Written evidence submitted by The Open Data Institute (DDA0042)

About the ODI: The ODI is a non-profit with a mission to work with companies and governments to build an open, trustworthy data ecosystem. Founded in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt, we work with a range of organisations, governments, public bodies and civil society to create a world where data works for everyone.
About this evidence submission: We're submitting evidence to this inquiry in order to share our expertise on trusted and trustworthy data use, and on effective data sharing across and between Government, industry, civil society and academia.

Key messages (1,000wds)

Data as infrastructure and the nature of data ecosystems

- **Data as a form of infrastructure is integral to our society**: it underpins operations, policy, and decision-making across government, businesses, and civil society, and so plays an essential role in our day-to-day lives.[1] Access to high-quality data also makes new kinds of

---

[1] The ODI (2017), 'ODI welcomes UK National Infrastructure Commission recognition that data is infrastructure'; and Dodds, L. & Wells, P. 'Issues in Open Data - Data Infrastructure', in *The State of Open Data* (2019)

products, services, and activities specific to data – such as advanced data analytics and digital technology like artificial intelligence (AI) – possible. **Because of this, major national data assets stewarded by government should be viewed as a kind of public good, and must be representative of our society and also benefit all society.**[2] [3] In the UK health and care domain, high profile examples of gaps in this key national data infrastructure include the fragmentation of social care data across multiple care providers across private, public and third sectors;[4] and the absence of adequate data infrastructure to support social prescribing.[5]

- **Data ecosystems flow across sectors and across domains,** cutting across distinctions such as private sector, public sector and civil society. A category such as **"health data" can be an artificial distinction:** patient records could potentially be used for non-health purposes, such as creating entertainment products and services; and data such as mobility data, about traffic or public transport use, can be used for public health interventions. For example, in our work with the World Health Organisation (WHO) we've **mapped a spectrum of social, economic and environmental data in the health data landscape**[6]; and the experience of Covid-19 has demonstrated that data such as **mobility data or internet search data can inform public health interventions**.[7]"

- Because of this, it is **important that Government's data practices are joined-up across across Departments for the full benefits for public health**. But the National Audit Office, reflecting on the role of weak data infrastructure in the Windrush scandal and the Carer's Allowance scandal, observed that 'despite years of effort and many well-documented failures, government has lacked clear and sustained strategic leadership on data'[8], and our research has found a proliferation of over 100 data initiatives across UK Government[9]. Government must also recognise that the loss of public trust in Government data practices in one domain, such as health data, may have **unintended consequences** for public trust in Government data practices in other domains.

- **Government is one of the largest and most powerful agents in a national data ecosystem**, collecting and stewarding significant datasets across sectors and communities.[10] Therefore, the data practices of the government matter for how a country's national data ecosystem develops overall, impacting other sectors and domains too. In our mapping of the landscape of health data policies across the European region (covering the EC and its member states, Israel, Norway, Switzerland and the UK), we found that **the UK was among the weakest in its approach to ethics**[11].

Re-thinking data governance and data skills

- **The concept of "data ownership" - such as "my data" - is not fit for purpose:** for example, genetic data about an individual is also genetic data about their biological family; and so a **better framing is around data rights and data responsibilities, such as data stewardship**[12]. The UK's current data protection regime is based on the General Data

---

[2] The ODI (2021), 'Inclusive data: perspectives from a roundtable discussion'

[3] The ODI (2021), 'The ODI responds to the UK Statistics Authority Inclusive Data Consultation'

[4] Future Care Capital (2019), 'Data that cares'

[5] The ODI & Frontier Economics (2021), 'The role of data in unlocking the potential of social prescribing'

[6] The ODI (2021), 'Data landscape and health data governance: pre-reads for World Health Organisation global data governance summit'

[7] *Nature* (2021), 'Public mobility data enables COVID-19 forecasting and management at local and global scales'; and *Health Information Research* (2020), 'Prediction of COVID-19 Outbreaks Using Google Trends in India: A Retrospective Analysis'

[8] National Audit Office (2019) 'Challenges in using data across government. Report by the Comptroller and Auditor General'

[9] The ODI (2021), 'Mapping data in the UK government'

[10] The ODI (2021), 'Spending Review 2021: Investing in data to build back better'

[11] The ODI (2021), 'Discover which European countries are ready for the secondary use of health data'

Protection Regime (GDPR) which originated as a consumer right for individuals; consequently, **the framework doesn't take into consideration collective rights (such as cultural identity), collective harms (such as environmental damage or harms caused by unintended consequences of AI systems), or collective benefits (such as public health)[13].**

- **The concept of data governance should go beyond box-ticking exercises of protecting data privacy, and also include an obligation to get the best social and economic outcomes from data.** This could include duties on data stewards to share data. An important mechanism for data stewardship is data institutions, such as data trusts and data intermediaries[14]: the European Commission's Data Governance Act[15] highlights 'data altruism organisations' to enable people to donate data for use towards public good purposes; recent expert recommendations in India[16] around the governance of non-personal data focus on data trust-like models; and Japan is experimenting with 'information banks' to help share personal data.[17]

- **Data skills need to be understood as holistic:** analytical data skills need to be balanced with other strategic skills such as community building and change management to support successful data use and data innovation.[18] We believe this broader data literacy of asking the right critical questions about a data project can also help mitigate risks of bias in data projects: the 2020 A-level fiasco is an example of a Government data project that was implemented without adequate data literacy, and so unfairly exacerbated existing socioeconomic inequalities[19]

Government intervention for long-term success

- **Policy interventions to improve data availability need to consider impact on public trust over the longer-term.** Interventions to strengthen data access that increase data availability in the short term but damage public trust in data ecosystems will reduce the potential benefits of data for the wider economy in the longer-term.[20] The public reaction to the GPDPR controversy is a good example of this.

- In autumn 2021 the ODI convened an expert roundtable with Wellcome Trust to explore access needs and opportunities for health data research relevant to potential reform of UK data protection. Feedback from participants placed a strong emphasis on the importance of adherence to established research ethics protocols as a matter of research integrity. Consequently, roundtable attendees reported that **seeking consent from informed and empowered data subjects** is a matter of professional practice for the research community, rather than being only a matter of legal compliance, and so weakening the legal requirements around consent would be unlikely to change research practice (but might weaken longer-term public and patient trust in research).

---

[12] The ODI (2020), 'Data 2020: data rights and ownership'; and the ODI, RSA and Luminate (2020), 'Data about us'

[13] Martin Tisné for Luminate / Stanford University Cyber Policy Center (2020), 'The Data Delusion: protecting individual data isn't enough when the harm is collective'

[14] The ODI (n.d.) 'Data institutions'

[15] European Commission (2020), 'Proposal for a Regulation on European data governance (Data Governance Act) | Shaping Europe's digital future'

[16] The National Law Review (2021),'India Issues Non-Personal Data Framework'

[17] Keidanren, Japanese Business Federation (2021), 'Ideal Protection and Utilization of Personal Data toward Society 5.0; also Nikkei Asia (2019), 'Japan's 'information banks' to let users cash in on personal data'

[18] The ODI (2020), 'Data skills framework'

[19] The ODI (2020), 'Fiascos and the future: why data literacy is essential'; and the ODI (2020), 'What can we learn from the qualifications fiasco?'

[20] The ODI/Frontier Economics (2021), 'The economic impact of trust in data ecosystems'

Inquiry topic: potential benefits and existing barriers (800wds)

Key potential benefits of data sharing

- Effectively using and sharing data between and across Government, other public bodies, research institutions and commercial organisations can bring **multiple social and economic benefits.**
- In 2020, a pilot study by ODI **in collaboration with Mumsnet and Barnardos** showed how combining data sources from **civil society organisations, the Government (DfE and DWP) and social media companies** could allow us to get a more holistic understanding of the **impacts of the Covid-19 pandemic on childrens' wellbeing**, beyond just one-dimensional assessments such as educational attainment.[21]
- In the health and care sector, research commissioned by the ODI to Frontier Economics has shown that **effective data infrastructure** to combine national and local data sources is key to realise the full impact of **social prescribing** initiatives.[22]
- A joint report by the ODI and the Bennett Institute of Public Policy (University of Cambridge) shows that data sharing and data use can also **contribute directly to improving economic outcomes**.[23] For example, a 2017 study conducted by Deloitte found that by sharing data in an open way, **Transport for London saves £1 million per year** in customer support costs that they would otherwise have to bear, while at the same time **saving passengers time worth £70m-£90m per year** indirectly by allowing third party organisations to develop applications that use TfL data.[24]
- In our commentary on the **Spending Review 2021**, we noted that Governments are one of the largest and most powerful agents in national data ecosystems, collecting and controlling significant datasets across sectors and communities.[25] Therefore, **the data practices of the government matter for how that country's national data ecosystem develops overall across sectors and domains**. By adopting good practices in publishing, leading, collaborating, supporting and stewarding data, the government could better coordinate with private sector and civil society stakeholders involved in delivering public services, ensuring data works for the wider economy and society.

Current barriers to data sharing

- We believe that by looking at how data becomes valuable, government interventions can be designed to create the right incentives and set up the right institutional frameworks to unlock the potential value of that data.[26] For example, several of the measures outlined in the DCMS consultation 'Data: A new direction' point towards reducing the costs incurred by organisations to collect, use and share data.[27] **However, cost-cutting is not the only way in which data sharing can be incentivised or supported,** and **measures that focus on reducing the burden of compliance risk introducing harms and undermining public trust**, as the backlash against the General Practice Data for Planning and Research (GPDPR) shows. Our data institutions programme is exploring how **the creation of independent bodies in charge of data stewardship can increase responsible data sharing and use**, and therefore unlock more value while bolstering public trust.[28]

---

[21] The ODI (2020), 'Data about children's lives in the pandemic'; and The ODI (2021), 'Data on teachers' lives during the pandemic'

[22] The ODI & Frontier Economics (2021), 'The role of data in unlocking the potential of social prescribing'

[23] The ODI & Public Bennet Institute for Public Policy (2021), 'The Value of Data'

[24] Deloitte (2017), 'Assesing the value of TfL's open data and digital partnerships'

[25] The ODI (2021), 'Spending Review 2021: Investing in data to build back better'

[26] The ODI (2021), 'Policy to unlock the economic value of data'

[27] DCMS (2021) 'Data: a new direction'

[28] The ODI (n.d.) 'Data institutions'

- Our work on mapping the data initiatives in UK government identified **over 100 bodies with some responsibility for 'data' across government,** with associated risks of duplication and inefficiency.[29] However, in a recent roundtable discussion hosted by the ODI and the Institute for Government, it was argued that the **lack of data foundations**, which the National Data Strategy defines as 'data that is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable' is **a major barrier.**[30] Some examples cited included the **absence of a consistent standard** for a citizen record across government (across just 10 government bodies, the National Audit Office found more than 20 different ways of identifying individuals and businesses, and no standard format for storing data such as name, address and date of birth)[31] and the **difficulty in knowing who holds what data across government**. The ongoing political focus on building 'data-driven technologies' and artificial intelligence tend to **neglect the importance of this foundational work** for other data initiatives.
- Lack of public sector standardisation of **data about protected characteristics**, such as race and ethnicity, is a barrier to data sharing because it means **datasets are not interoperable with each othe**r, and **analyses on different datasets can't always be usefully compared.** For example, data about race might refer to subtly different phenomena;[32] and as well as standardising data about race and ethnicity, we should also aim to standardise data about racism and bias.[33] In our project 'Monitoring equality in digital public services', we found that as more public services go digital, they still need to meet legal requirements around discrimination – in the same way as non-digital services; but **there is no accepted practice for collecting and publishing data about who uses digital public services,** which makes it hard to tell whether they discriminate or not.[34]

Inquiry topic: data issues in key Government data strategies and consultations (1,200wds)

Data issues and the National Data Strategy (DCMS, 2020)

In our response to the DCMS consultation on the National Data Strategy (autumn 2020), we made the following key points:[35]

- **Set out a clearer, more coherent and ambitious vision for data in the UK, that builds on our unique strengths.** The NDS recognises the critical role of data in the economy and the fact that the UK is already a leading digital nation. But there are many different ways in which the UK could lead around data, and it is not clear how the UK's approach to data will be distinct from that of, say, the EC, US, Canada or India. We believe that the UK should take advantage of its historic leadership in open data, data ethics, and data institutions to build a vision oriented around recognition of data as a public good.
- **Be more balanced in recognising both the opportunities and the risks around data. Currently, the NDS has a welcome tone of optimism about data but this can downplay the harms that it can bring.** The NDS must be clear-sighted about the risks of data, both to ensure that they are effectively mitigated and to build trust in the

---

[29] The ODI (2021) 'Mapping data in the UK government'
[30] The ODI and the Institute for Government (2021) 'Data sharing for public service delivery: a roundtable discussion on the UK government's data protection consultation'
[31] National Audit Office (2019) 'Challenges in using data across government. Report by the Comptroller and Auditor General'
[32] Eleanor Shearer for the ODI (2020), 'The dividing line: how we represent race in data'
[33] Jeni Tennison for the ODI (2021), 'Objective data? Reflections on the Commission for Race and Ethnic Disparities report'
[34] The ODI (2020), 'Monitoring equality in digital public services'
[35] The ODI (2020), 'The UK National Data Strategy 2020'

approach that the government will take around data. Data ethics and responsible use of data cannot be seen as a separable add on: it must be embedded into every aspect of data policy.

- **Think carefully about how to empower the devolved administrations, local government, regulators, arms-length bodies and government departments outside DCMS and the Cabinet Office, to improve data across the country and across sectors.** Making best use of data requires coordinated but distributed effort. The NDS needs to be clearer about which duties, responsibilities and powers around data are held by which public bodies, and how DCMS and Cabinet Office will support and coordinate these efforts.
- **Take a more holistic approach in considering the needs and opportunities around data skills.** The NDS currently has more detail on relatively advanced data skills in central and local government, and also in research and the private sector. But comparable capabilities in civil society will help with the development of a balanced economy. And a strong foundation of core data literacy across all society will be a key component of engaged citizenship - supporting the development of trust between citizens, government, businesses and civil society through meaningful accountability and transparency about the management, use, and governance of data and digital technologies.
- **Pay more attention to the role of civil society within the data ecosystem.** The NDS currently focuses on the role of government and of the private sector. But charities and other third sector organisations are also collectors, maintainers, users and reusers of data, and are important intermediaries in getting data and information into the hands of those who need it. They also play a key role in facilitating innovation with data - as well as being key potential innovators themselves. The NDS should recognise their crucial role, particularly in a difficult economic climate, and factor in the support they need to grow their data capability and participate in data ecosystems.
- **Give greater weight to the issues of equality, diversity and inclusion, particularly with reference to the range of opportunities offered for data related jobs and skills development, and other economic opportunities around data use** - for example, for start-ups and SMEs. We would also urge more consideration around the ways in which data use and digital technologies could be used to proactively counter or ameliorate existing societal biases, discrimination or exclusion.

Data issues and Data Saves Lives (DHSC, 2021)

In our response to the DHSC consultation on Data Saves Lives (summer 2021), we made the following key points[36]:

- We think it is important that the Secretary of State for Health and Social Care **recognises the scale of his task and is committed to acting on it**
- not skimp on creating a trustworthy and trusted data sharing culture in the NHS and social care: **building half a bridge doesn't get you halfway across - it's an expensive way to get nowhere.**
- build a **strong data infrastructure** and ensure it is **integrated into national strategy**
- ensure that the strategy **benefits those who are digitally excluded** and does **not compound bias on social or economic grounds**
- recognise the **huge challenges regarding the quality of social care data** and invest in the long term changes needed to improve data infrastructure in this area
- empower NHS staff, researchers, data infrastructure builders, developers and innovators both **nationally and locally** to use data effectively for public benefit

We also highlighted to DHSC the findings of a workshop we held jointly with the Ada Lovelace Institute and Understanding Patient Data to inform consultation responses:[37]

---

[36] The ODI (2021), 'The ODI responds to the UK government's proposed data strategy for health and social care'

- Consider **a collective approach to choice and control** over how data is used
- Ensure **clear and consistent communication** with the public.
- Be **honest about the limitations of data**, taking into account potential harms in order to mitigate them.

In our response to the DCMS consultation on Data: A New Direction (autumn 2021), we made the following key points[38]:

- We argue for **data reforms to be accompanied by a commitment to developing holistic data skills in organisations and communities**, to strengthen trust in data ecosystems and so that more people can benefit from data practices.
- We **critique the government's economic impact analysis paper** for its narrow focus on direct compliance costs to business of data protection, rather than on the long term benefits and harms data can deliver for the UK.
- We argue that **economic growth from greater data sharing and data use must also be sustainable, inclusive, and equitable for longer-term trust in data practices**.
- We argue that interventions to strengthen data innovation should also **widen participation in data innovation, and improve the distribution of the benefits of data innovation,** for longer-term trust in data practices.
- We argue that the **ICO and other regulators should be supported in building capacity in regulatory technology**.

We also highlighted to DCMS the findings of expert roundtables we held (including in collaboration with Wellcome Trust and the Institute for Government):[39]

- The importance of transparency and trust – it was also recognised that **transparency is necessary but not sufficient, and should be accompanied by accountability and mechanisms for redress.**
- That **harms and benefits from data policies and data practices can change over time**, and **some harms and benefits are collective rather than individual**. So a trustworthy data protection regime is one that protects communities as well as individuals, and that gives weight to the medium- and longer-term impacts of current actions.
- There were **questions about whether legislative change was necessary, or whether it was the right kind of intervention needed** to secure the benefits of a trusted data protection regime.

Inquiry topic: ethics and health and care data sharing (500wds)

Recent and current ethics challenges for UK health and care data sharing

The failure in **2013-14 of the care.data programme[40]** and the early **backlash in 2021 against the GPDPR programme**[41] are powerful examples of the public concern with data handling by the

---

[37] The ODI, Ada Lovelace Institute & Understanding Patient Data (2021), 'Workshop Summary: NHSX Data Strategy for Health and Care'

[38] The ODI (2021), 'The ODI responds to the UK government's consultation on proposed reforms to data protection'

[39] The ODI (2021), 'Expert perspectives on the UK data protection consultation'

[40] BMJ (2016), 'Controversial database of medical records is scrapped over security concerns'; and BBC (2014), 'Care.data: How did it go so wrong?'

NHS and health providers. A survey by the ODI also found that,**while the NHS is one of the most trusted organisations in the UK, only 59% of the population trust the NHS and healthcare providers to use data ethically**.[42]

Data about individuals in health and care contexts can be used to deliver healthcare and to make decisions about care of the individual from whom it was collected: this is called **primary use of health data**. Data can also be aggregated from population-level sources to improve personal care planning, medicines development, research, safety monitoring and policymaking: this is called **secondary use of health data**, and can play a crucial role in improving health systems. A recent report by the ODI evaluated how different countries across the European region rank in terms of policy readiness for secondary use of health data;[43] **the UK had one of the lowest scores for ethics**.

Building trust: practical tools

The ODI's [work on Trustworthy Data Stewardship](#)[44] shows that **trust and trustworthiness around data practices depend on a variety of factors.** The trustworthiness of an organisation and the data it collects, manages, uses or shares needs to be systematically examined in a way that accounts for this. **Data assurance products and services** – including but not limited to the **assurance of processes and technical standards; repeatable audits; advisory and training services; and certification schemes** – can help to mitigate against data risks and improve data governance more widely. We have produced several tools to help, including:

- A health domain **"data governance playbook"** directed at non-technical leaders, to support healthcare leaders and professionals in accessing, using and sharing data in responsible, ethical and equitable way [45]
- A guide for **applying data anonymisation techniques** that comply with several
- standards followed by industry and current legislative and regulatory requirements:[46]
- Guides on **publishing data and publishing scientific or computing models in a crisis** (developed as part of our work with Luminte on Covid-19 response).[47]
- Our **Data Ethics Canvas** helps identify and manage ethical issues by mapping questions that need to be answered and reflected upon throughout any data initiative.[48] This tool **considers data ethics in a holistic way**, taking into consideration not only questions about privacy and individual rights, but also the importance of public engagement with relevant stakeholders, openness and transparency, and clear communication.[49]

Inquiry topic: data safeguards and privacy (650wds)

Safeguards and privacy for health data research

The ODI collaborated with the [Wellcome Trust](#) to convene a roundtable discussion to explore the proposals in 'Data: a new direction', around data for research.[50] Health data is a particularly sensitive

[41] National Data Guardian (2021), '[National Data Guardian statement on the General Practice Data for Planning and Research (GPDPR) programme](#)'
[42] The ODI (2019), '[Nearly 9 in 10 people think it's important that organisations use personal data ethically](#)'
[43] The ODI (2021), '[Report scrutinising the policy readiness of Europe for the secondary use of health data](#)'
[44] The ODI (2021), '[Introducing the ODI's Trustworthy Data Stewardship Guidebook](#)'
[45] The ODI (2021), '[The softer side of data governance: a playbook for non-technical users](#)'
[46] The ODI (2020), '[Anonymising data in times of crisis](#)'
[47] The ODI (2020), '[Covid-19: Building an open and trustworthy data ecosystem](#)'
[48] The ODI (2021), '[The Data Ethcis Canvas](#)'
[49] The ODI (2021), '[The Data Ethcis Canvas](#)'
[50] The ODI (2021), '[Expert perspectives on the UK data protection consultation](#)'

category of data, in terms of personally identifiable characteristics; health data is also a particularly strategic category of data for both national and international health security (for example, pandemic response), and for national and international economic security. The roundtable participants were experts in managing research processes around health data, in both academia and the NHS, and so well placed to evaluate the possible impacts of the proposed reforms. Key points from the discussion were:

- A majority of roundtable participants were **sceptical about many of the proposals** and expressed a preference for **clearer, more joined up guidance rather than legislative change.**
- A few were open minded about changes to the law, provided it would provide **greater clarity and reduce administrative burden, especially for NHS organisations.**
- When it came to discussing the proposals surrounding lawful grounds in research, and the need for a separate provision for research, the consensus was that **currently, researchers use existing provisions effectively.** However, when it came to international collaborations, some **flexibility when identifying lawful grounds for data processes,** was identified as potentially helpful.
- A clear majority of attendees recognised that **most health and medical research does not at present rely on 'broad consent' for its lawful basis to conduct research.** Participants did not anticipate that the proposed changes to consent would benefit health and science research significantly. There was a strong emphasis on the **importance of adherence to established research ethics protocols as a matter of research integrity.** Consequently, roundtable attendees reported that **seeking consent from informed and empowered data subjects is a matter of professional practice for the research community,** rather than being only a matter of legal compliance, and so weakening the legal requirements around consent would be unlikely to change research practice (but might weaken public and patient trust in research).
- Many urged the Government to **recognise the wider context in which research - including re-using data - is conducted, including the role of research ethics.** It was felt that current proposals need to better complement existing research processes, at least when it comes to health-related and medical research.
- Participants were clear that **they did not want any of the changes to risk EU data adequacy,** which was perceived to significantly benefit UK research.

How data policy can support health data research

We asked participants whether there were any common data protection and management issues for researchers that were not responded to directly in the proposals, and also what kinds of Government intervention around data protection would be most useful for health research. The key needs and opportunities identified were:

- Updating guidance to clarify what constitutes **anonymised data.**
- Clarity on **transparency requirements**, especially when researchers no longer have contact with data subjects (eg if they change address or die).
- More ambition when it comes to data governance models. **'Data stewardship' for example, is a concept that is gaining salience in the research sector and provides more flexibility when compared to more traditional data management processes.** Deploying novel types of data stewardship, such as 'data banking' or more deliberative methods of engagement with data subjects could prove a long-term, sustainable way of using data for research more innovatively.

- Further work to **understand the reasonable expectations of data subjects when it comes to further processing for commercialisation purposes.** This could include conducting public attitudes research with the public.
- Coordinated **cross-sector guidance on a number of critical issues for research, such as confidentiality, anonymity, data controllers in international research,** and transparency requirements, should be prioritised over changes to legislation.
- Any changes to consent, including greater use of 'broad consent' for research, **should not dilute its meaning;** and **routinely collected health data must continue to require additional protections when re-used.**
- A new data regime must be **reinforced by transparency and public involvement.**

Inquiry topic: governance arrangements (650wds)

Key needs: independence and accountability

- In our response to the National Data Strategy consultation[51], we argued that one way in which the government could increase trust in its own handling of data and use of AI would be for there to be **an independent body actively monitoring and holding public sector bodies to account**. We believe there currently seems to be **a gap in organisations holding government accountable for its use of data and algorithms that none of CDEI, ICO or the National Audit Office are filling.** If CDEI were to take on a role that aimed to build trust by strengthening accountability around the ethical use of data and AI by the public sector, then it would need a change in statutory status to ensure its independence from government. In this case, CDEI could report to Parliament, in the same way as the NAO.
- There is substantial evidence for the **importance of institutional independence as a determining factor for trust in data ecosystems.** The ODI's report "Designing Trustworthy Data Institutions" discusses the importance of independence for some data institutions such as data trusts, to ensure trustworthy decisions about data access,[52] while our report on "Demonstrating and Assessing Trustworthiness" also emphasises the role of independent advisory processes.[53]
- It is also clear that **effective regulators help to build trust in data ecosystems.** In "Designing Trustworthy Data Institutions", we also argue that the trust ecosystem extends to actors beyond the immediate environment of a data institution, including regulators and certifying bodies. This latter point is reiterated in our recent report with Frontier Economics on the impact of trust on data ecosystems, which examines the role of regulators in building trust.[54]
- For these reasons, **we believe some of the proposals included in the DCMS consultation 'Data: A new direction', which give the Secretary of State for DCMS power over the Information Commissioner Office could affect trust in data ecosystems.**[55]

[51] The ODI (2020), 'The UK National Data Strategy 2020'
[52] The ODI (2021), 'Designing trustworthy data institutions'
[53] The ODI (2021), 'Demonstrating and assessing trustworthiness when sharing data'
[54] Frontier Economics / the ODI (2021), 'The economic impact of trust in data ecosystem'
[55] The ODI (2021), 'The ODI responds to the UK government's consultation on proposed reforms to data protection'

- We consider that data institutions - **organisations that steward data on behalf of others, often towards public, educational or charitable aims** - are essential to data infrastructure.[56] [57] **The UK already has some well-known data institutions:** for example, in the health domain, UK Biobank, set up by the Department for Health (in conjunction with partners including Wellcome) in 2006 to steward genetic data and samples from around half a million people, and continues to support their use for health research. INSIGHT Health Data Research Hub, founded in September 2019 and funded through Health Data Research UK (HDR-UK), which brings together and makes available anonymised data from eye scans and images, and advanced analytics, to NHS, academic and industry researchers. It aims to unlock new insights in disease detection, diagnosis, treatment and personalised healthcare. In the public sector, the Office of National Statistics, Ordnance Survey and NHS Digital each have responsibilities defined in legislation to collect, maintain, and share data.
- To meet the current challenges for data governance **our existing public sector data institutions must be supported through funding, capacity-building, and the development of sustainable business models**, to enable and encourage them to share data effectively while demonstrating leadership in good data governance[58].
- In our written evidence for the APPG on AI in February 2021[59], we suggested that **Parliamentarians should scrutinise how well these public sector bodies are supporting – or inhibiting – research and innovation with data**, and the appropriateness of their data governance regimes and business models.[60] We also suggested that **Parliamentarians should examine the support and incentives offered to organisations that collect, maintain and share data on behalf of others**, ensuring that *powers* to collect data are balanced with adequate *duties* to share it. And we suggested that **Parliamentarians should encourage investment in research and development of these new data institutions**; and should also ensure there is proper attention to the consequences of their introduction; examine how new data institutions or institutional models can be registered or monitored; and ensure there are protections in place to mitigate against possible harms.

February 2022

---

[56] The ODI (2021), 'The ODI's written evidence for APPG on Artificial Intelligence'

[57] The ODI (2021), 'Data Institutions Explainer'

[58] The ODI (2019), 'R&D: Sustainable data institutions'; and the ODI (2020-2021) 'R&D: Sustainable data access'

[59] The ODI (2021), The ODI's written evidence for APPG on Artificial Intelligence

[60] The ODI (2021), The ODI's written evidence for APPG on Artificial Intelligence