoting its wide take-up and usability in a growing number of (collective) service oriented IT applications; but in a more subtle way, data ‘commoditization’, i.e., its transformation into a tradable good at a price reflecting its value potential, has resulted in its further lock-up, rather than opening up, contrary to the original plans of the authors of this policy revolution.

More than 12 years after the first appearance of the celebrated 5-star deployment scheme for open data [13], we consider the scenario presented in this Introduction to be quite consolidated, resilient enough not to be influenced by a mere replication of current policy approaches. Therefore, in this paper, we propose to reverse the viewpoint and treat the ‘un-exploitable smartness of open data’ as a perfect example of ‘market failure’, in the

sense well known to economists: “the failure of a more or less idealized system of price-market institutions to sustain ‘desirable’ activities or to inhibit ‘undesirable’ activities” [14]. When such is the case, only a purposeful policy shift from non-intervention to intervention is likely to alter the framework conditions that create the observed impasse: just think of the classical example of public financial incentives to private R&D, lowering the threshold of convenience for this specific kind of investment. (We elaborate on this example in the next section.) Our thesis is that maybe the time has come to question the generalized and ‘ethically’ or ‘politically’ inspired promotion of government data disclosure undertaken so far in favor of a new and probably more pragmatic vision that considers open data as one of the many infrastructural investments that are normally attributed to a well-functioning government. Then, if a clear return on that investment exists and can justify the expenditure, this is acceptable. Otherwise, new incentives (or rules of behavior) are needed to make ‘the business case for open data’ far more similar to that of ‘big data’.

One caveat is in order at this point: despite using such expressions as ‘market failure’, ‘return on investment’, or ‘business case’, our proposal is not to obliterate the differences between the public and private sectors nor to forget about the vast literature on the so-called ‘barriers to government data disclosure’, which mostly delve into psychosocial—and thus emotional—rather than organizational—and therefore rational—factors (see e.g., [15]). (We develop this argument further in the conclusions.) However, the nesting of psychosocial and organizational with economic, financial, and maybe also legal aspects in the ‘black box’ of public administrations only makes things more complex to configure but does not weaken per se the necessity, and anticipated usefulness, of a different policy intervention.

The remainder of this article goes as follows: Section 2 describes the concept of market failure and the paradox of unexploitable data smartness in more detail; Section 3 briefly presents four inspiring cases, from two European cities (London and Barcelona) and two popular applications (Foursquare and Ofo); finally, Section 4 points at two areas where legislation and/or active policy making can contribute to improving the current scenario in the desired direction.

2. Open Data Release Policies as an Example of Market Failure

According to consensus in the economic literature, there are four possible sources of market failure: public goods, market control, externalities, and imperfect information. Public goods are those for which non-payers, also known as free riders [16], cannot be excluded from consumption, implying that no price equilibrium can materialize for voluntary market exchanges. Market control occurs because of too limited competition among buyers or sellers, which determines a situation where the price (due to monopoly or oligopoly) or the quantity (due to monopsony or oligopsony) of a certain good is determined rigidly instead of flexibly. As a matter of fact, the abuse of dominant positions has been explicitly incorporated in the competition legislation of various countries, including the European Union’s member states. Externalities (which can be either positive or negative) refer to the generation of extra benefits or losses that are not fully incorporated in the market price of a certain good. In fiscal policy, negative externalities are usually invoked as rationale for the introduction of tariffs and levies. Finally, imperfect information, or information asymmetries between buyers and sellers, make it such that the price paid or the quantity demanded for a certain good may be lower or higher than its ‘real’ value or amount—for example, because consumers do not perceive that value in full or due to advertising that creates an extra demand for that specific product or service.

Looking into the current data market, we can see that several of these sources of failure are indeed at work. For example, considering open data as a public good that should be freely accessible without restriction, as has been the main tenet of promotion policies so far, brings with it the involuntary consequence of zeroing the market price of the disclosed dataset, and the activity of maintaining and upgrading that dataset becomes a sheer cost for the public sector organization involved. Could this be the true motivation, in these times of government budget crises, for the reluctance and/or lack of continuity in public

data disclosure policies and strategies? We suspect this might be the case, although it is impossible to document in the absence of empirical evidence.

Another source of failure is market control. Evidently, this is operating in the case of data-based IT services, for instance in the advertising domain. Here, the lesson learnt is that the owner of the data can to some extent determine the sale price for those services, but if a given dataset is not in possession of a single player, market competition can have the power, or at least the potential, to keep prices low enough. We take this as an argument in favor of a more extended, or 'plural', but still selective release of public datasets, as distinct from approaches for no disclosure or fully opening up.

As concerns externalities, Duch-Brown et al. [17] examine three distinct types, related respectively to economies of scope in data analytics, the strategic behavior of data owners, and transaction costs in data exchanges. The first type of externality is well represented by the case of adding related data on a certain phenomenon, which is already (at least to some extent) described by a pre-existing dataset. One can predict that, possibly with diminishing returns, the investment in expanding the scope of raw data sources would be more than offset by the savings in time and cost derived from implementing e.g., computing or machine learning algorithms only once on the global dataset, instead of doing this repeatedly on separate ones. The second type is exemplified by the case of fragmented property of some related datasets, where the previous savings cannot be achieved since each owner uses their exclusive rights to prevent the others from realizing the potential benefits of data aggregation. The third type of externality is associated to the exercise of rights from data owners trying to maximize the benefits for themselves during data exchanges.

Finally, imperfect information characterizes virtually every market [18], including that of data. A policy area where the impacts of imperfect or asymmetric information have been deeply explored and analyzed is public co-financing of private R&D and innovation investments [19]. Put very simply, to the extent that every innovation starts with an act of creative invention, it is extremely likely that the inventor has better information regarding the nature and likelihood of success of the proposed idea than any investor (including the firm owner, if the innovation materializes in-house) asked to put money on it. This leads to an underestimation of the rate of return of long term, high risk R&D projects compared with short term, low risk ones [20]. Reducing information asymmetry via fuller disclosure of the innovation contents would be of limited efficacy, due to the ease of imitation of the underlying ideas. In addition to that, the high intensity of human work in R&D and innovation projects makes it such that the risk of (voluntary or involuntary) spill overs by, e.g., laid off staff is often a major deterrent from engaging the firm in complex, multiannual R&D projects, not to speak of changing market conditions during project execution. Indeed, we know that most firms, especially if already successful in their markets, would rather not engage in further R&D and innovation activities were it not for the existence of public co-financing [21].

This discourse on the efficacy of financial (or tax) incentives to promote R&D and innovation is helpful to document, by way of analogy, the likelihood of a similar policy approach aiming to act as a game changer in the data market. In fact, as the following Table 1 shows, the comparison between R&D and innovation project outputs on the one hand and data on the other hand, as the 'fuel' of a new economic surge, can be furthered to a much deeper level, showing the similarities but also some differences between the two.

As can be noted from the Table 1, the two major differences in this comparison between data on the one hand and R&D and innovation outputs on the other, are that the former cannot be patented and only the investments in the latter are currently co-financed by the public hand.

Table 1. Data vs. R&D&I outputs (Source: authors' elaboration of a suggestion coming from OECD [22]).

	Data	R&D&I Outputs
Intangible asset, resulting from the creative application of human knowledge and purposeful behavior	Yes	Yes
Can be patented	No	Yes
Is subject to access restriction measures for legitimate reasons	Yes	Yes
Can give a competitive advantage to sole owner or exclusive user	Yes	Yes
Is subject to spillover effects and positive externalities across the economy and society	Yes	Yes
Related investments are co-financed by the public sector when the market fails to incentivize them	No	Yes

Regarding the impossibility of patenting, some clarifications are needed. In fact, one could mention copyright law or even more generally the protection of intellectual property rights, as a surrogate of patenting for the commercial rights of data owners. Evidence contradicts this assertion. For example, Duch-Brown et al. [17] mention the European Commission's own evaluation of the Database directive (96/9/EC of 11 March 1996)—introducing exclusive rights in the production of electronic datasets—which came to the conclusion that no significant gap between the EU and the US or other countries had been filled after the approval of that directive. This may be due to the fact that the real source of market value is not the single data point or dataset, but the use that is made of it, e.g., to create new IT applications or services. That is probably also the reason why EU legislation, including, more recently and for different purposes, the GDPR, has moved away from the protection of legal ownership and aims to regulate access and trade instead.

We conclude that, in light of the previous discussion and due to the close similarities between data and R&D and innovation outputs, the need for an active policy intervention is even more compelling than one may think of, and should aim to establish a new system of incentives bringing a different equilibrium between data ownership, access, and value.

Such a policy intervention might take the form of a direct subsidy to governments to disclose and maintain their own datasets, covering the expenses related to offering them to the market as public goods, including the cost of their periodic maintenance and updating and the provision of appropriate metadata. Otherwise, it might involve new laws or regulations, defining use licenses that protect data mashups as well as data and text mining in a more indulgent manner towards the creative work of users rather than the owners of original datasets [23]. In that respect, the EU directive on Copyright (2019/790 of 11 April 2019) does not constitute a viable remedy to the proliferation of open data licenses that impede the seamless reuse of public sector information. Another possible intervention could be taxation, often recommended in the case of externalities (think of environmental legislation). This might be a solution for the (not too infrequent) case of privately-owned datasets that are neither shared nor exploited for their value generation potential.

3. From Theory to Practice: Exemplary Case Studies from EU Cities

Up to this point, our discourse has been eminently theoretical. In this section, we corroborate our thesis with some concrete examples, taken from our experience of EU cities and service providers. The selected cases are neither exhaustive nor comparable, but were chosen as examples to (1) illustrate the conditions making the value of open data fully exploitable (London and Barcelona) and (2) provide evidence of the distortive power of the data market (Foursquare and Ofo).

We start with a virtuous example of open data release, which is also supported by apparently uncriticized or non-controversial analyses: the case of TfL, the Transport for

London company owned by the City of London. In a 2017 report by Deloitte, made available on a dedicated TfL page, it is stated with pride that the release of open data by TfL generated annual economic benefits and savings of up to GBP 130 million a year [24]. The number of apps that are specifically powered by TfL's open data feeds exceeds 600, used by more than 4 million travelers. Since 2007 and until today, TfL has consistently followed an open data release policy that delivers more than 200 data feeds to over 14,000 registered developers through a free, unified API, ensuring that accurate, near real-time data are available. Major data types include rail, tram, bus, tube, and river boat arrival/departure times, the status of cycle hire docking stations, air quality, accessibility, and toilets information, as well as a journey planner [25].

However, if we look deeper into the business case, a more nuanced scenario materializes. First of all, it all started towards the end of the 2000s, when TfL found out that some developers were scraping information about its services from its website [26]. This is consistent with international evidence pointing at mobility and transport as a fertile environment for public data disclosure and its exploitation. A key decision taken at that time, which might not have been driven by budget considerations only, was to avoid the creation of an internal development team dedicated to the production of mobile apps using own data feeds, but to follow the opposite approach of making the raw data openly available to IT developers. Indeed, such a decision is not shared by a good number of comparably sized transport companies, such as RATP in France, DB in Germany, or ATM in Italy.

The budget impact of that decision must have been remarkable, however, if one considers that on the occasion of the 2015 Cycle hire scheme initiative, an app was realized in-house by TfL at the cost of GBP 118,898.06 for a number of downloads as high as 29,139 according to official estimates. To arrive at a comparable ratio with the millions of downloads recorded by the hundreds of private apps built with TfL data, an immense budget would have had to be set aside, with no guarantee of success [27]. The huge number of daily users of TfL services might also have played a role in supporting the decision, considering that the company is in fairly good economic and financial condition. However, no information is available on the organizational costs of setting up an open data release policy, based on the maintenance of real time data sources at the free disposal of registered users on the TfL portal.

Another related, also crucial decision was the development of a single, standardized API. This has evolved into a separate platform and company, TransportAPI, which redistributes data from a variety of official UK sources after validation and standardization. Service prices start from nil (a free personal account under the CC-BY license with daily limits on open access) to monthly charges between GBP 30 and GBP 500 + VAT depending on the number of hits. Data include journey planning, non-London live bus schedules, taxi pricing, and cycle routing information.

Somewhat the opposite to the London case is Barcelona. Starting in May 2015, with the election of a new Mayor, the city first developed and then gradually implemented its digitalization plan 2017–2020, entitled “Barcelona Digital City: a roadmap towards technological sovereignty” (<https://ajuntament.barcelona.cat/digital/en>; accessed on 30 April 2020). In coherence with a vision of Barcelona as an “open, fair, circular, and democratic city”, the Plan puts special emphasis on the promotion of citizens’ data ownership up to the design of a new policy framework entitled “Barcelona Data Commons” (<https://ajuntament.barcelona.cat/digital/en/digital-transformation/city-data-commons>; accessed on 30 April 2020). This is defined as “a shared resource that enables citizens to contribute, access, and use the data—for instance, about air quality, mobility, or health—as a common good, without intellectual property rights restrictions” (p. 1). As noted by Francesca Bria, then among the leading figures of that policy [28], “Barcelona is experimenting with socializing data in order to promote new cooperative approaches to solving common urban problems: tracking noise levels and improving air quality, to take

just two examples. This data is collected via sensors operated by citizens with the City taking the lead in aggregating and acting upon such data”.

Seen from a government perspective, this approach could not be more different than the one adopted by TfL in London: there, data are internally sourced, while here, they mostly come from the external world; there, the dilemma (to some extent) is between exploiting ‘data smartness’ directly or via the involvement of professional third parties, while here, it is not even clear what raw data actually exist, as their collection is planned and organized by the people, sensors, and devices that ultimately produce it.

The financial budget allocated by the City of Barcelona for the realization of its digitalization plan was huge: according to some estimates, EUR 85 million only in the period 2017–2019 [29]. Presumably, societal value was largely expected to exceed the economic one: this is witnessed by the reference made to ‘Data as the new Commons’, which, although not new [30], summarizes the policy goal of developing a strong data culture in the city, giving to people better control of the data generated both in their homes and the community at large and encouraging their participation in rule setting for, e.g., who can access it, for what purposes, and on which terms. This is seen as a bridge towards the creation of a new wave of social rights in the 21st century.

However, as ambitious and pathbreaking as this can be, the long run sustainability assessment of the Barcelona vision totally relies on continuing allocations of public funding. This is of course coherent with the systemic and open data prone logic of the City and its Plan, but also implies putting all bets on the capacity of digital transformation to create a new wave of businesses and jobs, providing an indirect return—or even a direct one, e.g., through taxation—to the city’s economy and the government’s budget.

It is therefore of little surprise that other public initiatives emerging in the domain of open data have adopted a far more nuanced and less financially demanding approach to government data disclosure, based on, e.g., prioritizing the datasets to be opened up in relation to their potential use [31], minimizing the costs associated with data cleansing and maintenance through innovative public private partnerships [32], or building synergies with civil society, the private sector, and other public agencies interested or involved in data utilization for new service creation [33].

For instance, the Interreg Med project Odeon (<https://odeon.interreg-med.eu/>; assessed on 30 April 2020) aimed to set up an innovative Mediterranean-wide open data cluster, composed of small and medium enterprises, start-ups, and research institutions interested in developing the economic and societal value of government data; this was done with the support of local data hubs engaged in triggering governments’ propensity to disclose and their willingness to maintain quality open datasets while at the same time assisting SMEs and start-ups to develop disruptive products and services for the Mediterranean area’s sustainable growth and quality of life.

Conceptualizing a data hub or, more generally, making room for an intermediary profile between the supply and demand of public sector information is quite recurrent in recent policy experimentation, signaling a true need for new institutions that act in support or even substitution of market forces. Compare also the H2020 Open4Citizens project (<https://open4citizens.eu/>; accessed on 30 April 2020), which was aimed at integrating more non-IT-savvy citizens in the complex process of city service redesign based on open data, as epitomized by a wave of urban hackathons. After a successful round of such events in five European locations, including Barcelona, the project launched the idea of a pan-European network of open data labs, i.e., public or private entities playing a similar role to what fablabs do for the ‘maker economy’: widening access to innovation and promoting exemplary data-based IT applications.

Even with the active contribution of a matchmaking service, however, the risk is always there (and very well perceived by TfL) that the newly developed services simply add to the plethora of IT solutions that never reach, or become successful in, the market.

Take the example of the Foursquare app, now known as Foursquare City Guide. When it was launched in 2008, it was an innovative local search-and-discovery mobile app, which

popularized the concept of real-time location sharing and checking in. Indeed, Foursquare grew rapidly and acquired millions of users in only a few years. At that time, its business logic was fully adherent to making innovative use of open data (coming from, e.g., urban points of interest and commercial activities) alongside its site-specific social networking capacity. In fact, already in 2009, Foursquare opened up access to its APIs, thus enabling developers to access on the fly the data generated during use and build new applications on top of that. However, something important must have gone wrong, if after few years, the following statement could be read on their corporate web page: “Where matters. Put the most trusted, independent location data and technology platform to work for your business”. This clearly suggests that they were commercializing data! Users no longer stand out, not even at the periphery of the value proposition and communication campaign. Foursquare’s business model now relies 100% on the market value of the legacy data built up from previous use of the app: an informational treasure left as a gift by its original users.

If Foursquare modified its business model, we are today increasingly seeing the birth of services whose scope is not in the service itself but in the gathering of citizens’ data. Take the story of the Ofo mo-bike service, no longer in operation, yet interesting for the purpose of our discussion. The main stakeholder of the Ofo company was the biggest Asian e-vendor Alibaba. What did Alibaba have to do with improving sustainable mobility in cities? In fact, the idea of Ofo was not born at Alibaba but came from a couple of students at a Chinese university looking for alternative ways to take and leave rented bikes while moving across the city to get to and from the university. Indeed, Ofo was an innovative and very local idea whose potential, recognized and financed by Alibaba, was going beyond its immediate social and environmental benefits. Citizens using Ofo bikes were providing their own data for commercial use [34]. Curiously, the realization of this fact was not the main reason for the failure of Ofo, which was apparently more related to the poor quality of the bikes themselves and the consequent need for expensive maintenance.

To summarize this section, and having in mind that one of the first hackathon campaigns launched in 2011 by the Mayor of New York was looking for “good, useful, innovative ideas” to transform the (just released) City datasets into services, utilities, or apps for the benefit of NY residents (<https://mashable.com/2011/08/01/nyc-website-hackathon/?europe=true>; accessed on 30 April 2020), there are more and more signs around the world of market trends in the opposite direction, which finds value in private and personal datasets rather than in public or disclosed ones. This is just another piece of evidence of the ‘market failure’ described above, and what mostly affects our discussion is the fact that its justification lies with the reduced profitability (if not also exploitability) of open data when used for the creation of new IT services and applications.

4. Discussion and Outlook

On 24 January 2018 at the World Economic Forum in Davos, Angela Merkel announced that “data will be the raw material of the 21st century” and then added that “the question ‘Who owns that data?’ will decide whether democracy, the participatory social model, and economic prosperity can be combined” (as reported by Calzada [35], p. 2). That data driven business models represent a sound opportunity for new economic growth is already proven. What the statement by Angela Merkel and the currently active debate on data sovereignty [36] add to the picture is that the way we shape future policies on data ownership and use will also be crucial for putting more or less democracy and equity into the standard model of a data driven economy. Furthermore, the discussion on data ‘extractivism’ and its applications [37] shows that the smartness of open data and its market potential are indeed at risk, exactly because of the unresolved issue of ‘who owns the data’. If things stay unchanged and the major market players keep moving their business propositions towards valorizing data possession and control instead of the generation of know-how for data-driven innovation, the smartness of open data will remain unexploited—and worse still, unexploitable.

Indeed, some inspired observers [38] have even started to reverse the angle of investigation, from asking how the private sector can create value from the information held by governments, to highlighting how data in possession of the private sector can be conveniently used to address unsolved social issues of public interest (see, e.g., the narratives in http://datalandscape.eu/sites/default/files/report/Story_1_New_format.pdf; accessed on 30 April 2020). This also includes the emerging movement called Citizen Science, whereby data relating to the natural world are collected by members of the general public, typically as part of a collaborative project with professional scientists. For instance, the H2020 project iSCAPE (<http://www.iscapeproject.eu/about/>; accessed on 30 April 2020) developed and tested a structured process in support of a diffused adoption and implementation of the Smart Citizen Kit (<https://smartcitizen.me/>; accessed on 30 April 2020), a low cost, sensor-based solution for gathering air pollution data on a massive scale and in real time. The first version of the solution was crowd-funded by a Kickstarter campaign led by IAAC and the Barcelona FabLab in 2013. The second version was made available to the six iSCAPE pilot cities (Bologna, IT; Bottrop, DE; Dublin, IE; Guildford, UK; Hasselt, BE; and Vantaa, FI) in 2018 and used to deploy a massive air pollution data gathering campaign in each of them, with a global engagement of 88 citizen scientists collecting 892,050 sensor readings over a period of 15 months [39]. The third and validated version of the supporting Kit is now commercially available, with over 500 sold in 2019 only. However, there are few signs that this innovation has been gaining momentum and a widescale take-up in the European public sector.

To solve or at least tackle this impasse, a number of research questions—most of which are not at all new [40–43]—present themselves as worth exploring further, under a policy- or purpose-oriented perspective:

1. How dependent is the slow-paced open data release process on the reluctant attitude shown by many public officials in charge of coordinating it?
2. How critical in determining this outcome is the low pressure of IT companies not seeing a clear business interest in developing services enabled by open data of prevailing collective and public value?
3. How helpful might it be to move away from an excessive focus on data ownership towards the promotion and diffusion of a local (open) data culture, leading to innovative governance models that treat data as a common good?

To rebuild the foundations of a truly innovative data economy, the cases presented in this paper suggest incentivizing the creation and maintenance of open datasets, thus transforming a selected and predefined sample of the information owned by public administrations into a source of authentic (market or community) value.

There are two main options for tackling market failure: (a) direct subsidies to governments engaged in disclosing and maintaining their own datasets clean and accessible over time, or (b) new laws or regulations that impose the establishment of more productive data ecosystems, rewarding knowledge creation rather than mere data ownership.

The first option, while appropriate to compensate for the reluctant attitude of some public data owners, would nevertheless be hard to sell at face value, as subsidies could appear as a way of paying twice for the same service. However, and as Table 1 above implies, data IPR (Intellectual Property Rights) are more difficult to protect legally than those of R&D and innovation results, and a key reason for subsidizing data creation would be the fact that, in some cases at least, the resulting data disclosure would bring relevant value to the economy and society, and therefore to the public commons.

Another reflection is that even the (not too) recent emergence of citizen data and user-generated data portals or georeferenced applications—such as the MappiNa service [44], born as a free and open source map of the City of Naples' points of interest (<http://www.mappi-na.it/#/>; accessed on 30 April 2020) or the aforementioned Citizen Science project iSCAPE—has hardly been accompanied by the establishment of new and financially viable business models. This implies that public policy should reconsider and partially adjust the direction of financial support to date, away from the application level (tools and apps

fueled by data, especially open data) and back to the data generation level, not only by the public but also the private sector and especially by creative individuals. This would also contribute to solving one of the main detected problems of Open Government portals: in order for them to be economically viable and technically usable, data need curation and continuous updating, while the quality and timeliness of the datasets stored in government portals are usually found lacking [45].

A related reflection is that cities are ultimately the platforms where most data-based business is created and developed. This is because it is mainly within cities, for city governments, and about citizens, their actions, and their behavior in the urban community, that most user data are being collected today. Cities and communities offer the socio-economic justification and are the key enablers of the prospective data-based IT services appearing in this market, so local governments may well ask to be paid for the supporting role they play. This can happen in two possible ways: either in the form of commission fees collected against the permission to implement and supply services (and collect data) in the territories they govern, or through financial remuneration for the contribution they give to increased data availability. Indeed, there could also be a third way, namely, to expect that companies share raw data collected during service usage with the cities, making data more and more public and reusable. However, evidence does not point in this direction, as some of the examples commented in the previous section document well enough.

The reference to what cities and companies should do or not brings us smoothly to the alternative option, which is about passing laws or regulations helping to solve the contradiction between data openness and socioeconomic value creation. We are naturally prone to thinking that additional rules do not make any significant step forward when issues such as ‘data as commons’ come to the forefront. Here, the above reference to the new Copyright Directive is evidence to that regard.

However, there are two areas where legislation—and/or active policy initiatives—can contribute to improving the current scenario in the desired direction. Both are to some extent related to capacity building and education, two areas where, at any rate, new rules have no impact if unaccompanied by additional funding.

The first area is data governance. Public authorities as the main service suppliers (especially in urban environments) are indeed significant (big) data owners. The more they innovate and digitalize their services, the more can they find ways to have their say in the data market. This, however, has two implications in terms of capacity building in the public sector. The first is more strategic: the pathway to the digitalization of public services generally needs a clear vision of the related potential to generate innovation. This aspect has been noted, among others, by Molinari and Concilio [12] dealing with a case where service digitalization is implemented by a third party on behalf of a public authority. The second implication is more tactical, namely related to the need to overcome the psycho-social but also legal and organizational barriers obstructing the way data are internally produced and managed in the public sector. For example, still today, different offices and units do not frequently share data and information between and within each of them, so that, as the most visible outcome, too often citizens are asked to provide the same information to the government more than once. More importantly, to the extent that data production and management are streamlined and made more effective inside the public administration, data disclosure, cleansing, and maintenance would also be facilitated and become more appropriate.

The second area is data culture. As the Barcelona Digitalization Plan and the H2020 Open4Citizens and iSCAPE projects demonstrate, the search for innovative solutions to urban problems can represent a genuine driver for activating processes of data collection, exchange, and usage. These can be reinforced by policies supporting citizens’ data literacy and stimulating broad and intense debate at the community level on issues such as data sovereignty and ownership. Obviously, these initiatives are effective only in the long run, but the reverse is also true, namely, that a persistent lack of awareness on the value of data

and their relevance as a market resource plays a role in hampering the generation of greater and better exploitation opportunities from the various data used.

Author Contributions: Conceptualization, G.C. and F.M.; methodology, G.C. and F.M.; investigation, G.C. and F.M.; resources, G.C. and F.M.; writing—original draft preparation, G.C. and F.M.; writing—review and editing, G.C. and F.M.; funding acquisition, G.C. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by European Union’s Horizon 2020 research and innovation program under grant agreement No 769608, Polivisu project and reflects only the authors’ view. However, the present article is not the product of one specific study, it rather presents reflections matured throughout several research projects in the domain of open-government and data management and is aimed at discussing the necessity for more effective decisions by public authorities towards data disclosure in order to make its smartness more exploitable, contrasting the current trend that clearly shows greater commercial value assigned to locked than open data.

Acknowledgments: Authors are grateful to Nicola Morelli and Jesse Marsh for their support and suggestions while revising this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Vickery, G. Review of Recent Studies on PSI Re-Use and Related Market Developments. Available online: <https://ec.europa.eu/digital-single-market/en/news/review-recent-studies-psi-reuse-and-related-market-developments> (accessed on 30 April 2020).
- The World Wide Web Foundation. *Open Data Barometer*, 4th ed.; Global Report of the World Wide Web Foundation: Washington, DC, USA, May 2017. Available online: <https://opendatabarometer.org/4thedition/report/> (accessed on 30 April 2020).
- Batini, C. Dacity, Perché una Scienza per Studiare L’impatto dei Dati Sulla Società. Available online: <https://www.agendadigitale.eu/cittadinanza-digitale/dacity-perche-una-scienza-per-studiare-limpatto-dei-dati-sulla-societa/> (accessed on 30 April 2020).
- Davies, T.; Walker, S.B.; Rubinstein, M.; Perini, F. (Eds.) *The State of Open Data. Histories and Horizons*; African Minds: Cape Town, South Africa, 2019.
- Turning Fair into Reality*; Final Report and Action Plan from the European Commission Expert Group on FAIR Data; European Commission: Brussels, Belgium, 2018. Available online: https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_0.pdf (accessed on 31 May 2021).
- OnAudience.com. Global Data Market Size 2017–2019. Second Report. 2018. Available online: https://www.onaudience.com/files/OnAudience.com_Global_Data_Market_Size_2017-2019.pdf (accessed on 30 April 2020).
- Schwenk, H. Western Europe Big Data and Analytics Forecast 2018, 2018–2022. Available online: <https://www.idc.com/getdoc.jsp?containerId=EMEA43323118> (accessed on 30 June 2019).
- Claros, E.; Davies, R. *Data Employment and Market Value in the EU and other Major Economies*; European Parliamentary Research Service: Brussels, Belgium, 2016. Available online: <https://epthinktank.eu/2016/09/29/economic-impact-of-big-data/big-data-employment/> (accessed on 30 April 2020).
- IDC and Open Evidence. European Data Market. SMART 2013/0063. Report for the European Commission (Directorate-General for Communications Networks, Content and Technology). 2017. Available online: <http://datalandscape.eu/study-reports> (accessed on 30 April 2020).
- Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “Building a European Data Economy”*; European Commission: Brussels, Belgium, 2017. Available online: <https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy> (accessed on 30 April 2020).
- Einav, L.; Levin, J. Economics in the age of big data. *Science* **2014**, *346*, 1243089. [CrossRef] [PubMed]
- Molinari, F.; Concilio, G. Culture, Motivation and Advocacy: Relevance of Psycho Social Aspects in Public Data Disclosure. In Proceedings of the ECDG2017 Conference, Lisbon, Portugal, 12–13 June 2017; Rouco, J.C.D., Borges, J.V., Eds.; Acpil: Lisbon, Portugal, 2017; pp. 86–95.
- Berners Lee, T. Linked Data. 2006. Available online: <https://www.w3.org/DesignIssues/LinkedData.html> (accessed on 30 April 2020).
- Bator, F.B. The Anatomy of Market Failure. *Q. J. Econ.* **1958**, *72*, 351–379. [CrossRef]
- Zuiderwijk, A.; Janssen, M. Barriers and development directions for the publication and usage of open data: A socio-technical view. In *Open Government*; Gascó-Hernández, M., Ed.; Springer: Berlin, Germany, 2014; pp. 115–135.
- Asch, P.; Gigliotti, G.A. The Free-Rider Paradox: Theory, Evidence, and Teaching. *J. Econ. Educ.* **1991**, *22*, 33–38. [CrossRef]
- Duch-Brown, N.; Martens, B.; Mueller-Langer, F. The Economics of Ownership, Access and Trade in Digital Data. JRC Digital Economy Working Paper 2017 (01), JRC Technical Reports. 2017. Available online: <https://ec.europa.eu/jrc/en/publication/euro-scientific-and-technical-research-reports/economics-ownership-access-and-trade-digital-data> (accessed on 30 April 2020).

18. Stiglitz, J.E. Contributions of the economics of information to twentieth century economics. *J. Q. Econ.* **2000**, *115*, 1441–1478. [CrossRef]
19. Hall, B.H.; Lerner, J. *The Financing of R&D and Innovation*; UNU-MERIT Working Paper Series #2010-012; United Nations University-Maastricht Economic and Social Research and Training Centre on Innovation and Technology: Maastricht, The Netherlands, 2010. Available online: <https://www.merit.unu.edu/publications/working-papers/abstract/?id=3904> (accessed on 30 June 2020).
20. Leland, H.E.; Pyle, D.H. Informational Asymmetries, Financial Structure, and Financial Intermediation. *J. Financ.* **1977**, *32*, 371–387. [CrossRef]
21. Edler, J.; Cunningham, P.; Gök, A.; Shapira, P. *Handbook of Innovation Policy Impact*; Edward Elgar Publishing: Cheltenham, UK, 2016.
22. Mockus, M. Open Government Data Licensing Framework: An Informal Ontology for Supporting Mashup. Ph.D. Thesis, University of Bologna, Bologna, Italy, 2017.
23. OECD. *Data-Driven Innovation: Big Data for Growth and Well-Being*; OECD Publishing: Paris, France, 2015. [CrossRef]
24. Deloitte. Assessing the Value of TfL’s Open Data and Digital Partnerships 2017. Available online: <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf> (accessed on 30 April 2020).
25. Shah, R. Innovation through Open Data at Transport for London. In *EC Workshop on Data Access and Transfer*; European Commission: Brussels, Belgium, 8 June 2017. Available online: <https://ec.europa.eu/digital-single-market/en/news/innovation-through-open-data-transport-london-presentation-r-shah-ec-workshop-data-access-and> (accessed on 30 April 2020).
26. Parkes, E.; Karger-Lerchl, T.; Wells, P.; Hardinges, J.; Vasileva, R. Using Open Data to Deliver Public Services. Report for Open Data Institute 2018. Available online: <https://theodi.org/article/using-open-data-for-public-services-report-2/> (accessed on 30 April 2020).
27. Hogge, B. Open Data’s Impact. Transport for London, Get Set Go! 2016. Available online: www.odimpact.org (accessed on 30 April 2020).
28. Bria, F. *Our Data Is Valuable. Here’s How We Can Take That Value Back*; The Guardian: London, UK, 2018. Available online: <https://www.theguardian.com/commentisfree/2018/apr/05/data-valuable-citizens-silicon-valley-barcelona> (accessed on 30 April 2020).
29. Ajuntament de Barcelona. Mesura de Govern: Transició cap a la Sobirania Tecnològica: Pla “Barcelona Ciutat Digital”: Mesura de Govern 2016. Available online: <https://bcnroc.ajuntament.barcelona.cat/jspui/handle/11703/98713> (accessed on 30 June 2020).
30. Grossman, R. How Data Commons Are Changing the Way We Share Research Data and Make Discoveries: The Open Commons Consortium Perspective. Speech Delivered at the NSF Data Science Seminar. July 2016. Available online: <https://www.nsf.gov/attachments/139105/public/grossman-data-commons-NSF-16v5p.pdf> (accessed on 30 April 2020).
31. Wyns, B.; Bargiotti, L.; Loozen, N.; Loutas, N.; Dekkers, M.; De Keyser, M.; Goedertier, S. Good Practices for Identifying High Value Datasets and Engaging with Re-Users: The Case of Public Tendering Data. 2013. Available online: https://www.w3.org/2013/share-psi/wiki/images/3/31/Share-PSI_Submission_Paper-PwC_v0.03.pdf (accessed on 30 April 2020).
32. Kostura, A.; Castro, D. Three Types of Public-Private Partnerships That Enable Data Innovation. Post for the Center for Data Innovation 2016. Available online: <https://www.datainnovation.org/2016/08/three-types-of-public-private-partnerships-that-enable-data-innovation/> (accessed on 30 April 2020).
33. Concilio, G.; Molinari, F.; Morelli, N. Empowering Citizens with Open Data by Urban Hackathons. In Proceedings of the Conference on e-Democracy and Open Government (CEDEM2016) Conference, Krems, Austria, 17–19 May 2017; Parycek, P., Edelman, N., Eds.; IEEE Computer Society: Washington, DC, USA, 2017; pp. 125–134.
34. Rociola, A. Storia e Strategia Della Startup che Cresce Più Velocemente al Mondo: Le bici di Ofo. *AGI Economia*, 15 March 2018. Available online: https://www.agi.it/economia/startup_ofo_bike_sharing_alibaba-3628566/news/2018-03-15/ (accessed on 30 April 2020).
35. Calzada, I. (Smart) Citizens from Data Providers to Decision makers? The Case Study of Barcelona. *Sustainability* **2018**, *10*, 3252. [CrossRef]
36. Ekkehard, E. Big Data and Its Enclosure of the Commons. 2019. Available online: <https://www.socialeurope.eu/big-data-and-the-commons> (accessed on 30 April 2020).
37. Morozov, E. There is a Leftwing Way to Challenge Big Tech for Our Data. Here it is. *The Guardian*, 19 August 2018. Available online: <https://www.theguardian.com/commentisfree/2018/aug/19/there-is-a-leftwing-way-to-challenge-big-data-here-it-is> (accessed on 30 April 2020).
38. IDC Italia and the Lisbon Council. European Data Market Study. Opening Up Private Data for Public Interest. Story 1. Report for the European Commission, Directorate-General Communications Networks, Content and Technology Unit G1–Data Policy and Innovation. 2016. Available online: http://datalandscape.eu/sites/default/files/report/Story_1_New_format.pdf (accessed on 30 June 2020).
39. Gonzales, O.; Camprodon, G. *Air Quality Sensing Experiments. The iSCAPE Sensing Tools*; University of Hasselt Summer School: Hasselt, Belgium, 2019. Available online: <https://docs.google.com/presentation/d/1MPvRuPvP9vKRDUhleZfvJpMnhQXuLeP7jc-VCzA7IQM/edit#slide=id.p1> (accessed on 30 April 2020).
40. Janssen, M.; Charalabidis, Y.; Zuiderwijk, A. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Inf. Syst. Manag.* **2012**, *29*, 258–268. [CrossRef]

41. Welle Donker, F.; van Loenen, B. Sustainable Business Models for Public Sector Open Data Providers. *JeDEM eJ. eDemocr. Open Gov.* **2016**, *8*, 28–61. [[CrossRef](#)]
42. Beno, M.; Figl, K.; Umbrich, J.; Polleres, A. Open Data Hopes and Fears: Determining the Barriers of Open Data. In Proceedings of the 2017 Conference for E-Democracy and Open Government (CeDEM), Krems, Austria, 17–19 May 2017; pp. 9–81. [[CrossRef](#)]
43. Hardy, K.; Maurushat, A. Opening up government data for Big Data analysis and public benefit. *Comput. Law Secur. Rev.* **2017**, *33*, 30–37. [[CrossRef](#)]
44. Concilio, G.; Vitellio, I. Co-creating intangible cultural heritage by crowd mapping: The case of mappi[na]. In Proceedings of the IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), Bologna, Italy, 7–9 September 2016.
45. TALIA Project. Data Driven Innovation: Leveraging the Creative and Social Dimensions. Policy Brief #6 of the TALIA Interreg Project 2018. Available online: <https://social-and-creative.interreg-med.eu/> (accessed on 30 April 2020).