"Creating and governing social value from data"

Diane Coyle

(University of Cambridge)

Stephanie Diepeveen

(University of Cambridge)

Paper prepared for the IARIW-ESCoE Conference
November 11-12, 2021
Session 4
Time: Friday, November 12, 2021 [9:30-11:00 GMT+1]

Creating and governing social value from data

By Diane Coyle & Stephanie Diepeveen[1]

Abstract

Data is increasingly recognised as an important economic resource for innovation and growth, but its innate characteristics mean market-based valuations inadequately account for the impact of its use on social welfare. This paper extends the literature on the value of data by providing a framework that takes into account its non-rival nature and integrates its inherent positive and negative externalities. Positive externalities consist of the scope for combining different data sets or enabling innovative uses of existing data, while negative externalities include potential privacy loss. We propose a framework integrating these and explore the policy trade-offs shaping net social welfare through a case study of geospatial data and the transport sector in the UK, where insufficient recognition of the trade-offs has contributed to suboptimal policy outcomes. We conclude by proposing methods for empirical approaches to social data valuation, essential evidence for decisions regarding the policy trade-offs. This article therefore lays important groundwork for novel approaches to the measurement of the net social welfare contribution of data, and hence illuminating opportunities for greater and more equitable creation of value from data in our societies.

JEL codes: D02, D80, O30

Keywords: Data, valuation, social welfare, governance

[1] Bennett Institute, University of Cambridge. We are grateful to the Nuffield Foundation for funding the research on which this article is based (grant number WEL/43956). We thank Lawrence Kay, Julia Wdowin, Maha Rafi Atal, and particularly Jeni Tennison for helpful discussions and comments, and are grateful to the anonymous reviewers for their constructive comments on an earlier version. Responsibility for any errors and infelicities is all ours.

1. Introduction

A growing number of companies are accumulating large amounts of data, often freely provided by users of their zero-price services. More data is being produced by sensors and equipment in the 'Internet of Things'. Businesses increasingly regard their data as a valuable intangible asset helping increase sales and profits. Individuals and organisations of many kinds derive value from non-marketed data such as official statistics or free maps, while public bodies collect and use more and more data. There are also markets (legal and illegal) in certain types of data.

Yet, although data is evidently of large and growing importance, and there is a lively debate about policy, there are no agreed methods for determining the economic value of data of different types, including public data sets. Nor, importantly, is there a well-developed and integrated analysis of the non-market, economic welfare aspects of all data. The presence of both large positive and negative externalities drives a wedge between market value and social value. Understanding how this social value can be created and governed will, however, determine the returns to any investment in data collection and analysis by organisations. It thus will be essential to public policy decisions ranging from competition policy concerns about data hoards as barriers to entry, to innovation policy.

A growing literature (e.g. Corrado, 2019; Jones & Tonetti, 2019; Savona, 2019) sets out the fundamental economics of data. Data varies considerably in type and provenance. Moreover, the economic value of data depends on its information content and use (rather than any physical measure such as data records or bytes), and may therefore be highly heterogeneous. There are in addition significant data externalities. Much of the academic literature focuses on negative externalities especially privacy breaches (Acemoglu et al., 2019) or surveillance (Tirole, 2021). There is a separate literature, largely policy-focused, considering the potential of open data. Our contribution is to combine consideration of the negative and positive externalities, and to describe the potential for public policy to encourage innovation through enhanced access to data. In particular, we suggest what empirical methods might be appropriate for considering the value data could contribute to aggregate social welfare, as the contrasting positive and negative implications of data use cannot be evaluated in practice without a sense of scale.

We argue that an incomplete picture of the characteristics of data through the non-integrated focus in the existing literatures means current policy debates do not fully account for the contexts and conditions through which data generates net social welfare. We extend the analysis through a case study looking at geospatial data and the transport sector in the UK, uncovering how insufficient recognition or understanding of the positive and negative externalities in shaping policy has contributed to suboptimal outcomes. We finally propose potential methods for empirical approaches to data valuation in terms of net social welfare.

The paper begins with a discussion of the economic approach to the social welfare value due to data. Building on the existing literature, it develops a framework that introduces the informational context of data, and hence use values, accounting for variation between different provenances, types and users of data. Although some data markets are well established (for example, for credit scoring based on personal data), others are thin and the data sets involved are either non-standardised, or non-transparent, such as Amazon's sale of analytics services based on its accumulated user data (Li et al., 2019), or the automated online advertising market. Even large, highly reputable companies offering data on consumers or businesses do not post prices, but invite would-be purchasers to inquire. Moreover, as private and social value diverge, market prices will in any case not reflect the full economic (social welfare) value, so absent appropriate policies, private actors will have an incentive to either over- or under-use data, or both.

The paper then illustrates the key trade-offs resulting from this divergence with a case study of the transport sector and geospatial data in the UK. The study is based on analysis of a substantial body of policy documents over a 20-year period and 13 stakeholder interviews with key individuals involved in commissioning, analysing and using data in the UK throughout the period. Interviewees were selected to provide additional insights into policy debates, as well as a variety of perspectives on data and its value at play. Documents and interviews were coded and analysed around the language used around data, presentation of trade-offs in value creation, and identification of externalities.

Finally, as implementing policies to create value for society from data will require empirical estimates, we propose as future avenues for research two methods: contingent valuation methods including conjoint analysis, and a real options approach. Our framework for assessing the scope of data externalities and estimating values leads to an approach to data policy taking account of some key trade-offs, so we conclude with a discussion of the implications for governance and data policies.

2. Key economic characteristics of data

The growing literature on the economics of data makes clear that as an intangible and non-rival (although often excludable) asset characterised by externalities, it is distinctive as compared with many other goods in the economy. The public good or club good aspect implies that models for investing in data and distributing the value thereby created may be contested: financing the fixed costs of such goods is always challenging. The existence of externalities implies that market exchanges for data in general will not deliver the maximum social welfare. For example, if there are negative externalities (loss of privacy) in the case of data collected about individuals, there might be too much accumulation of data by organisations using it. On the other hand, data generally gains value from being combined with other data – such as data about multiple individuals' health status during a pandemic, or the typical purchasing patterns of others who have bought the same good, or the best travel route to take at a certain time of day. In other words, adding new fields to a dataset can add information and create network effects. Where there are positive externalities such as these, there may be under-production of data, to a variable extent depending on whether there are increasing or diminishing returns to additional data points.

Much of the economic (and legal) literature to date has focused on personal data (explicitly or implicitly defined in somewhat different ways) and has largely emphasised the negative aspects of data sharing. Acemoglu et al. (2019) focus on the loss of privacy from data sharing both to an individual and to others whose data is correlated with theirs. Jones and Tonetti (2020) consider as well as the loss of privacy from the accumulation of market power by firms accumulating individual data, therefore arguing for individuals to hold property rights in their data as a Coaseian (1960) solution to the market inefficiencies. Arrieta-Ibarra et al. (2018) similarly argue for 'data as labour' individual rights to sell data, whereas Savona (2019) posits 'data as intellectual property' whereby individuals can license use of their data. The implications of these authors' conclusions depend on their definitional approach – for example, whether the negative externality of privacy loss arises due to targeted use of information about them, or whether it occurs even when other people's online behaviour contains predictive information about an individual.

In general, though, any approach in which individuals hold property rights in 'personal' data will imply that transactions determine an amount of data sharing dependent on the relative valuations of the individual (concerning their privacy) and the purchaser (concerning the return to their use of the aggregated data) (Noam, 1997). The allocation of the ownership rights will determine the distribution of the gains (Acquisti et al., 2016). At present, legal and regulatory regimes allocate ownership to the platforms or organisations accumulating the data, hence the interest in creating countervailing individual rights, either alienable (data as labour) or not (data as an IP asset). The EU's GDPR safeguards against misuse but does not change the property allocation.

Yet, as Jones and Tonetti (2019) note, there are parallels between the role of data and the role of ideas in economic growth. In endogenous growth models such as those building on Romer (1990), ideas are non-rival and not wholly excludable, leading to positive spillovers across the economy driving productivity and growth. Ideas organise data, and may be easier to copy, but data itself can similarly generate positive spillovers. This is easiest to appreciate at the consumer level: there are many examples of services whose user value increases the more other users' data is accessed by the service. These include, for instance, apps measuring traffic congestion, search results, recommendation algorithms on streaming services and shopping sites, and matching two-sided platforms such as those matching diners and restaurants. The same will be true of the growing number of 'internet of things' data sources, in urban 'smart city' environments or supply chains, for example. The issues concerning non-personal data have been less widely examined in the law and economics literatures.

It is possible to model people's preference for privacy, the focus of this literature, by including it in the individual's utility function. Depending on the specifics of the model, firms may still overuse personal data; the social welfare implications vary, as does the equilibrium level of data sharing, depending on specific assumptions and parameterisation. However, such models show that there may be suboptimal over-collection and aggregation of data if people have an intrinsic preference for privacy, and if data access and use is controlled by firms which do not internalise privacy concerns. On the other hand, if aggregating data increases knowledge spillovers, and ultimately innovation and economic growth in an endogenous growth context, there may be suboptimal aggregation or data sharing, particularly if those investing in creating, maintaining and securely storing data are unable to capture an adequate return.

The parameters determining the actual terms of the trade-off implied by the presence of both negative and positive externalities are not technical primitives but are affected by policy choices as well as consumer preferences: to what extent is excludability legally or technical enforceable; how effective is data security and anonymisation; what legal rights do individuals have over data collection and usage; what licensing arrangements exist; and so on. Hence the social optimum will not be delivered without policy intervention, and at the same time the design of policy intervention depends on an assessment of the size of the externalities; this is a world of complex second-best outcomes (Lipsey, 2007).

There are other policy-related issues concerning data. One is the existence of substantial asymmetries of information. Regulators, and indeed individuals, will not know the value to firms of their privately-held data sets. Thus, contracts for data use and sharing are bound to be incomplete, while monitoring by regulators will be difficult. Application of the insights of regulation of network industries under information asymmetries to the data context is an important future research agenda (Tirole, 2020). This includes sectors where data sharing will have important societal, non-economic efficiency implications, such as transport and health.

There is also a significant competition policy issue. In many markets where data is used there are one or a few very large incumbents holding vast amounts of data and with substantial market power. Basu et al. (2020) argue that the effective natural monopoly dynamics of many digital platform markets ultimately reduce social welfare by eliminating the value in the outside option of not participating in the platform. Furman et al. (2019) argue that the data holdings of digital companies constitute a formidable barrier to entry in a range of digital markets, and inhibit investment including in upstream markets and innovative new entry: large data holdings extend economies of scope and scale created by network effects by allowing the incumbents to improve their services to users both through better use of information and by earning more advertising revenue (the ‘data loop’). We do not consider competition policy further here, but there is substantial work on data’s role in digital market competition (for example, Furman et al., 2019; Crémer et al 2019, and for countervailing views Bork & Sidak, 2012; Sidak & Teece, 2009).

Alongside the growing academic literature, a separate literature has developed focusing on the potential for public open data, or, in other words, on the potential positive externalities. Pollock (2008) influenced the UK’s Open Data White Paper (2012). The open data movement has gained traction more broadly among policymakers (e.g. European Commission, 2015;

APO, 2014). However, this strand of literature has been largely disconnected from the more recent academic interest in the economics of data.

Our focus is the core trade-off between positive and negative externalities, the distribution of costs and benefits, and empirical strategies to inform policy decisions. In other words, what policies will maximise net social welfare? How should policy-makers evaluate the potential benefits of exploiting positive externalities through enhanced or open data access enabling innovation and new services, as compared to the potential costs of privacy loss, or the financial cost of maintaining data, or the financial opportunity cost of not selling it at commercial rates? These dilemmas are at the heart of data policy discussions, as our transport case study below makes clear.

Before turning to the case study, it is important to note that the policy trade-offs will be shaped by the application context. The contribution of data to social welfare depends critically on its information content, which is context specific. Data's use value is therefore highly heterogeneous. The now-classic information pyramid places data at the bottom, and envisages value being added at each successive stage, but the value of the data itself at the bottom of the pyramid is not independent of the end-value of the insights gained, generally labelled 'understanding' or even 'wisdom'. In general, the value of intangible assets is more context-specific than that of many tangible assets, as a unit of 'information' does not translate well into standardised units such as bytes or data records (Savona, 2019).

The value of data will thus depend on how it can be used by the organisation holding it, as well as by the analytical framework provided by the economic lens. Part of this use value will be determined by the use context (such as improving a production process or traffic flow, or targeting an ad to a potential consumer). It will also partly depend on a number of characteristics of the data itself that are themselves contextually specific. These include the type and provenance of the data, and its accuracy, completeness and timeliness. Some characteristics are especially relevant to the scope for data use by multiple users and data sharing. One example is general reference data such as geospatial location codes or classification codes for economic statistics. Similarly, data is easier to use because it is in standard interoperable formats and accessible beyond those holding the data; although, this might make it less valuable to an individual organisation with commercial objectives controlling it, as excludability is reduced. Data that is about identifiable individuals and consequently sensitive because of the potential privacy harms may be more valuable to its users, but also more costly to collect and store securely. Table 1 sets out the types of factors affecting the value of data in any given use context.

Table 1: Two lenses on the social value of data

Analytical (economic) lens	Contextual (information) lens
Positive and negative externalities	Provenance
(Non-)excludability	Data type
Increasing/decreasing returns	Data subject/sensitivity
Depreciation	Generality (reference data)
Fixed and marginal costs	Accuracy
Complementary investments	Interoperability/accessibility

Source: authors' own

One natural question is whether the complexity due to heterogeneous use values can nevertheless be bypassed by taking a simple approach to measuring social value at the aggregate, macroeconomic level. After all, heterogeneity characterises many economic contexts. There are significant practical challenges to doing so, however. Very little data on data stocks and flows is gathered at all. Furthermore, GDP is largely a measure of market transactions, and by construction excludes the economic welfare effects of externalities, which is our interest here (Coyle, 2014). At present, certain investments in databases are included as an intangible investment in GDP, and their value is measured as the cost of preparing the data in an appropriate format for use (but not the cost of capturing or acquiring the data). The formal definition states: “databases consist of files of data organized in such a way as to permit resource-effective access and use of the data” (SNA, 2008). An extension of the definition is likely in the revision of the System of National Accounts currently under way but in any case the national accounts will likely, as now, use market prices for purchased data and a sum-of-costs approach for own-produced data (Statistics Canada, 2019). However, these aggregate approaches by definition do not capture the social value of interest here, and still less can they illuminate the trade-offs regarding the potential for future innovation from more extensive data use.

To understand better how to inform policy choices through taking into account the specificity of social value creation, we next turn to a case study concerning geospatial data and UK transport policy. We then turn to the question of empirical methods for estimating net social welfare impacts of data.

3. Policy trade-offs in practice: the social value of data in the UK transport sector

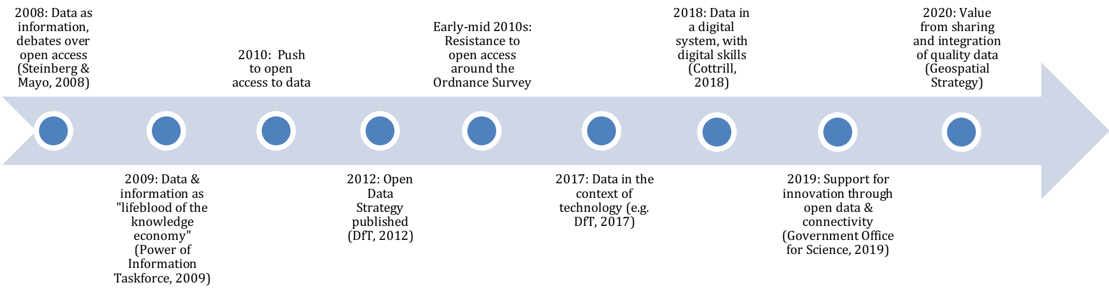
To provide new evidence regarding the policy trade-offs in a specific context, we conducted a case study of the UK transport sector. Debates over data use in the UK transport sector in recent years illustrate the social, public good value of data, and the presence of positive externalities that would not be captured in prices set by market transactions. Informed by the case study, in the next section we propose empirical avenues for weighing alternative ways of measuring value that better account for social welfare contribution of data across changing contexts and use.

This section focuses on policy developments throughout the 2010s, a period framed by an overall recognition in UK government policy of the public or social value of data. In transport, and more widely, there was a government-wide push to open access to certain data thereby realising its value for citizens. Yet, progress in opening access to data was uneven in transport and geospatial datasets, shaped by the changing incidence of positive and negative externalities.

First, we look at early efforts to open access to transport and to geospatial data, in which processes designed to realise the positive externalities of open transport data come into tension with policy toward geospatial data, and its associated costs. Specifically, the Ordnance Survey (OS) pressed for the costs of making its reference geospatial data accessible to be covered by financial charges. Tensions emerged over the capture of positive externalities and the distribution of benefit, and the costs associated with ensuring these benefits can be realised, which appeared to fall on a few public sector actors.

Second, we examine how debates over the distribution of costs and value evolved in a second stage as the conditions for value creation from data changed alongside technological developments. Firms' ability to process data and combine datasets grew, and became increasingly important in their value creation. By the end of that decade, the green economy and autonomous vehicles had become greater priorities, partially driven by changed technological possibilities (Figure 2).

Figure 2: changing discourse around data in transport, 2008-2020



Three features of the value of geospatial data become apparent through these changing policy debates, which are not accounted for in prices determined through market transactions and are tied to the public good aspects of the data.

First, the scope for positive externalities in the use of data that become increasingly apparent over the course of the decade, in combining public and private sector data. The value of combining datasets for wider use was linked to its informational characteristics and to the decisions it was intended to inform. Access and combination of data enable the provision of innovative new services.

Second, there has been unrealised value in the use of OS geospatial data, resulting in likely suboptimal outcomes. This was partially the result of the distribution of custodians, contributors and users of OS data: in which there was a single custodian that was required to cover costs (OS) but a multiplicity of those who could benefit, with the OS also competing with its customers in selling paid-for, value-added products.

Third, linked to this distribution of costs (borne by one entity) and benefits (accruing to many), custodians of geospatial data imposed access restrictions, resulting in likely suboptimal outcomes by preventing positive spillovers from more accessible data and combining data. This was partially driven by the anticipated costs to them of providing the data, with no direct return. The full realisation of the social value of data in the transport sector has been complicated by the question of how to cover the costs in making data useable and accessible, and its disjuncture from the distribution of realised value.

3.1 Early 2010s: Tensions around the distribution of value

The early 2010s were marked by a recognition in UK government policy that transport data had broad social value, and access should be opened up to help expand the extent of benefit. Early on, the Department for Transport took several steps to increase access to data around transport services, as part of a wider open data push by the Coalition government, introducing, for example, the London Data Store, Ordnance Survey Open Data Platform, and the National Public Transport Access Nodes dataset. The focus was on broader access to operations data held by private bus and rail companies, as well as reference data held by public sector bodies like the Ordnance Survey (OS).

Policy discourse emphasised that opening up data access would lead to improved services for consumers. There was public pressure on private contractors for public transport (such as bus company franchisees) to open access to their timetable data to third party applications. Transport Direct was formed in the Department to spearhead these efforts, using contract renewals as an opportunity to require private sector providers to publish their data.

Here, the basic trade off seemed to be potentially less revenue accruing to existing private transport providers, but improved benefits to citizens and other businesses around accessing and providing for more efficient and informed transport services. Transport Direct led in helping to create conditions for these changes in the realisation and distribution of value to take place. Private providers initially resisted open access to timetable data, which would enable others to track the quality of their service provision and foreclose earning revenue from selling the data. Still, by 2012, Transport Direct had acquired data and permissions for reuse from more than 100 providers, not only pushing for open access but also ensuring data was in a readable and useable form. The Shakespeare Review (2013) estimated £15-£28 million in the value of saved time for users of public transport through the availability of live information about transport systems on Transport for London (p. 6). Access to timetable information reduced the overall cost for transport users by providing more certainty about when a service would arrive. At the time, the primary trade-off the policy had to negotiate was presented as being between receiving improved services due to sharing and combining data versus the costs associated with publishing data. The policy focus was on the benefit to transport users through improved public services. While existing transport firms might lose out on revenue from selling data, the overall benefits for the citizen seemed clear, alongside new opportunities for third party applications to gain by providing real time transport service information. By the end of the decade, additional efforts to open access to bus operations data across England were underway. The Bus Services Act 2017 introduced provisions to require bus operators and local transport authorities to publish data on bus services (e.g. timetables, fares, location), and has been followed by a phased implementation process (Bus Services Act, 2017; Department for Transport, 2020).

The policy position was less clear when it came to debating the public good value of opening up access to reference data, specifically geospatial data held by the OS. Here, the loss of revenue from increasing access to the data was borne by public sector bodies, not private franchisees. The price and access model around the OS posed a challenge. From 1999, OS operated as a trading fund and was based on revenues, transitioning to a government owned limited company in 2015. It had to pay tax and an annual dividend to government as well as self-finance its operations (Rogawski et al., 2016). OS both licensed access to use its datasets, and created products that could complete with its customers' offerings. Therefore, commercial value was the driver: the OS charged users an average cost, rather than far lower marginal cost, which maximised its own business return, but not overall social welfare (Power of Information Taskforce, 2009).

A tension emerged between a view that the social welfare value created from data use would increase when access was opened up, and the OS's commercial interests underpinned by a requirement to meet the costs associated with producing, cleaning and maintaining data. Pressure to increase access to Ordnance Survey datasets came from wider society and government policy. In 2006, the Office of Fair Trading published a report on the commercial use of public information; it recognised the OS as obtaining nearly all of its income from selling and licensing information. It also identified problems with how the OS licenses use of its data, recommending a review of the extent to which its licensing facilitated use and re-use of information (2006, p. 137). Also that year, a *Guardian* newspaper Free Our Data Campaign identified the Ordnance Survey when critiquing the inaccessibility of data collected on behalf of the British public, and in 2009, a government-commissioned task force recommended freeing up geospatial data.

Opening up OS data had the potential to contribute to an increase in economic growth through dispersed and incremental productivity gains and higher tax revenues (Carpenter & Watts, 2013). The Department for Transport intended to remove restrictions on commercial use of geospatial data (DfT, 2013), but progress was incremental. OS's business strategy in April 2009 suggested a move to greater data use and sharing, but also reaffirmed a user-pays model. November that year, OS announced it would release some products for free (DotEcon, 2015, p. 75). Progress to opening data access started to come about in 2010 with the launch of OS OpenData.[2] Still, more detailed data remained available for commercial users to purchase or licence (via OS MasterMap).[3] Progress was slowed down by calculations concerning the distribution of cost and value capture within the OS. OS argued it needed to make a profit to cover the costs of maintaining a high quality and useable dataset, as the value derived through the use of OS data was realised by users, from individual citizens to companies. Private sector companies using OS data also indicated dissatisfaction. OS received government payments to compensate for its role maintaining open data, but as a user of its own data to create products, these payments were seen to give an unfair advantage over other data users (DotEcon, 2015, p. 147).

As a result, overall outcomes during this period appear likely to have been suboptimal. In transport services, public backing to Transport Direct pushed forward opening up of data held by private providers, enabling third party applications to use the data for more efficient and informed services. This was not paralleled by a similar strong push for change around geospatial data. The disjuncture between dispersed welfare gains by firms and citizens, including through overall economic growth, and the cost of maintaining the reference dataset anticipated by the OS, stalled changes to data access that could facilitate its wider use and application.

3.2 Late 2010s: The effect of technological change on value creation and capture

In the early part of the decade, debates about the social welfare value of data in the Department of Transport were tied to opening up access. By the late 2010s, this shifted with new technological capabilities and application of data. The social and economic value created

---

[2] OS OpenData is a collection of eleven Ordnance Survey digital datasets that were made available for free use and re-use from 1 April 2010 (Carpenter & Watts, 2013).

[3] Higher quality data was used more often by private sector firms, government agencies and researchers (though low cost or free agreements usually agreed for academic purposes).

from the use of geospatial data was increasingly attributed to technological drivers (Cabinet Office & Geospatial Commission, 2020). Developments in, for example, connectivity, edge computing and machine learning created new demands for high-quality, granular and real time data, which was interoperable to be applied to innovations in transport. The identity of who produced geospatial data was also changing. The data concerned was no longer primarily or solely held by the public sector; the advent of smart phones informed alternative coordinate systems to the OS.

Technological change altered the costs associated with value creation from data, and the distribution of value. Data was becoming less costly to gather – both as a side effect of other activities and also in a move away from the use of conventional methods of data collection, e.g. OS used fewer surveyors and more aerial photography. Cleaning and structuring of data sets accounted for a relatively greater share of the costs. Therefore, the earlier focus on the costs of maintaining the public dataset lessened.

In existing transport services, private firms' relative capacity to collect and utilise data for transport services has become increasingly apparent. For example, a company like Uber can now accumulate data on passenger preferences and travel patterns in real time. They can use this data to apply price differentials to internalise congestion fees, while also taking into account passengers' preferences. This private provision of a public good by companies, with the incentive and capacity not only to use the data to create private value but also to limit access for reasons of commercial advantage, has significant implications for the scope for others to realise value from access to and use of the transport data.

Who captures the potential value from data thus emerges as a key issue, with some dominant private sector actors able to limit value from the use of data, in contrast to the public sector, despite the latter continuing to contribute by providing basic transport infrastructure. Both transport infrastructure and the private providers' services inform well-run transport services, for example, in the public sector through investment into the underlying conditions for data generation and use (e.g. road networks, maintenance). However, the benefits from the application of data may lie predominantly with the private providers, not only by limiting others' access, but also from capturing consumer surplus through their greater ability to price discriminate or otherwise exercise market power. This is influenced by the relative capacity of certain firms to collect and utilise data. This scenario indicates a shift in the implications of externalities tied to value creation, with dominant private sector actors incentivised to maintain market dominance by capitalising on real-time data from customers' behaviours. This indicates potential need for governance and regulatory change to ensure the value of data concerning the behaviour of the public, using publicly-funded infrastructure, is not wholly captured by a few private companies, and consequently also limiting the potential for future innovation.

In addition to the distribution of value creation and capture, technological developments have reshaped the positive and negative externalities that must be taken into account in determining data value, alongside new possibilities for use. Intelligent mobility involves using data-driven devices to enable more "integrated, efficient and sustainable transport systems", e.g. autonomous vehicles, or the optimisation of network resilience (Catapult Transport Systems, 2017, p. 8). New technological applications change data requirements and the real and potential associated costs. For example, these dynamics are visible in developments in autonomous vehicles, where there are larger potential positive externalities through the combination of datasets, but their realisation requires highly granular and interoperable data available instantaneously. Autonomous vehicles require real time data to operate safely. Timely access to all relevant data for the public benefit becomes a concern alongside who holds data and where in service provision. Publicly-held geospatial data held by OS is useful but not sufficient for autonomous vehicles, which require much greater geospatial granularity as well as dynamic, real time data on road conditions, gathered from sensors on vehicles as well as in roadside infrastructure.

Reflecting such changes, the policy language around the contribution of OS geospatial data to value creation has shifted to identifying its role in facilitating positive externalities through data interoperability. OS is shifting from being presented as a data provider to increasingly having a role in setting standards for data quality and interoperability. It is seen as part of an "ecosystem" of reference data underpinning autonomous vehicle technology (Ordnance Survey, 2019). This policy language arguably reflects wider conditions for value creation through data use, beyond the costs of data production, cleaning and use.

New considerations have also arisen in public sector discussions, linked to concerns over the risks that people will be willing to take in using autonomous vehicles. Confidence in the quality of data is critical as public concerns about safety remain high (YouGov, 2021; Ordnance Survey, 2019). This reflects a policy concern relating to the potential economic growth contribution of the new technology. It too has helped to push discussions of a changing role for OS and the Department of Transport toward standardisation for interoperability and quality of data. Interviewees indicated that the value of the data, and its distribution, is contingent on regulation, which will have a role to play in shaping the distribution of value, such as mandating data sharing from sensors on autonomous vehicles or which base maps the vehicles can use. Therefore, again, price mechanisms are insufficient for capturing the positive and negative externalities associated with data use.

Throughout the past decade, the distribution of externalities limited the realisation of the potential value created through the use of geospatial data in transport. Changes in transport technology and investments, and the roles of private and public sector actors, are likely to continue to inform debates over the trade-offs in data investment and use, and the creation and distribution of value. There are consequently two areas where an improved account of the value of data to private entities on the one hand and to society as a whole on the other is likely to be of increasing policy importance.

First, what role is the public sector role in data governance, such as setting standards or mandating access, for example through licence conditions or in the terms of procurement contracts? Early on, pressure from Transport Direct facilitated the opening up of private transport providers' data to facilitate more innovation and more efficient services. In contrast, conflicting interests around opening up access to geospatial data, prevented similar changes from taking place with the OS. How does regulation inform the value to citizens, and how can its contribution to value creation be accounted for? And, second, what are the, trade-offs between the total social welfare created through the application of data in transport, and value capture by particular organisations? Linear and/or static value chains are unlikely to be appropriate to geospatial data, given the likely future changes in context for its use – driven by public views of acceptability, technological developments, and public and private sector arrangements around service provision.

Policy debate about transport to date has focused on the question of how to compensate data holders as access is increased. Commercial returns are important to fund investment in data by the private sector and to enable future economic growth. But, as our examination of externalities around data use in transport services, and how they affect value creation, shows, this is insufficient. Externalities need to be considered and evaluated, as these drive innovation, market entry and growth. Policy also needs to consider the tensions regarding the distribution of this social value given positive and negative externalities – which will vary depending on context, and technological and business innovation. Furthermore, the value of data for private data holders and in terms of social welfare is a political calculation, as these are endogenous to regulation and governance. The next section outlines alternative approaches to measuring the value of data to better account for the distribution of value across stakeholders around opening access, given both existing contexts and uncertain futures of data use. We then return to the policy choices.

4. Potential methodologies for estimating the social value of data

The transport case study indicates the need for empirics in order to advance assessments of the policy trade-offs regarding data. There is little consensus on how to measure the social value of a dataset and yet without methods to do so, crucial evidence for policy decisions is missing. The empirical literature to date has focused on techniques for estimating the private value of data for its holder. Market-based valuations are seen as preferable in the economics literature. One approach is to use stockmarket valuations to compare how data-intensive companies are valued compared to others (e.g. Li & Hall, 2020; Coyle & Li, 2021). Koutroumpis et al. (2020) suggest a matching market design framework could establish market values for data whose provenance (quality) has been confirmed and where ownership of intellectual property rights is established. Arrieta-Ibarra et al. (2020) is an example of the use of counterfactual simulations to determine the effect of changing the amount of data available to its ML algorithms on a data-using company's objective function such as profit (see also Bajari et al., 2019), as more generally, the revenue productivity of data use can measure its economic value to the organisation controlling it.

However, there are no established methods specifically for estimating the potential social value of data, or the size of the wedge between private and social value. The scale of the externalities involved is a key issue, both to understand the payoff to policies to mandate or encourage data sharing, and to address the familiar public goods dilemma of how much to invest in the fixed costs of data-gathering infrastructure, data updating, and secure storage. Some studies suggest there is a large welfare gain from the use of 'free' digital goods whose zero price is enabled by data collection (Brynjolfsson et al., 2019; Coyle & Nguyen, 2020), so it is possible that the scale will not be insignificant in the case of data itself.

Economists have a strong preference to consider market prices the most reliable estimates of value. However, our analysis and empirical case study of data in transport services demonstrate why market-based approaches will not address either the potential social value of data, reflecting its non-rival nature, or the distributional questions due to the need to cover fixed costs when the marginal costs of use are low. Therefore, here we propose for future research two potential approaches to the empirical estimation of the social surplus associated with data. One is the contingent valuation methodology widely used in environmental and cultural economics where there are similarly large externalities, including conjoint analysis; and the second is real options modelling. These two approaches provide more complete ways of reflecting the scope to create value across stakeholders, taking into account uncertain future value around technological change and changes in use – two issues illustrated by the case of the transport sector.

4.1 Contingent valuation

Contingent valuation (CV) methods are widely applied in some areas of economics when market prices are either unavailable or are known to fail to capture full economic welfare, such as environmental and cultural economics. Also known as stated preference approaches (as opposed to the revealed preference approaches using market prices and volumes), CV methods are based on surveys asking hypothetical questions to elicit statements of monetary value (see Carson, Flores & Meade, 2001 and McFadden & Train, 2017 for surveys). In the case of any marketed good, consumers gain a surplus above the market price they actually have to pay whenever they would have been willing to pay more. The elicited 'contingent' values from stated preference surveys are a measure of full consumer surplus, that is, the full amount each respondent would have been willing to pay for the good in question. They are a measure of compensating variation in Hicksian consumer theory – how much would someone need to keep them at their original level of utility when quantity or price of a good changes – or alternatively of equivalent variation – what amount would keep someone at their post-change level of utility, if the change had not happened. These correspond to willingness to pay (WTP) and willingness to accept (WTA) in CV surveys.

The approach has well-known limitations, and indeed there is a vast literature. Some economists (e.g. Hausman, 2013) have concluded this approach is inherently flawed. They determine the framing of survey questions affects responses, and there are various potential biases such as strategic responses when respondents desire a particular outcome, or hypothetical bias when there are no anchors in experience to pin down stated values. Furthermore, there can be wide gaps between willingness-to-pay (WTP) measures for services and willingness-to-accept (WTA) measures for their absence or loss in public good contexts. Nevertheless, there are no other methods available to attempt the quantification of the social welfare wedge, or consumer surplus, absent any market prices and quantities (Blinder, 1991). The CV literature has also developed a range of methods to address pitfalls such as potential biases.

A test of feasibility in the context of data is provided by the nascent literature on valuing free digital goods. User's valuations are implicitly reflected in their many hours of use of online digital services. Although the valuations of these goods are unlikely to equal the value of the resulting data sets to the digital platforms, as the platforms capture most of the social externality value, these studies illustrate the potential and also the challenges for the methodology. The results to date are variable. Some studies (e.g. Brynjolfsson et al., 2019; Mosquera et al., 2019) find large annual WTA valuations of thousands of dollars for some 'free' digital services such as Facebook. Using an incentive-compatible experimental method (with payments to participants), Allcott et al. (2020) found that the median WTA for Facebook in an incentive-compatible experiment started at $100 for an initial month, though it declined by up to 14% after participants had experienced doing without the service. By contrast, using a survey approach, Sunstein (2019) found the median WTP for Facebook was much lower at $5 a month (with 46% of his respondents choosing $0) while the WTA was just under $64 a month. This gap is larger than the disparities commonly found between WTP and WTA for environmental goods. Further confirmation of the gap comes from the actual average revenue per user for some services; Coyle and Nguyen (2020) found these to be similarly an order of magnitude lower than stated WTA valuations. Sunstein (2019) speculates that reasons for the WTA/WTP gap in this context could include 'protest voting' by some participants against a company they do not like, the loss aversion/endowment effect of being asked about paying for something that has hitherto been free, or the difficulty of projecting forward how they feel about an experience good (although this latter feature does not prevent people purchasing many market goods).

Nevertheless, despite the range of estimates and open questions about the interpretation of results, the contingent valuation studies of free digital goods provide results consistent with economic theory, such as a decline in the proportion of people willing to give up a good as amounts offered decrease, an increase in WTA when the mooted loss concerns a longer time period, and a similarly intuitive relationship with income levels of respondents. Coyle and Nguyen (2020) also show that WTA measures for free digital goods correlate positively with usage, relate sensibly to other free goods (such as parks) and to market substitutes (such as going to the cinema), and show a well-defined ranking. In addition, the WTA valuations change in response to external conditions – in this case the introduction of lockdown in the UK in March 2020 – in directions consistent with economic intuition. This recent literature concerning other intangible quasi-non-market goods suggests contingent valuation is a promising methodology in the data context.

Contingent valuation estimates of the value of datasets have been applied in a few existing studies (BEIS, 2018; Deloitte, 2017; Miller et al., 2013). They suggest that a survey-based WTP approach could be applied to datasets to help ascertain the value of opening them. In these cases, the good (the dataset) is quantity rationed; the shadow prices attached to changes in quantities may be large, even if own-price and cross-price elasticities would not be, absent quantity limits (Carson et al., 2001). Given a well-constructed survey of the relevant user group, this could help policymakers estimate the potential social welfare value of making public data sets more accessible rather than charging for them to help recover costs; or of mandating greater access to some privately-controlled, inaccessible data sets – issues that posed challenges around opening up data access in transport.

One specific type of survey-based method is conjoint analysis, used more often in marketing or operations research than in economics; the typical use is to elicit consumer preferences for specific product features. The method allows the construction of choice rankings including price variations, and the relative utilities make it possible to construct a willingness to pay measure. A pioneering pilot study by the Office for National Statistics has tested whether the technique could be used to value official statistics, a vast – and free – series of data sets developed using rigorous methodologies that clearly brings great value to users (Williams, 2021). It concluded that the results from a small sample warranted a larger-scale test. The method could be particularly useful for geospatial data and transport information for example, and would help establish benchmark values, especially for reference data. This could help to give clarity to the distribution of value across different stakeholders, and the conditions that give rise to its usefulness. The methodology enables the testing of a range of characteristics, and analysis by different groups of respondents. Conjoint analysis could also be useful for testing the social welfare value of greater access to specific datasets, such as those currently privately-held by businesses. There is a promising future research agenda in applying contingent valuation methodology to data.

4.2 Real options

The interviews underpinning the case study indicated that organisations sometimes invest in datasets for currently unknown future uses; they therefore have an option value. The past decade of data use for transport services already indicate this potential, as technological change indicates new possibilities for combining reference and real time data for new users, e.g. autonomous vehicles. This suggests that real options methodology is worth exploring as an empirical method, as these data investments are made for the same reason that organisations undertake any other strategic investments. Real options were devised as a parallel to financial options, with Myers (1977) describing an opportunity to invest as a real-asset application of the Black-Scholes pricing model for call options in finance. Subsequently, the concept has been applied more intuitively as a framework for systematically assessing the strategic value of corporate decisions to invest in certain activities (or contract them). The toolbox of empirical techniques has expanded beyond the original Black-Scholes equation to include numerical methods such as Monte Carlo simulations (Luehrman, 1998; Rigopolous, 2015). Applications include valuation of the option to use natural resources such as mineral reserves (Damodaran, 2005). As with financial options, the more uncertainty there is about the future, the greater the option value. The parallels between real and financial options are inexact, especially in our current context: real options are unlikely to be freely tradable, may have lumpy rather than continuous prices, and may have unknown or time-varying risk profiles and variance. Nevertheless, they could be useful for estimating valuations for data sets in some contexts.

The simple Black-Scholes formula gives the value of a call option as:

$$S.\ N(d_1) - K.\ e^{-it}.\ N(d_2)$$

where

$$d_1 = [\ln(S/K) + (i + \sigma^2/2)t] / (\sigma.\sqrt{t})$$

$$d_2 = d_1 - \sigma.\sqrt{t}$$

and

S = current value of the underlying asset

K = option exercise price (present value = K. e^{-it})

t = time to expiry of the option

i = risk-free rate over same time horizon

σ^2 = variance of (ln) underlying asset price

N denotes cumulative normal distribution function.

$N(d_1) > N(d_2)$ represent the range of probabilities that a call option will be in the money at its expiry.

As a concrete example, consider a database being accumulated by a digital platform from traffic and location sensors. Suppose a firm is considering how much to pay for access to this data if it is offered, or a regulator is considering the value of enforcing open access. Investing in the data will have a similar payoff structure to a call option. A number of variables will be needed, including an initial estimate of value, cost of developing or managing the data for own-use, time horizon for accessing it (for example, a 20-year licence), and variance in the underlying value of the database. In this example, some of these are straightforward, including the risk-free 20-year rate (typically on government bonds, 0.5% in the UK at present). The potential purchaser will have internal estimates of (the NPV of) how much it will cost to acquire and use the data (K).

The challenge is in finding an estimate of underlying current value of the database and its variance, and there are several possible approaches. The former could for example be estimated as the present value over 20 years of the annual profit the current incumbent derives from its use, and the variance as the variability of that profit stream. An alternative (used in some studies e.g. PWC, 2019) is to compare stockmarket valuations of data-driven companies versus others in the same sector. A third approach is to use an estimate by the organisation holding the data of how much it would increase its revenues and profits if it used it differently. For example, BP paid Palantir $1.2bn for 10 years to integrate data across its businesses, increasing oil production by 30,000 barrels a day, an increase which could be valued at prevailing oil prices and was presumably expected by the company to exceed the fee by some margin (Raval, 2019). In a similar vein, the simulation approach in Arrieta-Ibarra et al. (2020) provides upper bound estimates of the impact of data use on company profits, and there is a growing strand of work using such machine learning techniques to explore the space of possible data uses. The inherent difficulty is that the potential future uses of data sets are more than usually unknowable because they depend on the rapid evolution of a frontier technology, machine learning and AI; however, this is the case in other contexts in which real options have been used. Here too there is a rich potential strand of future empirical research.

5. Discussion: policy implications

This paper has argued that measuring the value of data requires considering its social welfare value, reflecting its economic and informational characteristics in specific contexts. This requires looking outside of market based valuations. Data markets, even if they exist, will deliver suboptimal outcomes from the perspective of social welfare. We have demonstrated how this approach helps to make sense of policy dilemmas and governance decisions around data, given both negative and positive externalities, in a case study of the UK transport sector.

Working from analysis of the inadequacy of market based valuations to represent the social value of data, we proposed some empirical approaches to estimating the potential social value of specific datasets, which could for example inform the issues arising in the transport case study, as avenues for future research. Empirics will be needed if governments are to be able to assess questions such as how much they should invest in public data, how far to require open data provision by private entities, or even how much public sector bodies should charge if they sell data to private sector companies. In the transport case, for instance, empirics could usefully inform a decision about whether OS should cover its full average cost by selling data, or instead be entirely taxpayer funded and make all its data open, or opt for an intermediate approach of charging large commercial users more than unit cost of supply to subsidise innovation by smaller users. The competing arguments over time about the best policy approach embed implicit – and different – views about the size of the potential for economic value creation from more open access, and about the distribution of likely benefits.

The conceptual and empirical exploration of how value is created from data provided here takes as its premise that this occurs in dynamic, context-specific ways. We have shown how this framework has been reflected in approaches to policy analyses and decision-making in the UK transport sector. This final section discusses the implications for regulation and governance of data from our expanded conceptualisation combining both the contextual and economic characteristics of data. We consider that the prevailing framing of 'personal' data and property rights in data will not permit the realisation of the full social welfare value of data, as it inherently omits substantial externalities and indeed the relational nature of much data (Viljoen, 2020). A language of individual property rights obscures the externalities and dynamic processes through which data can add to social welfare, and the potential capture of this social value by private companies. This is not least because much data (including much data provided *by* individuals) does not relate to solo individuals. Thus, for instance, payment by digital platforms to individuals for providing personal data may still allow the platforms to retain the positive externality benefits from aggregated data.

We therefore consider there is a case for a spectrum of access rights that would enable increased societal value and broader distribution of the value of data. This would include a commercial return to private data holders, as the creation and maintenance of datasets requires significant investment and skill, but would also enable the creation of greater total social welfare than is currently the case. For example, rather than posing a confrontation between conflicting rights or values – such as innovation versus privacy – considering a spectrum of appropriately governed access rights may allow the generation of multiple different insights and thus sources of value.

Public and policy debates about data governance have become increasingly marked by concerns about the appropriation of value, often pertaining to individuals, and enabling the dominance of a few corporations, illustrated in some of the recent policy debates in the transport sector around open access to data. The issues are visible around geospatial and other data in the UK transport sector, where concerns about the distribution of costs and risks fed into policy decisions around data access. Progress in the policy debate will require the further development and use of methods to estimate the social value of datasets in specific applications; at present it is impossible to say which policy choices would have the best outcomes for social welfare. We have therefore proposed some possible empirical approaches. There is a significant research agenda in testing these (and no doubt other) methodologies in different applied contexts.

There are important policy decisions to be made around determining value capture and distribution in contexts where the wedge between private and social value may be high, with highly context-specific issues. Distributional and growth outcomes will come about by default without an explicit discussion not only of potential commercial opportunities but also of wider social value. These possible future implications for the realisation of the value of data affirm the importance of this article's contribution in laying the conceptual groundwork from which to build approaches to analysing and measuring value creation. By accounting more fully for the ways that data creates value, by whom and in what conditions, policymakers might better plan for more optimal and equitable futures in value creation and distribution.

References

Acemoglu, D, et al. (2019). "Too Much Data." NBER Working Paper 26296, September.

Acquisti, A., C. Taylor, and L. Wagman. (2016). "The Economics of Privacy." *Journal of Economic Literature*, 54(2): 442-92. DOI: 10.1257/jel.54.2.442

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow. (2020). "The Welfare Effects of Social Media." *American Economic Review*, 110(3): 629-76. doi: 10.1257/aer.20190658

APO. (2015). "Open for Business: How Open Data Can Help Achieve the G20 Growth Target." Accessed at: https://apo.org.au/sites/default/files/resource-files/2014-06/apo-nid40141.pdf on 6 June 2021.

Arrieta-Ibarra, I. A., L. Goff, D. J. Hernández, J. Lanier, and G. E. Weyl. (2018). "Should We Treat Data as Labor? Moving Beyond 'Free'." *American Economic Association Papers & Proceedings*, 1(1). doi: 10.1257/pandp.20181003

Arrieta-Ibarra, I., Caruso, T., & Hernández, D.J. (2020). "An empirical study of the value of data." Mimeo.

Bajari, P., V. Chernozhukov, A. Hortaçsu, and J. Suzuki. (2019). "The Impact of Big Data on Firm Performance: An Empirical Investigation." *AEA Papers and Proceedings*, 109: 33-37.

Basu, K., A. Caspi, and R. C. Hockett. (2019). "The Law and Economics of Markets with Digital Platforms." Cornell Legal Studies Research Paper No. 20-32, 16 October. http://dx.doi.org/10.2139/ssrn.3566590

BEIS. (2019). "Companies House data: valuing the user benefits." Accessed at: https://www.gov.uk/government/publications/companies-house-data-valuing-the-user-benefits on 21 August 2020.

Blinder, A.S. (1991). "Why Are Prices Sticky? Preliminary Results from an Interview Study." *American Economic Review*, 81: 89-100.

Bork, R. H. and G. J. Sidak. (2012). "What Does the Chicago School Teach about Internet Search and the Antitrust Treatment of Google?" *Journal of Competition Law and Economics*, 8(4): 663-700.

Brynjolfsson, E., A. Collis, W. Erwin Diewert, F. Eggers, and K. J. Fox. (2019). "GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy." NBER working paper 25695.

Bus Services Act. (2017). *UK Public General Acts, c. 21*. Accessed at: https://www.legislation.gov.uk/ukpga/2017/21/contents on 6 June 2021.

Cabinet Office and the Geospatial Commission. (2020). *Unlocking the power of location: The UK's geospatial strategy*, 16 June 2020, updated 14 July 2020.

Carpenter, J. and P. Watts. (2013). "Assessing the Value of OS OpenData to the Economy of Great Britain – Synopsis." Prepared for the Department of Business, Innovation and Skill by ConsultingWhere Ltd., and ACIL Tasman.

Carson, R. T., N. E. Flores, and N. F. Meade. (2001). "Contingent Valuation: Controversies and Evidence." *Environmental and Resource Economics,* 19: 173–210. https://doi.org/10.1023/A:1011128332243

Catapult Transport Systems. (2017). "The case for government involvement to incentivise data sharing in the UK intelligent mobility sector." Briefing Paper, March. Accessed at: https://ts.catapult.org.uk/intelligent-mobility/im-resources/opendata/ on 6 June 2021.

Corrado, C. (2019). "Data as an Asset." Presentation at EMAEE 2019 Conference on the Economics, Governance and Management of AI, Robots and Digital Transformation.

Cottrill, C. D. (2018). *Data and Digital Systems for UK Transport: Change and its Implications*. Review commissioned by the Government Office for Science, December.

Coyle, D. (2014). *GDP: A Brief but Affectionate History*. Princeton: Princeton University Press.

Coyle, D. and D. Nguyen. (2020). "Free Goods and Economic Welfare," ESCoE Discussion Paper 2020-18. Accessed at: https://www.escoe.ac.uk/publications/free-goods-and-economic-welfare-escoe-dp-2020-18/ on 6 Jun 2021.

Coyle, D. and W. Li (2021). "The Data Economy: Market Size and Global Trade." Forthcoming as ESCoE Discussion Paper.

Crémer, J, Y-A de Montjoye and H Schweitzer. (2019). "Competition Policy for the Digital Era." European Commission. Accessed at https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf on 23 August 2020.

Damodaran, A. (2005). "The Promise and Peril of Real Options." NYU Working Paper No. S-DRP-05-02.

Deloitte. (2017). "Assessing the value of TfL's open data and digital partnerships." Accessed at: http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf on 21 August 2020.

Deloitte. (2019). "Data Valuation: Understanding the Value of Your Data Assets." Accessed at: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Finance/Valuation-Data-Digital.pdf on 23 August 2020.

Department for Business, Innovation and Skills (BEIS). (2013). *Shakespeare Review: An Independent Review of Public Sector Information,* May.

Department for Transport (DfT). (2012). *Open Data Strategy*. Accessed at: http://data.gov.uk/sites/default/files/DfT%20Open%20Data%20Strategy.pdf as of 16 November 2012, via wayback machine.

DfT. (2013). *Door to Door: A strategy for improving sustainable transport integration*, March.

DfT. (2017). *Transport Investment Strategy: Moving Britain Ahead*.

DfT. (2020). *Guidance: Bus open data implementation guide.* Accessed at: https://www.gov.uk/government/publications/bus-open-data-implementation-guide/bus-open-data-implementation-guide on 6 June 2021.

DotEcon. (2015). "Independent evaluation of the OFT's 2006 market study into the commercial use of public information (CUPI)." A report by DotEcon for the CMA, March. Accessed at: http://www.dotecon.com/assets/images/Evaluation_of_CUPI_study.pdf on 6 June 2021.

European Commission. (2014). "Creating Value Through Open Data." Accessed at: https://data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf on 6 June 2021.

Furman, J, et al. (2019). "Unlocking Digital Competition: Report of the Digital Competition Expert Panel." Accessed at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf on 27 July 2020.

Government Office for Science. (2019). "A time of unprecedented change in the transport system. The Future of Mobility." Foresight, January.

Hausman, J (2013). "Contingent Valuation: From Dubious to Hopeless." *Journal of Economic Perspectives,* 26(4): 43-56.

Jones, C. I. and C. Tonetti. (2019). "Non-Rivalry and the Economics of Data." NBER Working Paper No. 26260. Accessed at: https://www.nber.org/papers/w26260.

Koutroumpis, P, Aija Leiponen, Llewellyn D W Thomas. (2020). "Markets for data." *Industrial and Corporate Change*, 29(3): 645–660. https://doi.org/10.1093/icc/dtaa002

Li, W.C.Y. and Hall, B.H. (2020). "Depreciation of Business R&D Capital." *Review of Income and Wealth*, 66: 161-180. https://doi.org/10.1111/roiw.12380

Li, W., C.Y., Makoto Nirei, and K. Yamana. (2019). "Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy." BEA working paper.

Lipsey, R.G. (2007). "Reflections on the general theory of second best at its golden jubilee." *International Tax and Public Finance,* 14: 349–364. https://doi.org/10.1007/s10797-007-9036-x

Luehrman, T. A. (1998). "Strategy As A Portfolio of Real Options." *Harvard Business Review,* 76(3): 89-99.

McFadden, D., & Train, K. (2017) *Contingent Valuation of Environmental Goods: A Comprehensive Critique*. Cheltenham, UK: Edward Elgar.

Miller, H. M. et al. (2013). "Users, uses, and value of Landsat satellite imagery—Results from the 2012 survey of users." *U.S. Geological Survey Open-File Report 2013–1269,* 51. https://doi.org/10.3133/ofr20131269

Mosquera, R., M. Odunowo, T. McNamara, et al. (2020). "The economic effects of Facebook." *Experimental Economics,* 23: 575–602. https://doi.org/10.1007/s10683-019-09625-y

Myers, S. C. (1977). "Determinants of Corporate Borrowing." *Journal of Financial Economics*, 5(2): 147-175. doi:10.1016/0304-405X(77)90015-0

Noam, Eli M. (1997). "Privacy and Self-Regulation: Markets for Electronic Privacy." In *Privacy and Self-Regulation in the Information Age*. Washington, DC: US Department of Commerce, National Telecommunications and Information Administration. Accessed at: https://www.ntia.doc.gov/report/1997/privacy-and-self-regulation-information-age on 6 June 2021.

Office of Fair Trading. (2006). "The commercial use of public information (CUPI)." December. Accessed at: https://joinup.ec.europa.eu/sites/default/files/document/2014-12/media2562.pdf on 6 June 2021.

Ordnance Survey. (2019). *Geodata Report – analysis and recommendations for self-driving vehicle testing*. Report produced for Zenzic-UK. Accessed at: https://zenzic.io/content/uploads/2019/07/Geodata-Report.pdf on 21 August 2020

Pollock, R. (2008). "The Economics of Public Sector Information." Working paper. Accessed at: https://rufuspollock.com/papers/economics_of_psi.pdf on 6 June 2021.

Power of Information Taskforce. (2009). *The Power of Information Taskforce Report,* February.

PWC. (2019). "Putting a value on data." Accessed at: https://www.pwc.co.uk/dataanalytics/documents/putting-value-on-data.pdf on 9 October 2019.

Raval, A. (2019). "BP's Bernard Looney takes oil major into energy transition." *Financial Times*, 6 October.

Rigopolous, G. (2015). "A Primer on Real Options Pricing Methods." *International Journal of Economics and Business Administration*, 1(2): 39-47.

Rogawski, C., S. Verhulst, and A. Young. (2016). "'Open Data's Impact: Great Britain's Ordnance Survey: A Clash of Business Models'." GovLab and Omidyar Network, January.

Romer, P. (1990). "Endogenous Technological Change." *Journal of Political Economy*, 98(5): S71-102. Accessed at: https://EconPapers.repec.org/RePEc:ucp:jpolec:v:98:y:1990:i:5:p:s71-102 on 6 June 2021.

Savona, M. (2019). "The Value of Data: Towards a Framework to Redistribute It." SPRU Working Paper Series (SWPS), 2019-21: 1-22. Accessed at: www.sussex.ac.uk/spru/swps2019-21

Sidak, J. Gregory and Teece, David J. (2009). "Dynamic Competition in Antitrust Law." *Journal of Competition Law and Economics*, 5(4): 581-631. http://dx.doi.org/nhp024

SNA. (2008). *System of National Accounts*. Accessed at https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf on 23 August 2020.

Statistics Canada. (2019). "Measuring Investment in Data, Databases, and Data Science: Conceptual Framework." *Latest Developments in the Canadian National Accounts*.

Steinberg, T. and E. Mayo. (2008). "Interim Progress Report on implementing the Government's Response to the Power of Information Review (Cm7157)." Cabinet Office. Accessed at: https://webarchive.nationalarchives.gov.uk/20081229162456/http://www.cabinetoffice.gov.uk/media/cabinetoffice/corp/assets/publications/reports/power_information/poi_interim.pdf.

Sunstein, C. R. (2019). "Valuing Facebook." *Behavioural Public Policy,* 1-12. doi:10.1017/bpp.2018.34

Tirole, J. (2020). "Competition and the Industrial Challenge for the Digital Age." Toulouse School of Economics mimeo, April. Accessed at: https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/by/tirole/competition_and_the_industrial_challenge_april_3_2020.pdf on 28 July 2020.

Tirole, J. (2021). "Digital Dystopia." *American Economic Review*, forthcoming. Accessed at: https://www.aeaweb.org/articles?id=10.1257/aer.20201214&&from=f on 6 June 2021.

Viljoen, S. (2020). "Democratic Data: A Relational Theory For Data Governance." Available at SSRN: https://ssrn.com/abstract=3727562 or http://dx.doi.org/10.2139/ssrn.3727562.

Williams, S. (2021). "Valuing Official Statistics with Conjoint Analysis." Office of National Statistics (ONS). https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/valuingofficialstatisticswithconjointanalysisapril2021

YouGov. (2021). "Are Brits comfortable with self-driving cars?" YouGov Direct Survey, conducted 21 April. Accessed at: https://yougov.co.uk/topics/transport/articles-reports/2021/04/28/car-manufacturers-still-some-way-convincing-brits- on 8 June 2021.