Independent assessment of the Open Data Institute's work on data trusts and on the concept of data trusts

Report to the Open Data Institute

DTE-005

LE
London
Economics

April 2019

About London Economics

London Economics is one of Europe's leading specialist economics and policy consultancies. We advise an international client base throughout Europe and beyond on economic and financial analysis, litigation support, policy development and evaluation, business strategy, and regulatory and competition policy.

Our consultants are highly-qualified economists who apply a wide range of analytical tools to tackle complex problems across the business and policy spheres. Our approach combines the use of economic theory and sophisticated quantitative methods, including the latest insights from behavioural economics, with practical know-how ranging from commonly used market research tools to advanced experimental methods at the frontier of applied social science.

We are committed to providing customer service to world-class standards and take pride in our clients' success. For more information, please visit www.londoneconomics.co.uk.

Head Office: Somerset House, New Wing, Strand, London, WC2R 1LA, United Kingdom.

w: londoneconomics.co.uk   e: info@londoneconomics.co.uk   🐦: @LondonEconomics
t: +44 (0)20 3701 7700   f: +44 (0)20 3701 7701

Acknowledgement

Authors

Moritz Godel, Dr. Ashwini Natraj

Table of Contents

1 Background & context

1.1 Economic benefits of making data accessible across organisation boundaries

A data trust is defined by the ODI as 'a legal structure that provides independent stewardship of data'[1], where stewardship involves deciding who has access to data, the conditions of access and who benefits from data access. Data trusts allow for collection, maintenance and sharing of data, whilst protecting the privacy of parties which could be affected by opening up this data.

Data trusts therefore allow data to be shared which otherwise would have remained locked in organisational silos[2], or not used for productive or innovative purposes at all. As such, data trusts can enable economic benefits by allowing data to be used where it is most useful, whilst also minimising potential market failures, including the loss of trust from data subjects (or individuals subject to data-driven decision making).

Crucially, data trusts make it easier for organisations to access data which they have not collected themselves, which reduces the overall cost of data usage. Previous work by LE[3] has mapped the economic effects of opening data to new uses and users by removing frictions that impede data mobility:

■ external benefits arising from increased use of data (i.e., more data being provided, which may have value that is not reflected in the benefit received by data subjects themselves);

■ Higher productivity (i.e., making it easier to combine data from different sources lowers the cost of producing data-enabled products and services); and

■ innovation from combining data in new ways across organisations and industry silos.

The government is pursuing a target for the "UK to be the world's most innovative economy". Innovation in the field of AI, which is seen as a key source of future innovation, is part of this. This means more and better AI-related R&D being done in the UK, which implies lower cost of AI-related research (including easier access to data) and more and better data to be used as the basis for such R&D.

The assessment of the concept of data trusts focuses on their economic function. The rationale for government support for data trusts as a new data sharing mechanism is framed in terms of their contribution to productivity, and development of AI and data-driven innovation more broadly.

---

[1] https://theodi.org/article/defining-a-data-trust/

[2] "Data Silos refers to situations where data rarely leaves the organisation that collects it and is rarely used for purposes other than what it was first collected for…" LE (2018)

[3] In particular, this report draws on two pieces of LE research:

- Personal Data Protection Commission and Consumer Commission of Singapore (25 February 2019). Discussion paper on Data Portability. Retrieved from https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/Data-Portability/PDPC-CCCS-Data-Portability-Discussion-Paper---250219.pdf

- London Economics (2018). Data mobility economic considerations (Appendix 4). In Ctrl-Shift (2018). Data mobility: The personal data portability growth opportunity for the UK economy. Retrieved from https://www.ctrl-shift.co.uk/reports/DCMS_Ctrl-Shift_Data_mobility_report_full.pdf

Other key aspects are:

- increasing ease of access to data (for use in AI, both R&D and applications);

- facilitating the sharing of data between organisations holding data and organisations looking to use data to develop AI;

- enabling data use (for AI) to be done more easily and frequently; and

- allowing data transactions to proceed with confidence and trust.

A data trust fulfils its intended function if it leads to more data being accessed/used and data from different sources being brought together; if this data results in innovations in AI (or enabled by AI), if the resulting benefits are shared equitably between the stakeholders; and if trust in the use of data under these conditions is increased.

Data trusts are data sharing mechanisms, and as such can be compared with alternative data sharing mechanisms, with whom they share basic economic characteristics. The distinguishing feature of data trusts is independent stewardship of data.

1.1.1 The ODI's Data Trust pilots

While the General Legal report cautions that trust law is not the most appropriate framework to think of data trusts, it is however possible for data trusts to adapt legal forms and structures from trusts in other contexts e.g. community land trusts and trust ports. Data trusts are intended to be a replicable and scalable structure of arrangements that can be adapted to different situations.

The ODI were commissioned to implement three pilot projects:[4]

- The first had the aim of tackling illegal wildlife trade. This project involved making image and invoice data of endangered species more accessible with the aim of training 'recognition algorithms for border officials, helping them to identify illegal animals and animal products.' This project also looked at sharing 'image and acoustic data' to train algorithms to create real-time alerts, identify animals and people travelling through restricted areas, with the goal of 'increas[ing] access while protecting privacy.'

- The second was intended to help reduce 'global food waste'. Its main objectives were assessing present UK processes to track food waste in the supply chain, identify incentives and barriers for data sharing, and design and review replicable legal and governance structures which could be used to measure food waste.

- The third involved the delivery of public services. Working with the Royal Borough of Greenwich and the Greater London Authority to determine whether data collected through the 'Sharing Cities Programme' could be made available in a data trust. Data collected included data about energy consumption, 'collected by sensors and devices in buildings', and 'information on parking spaces and charging bays for electric vehicles.' The ODI intended for this data sharing to help 'people, communities and organisations make better and more timely decisions', in contexts such as choosing an energy supplier, or deciding on which bus route to take.

---

[4] https://theodi.org/article/uks-first-data-trusts-to-tackle-illegal-wildlife-trade-and-food-waste/

An early-stage review of these projects provided insight into the potential long-term effectiveness of data trusts and the associated economic and social benefits.

2   Parameters for assessing data trusts

The government's support for data trusts must pass two basic tests: that there is a clearly identified need; and that the support is likely to be worth the cost[5]. The pilots provided initial evidence for an economic rationale for support for data trusts, but more evidence is needed.

2.1.1   Market failure

The sharing of data for economic benefit is hindered by a number of market failures, which data trusts can help to mitigate.

- Imperfect and asymmetric information: different parties who may benefit from better access to data do not have the information necessary to access it. Moreover, data sharing often goes against ingrained organisational instincts/practices. Organisations are used to treating data as an asset to be protected, not to be shared. Organisations outside the core 'data sector' may be unaware of the value of data and AI, or don't have the skills and infrastructure to use it productively. (M&A in the data sector may be a sign of savvy organisations both knowing where to find complementary data (held by their competitors) and being averse to open/sharing structures). Finally, data users (firms) typically have better information about the value of data than individual data subjects.

- External costs and benefits: positive externalities might lead to an amount of data sharing that is less than optimal for society at large. Innovation often has an element of positive externality. Conversely, insufficient data-sharing imposes external costs, as new products and services (some of which are free/have spillover effects, e.g. in the health sector, in planning and transport) do not come online as fast as they could. There is also an effect on trust, as more participants share data through the data trust, overall trust in responsible data sharing grows.

- Economies of scale and scope: AI development often benefits from economies of scale (more data = more precise predictions, better trained algorithms etc.) and scope (more inputs (sensors etc) can predict more outcomes).

- Network effects: participants in a sharing framework benefit more the more other participants there are (access to more/different data)

- Missing markets: some types of data are not traded/available to parties other than those who collect them.

- Barriers to entry: innovation in AI is made more difficult by proprietary access to data, which benefits organisations that collect data and disadvantages AI innovation outside the core 'data sector'.

2.1.2   Inefficiency

In addition to market failures (in the strict sense that they prevent markets from operating efficiently, thereby providing a strong rationale for remedial action by government), there are other

---

[5]   HM Treasury (2018). The Green Book: appraisal and evaluation in central government. Retrieved from https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent

frictions that data trusts can alleviate. Note that the dividing line between market failures and other frictions are not always very precise.

- High transaction costs: discoverability, lack of standardisation in data formats and access modes (this could be described as a coordination failure, a subset of imperfect information), compliance cost, lack of quality standards/certification.
  - □ Low discoverability and multiple potential competing data sources can create 'choice paralysis'[6] i.e. when people are faced with too many options, they 'freeze' and avoid making a decision at all. Therefore, users/ innovators may end up making less use of data than they should if they are confronted with too many data sources which are too difficult to rank or distinguish between.

- Trust deficit: the intangible nature of data, the complexity and opacity of data processing in different settings, evidence of risk of harm, together with cultural context and public discourse have resulted in an environment where individuals and organisations are often hostile to data sharing.

- Smaller organisations such as SMEs face a number of operational and structural obstacles that prevent them from making the best use of data, including the demands of data protection compliance and data security and access to hardware, software and skills. For many organisations outside the core data sector, accessing and using data for value creation does not come naturally and the organisational preconditions are often lacking. This may partly explain the lack of evidence for pent-up demand for data sharing from SMEs.

2.1.3 Evidence from the pilots

If a data trust doesn't effectively address barriers and inefficiencies, this would undermine the economic case for government support. Initial evidence is promising, especially on information asymmetries. Further research should seek to compare the effectiveness of data trusts in addressing market imperfections to other data sharing mechanisms.

Below we summarise the market failures or inefficiencies identified by the data trust pilots, which data trusts may help to address. There was no clear evidence of missing markets (demand for data to be shared through the trust) or direct external benefits (although a case could be made for external benefits from the application of some of the data included in the pilots, for example benefits from renewable energy procurement by councils or protection of ecosystems).

Table 1   Market failures/inefficiencies identified by pilots

	GLA/Greenwich pilot	**Food Waste pilot**	**Illegal Wildlife Trade pilot**
Market failures			
Imperfect or asymmetric information (e.g. inadequate information on how to access data or its value)	√ (stakeholder workshop participants pointed out that data trusts could help to co-ordinate	√ (users and researchers have an idea of the value of food waste data, but often cannot	

---

[6] Iyengar, S. and Lepper, M. (2000), When Choice is Demotivating: Can One Desire Too Much of a Good Thing? Journal of Personality and Social Psychology, 2000, Vol. 79, No. 6, 995-1006

	users/holders of data and maximise the value of data)	communicate this value to holders of data)	
Economies of scale			√ (more data enables more robust analysis and predictions)
Network effects		√ (the value of sharing data increases with the number of users/sharers)	√ (the value of sharing data increases with the number of users/sharers)
Barriers to entry		√ (innovation is impeded by proprietary access to data from e.g. manufacturers)	
Inefficiencies			
High transaction costs (e.g. low discoverability of data, lack of standardisation, high search costs)		√ (while e.g. retailers understand where data is held, lack of standardisation, quality controls and commonly-understood definitions were perceived to be a problem)	√ (data is produced by multiple agencies which often do not have adequate means of co-ordination)
Trust deficits where individuals and organisations may be averse to data-sharing	√ (there were concerns about the ability of organisations/corporations to protect privacy)	√ (retailers' perception was that manufacturers did not share their data)	
Obstacles disproportionately faced by SMEs or other smaller organisations/units		√ (SMEs or smaller organisations/researchers may face disproportionate difficulty negotiating for the use of data from holders)	√ (Given that data is held in multiple places, SMEs /individual researchers may face disproportionately high search/transactions costs)

Identifying clear market failures that a data trust can address is an important next step in building a case for continued government support, especially identifying if data trusts are uniquely or better suited to deal with these market failures than alternative data arrangements.

3 Effects of data trusts

The effects of data trusts can be described in terms of their effect on the inhibitors of data use and AI innovation.

3.1 Data trusts as solutions to principal-agent problems

Data sharing for social and economic benefit can be hampered by the agency problem, which can arise when an economic actor (the "agent") undertakes actions or decisions on behalf of the "principal": an organisation (or organisations), person or persons etc. (Jensen et al., 1976)[7]. For example, platforms or intermediaries may act as agents to facilitate data sharing on behalf of organisations, individuals or entities with an interest in ensuring that data can be made more useful.

The "principal-agent problem" arises when:

■   The interests of the principal and agent cannot be perfectly aligned; and
■   the principal cannot perfectly monitor the agent's actions, or the agent has information that the principal does not.

In these situations, the agent has an incentive to act in their own best interests rather than that of the principal (also known as "moral hazard"[8]).

The literature suggests a number of ways to deal with agency problems, but these solutions may be infeasible for the case of data sharing. For example, principals could incentivise agents, linking rewards to observed outcomes (Grossman and Hart, 1983)[9]. However, for incentives to be effective, the principal should have a clear understanding of the link between the agent's actions and observed outcomes. If the link is not clear, steep incentives can lead to unintended consequences such as the agent directing effort towards measurable outputs and neglecting other activities which the principal may actually value more. The phenomenon is known as "crowding out" of intrinsic motivation (Holmstrom and Milgrom, 1991)[10].

The literature on trust law suggests that trusts may minimise the agency problem by helping to ensure alignment of interests/motivation between principal and agent. For example, Halbach (1992) says that modern trust law has "flexibility and efficiency in the pursuit of the best interests of the beneficiaries within the settlor's legally permissible objectives"[11]. However, Sitkoff (2003) points out that care needs to be taken to ensure continued alignment of the objectives of principal and agent, as well as resolving any potential conflicts between the actions and objectives of the agent, the original intentions of the principal, and the best interests of the beneficiaries[12].

---

[7] Jensen, Michael C. and William H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics (October), 3(4): 305–360.

[8] Holmstrom, B. (1979), "Moral hazard and observability". Bell Journal of Economics, pp. 74–91

[9] Grossman, S., & Hart, O. (1983). An Analysis of the Principal-Agent Problem. Econometrica, 51(1), 7-45.

[10] Holmstrom, B., and Milgrom, P. (1991), "Multitask principal-agent analyses: incentive contracts, asset ownership and job design", Journal of Law, Economics and Organisation 24

[11] Halbach, supra note 21, at 1881.

[12] Sitkoff, R. (2003), An Agency Costs Theory of Trust Law, Law & Economics Working Papers: Law & Economics Working Papers Archive: 2003-2009

Therefore, data trusts (which, as explained by the ODI take the principles learned from trust law and apply them to data using other legal structures) may reduce the agency problem if independent stewardship of data more closely aligns the interests of principals and beneficiaries. However, contractual agreements between participants need to be carefully designed to ensure that the "rights and interests of the data trust, data providers and data subjects are respected.", as outlined in the General Legal Report. This includes ensuring that contractual agreements comprehensively cover re-use and follow-on use, which is not always easy or even feasible (for example with open data).

3.2 Increased use of data, productivity and efficiency

Data trusts can lower transaction costs for acquiring data for use in AI, making data more available and accessible. Therefore, organisations involved in AI R&D can use more data. For organisations that do not collect their own data, this effect can be large[13]. In addition, data trusts facilitate combining different data sources, which enables organisations to use the most appropriate data in their production/discovery processes.

Over a certain output range[14], any production process in which data is used as a factor of production (e.g. AI development) will benefit from more data being used. Being able to use the right data – type and quantity – makes the process more efficient. This benefit is particularly relevant to organisations that don't collect their own data directly from users, which typically includes start-ups.

3.3 Positive externalities

The aggregation of data for the purpose of AI development has the potential to benefit the population at large, rather than just the data subjects/parties who contributed data. For instance, in public health applications, the use of more data may improve quality of diagnoses (including for patients that did not provide data themselves).

3.4 AI innovation from combining data

The effect on innovation of combining data currently held in separate 'silos' is possibly very significant. There are clear potential benefits to combining different types of data in novel ways. For instance, this type of innovation can lead to superior diagnostics and treatments in the healthcare sector. Through data mobility, traditional healthcare information, such as medical histories, can be combined with emerging healthcare information, such as lifestyle monitoring. These two separate types of data provide a different picture of one's health. Therefore, combining them provides a more complete picture, which can feed into the health services provided to the individual.

Recombinant innovation can also lead to the rise of new products. A case study, in this regard, is the invention of satellite navigation. This invention was only possible when map data could be combined with location data.

---

[13] Note that using more data has an ambiguous impact on individuals as an increase in the amount data being used may increase privacy risks. See Wohlfarth (2017), Data Portability on the Internet: An Economic. 28th European Regional Conference of the International Telecommunications Society (ITS): "Competition and Regulation in the Information Age", Passau, Germany, July 30 - August 2, 2017.

[14] In practice, in any given process, the benefit of adding more data is likely to diminish at some point.

It is likely that certain data types will be more widely useable than others. Location data, for example, is likely to have beneficial applications across many different sectors.

By its very nature, it is difficult to model the economic effects of recombinant innovation. Because different data sources can be combined in many different ways, it is difficult to predict the way innovation will go. However, the potential of combining data sources makes it undoubtedly an important driver for future growth.

The flow of data between private and public sector organisations (in both directions) is likely necessary for unlocking innovation.

3.5   Competition

Data trusts may enable access to bulk data by organisations that don't collect their own data. This could increase competition for businesses with access to their own proprietary data sources (user data, monitoring data from sold devices, etc.)

It is an open question whether the current market structure (the presence of firms with market power, who have typically chosen M&A, licensing and similar approaches to gain access to data) would enable data trusts to become established alongside other approaches to data access unaided (i.e. without some form of compulsory participation by parts of the market).

3.6   Increased trust in data sharing

Confidence when engaging in data transactions depends on many factors, among them personal experience, the reputation of the counterparty as well as the strength of the regulatory framework. Data trusts provide a combination of better regulation and reduced counter-party risk.

From a rational choice perspective, institutions such as data trusts matter in this context because they are parameters in the calculation of whether or not to trust. "Legal arrangements which lend special assurance to particular expectations and make them sanctionable (…) lessen the risk of conferring trust."[15] "The apparent basis on which trust is given is necessarily various, including, for instance, the other's reputation, appearance, past performance, expert qualification or certification, as well as situational rule governance, availability of negative sanctions and so on"[16]. "Claims to trustworthiness are part of the context in which trust is given, not its basis"[17].

Experimental studies have shown that transacting based on confidence/trust (the counterparty's "tendency to keep promises") can increase social surplus if the alternative, relying on complete contracts, is costly[18].

---

[15] Luhmann, N. (1979). Trust and power. Chichester: Wiley.

[16] Barbalet, J. (2006). A characterisation of trust and its consequences. Social Contexts and Responses to Risk Network (SCARR) Working Paper 13-2006. Retrieved from: https://www.kent.ac.uk/scarr/publications/Barbalet%20Wk%20Paper(2)%2013.pdf

[17] Ibid.

[18] Chen, Y. (2000). Promises, trust, and contracts. Journal of Law, Economics, & Organization, 16(1), 209-232. doi:https://doi.org/10.1093/jleo/16.1.209 . Also, as Beresford et al. (2010) point out, the enforcement of contracts runs into problems because "many contracts involving personal data are incomplete or highly opaque, as they typically lack clear-cut information about secondary uses and sharing of personal information". See Beresford et al. (2010), Unwillingness to pay for privacy: A field experiment, Economics Letters 117 (2012) 25–27

However, making 'trust' salient through the introduction of a 'data trust' as an additional or new layer in the relationship between data providers (including individuals) and users is also problematic. It is well documented that individuals' reported preferences around data protection are often at odds with revealed preferences (observed behaviour)[19]. Developing data trusts based in part on the views of data subjects (and including indirect stakeholders, such as people who may be affected by decisions that are made with the help of AI) is likely to inflate the cost of the structure in response to reported, rather than proven concerns.

An issue that needs to be explored from a legal and economic angle is whether data trusts can reduce the friction arising from 'incomplete contracts' between data subjects/contributors and data users. Incomplete contracts have a risk of a 'hold-up problem' where it might be optimal for parties to co-operate, but they do not do so out of fear that they will give up their payoffs from a transaction by giving the other party more bargaining power.[20] Considerable benefits could result if data providers do not have to judge every possible data use, but leave this to the trustees within the parameters set by the data trust.

3.7 Data trusts as multi-sided markets

Data markets are often characterised by network externalities i.e. the value of participating increases with the number of participants[21]. However, neither data producers nor users can confidently predict the market size and so may not understand the value in sharing data. In other words, data producers only have an incentive to make data available if they think there are enough users, and users need available data to get value from it. But data producers don't know how many potential users there are, and users don't know the amount, varieties or quality of data that is available. This double-sided uncertainty impedes data sharing.

Data trusts can facilitate data sharing by identifying and linking data producers and users. Therefore, data trusts may be valuable when:

- Data production and use happen in different places; and
- Users and producers lack other means of identifying and co-ordinating with each other.

Therefore, data trusts can play a key role in increasing the visibility of data in the trust, and link organisations that would ordinarily not be in a position to interact or exchange information e.g. public and private sector, old and new economy etc. Data trusts may also circumvent some potential competition concerns of data platforms: specifically, that network effects allow companies to subsidise services to more price-sensitive customers, collecting data which is monetised elsewhere[22].

On the other hand, data trusts may be less useful if data producers and users are already visible to each other, understand the value of the data and/or have ways to communicate, bargain and co-ordinate with each other.

---

[19] See Barth et al. (2017), The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review, Telematics and Informatics Volume 34, Issue 7, November 2017, Pages 1038-1058

[20] See Rogerson, W.P. (1992). Contractual Solutions to the Hold-Up Problem. The Review of Economic Studies, 4(59), 777-793.

[21] Rochet, J-C., Tirole, J. (2003), Platform competition in two-sided markets, European Economic Association

[22] OECD (2016), Big Data: bringing competition policy to the digital era, retrieved from https://one.oecd.org/document/DAF/COMP/M(2016)2/ANN4/FINAL/en/pdf

Potentially the Illegal Wildlife Trade pilot is an example of circumstances where data production and use happen in different places and there is a need for some co-ordinated way to put data producers and users in touch, while the Food Waste pilot is an example where there is an existing repository of accessible, available data.[23]

Note, however, that in order to function effectively the data trust needs to 'get people on board' from both the producer and user side of the market. This could be done by, for example, acquiring (potentially purchasing) data from data producers, and (at least initially) charging users a licence fee. In turn, this implies that the data trust requires an ability to forecast or estimate the likely demand for data in order to select or target its investments.

4   Relevant benchmarks

Organisations have found other solutions to the problem of inadequate access to data. In addition to the approaches discussed in the ODI's synthesis report (data cooperatives, data commons, personal data stores and research partnerships), it is useful to compare data trusts to other approaches that are more common in the private sector, for example, acquiring data, merging organisations or project-specific data functions, and intermediaries.

To succeed in bringing together private and public sector data providers and users, data trusts need to offer benefits over and above what these approaches offer to their current users. Alternatively, making participation in data trusts mandatory for certain types of data would be necessary to achieve take-up. While this sets a high bar in terms of evidence (compulsory licensing as an anti-trust remedy), there may be cases where it is justified by the effects on competition and innovation, notably by smaller competitors, which would otherwise be the targets of takeovers by larger rivals.

Policy discourse in this area has recently focused on blocking or even unwinding acquisitions in the tech sector[24], However, economists have advanced arguments[25] that "in many cases, compulsory licensing is much better than blocking acquisitions"[26] Data trusts may be used as a vehicle for 'compulsory licensing' in the sense of making data that is needed for innovation available to a larger pool of potential innovators.

4.1   M&A, joint ventures, partnerships

4.1.1   Personal mobility

Acquisitions and project-based partnerships can come up with novel ways to share data and make it more useful. For example, BMW, Mercedes and Audi recently announced a joint venture in which they would share real-time driving data to develop app-based services like car-sharing in larger

---

[23] http://www.wrap.org.uk/content/food-waste-atlas

[24] Warren, E. (08 March 2019). Here's how we can break up Big Tech. Retrieved from https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c

[25] Bryan, K. and Hovenkamp, E. (2019). Antitrust limits on startup acquisitions. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3350064

[26] Gans, J. (13 March 2019). Compulsory licensing is better than blocking acquisitions (blog post). Digitopoly. Retrieved from https://digitopoly.org/2019/03/13/compulsory-licensing-is-better-than-blocking-acquisitions/

cities[27]. The automakers purchased mapping firm Here from Nokia to launch the project[28]. The acquisition will help the automakers pull together data from a variety of sources including forward-facing cameras on cars, probe and sensor data from connected cars, weather conditions and parking data from additional sources to build a picture of traffic conditions, parking space availability and information relevant to ride-sharing.

The acquisition is part of a landscape of partnerships between automotive manufacturers and technology services, for example BMW and Mercedes developing autonomous car technology together[29]; or Microsoft Azure's partnership with the Renault-Nissan Alliance to develop customer services and monitoring and security services to bolster autonomous driving capability[30].

Such arrangements substantially reduce frictions to acquiring/generating processing data, and have considerable efficiency benefits: pooling real-time driving data across users of multiple brands, for example, enables more robust analysis and predictions[31], benefitting more users than if data were kept in silos. The BMW-Mercedes-Audi partnership has announced that it plans to expand access to its mobility services to users from other automotive manufacturers, which would further increase the quality of the app's predictions[32]. Gains are even more compelling for large-scale undertakings developing autonomous driving technology, since combining resources helps to ensure uniform technical standards[33]. The technology could be used in a range of situations calling for granular disaggregated data from a number of sources: for example, town planning, circulation management[34], enabling individuals with mobility restrictions to plan routes etc.

A concern is that such data sharing arrangements may be anti-competitive. Investment-intensive technologies and services already have barriers to entry for new entrants and potential competitors, which present competition issues especially if the market is already relatively concentrated (Autorité de la concurrence and Bundeskartellamt, 2016[35]). Joint ventures such as the BMW-Mercedes-Audi partnership further reduce the number of competitors. While generating and processing large data quantities is not itself anti-competitive, the OECD points out that the infrastructure required to access and process large data volumes are "subject to economies of scale and scope"[36]. Furthermore, data sharing can intensify a "feedback loop" where users of services produce data which can be used to further tailor or improve the quality of services, which tends to benefit incumbents who already have access to a pool of service users/data producers[37]. Price

---

[27] https://www.boston.com/cars/car-news/2019/02/25/bmw-daimler-you-can-trust-us-with-car-sharing-data

[28] https://www.forbes.com/sites/federicoguerrini/2016/09/27/mapping-firm-here-launches-real-time-sharing-of-driving-data-among-connected-vehicles/#30f26f5c3c62

[29] https://www.motoringresearch.com/car-news/bmw-mercedes-benz-autonomous-self-driving-car-technology/

[30] https://www.forbes.com/sites/alexkonrad/2016/09/26/microsofts-partners-with-adobe-and-renault-nissan-to-boost-azure/#3173186911ff

[31] https://www.forbes.com/sites/federicoguerrini/2016/09/27/mapping-firm-here-launches-real-time-sharing-of-driving-data-among-connected-vehicles/#30f26f5c3c62

[32] Ibid.

[33] https://www.motoringresearch.com/car-news/bmw-mercedes-benz-autonomous-self-driving-car-technology/

[34] See also: Sidewalk Labs, which partnered with Waterfront Toronto and the local community in Toronto's Eastern Waterfront to design a district.

[35] Autorite de la concurrence and Bundeskartellamt (2016), Competition Law and Data, retrieved from http://www.autoritedelaconcurrence.fr/doc/reportcompetitionlawanddatafinal.pdf

[36] OECD (2016), Big Data: bringing competition policy to the digital era, retrieved from https://one.oecd.org/document/DAF/COMP/M(2016)2/ANN4/FINAL/en/pdf

[37] Ibid.

discrimination is a related issue that may arise when data sharing arrangements create or cement anti-competitive environments: for example, OECD (2016) points out that in multi-sided markets, dominant players have the ability to subsidise services to relatively price-sensitive customers, to the potential detriment of customers who are less able to search or advocate for themselves in transactions[38].[39]

However, the model of data acquisition and sharing described above is not restricted to large enterprises. Small and medium enterprises could form partnerships to pool capability and user information. Such partnerships could overcome entry barriers, exploit scale economies and increase competition.

A data trust might be a way to address the anti-competitive tendencies evident in the M&A activities of large corporates.

Therefore, data trusts could adopt features of already-existing data-sharing arrangements, allowing organisations to pool resources and capability to exploit economies of scale. However, as with all data-sharing arrangements, data trusts need to be careful to protect producers' and users' privacy, and as with all partnerships/market consolidation, competition may be a concern.

A contract structure seems appropriate in this setting, as NDAs, or standard contractual clauses under GDPR are already in widespread use in the corporate sector.

4.1.2   Fashion retail

In 2016, Launchmetrics[40] was created by merging Augure (a French-based marketing software firm) and Fashion GPS (a fashion-focussed ticketing and event management startup) to integrate their services and provide media intelligence and launch-to-market services to 1,100 clients from the fashion, luxury and cosmetics industries including Gucci, Topshop, Vogue, Louis Vuitton and Adidas. The merger has enabled the two firms to integrate not only technology platforms, but also data and customer insights and 'cross-pollinate' demand information from launch events to make more robust recommendations for launch-to-market strategy services.

Mergers and acquisitions within the fashion industry tend to unite complementary technologies, markets and skillsets as well as data. For example, on the wholesale side, Fashion Cloud merged with Nyon[41] to expand their product range and integrate ordering, product information and stock management capabilities for retailers. On the retail side, Nielsen's services integrate data from multiple sources including electronic point-of-sales data, sales and turnover data, primary research conducted in stores or through panels of shoppers, to provide robust business insights[42].

Insights from mergers in the fashion industry seem useful for data trusts organised to address complex and multi-dimensional problems requiring insights and data from multiple angles e.g. planning energy, traffic and structural needs for local communities.

---

[38] Ibid.

[39] Note that this practice does not necessarily lead to overall lower consumer welfare, but rather that some consumers may benefit from greater personalisation or targeting, while others experience harm

[40] https://www.launchmetrics.com/

[41] https://fashion.cloud/en/nyon-merger/

[42] https://portal.developer.nielsen.com/sites/default/files/Retail_Measurement_Data_for_System_Integrators.pdf

4.2   Statistical data repositories

One frequently-observed barrier to unlocking the value of data is its lack of discoverability, quality assurance and ease of access. Users often do not know what data is held and where, or what quality it is. A number of intermediaries in the public and private sector increase the visibility and searchability of data. Relevant examples here are national and international statistics agencies, or private initiatives that allow mixing private and public data sources on the same platform.

For example, Eurostat[43] presents data compiled from national statistical agencies[44], arranged by theme (e.g. economy and finance, environment and energy, science, technology and digital society etc.). The Wolfram Data Repository Ecosystem[45] presents a public data repository for access by both users and contributors; data repositories for both individual users and enterprises; and resources for research or other communities to partner with Wolfram to create a Community Data Repository.

These intermediaries reduce frictions by making data more discoverable. In addition, the intermediaries can often take responsibility for quality assurance, updating and maintaining metadata. However, this additional step requires substantial knowledge of the relevant standards, infrastructure to host and maintain large quantities of data, and capability to bring data producers on board.

A directory approach seems useful for data trusts organised along a 'big societal challenges' principle, where research and public benefit are in the foreground, and could build on the expertise of existing statistics agencies. This setup would require a more costly/institutional structure (likely the corporate model highlighted in the General Legal report).

5   Recommendations

Assess economic benefits as part of the rationale for government support

The review of the ODI's data trust pilots showed that the concept requires further testing to assess whether it is a viable new approach to data sharing that meets the requirements for further (perhaps ongoing) government support. A clear view of the economic effects of data trusts is thus required.

To achieve buy-in from different stakeholders the benefit minus the cost of the data trust needs to be greater than the benefit from data use available without the data trust. This has to be true for data contributors, data subjects and data users alike even though the difference can be marginal for some stakeholders, given the often very low cost involved (for a data subject the cost difference between having one's data in a data trust or in a non-trust database is likely to be more or less zero in many cases).

Crucially, the feature that distinguishes data trusts from other data sharing mechanisms, the independent stewardship function, also has to pass a cost-benefit test: does this 'additional functionality' of the data trust lead to greater benefits compared with the – presumably greater – cost.

---

[43] https://ec.europa.eu/eurostat

[44] For example, the UK's Office for National Statistics.

[45] http://www.wolfram.com/data-repository-ecosystem/

Evidence of net economic benefit is needed to provide a robust rationale for government support to develop data trusts. Data trusts that create additional burden for businesses and limit beneficial data flows would be difficult to reconcile with the government's vision for data trusts.

Recommendation 1:	Analyse the data trust pilots for additional evidence of market failures that can provide a rationale for government support. Design future pilots to incorporate an assessment of market failures.
Recommendation 2:	Incorporate a cost-benefit appraisal in the design of further exploratory work on data trusts.

Compare data trusts with current private sector approaches to data sharing and access

Evidence of (potential) economic benefit would be a useful guide to selecting/prioritising sectors/data types in which data trusts have the greatest promise. Relevant criteria are: data that is scarce or exclusive, has proven value, has many different potential applications. Mobility data seems a promising example: recent M&A activity and joint ventures by leading automotive firms provide evidence of demand.

M&A activity is an instructive benchmark for the data trust concept. This is where there is the most compelling evidence of economic benefit of access to data that is collected by other entities. Can data trusts enable superior benefits by widening data access, sharing benefits, and promoting consensus around data use?

Recommendation 3:	Explore if data trusts offer an alternative to approaches to data sharing/access used in the private sector (M&A, joint ventures).

Test whether trust is a competitive advantage over other data sharing approaches

A potentially important competitive advantage of data trusts over other forms of data sharing is that they engender greater trust in participants, thereby enabling more data to flow.

Recommendation 4:	Explore if data trusts can facilitate greater user trust without the need for narrow purpose limitations and burdensome consent mechanisms.

Consider implications for the setup of data trusts

In general, a focus on economic benefit sets a high bar for the cost-efficiency of a data trust. For a data trust that delivers benefits for small users (data contributors and data users, including SMEs and individuals), costs will have to be very low (likely zero for at least some user categories).

Robust evidence of benefits of a sufficient size would thus be needed to justify 'heavy' organisational structures involving specialist staff and dedicated infrastructure. However, for a data trust with a clear public purpose (the 'societal challenges' paradigm), such a structure might be justifiable.

Moreover, the cost profile and demand and pricing structure of a data trust needs to be analysed in greater detail. A corporate structure that integrates a number of technical functions in addition to the core stewardship function may be justified if the number of (paying) users is large enough or if alternative approaches (such as contracts) are relatively more costly due to the typical usage pattern (contracts for example may be less appropriate if participants move in and out of the data trust frequently). The economics of multi-sided markets shows that free access for certain users is

compatible with overall cost-recovery for a platform that uses differential prices to balance supply and demand from different types of users.

Recommendation 5:	Assess the demand and cost structure of data trusts to design viable funding models.

Consider the needs of SMEs

There is little evidence of pent-up demand for data sharing arrangements from SMEs. However, there is plenty of evidence that large corporations, including in traditional (non-tech) sectors are increasingly data-hungry. This might mean that SMEs may have other priorities besides access to data, or that SMEs still do not have the awareness, skills and capabilities to articulate their data needs.

This issue needs to be explored, given the importance of documenting the need for new data sharing arrangements to justify government support for data trusts. A data trust to which SMEs have access only in theory, because they lack the necessary skills, awareness and technology, might distort competition even more in favour or large, data-savvy businesses.

Whether a data trust is a model that can counteract the potential anti-competitive effects of these developments depends on take up by users who otherwise would not have access to the data (including SMEs, see above).

From a competition perspective, there may be a case for additional support, e.g. designation of certain data as an essential facility, mandatory licensing, etc.

Recommendation 6:	More research is needed into the ranking of obstacles to data access, especially for SMEs. A directory model for data trusts, with a focus on ease of access and discoverability may be appropriate to maximise SME participation
Recommendation 7:	Explore the legal and economic basis for using data trusts as a competition remedy in markets characterised by incumbent market power.

Forward Look: Ensure data trusts are useful for AI

The specific purpose of data trusts in contributing to the development of AI in the UK deserves closer attention in future work in this area. While the concept might have benefits unrelated to AI, the potential of AI as a source of future economic growth justifies greater risk-taking on the part of funders. An important question is if human stewards are equipped to deal with AIs as autonomous users of data. Potential issues are speed (queries submitted at high frequency), volume and complexity of data requests. It is easy to imagine AIs being involved in real-time problem solving (e.g. in emergency situations), where a bottle-neck caused by the need for human-speed decision-making may be very costly, to the point of erasing the benefit of AI. More broadly, as AIs become more sophisticated, their requirements for a data trust may diverge from those of humans (e.g. reduced need for data curation).

Recommendation 8:	In the light of the government's objectives for AI development in the UK, the setup of data trusts should be tested in relation to emerging and future AI capabilities.