



UNIVERSITY OF
CAMBRIDGE



Bennett Institute
for Public Policy
Cambridge

POLICY BRIEF

The National Data Library and public benefit

Jeni Tennison

**CONNECTED
BY DATA**

Author

Jeni Tennison OBE, Connected by Data

Suggested citation

Tennison, J. (2025). 'The National Data Library and public benefit.' Bennett Institute for Public Policy, University of Cambridge, & Connected by Data.

Date

March 2025

Acknowledgements

Thanks to Diane Coyle and Peter Wells for their comments on this policy brief,

About the Bennett Institute for Public Policy

The Bennett Institute for Public Policy at the University of Cambridge is committed to interdisciplinary academic and policy research into the major challenges facing the world, and to high-quality teaching of the knowledge and skills required in public service. Our research connects the world-leading work in technology and science at Cambridge with the economic and political dimensions of policymaking. We are committed to outstanding teaching, policy engagement, and to devising sustainable and long-lasting solutions.

bennettinstitute.cam.ac.uk

About Connected by Data

Connected by Data is a campaign to give communities a powerful say in decisions about data and AI to create a just, equitable and sustainable world. We bring together data experts, community facilitators and campaigners to help communities advocate for how the data affecting them should be used and governed in their best interest.

connectedbydata.org

Introduction

A National Data Library (NDL) will be developed over the next few years; it was announced [in the Labour manifesto](#) and again in the recent [AI Opportunities Action Plan](#) and [A Blueprint for modern digital government](#). Precisely what it looks like is yet to be seen. This policy brief is about what one particular feature of the initiative might (and should) look like, to satisfy public expectations about sharing data with commercial organisations, namely about how sharing data through the NDL should return value to the public.

The National Data Library

The [Labour manifesto](#) says “we will create a National Data Library (NDL) to bring together existing research programmes and help deliver data-driven public services, whilst maintaining strong safeguards and ensuring all of the public benefit”. This is written in the context of a section titled “Driving Innovation”, in a paragraph about supporting the development of the artificial intelligence (AI) sector.

The [AI Opportunities Action Plan](#) focuses on what the NDL should do to support innovators, recommending making five (unspecified) high-impact public datasets immediately available to the private sector and proactively collating data that is useful to innovators, such as “a copyright-cleared British media asset training data set, which can be licensed internationally at scale.”

[A Blueprint for a modern digital government](#), on the other hand, says the NDL will make it “easier to find and reuse data across public sector organisations; this supports better prevention, intervention and detection, and opens up data to industry, the voluntary sector, start-ups and academics to accelerate AI-driven innovation and boost growth.”

This gives at least three possible interpretations of what the NDL is supposed to achieve, all of which are also reflected in the Tony Blair Institute’s recent report, [Governing in the Age of AI: Building Britain’s National Data Library](#):

1. To help deliver data-driven public services by building the [public data infrastructure](#) that underpins them, perhaps building on previous “[Government as a Platform](#)” work carried out by the Government Digital Service (GDS) in the mid-2010s or the way Richard Pope talks about data infrastructure in his book, [Platformland: An Anatomy of Next Generation Public Services](#).
2. To help researchers and statisticians support evidence-based policymaking through cross-departmental data analytics, building on existing work on the [Integrated Data Service](#), as well as initiatives such as Administrative Data Research UK ([ADR UK](#)) and Health Data Research UK ([HDR UK](#)), and the Central Digital and Data Office’s (CDDO) more recent [Data Marketplace](#).

3. To help AI developers create new tools and services by providing large datasets – including content such as documents, not usually thought of as data – on which to train machine learning algorithms.

Whichever direction it takes, the NDL is going to have to work out, as a priority, how to navigate sharing public data with private sector organisations.

Private and public benefits

Frequently phrased as “unlocking the value of public data”, reducing existing barriers to access to public sector data – whether to support the delivery of public services, research, or AI development – promises several benefits.

First, there is the promise of innovation and economic growth – the primary mission of this Government – as companies carry out research and develop new products and services.

Second, there is a hope that at least some of those products and services could be used in the public sector to support productivity enhancements and other efficiencies to reduce the costs of delivery of public services, and/or to improve their quality. For example, the recent [£4m investment](#) in creating a dataset to support the training of AI systems that can mark children’s schoolwork aims to catalyse the creation of services that schools can then use to reduce their teaching costs, and improve teacher retention by reducing the amount of time teachers spend on marking.

Finally, there is a recognition that the public sector alone does not have the capacity to fully exploit the data it holds. Government cannot – and should not – develop all the possible apps, tools or research that society might benefit from. Enabling greater access to public data ensures we are not losing out on the public value that could be generated from it.

These public value benefits depend on commercial actors being important users of the NDL, whatever form it takes:

- As well as supporting commissioned delivery of public services, “Government as a Platform” makes it easier for the private sector to create new tools and services that use publicly held data: the classic examples being [Citymapper using Transport for London data](#), or the [start-ups supported by Geovation](#) to build services using Ordnance Survey data.
- Researcher access to public data includes access for R&D groups within commercial companies, such as enabling pharmaceutical firms to develop new treatments and medicines using health data held by the National Health Service (NHS).
- While there are [AI developers within government](#), most AI development is carried out by companies who are able to pay for talent in a highly competitive marketplace, such as [Faculty AI](#), [DeepMind](#), or various specialised start-ups and scale-ups.

In other words, the NDL cannot achieve its economic and social value goals – and society will not benefit as much as it might from AI – without commercial exploitation of public data.

Public concerns

While commercial exploitation is very likely to be an essential part of the NDL, one of the most consistent findings from [public attitudes research](#) is that the public do not feel comfortable with the private sector using public data. This is for three main reasons: privacy, ethics and exploitation.

Privacy

The first objection to commercial use of public data arises from privacy risks and concerns. Many people worry that public data that includes data about them and is made available to the private sector might end up in the hands of companies such as insurance providers, credit reference agencies and potential employers, who then use it to set prices, curtail access to services, or make hiring and firing decisions without an individual's knowledge or consent.

These concerns can be and are being largely addressed through legal, governance and technological controls implementing the [five safes principles](#) of safe data, safe projects, safe people, safe settings and safe outputs. Legal frameworks such as the [Data Protection Act 2018](#) already place constraints on how personal data can be shared and used. Data access agreements, [researcher accreditation](#), governance panels and so on all provide governance controls over who gets access, and for what purpose, to minimise risks. A well-designed NDL will also encourage the use of technological controls by making the best use of [privacy-enhancing technologies](#), including secure research environments, rather than passing personal data between organisations, and could also provide information and controls to citizens about when and with whom obviously identifiable personal data is shared.

In public deliberations, providing information about these kinds of governance measures tends to alleviate most privacy concerns, at least to the extent that, on balance, most people think the collective benefits of providing greater access to data are worth the privacy risks that remain.

The NDL will require a sophisticated data governance framework, as well as proactive communication and ongoing engagement with the public and civil society, to navigate these privacy concerns. We can see the beginnings of this in the way that ministerial technology proponents such as Wes Streeting emphasise protection and anonymisation when he talks about the [move from analogue to digital NHS services](#). But public concerns do not stop at privacy.

Ethics

A second reason that the public may be wary of companies using public data is that they are worried it might be used for unethical purposes.

The public is generally supportive of the use of public data for public good, even if they are not direct beneficiaries themselves. However, the specifics matter a lot. No category of data user is guaranteed to use data in ways the public would classify as public good – governments may be authoritarian; [researchers may engage in race science](#) – but the underlying profit motive of commercial companies means they do not have an underpinning public good incentive for their work, leading to particular suspicion.

As with privacy, the NDG can only address these concerns with a strong data governance framework that considers the ethics of proposed uses of data, and the ethics of the companies accessing data more generally.

The [people and communities affected by the uses of data](#) may be different from those exposed to privacy risks. The participation and judgement of both those most likely to be affected by particular uses of data, and those the data is about, is vital in making decisions about ethical acceptability of both uses and companies.

Exploitation

The final challenge around commercial use of public data is that the public is generally uncomfortable with private companies profiting (excessively) from exploiting public data. A recent salient example is the findings from [public attitudes research about the use of AI in education](#) where “there was widespread consensus that, if profit were to be generated through the use of pupils’ work to enhance AI in education, schools would be the preferred beneficiaries, and resistance to the idea of tech companies being the sole profiteers”.

People feel they have a stake in public data. Its collection and maintenance are paid for by taxpayers. Public data may be about them as individuals (such as health data or education data), but even where it isn’t, it is about our nation and society: our land, our neighbourhoods, our lives. People have a feeling of sovereignty and collective ownership of a national asset.

Across public attitudes research, we see that while the public is generally enthusiastic about public data being used for public benefit, the privatisation of value arising from this national asset feels unfair. This may be due to a zero-sum mentality that assumes that private sector benefit comes at a cost to the public sector. It may also be that value returned to the public purse through general taxation on increasingly profitable companies, or savings through improved public services, feels too distant and uncertain, compared to more immediate private sector profits.

Public attitudes research shows these concerns come to the fore when profit is excessive, the company is large, or it is owned by billionaires. In our research on [attitudes to sharing health data](#) we found that for some, the location of the company matters. For example, an organisation based in the UK, paying corporation tax and providing employment in the UK is more acceptable than one based in Silicon Valley.

The nature of the ongoing relationship between companies and the public sector also matters. The public is sensitive to [dodgy deals, non-competitive contracts, and vendor lock-in](#). Companies that exploit public data to create products and services that are then sold back into the public sector at a profit are regarded by some as particularly exploitative. For example, this can apply to pharmaceutical companies that use public data to develop life-saving treatments which are then sold at a profit back into the NHS. Another example would be EdTech companies that benefit from government investment in the creation of training datasets and then profit by selling their tools into state schools.

To address these kinds of concerns, the NDL will need to put in place a value sharing framework, and clear monitoring and communications, to ensure the public see that value is being returned to the public sector when the private sector gets access to public data.

Note: Political choices

There is frequently a substantial minority for whom government reassurances about privacy, ethics and fairness are insufficient. Civil society groups will rightly point to patchy enforcement of either contractual or data protection requirements; lack of penalties for those who break them; lack of proactive audits of data usage by third parties; and the insufficiencies of anonymisation techniques. Public attitudes around data sharing may also signal a deeper distrust around corruption and collusion between the government and the private sector, or concerns that (particularly non-legislative) rules and controls put in place by one government can be readily overridden by the next. They may also reflect longstanding inequalities and injustices experienced by minoritised communities, leading to scepticism and fear.

Most of the public do not have the benefit of the additional information and discussion that is shared during the deliberative exercises that public authorities use to shape and legitimise their approaches to data sharing. The public as a whole remains largely unaware of the legal, governance and technological controls that are put around public data access, or how value is returned to the public sector when companies use public data. The decisions and actions most of the public take – which may include withholding consent for data sharing; providing misleading information; or avoiding public services altogether – are based on their fears and concerns, influenced by public narratives, rather than a deep consideration of the risks and benefits.

Whatever governance or value sharing frameworks are put in place, then, it is of vital importance both that they operate effectively *and that they are seen to do so*. For example, there should be proactive transparency around requests for data access, the checks made on requestors, and the agreements entered into. Similarly, communications should highlight when organisations are refused access, to illustrate the presence of, and effectiveness of, governance controls and value sharing conditions. The NDL should proactively work with critical stakeholders, support routine independent reviews, and pre-bunk likely concerns and misconceptions.

Value sharing approaches

The remainder of this brief will focus on value sharing approaches that aim to ensure that the public gets a fair return from organisations that use public data, as this is one of the least-discussed aspects of the NDL. An effective value sharing approach should:

- **Encourage use of public data**, including by being easy for organisations to understand and administer
- **Provide a return on investment for the public sector**, offsetting at least some of the costs of supporting the NDL infrastructure and minimising administrative costs

- **Promote equitable innovation and economic growth in the UK**, which might mean particularly encouraging smaller, home-grown businesses and a dynamic economy, rather than favouring large US-based companies
- **Create social value**, particularly towards this Government's other missions, such as achieving Net Zero, or unlocking opportunity for all
- **Build public trust** by being easily explainable, avoiding misaligned incentives that encourage the breaking of governance guardrails, and feeling like a fair exchange

With these criteria in mind, it's important to explore different approaches that could be adopted.

Financial returns

There are six models through which the government can get a financial return: (a) pay to access; (b) discounts; (c) royalties, (d) equity; (e) levies; and (f) taxation.

a. Pay to access

Pay-to-access models charge data users for access to data. These are the most obvious models to use, not least because they cover at least some of the cost to the taxpayer of creating and maintaining data sets, but have a range of drawbacks.

First, it is hard to set the right price for access to data with no established methodologies for the public sector to do so ([Coyle & Manley, 2024](#)). You can base price on costs, but most costs are fixed and the marginal cost of supply (how much each additional data user costs the data supplier) tends to zero, which means it's hard to know how to spread fixed infrastructure costs across users without knowing how many there will be. Basing charges on what companies are prepared to pay can also be difficult: value of data to an organisation can be hugely variable depending on the type of use, and it can be hard for businesses to know how valuable data will be without having tried to use it.

Second, charging for data naturally creates a barrier to access to data, rather than encouraging it. What's more, it tends to favour larger and more liquid organisations over smaller and cash-poor start-ups, who may be priced out of access. The Tony Blair Institute recommends a "tiered data-access fee structure ... based on the type of organisation or individual accessing the data". However, in practice schemes that set lower costs for start-ups (such as [Ordnance Survey's Data Exploration Licence](#), which provides free access for a year) can have complex qualification criteria; present companies with a cliff edge when higher fees kick in; and be costly for data providers to administer. Pay-to-access models also incentivise data providers to place restrictive conditions on (or charge higher prices for) the onward sharing of derived data, as it reduces the number of direct customers, and therefore their financial returns, even when allowing such products would support other goals of data sharing, such as promoting innovation or increasing social value.

Finally, while pay-to-access models have the advantage of being easy for the public to understand, and provide a very clear and direct link between accessing data and providing financial value back to the public sector, they can also easily be characterised as “selling access to data”. This narrative can make members of the public think that the government is sharing more data simply to raise more money, rather than to create social value and for public benefit. It can also make them question whether this financial return incentivises the public sector to share more data than it would otherwise – perhaps loosening governance structures to enable data access in return for some cash.

In short, pay-to-access models have a number of flaws that undermine the goals of a ND. These same reasons formed some of the [rationale](#) for the government policy shift to free-to-access open data for non-personal data and the introduction of the [Open Government Licence](#) in 2010, and are reflected in Annex 6.2 of The Treasury’s [Managing Public Money guidance](#). So what are some other options?

Note: Paying for services

We are making a distinction here between addressing feelings of exploitation by paying for access to data, and providing sustainability through a business model that may include paying for services associated with data.

Several public data holders, in particular those whose primary task is stewarding data, have other revenue streams. Some, such as Companies House and Land Registry, charge fees to those who update the data they hold, for example. Public sector bodies may charge to provide additional added-value services around data, such as converting data into different formats, enabling subscriptions to feeds containing changes to data over time, or providing Application Programming Interface (API) access with reliability guarantees. The ND might charge administration fees for processing requests to access data, separately from any additional value exchange required if the request is granted.

The prices for these kinds of services can be relatively straightforwardly set based on the work it takes to service them.

b. Discounts

In some cases, the public sector may be able to secure discounts for (or even free access to) products and services created using public data. For example, if a pharmaceutical company develops a new drug using NHS data, part of the deal could be that it provides access to that treatment at a discount, perhaps set by limiting its profit margin on that drug when sold to the NHS. This would prevent businesses from profiting excessively from public procurement when they have already benefited from access to public data, one of the concerns expressed in public attitudes research.

There is some precedent for this kind of requirement: in its [standard grant agreement](#), the Advanced Research + Invention Agency (ARIA) requires that grant recipients “ensure that any Crown Body is able to procure any products and services that are covered by, use, or incorporate the relevant Foreground IP [Intellectual Property] within a reasonable time period at prices that are no higher than those offered to or agreed with any other customer for equivalent quantities, and in suitable quantities for the Crown Body’s requirements.”

The biggest challenges with this approach are first that it only applies in situations where the public sector is a customer; and second, the administrative burden it places on both companies and the government, who would need to enforce the restriction. Large companies might find it hard to pinpoint the profit margin of (digital) products or services – as opposed to fair returns on upfront research and investment costs. They may consequently seek alternative data sources that would not incur a restriction in that profit margin. Organisationally, the NDLP might find it hard to police where services fuelled by the data shared by one public body are procured by other public sector bodies, for example by individual schools or NHS trusts, to ensure a restriction on profits was being honoured.

c. Royalties

A royalty scheme would provide for payments back to the public sector when services are created using data shared with a company. For example, if an agency used administrative data from the Department for Work and Pensions (DWP) to identify vulnerable consumers for utility providers such as electricity and water companies, it could be required to return a percentage of what it charged for supplying that service to those companies, as a royalty fee.

The Tony Blair Institute report, [Governing in the Age of AI: Building Britain’s National Data Library](#), suggests that a royalty scheme could be applied to non-UK entities, based on the provisions in the [ARIA grant agreement](#), which require a 25% royalty fee when grant-funded products and services are either developed by non-UK entities, or sold to them. There’s no particular reason to restrict royalties to non-UK entities, though like other models, differentiating returns based on whether the beneficiary is UK-based or not, is supported by some public attitudes research.

Royalties have the advantage of applicability to all kinds of organisations (unlike discounts), and scaling with sales. However, they have similar disadvantages as discounts: they may be hard to administer and hard to enforce. They may have to be negotiated on a case-by-case basis, based on the contribution of public data to the development of the product or service. They also would not apply if public data were used within a company, or to provide free products and services rather than commercialised ones.

d. Equity

Taking equity can be another option for providing a financial return to the public sector. This is most suited to deals with start-ups whose primary product is built around access to public data, such as AI products trained on education data. Taking equity doesn't require the company to

have the up-front cash that might be needed under a pay-to-access model, nor incur the cliff-edge charges those models can create. If the company succeeds, the government can sell the equity at a profit.

There are precedents for the government taking equity in small companies. The British Business Bank – fully owned by the government – holds about £700m equity in just over 700 companies in its [Future Fund portfolio](#). The creation of the [National Wealth Fund](#) demonstrates the new Government looks favourably on these kinds of strategic, long-term investments.

From a public perception perspective, taking equity in UK start-ups using public data might be seen positively as an investment in the UK economy as well as a source for future funds. Constraints on which companies the government is prepared to invest in (see for example the [list of exclusions for Start Up Loans](#)) may provide a welcome additional governance check on the ethics of such businesses.

On the other hand, using equity as a mechanism for returning value to the public sector lacks the straightforward immediacy of cash that can directly support public sector costs. Returns are likely to take time to mature, and the link between public data access and those eventual returns may be difficult to explain to the public.

e. Levies

Levies are a mechanism for targeting taxes on particular categories of business with the explicit purpose of paying for the development or maintenance of infrastructure they rely on.

In the case of data or data-sharing infrastructure, organisations targeted by the levy would be those who might want to (or who already) access public data. For example, pharmaceutical companies operating in the UK could be required to pay a levy to support infrastructure that enables them to access health data for medical research. Lawtech companies might be charged a levy for the maintenance of legislation and case law databases, and so on.

One of the nice features of levies is that they can encourage the use of the thing being supported by the levy. The [Apprenticeship Levy](#) is an example of this: large organisations pay a levy, but any apprentices they take on can be paid for from their levy contributions. This provides a “use it or lose it” motivation for large businesses to take on apprentices. A data infrastructure levy could similarly motivate organisations who paid it to make use of the data infrastructure they are contributing to.

Levies also provide a scheme that is relatively easy to administer, with predictable costs to businesses and revenue for the government, while enabling data to be “free” at the point of access. They can be designed in ways that are progressive, such as only charging companies over a certain size. They do not introduce misaligned incentives: public authorities will not lose money if they refuse access to data for particular purposes, so will not be motivated to loosen governance controls. They are also relatively easy to explain.

The biggest challenge to using levies is likely to be finding ways of defining the categories of organisations targeted by them. Some datasets may have specific types of users: EdTech companies using curriculum data, for example. But other datasets are used across industries – core national data infrastructure such as geospatial information or national statistics, for example – and some of the largest tech firms operate across multiple sectors. Existing formalised approaches for digital companies include the [Digital Services Tax](#) (a 2% tax on UK revenues placed on large search engines, social media companies and online marketplaces) and [Strategic Market Status](#) under the [Digital Markets, Competition and Consumers Act 2024](#), which may be assigned to large digital firms with entrenched market power a position of strategic significance. It may be that these designations could be reused, or similar processes applied to identify which organisations should be targeted by levies for particular kinds of data.

f. Taxation

General taxation returns funds from profitable businesses to the public sector, without the same targeting towards a particular industry, project or infrastructure as a levy.

As mentioned previously, general taxation has several advantages such as being easy to administer for both businesses and government, and progressive. Recognising data as a new form of infrastructure, and understanding that its provision is a public sector duty that underpins our economy, just like roads do, makes general taxation the obvious way to pay for it.

However, using general taxation to fund data infrastructure depends on the government using tax revenue raised on that data infrastructure. In practice, tax revenue has frequently been targeted on the urgent needs of delivery of public services instead. Even as the current government targets capital investment in its spending plans, this is more likely to favour crumbling schools and hospitals than infrastructure to support data access. This is only likely to shift with legislative changes that recognise the maintenance of this infrastructure as a public sector duty.

Its untargeted nature also means general taxation might not satisfy the public's desire to see reciprocity for the use of public data. Corporation tax may be seen as something companies have to pay anyway, not an additional contribution in return for data access. Targeted consultation with the public would be necessary to determine public attitudes on this question.

Non-financial returns

While returning financial value is important, there are other kinds of non-financial conditions that the NDG could place around access to public data by the private sector, in ways that return value to the public sector.

In her paper [Rethinking the social contract between the state and business](#), Mariana Mazzucato recommends introducing **conditionalities** when governments invest in businesses:

Conditionalities are one powerful tool that governments can use to co-shape investment and cocreate markets with the private sector. When companies benefit from public investments in the form of subsidies, guarantees, loans, bailouts or procurement contracts, conditions can be attached to help shape innovation and direct growth so that it achieves the greatest public benefit. For example, procurement can be made conditional on greener supply chains, reinvestment of profits and better working conditions.

Importantly, there are different types of conditionalities governments can leverage and this working paper addresses four of them: (1) conditionalities related to **access**, where equitable and affordable access to the resulting products and services is ensured; (2) conditionalities related to **directionality**, notably related to the **green transition**, where firms' activities are directed towards climate-friendly goals, and they intentionally use green options and reduce negative environmental impacts, and to the improvement of **labour conditions**, where productive employment opportunities are created by firms, measured not just in quantity but also in quality, and diversity and equity are embraced; (3) **profit-sharing** conditionalities, where profitable firms share royalties or equity with government and may be incentivised to leverage their profits through acquisition of government shares; and (4) conditionalities related to **reinvestment**, where profits gained are reinvested into productive activities and R&D for longer term benefit, avoiding financialisation; (Laplaine and Mazzucato 2020). In the context of the revival of industrial policy across nations, it is critical to reflect on how best to build accountability and alignment with key policy goals into government support.

Any of these can apply to organisations accessing public data as well as money from the public purse.

There are three data-specific conditions that could be usefully applied to help meet some of the requirements the [AI Opportunities Action Plan](#) places on the NDLP, namely enhancing the quality of high-value datasets; promoting access to privately held data; and providing coupled compute and data to researchers and start-ups.

Data quality improvements

A lot of public data is in a poor state. It may be out of date, lack standardisation, be incomplete, unstructured and so on. It takes investment to improve the quality of public data, and it's the kind of up-front and behind-the-scenes cost that has proved hard for the public sector to justify.

The end result of poor-quality public data is not only that it's hard for the public sector to use, but that third party users accessing it each have to spend effort cleaning and tidying it. This is duplicative and wasted effort. It also leads to complex and potentially buggy data cleansing workflows when source data is updated.

One way in which a company could return value to the public sector would be to share back the results of any data cleansing that it carries out on public data. This would both reduce the

duplication of effort by other data re-users, and provide higher quality data for the public sector to rely on.

Making this form of reciprocity work isn't as simple as incorporating any and all changes or additions made to data back into the original public data. As with any such operations, there would need to be rigorous validation of the process and the results. This validation is also beneficial for the company supplying cleansed data.

In some cases, the risks of introducing errors may be too high, or the validation too costly. Even then, being able to share, with appropriate warnings and caveats, an unofficial cleaned version may spare other data users' time.

By way of example, this is the model we pursued to get legislation up to date on legislation.gov.uk. Rather than the government and commercial legislation publishers all working individually to apply changes to legislation and keep the data up to date, the [Expert Participation Programme](#) was designed to enable organisations to share that effort and create a single official source of high quality data. (In this case, the data was available openly rather than being restricted, so companies weren't massively incentivised to take part.)

Aside from the cost of validating corrections to public data, public sector organisations would need to have some auditing capacity to ensure that companies are passing on all the quality improvements they make to data. Whether the additional costs of these services (over and above those put in place to provide access to data in the first place) are worthwhile depends on the importance and value of the data to the public sector.

Private sector data sharing

Organisations that want to get hold of public data sometimes have data themselves that others could benefit from. So, another form of reciprocity is to use access to public data to unlock access to privately-held data. Public sector bodies could say that only those companies who are prepared to share some of their own relevant data can access public data.

The Greater London Authority has some nice examples of this working. The [Infrastructure Mapping Application](#), for example, brings together data from local councils and utility suppliers to help all these stakeholders coordinate when and where they upgrade their infrastructure. This reduces roadworks, saving everyone time, cost and inconvenience.

The public sector does already have other levers through which it can gain access to private sector data. [Section 80 of the Digital Economy Act](#) in principle enables the UK Statistics Authority to get access to data held by larger companies (and charities) to enable the creation of statistics. Regulators frequently have powers to require organisations to share data about their activities. Smart Data schemes such as Open Banking require that companies publish or share data, such as about their products or branches. [Part 3](#) of the new [Data \(Access and Use\) Bill](#) includes provisions that enable the construction of a National Underground Asset Register

(NUAR), such as requiring organisations that own street-level infrastructure to maintain records about it within the NUAR.

Given these examples, if access to privately held data were shared in return for access to public data, members of the public might reasonably ask why companies need to be given that access when public sector bodies have these other mechanisms to get hold of private data without giving public data away. (The answer may be that access is too hard to enforce and monitor.) This kind of data-for-data trade could also mean that incentives around the sharing of more sensitive public data are misaligned, with the government sharing more than it would otherwise, in order to access private sector data.

Members of the public might also be concerned about the public sector getting hold of privately held data, particularly administrative data about individuals. For example, there have been [outcries over plans for the Department of Work and Pensions to access data about the bank accounts of benefit claimants](#). These concerns mean the NDL should avoid this form of reciprocity for personal data.

Access to compute

A final form of reciprocity that private sector organisations may be able to offer is public compute.

Compute refers to the hardware, software and infrastructure that enables data storage and processing. [The Ada Lovelace Institute](#) defines compute as “the public provision of compute, either by public entities or by private entities receiving public funding or investment”. While this definition specifies that public compute from private entities might occur in exchange for public funding or investment, access to public data could be another trigger.

In other words, tech companies that run or otherwise have access to data centres could be required to grant public sector organisations, academics, civil society and start-ups free data processing and/or storage in return for access to public data.

This form of reciprocity is most applicable to the kinds of larger AI companies that the public wants to see the most from. Companies could be required to provide free data processing capacity over just the data they get access to, enabling others to also use it to train AI models. This would be particularly applicable to large and unwieldy datasets that smaller organisations might otherwise lack the resources to be able to usefully or securely handle. As long as this access did not circumvent governance controls (i.e. the public sector steward of the data was still the one making decisions about who could access it), this could provide several benefits, including making it cheaper for the public sector to provide good access to the data.

Alternatively, companies could be required to provide more general public compute, not constrained to the specific datasets they access. This is potentially more useful, because of the additional flexibility in the capability that would be made available, but may also be challenging to set limits around. Answering how much compute should be made available in

return for access to a particular public dataset returns us to the tricky question of how to value data.

Conclusion

Sharing public data should bring benefits back to the public sector, but there are plenty of alternative ways to do it than getting companies to pay for access.

There is a sweet spot between different kinds of value exchange. Too distant, such as general taxation, and the public doesn't see the return happening. Too direct with the organisation sharing data, such as payments or reciprocal data sharing, and measures may be distrusted on the basis of misaligned incentives: there will be suspicion that governance controls may be loosened to enable the public sector to benefit.

Different forms of value exchange are appropriate for different kinds of organisations. Taking equity only works for early-stage businesses; requiring the provision of public compute can only be asked of large tech companies; discounts can only be required from those who sell services back to the private sector. Different kinds of data can also open up the possibility of different approaches: collaborative maintenance for low quality or rapidly changing data; public compute over large and unwieldy datasets. A value exchange framework should treat different kinds of organisations and data differently, and enable experimentation, rather than adopting a single approach across all public sector data sharing.

Finally, to reiterate, regardless of the value exchange framework the NDLP ultimately puts in place, it will need to be able to communicate how access to public data provides value back to the public. Promises of growth through innovation might not be persuasive to the public, who worry about their privacy, the ethical use of our data, and who ultimately gets value from it. Checking different approaches to value exchange with relevant publics will be essential to avoid backlashes and damaging trust.



UNIVERSITY OF
CAMBRIDGE



Bennett Institute
for Public Policy
Cambridge

bennettinstitute.cam.ac.uk

CONNECTED
BY DATA