Bottom-up data institutions: mechanisms for government support

Contents

About

This report has been researched and produced by the Open Data Institute, and published in March 2022. Its lead authors are Sara Marcucci and Matt Davies. To share feedback by email or to get in touch, contact the Data Institutions programme at datainstitutions@theodi.org

Executive summary

Bottom-up data institutions enable people to take a more active role in decisions about data. At the Open Data Institute (ODI), we think they have the potential to address data harms while unlocking new uses of data for public good. This report identifies four areas that need to be improved if this is to happen.

Knowledge

The existing evidence base on bottom-up data institutions and the market opportunity they represent is limited. This could be improved by:

- Joining up existing government support for data institutions.

- Commissioning research on the market for bottom-up data institutions.

- Maintaining the UK's rich institutional knowledge on data stewardship.

Trust

There is evidence that users are likely to be wary of unproven approaches to data stewardship. This could be improved by:

- Providing legal definitions of data intermediation and data altruism.

- Introducing a register for bottom-up data institutions.

- Supporting the development of an assurance sector for data institutions.

Rights

Bottom-up data institutions often rely on the exercise of data portability rights, but these are limited in UK and European law. This could be improved by:

- Enhancing rights to data portability.

- Providing a right to data reciprocity for bottom-up data institutions.

- Loosening data processing restrictions for certified organisations.

Infrastructure

The technologies and infrastructure needed aren't always available to facilitate participatory data stewardship. This could be improved by:

- Supporting the development of shared technological infrastructure.

- Experimenting with data institutions as providers.

- Mandating interoperability and the development of open standards.

1. Introducing bottom-up data institutions

1.1 What is data stewardship?

At the ODI, we want those who collect, maintain and share data to act in ways that lead to the best economic and social outcomes for everyone. Our Theory of Change contrasts this positive data future against two negative alternatives:[1]

- A *data hoarding* future where organisations restrict access to data, using inappropriate business or funding models for data that limit the value that we, as a society, get from it.

- A *data fearing* future where unaddressed fears arising from legitimate concerns – such as who has access to data and how it might be used – prevent us from realising its full benefits.

The concept of data stewardship is a response to data hoarding and data fearing. It is a 'responsible, rights-preserving and participatory concept [which] aims to unlock the economic and societal value of data, while upholding the rights of individuals and communities to participate in decisions relating to its collection, management and use'.[2]

The Ada Lovelace Institute defines stewardship as a concept that aims to achieve the responsible and accountable use of common resources.[3] Data has particular economic characteristics that make it unlike traditional examples of common resources such as oil and forests,[4] so new paradigms of stewardship need to be developed and implemented.

We define data stewardship as the process of deciding who has access to data, for what purposes and to whose benefit, to realise the value and limit the harm that use of data can bring.[5] We see stewarding data as the foundational activity in the lifecycle or value chain of data – collecting, maintaining and sharing it. How data is stewarded

---

[1] OD,I (2018), <u>Our theory of change – The ODI</u>

[2] Aapti Institute, Open Data Institute and Global Partnership for AI (2022), <u>Enabling data sharing for social benefit through data trusts</u>

[3] Building on Elinor Ostrom's work on the different principles for governing the commons (Ostrom, E [2015], Governing the Commons. Cambridge University Press). Ada Lovelace Institute (2021), <u>Participatory data stewardship - The Ada Lovelace Institute</u>

[4] Bennett Institute for Public Policy (2020, <u>The Value of Data - Summary report</u>); Towards Data Science (2019), <u>Data is not the new oil: towards data science</u>

[5] Open Data Institute (2021), <u>What are 'bottom-up' data institutions and how do they empower people?</u>

ultimately affects what types of products, services and insights it can be used to create, what decisions it can inform and which activities it can support.

1.2 What are data institutions?

Data institutions are organisations that steward data on behalf of others, often towards public, educational or charitable aims.

Data institutions steward data in different ways, including:

- **Protecting sensitive data and granting access under restricted conditions**. UK Biobank, for example, was set up in 2006 to steward genetic data and samples from around half a million people, and continues to support their use for vital health research.

- **Combining or linking data from multiple sources**, and providing insights and other services back to those that have contributed data. In the maritime sector, HiLo takes data generated by around 3,500 ships globally to generate vital risk and safety analyses related to lifeboat accidents, engine room fires and other incidents.

- **Creating open datasets that anyone can access, use and share** to further a particular cause. OpenCorporates, Open Apparel Registry and 360Giving collate and make accessible open data about companies, factories and grant-making, respectively.

- **Acting as a gatekeeper for data held by other organisations**, ADR UK ensures important administrative data held by the UK Government can be accessed safely and securely, and in the US, Social Science One plays a similar role to make Facebook data available for new research.

- **Developing and maintaining identifiers, standards and other infrastructure** for a sector or field, such as by registering identifiers or publishing open standards. Open Banking Limited was established in the UK in 2016 to develop standards and guidelines to drive competition and innovation in the retail banking sector.

- **Enabling people to take a more active role in stewarding data about themselves and their communities**. Open Humans enables people to explore, analyse and share data about their health, and make it available for citizen science projects.

Although the term is new, data institutions have existed in the private, public and third sectors for as long as hundreds of years. Public organisations like national mapping agencies, statistics agencies and archives are perhaps the oldest data institutions,

while civil society organisations, trade unions and other organisations are increasingly taking on new data duties.[6]

This report focuses on this last role: enabling people to take a more active role in stewarding data.

Box 1: Data institutions, data intermediaries and other terms

The ODI defines **data institutions** as:

> "organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims."

We take 'institution' to mean an organisation, establishment, foundation or society, especially those that are public, educational or charitable.

The term might also be used in other contexts to mean different things.

For example:

- To describe organisations that are essential components of the data ecosystem but tend not to actively steward data themselves, such as advocacy organisations, funders, auditors and so on. We think of these as simply 'institutions'.

- To convey the need for a richer institutional environment around data to ensure it works for everyone, which, as well as organisations, would include laws, codes of conduct, norms and informal behaviours. Using this wider lens, data protection or data portability can be seen as institutions related to data.

Other organisations use different terms to describe similar or related concepts.

For example:

- The UK government used the term '**data intermediaries**' in its National Data Strategy and reports from the Centre for Data Ethics and Innovation (CDEI).

- The European Commission uses the terms '**data spaces**', '**data sharing services**' and '**data altruism organisations**'.

- GovLab uses the term '**data collaboratives**', to describe initiatives that enable data to flow from the private to public sector.

- The Oxford Martin School is focused on '**institutional and legal constructs**' that 'could allow data to be shared in more flexible and innovative ways, respecting individual autonomy while generating wider societal benefits'.

These terms are roughly the same. This report uses 'data institutions' throughout, but our findings will be relevant to organisations and practitioners that use other terms.

---

[6] Open Data Institute (2021), What are data institutions and why are they important?

1.3 What are bottom-up data institutions?

In recent years, data institutions have emerged to empower individuals and communities to take a more active role in the stewardship of data about them or that they have a vested interest in. At the ODI, we refer to these as **bottom-up data institutions**. These data institutions use participatory methods to enable individuals and communities – usually those that have generated the data or that the data is about – to exercise greater control over the collection, maintenance and sharing of data.

One can observe bottom-up data institutions in the wild that enable groups to generate or repurpose data about them, and exert collective control over it for a common purpose. For example:

- Variant Bio works with historically marginalised populations, where communities are engaged prior to the beginning of research projects, and the data is collected and used within a framework that takes into account community concerns.

- Driver's Seat is an independent, driver-owned cooperative where members' data is used to derive insights that help them optimise their performance.

- Open Humans empowers individuals and communities to explore and share their personal data for the purposes of education, health and research.

- MIDATA enables users to contribute to medical research and clinical studies by granting selective access to their personal data.

- Gyeonggi Data Dividend ensures that any financial profits generated by selling access to data about transactions using the local currency are returned to citizens in the form of a dividend.

In addition to the organisations themselves, an ecosystem of research and practice has emerged around the concept of bottom-up data stewardship. For example:

- The MyData Global community, which aims 'to empower individuals by improving their right to self-determination regarding their personal data'.[7]

- The Mozilla Data Futures Lab, which was launched in 2021 to support experimentation around 'new approaches to data stewardship that give greater control and agency to people'.[8]

- The Ada Lovelace Institute, which advocates for 'participatory data stewardship', where people whose data is used or about which data decisions are taken are meaningfully involved.[9]

---

[7] https://mydata.org

[8] https://foundation.mozilla.org/en/data-futures-lab/

[9] https://www.adalovelaceinstitute.org

- The Aapti Institute's work at the Data Economy Lab, which aims to empower individuals and communities to play a bigger part in data governance, and it has documented numerous examples of this in practice.[10]

- Connected by Data, which is a research and advocacy organisation aiming to put community at the centre of data narratives, practices and policies.[11]

At the ODI, our data institutions programme is designed to bring about new data institutions and improve the practices of existing ones, including bottom-up data institutions.[12]

1.4 What are the different types of bottom-up data institutions?

There are many different forms of bottom-up data institution, and lenses to use to see them through.

Our use of the term 'bottom-up' refers to the existence of processes that involve people in decision-making (and less an indication of how the institutions came into being). We observe three main approaches taken by bottom-up data institutions to enable this participation in practice:

- **individual** decision-making; people making decisions individually about the data about them

- **collective** decision-making; people making decisions as part of a larger group

- **delegated** decision-making; people delegating decision-making authority to another party.[13]

Examples of bottom-up data institutions and the nature of involvement they support can be seen in Figure 1.

---

[10] https://www.aapti.in

[11] https://connectedbydata.org

[12] https://theodi.org/project/rd-data-institutions/

[13] Open Data Institute (2021), What are 'bottom-up' data institutions and how do they empower people?
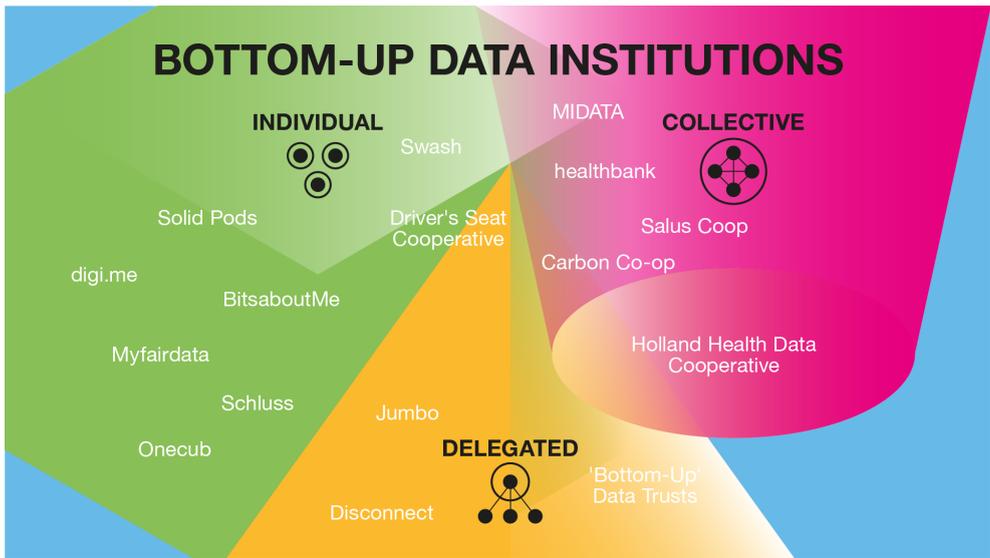
Figure 1: Examples of bottom-up data institutions

Bottom-up data institutions also differ in the nature of the decisions they involve people in. Some enable people to take part in governing access to data (for example, making decisions about whether to provide access to data to a particular project or group of researchers) or governing the data institution itself (for example, making decisions around the goals and principles of the organisation, plans for future investment, and distribution of profits).

A detailed exploration of the different models for participation in data stewardship has been carried out by the Ada Lovelace Institute.[14]
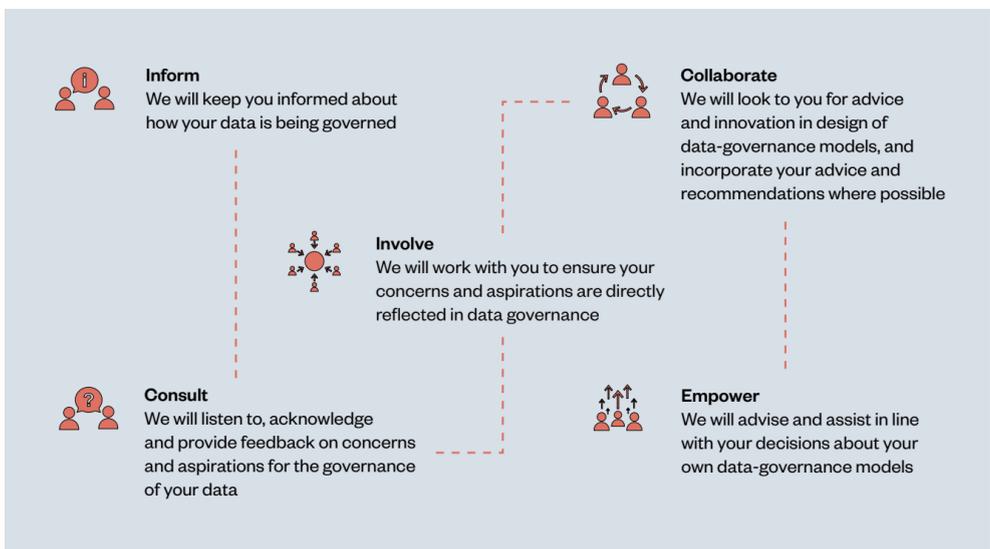

Figure 2: Framework for participation in data stewardship, reproduced from Ada Lovelace Institute (2021), _Participatory data stewardship_.

---

[14] Ada Lovelace Institute (2021), _Participatory data stewardship_

Based on Sherry Arnstein's 'ladder of citizen participation',[15] the framework produced by Ada (in Figure 2, above) sets out the possible outcomes of informing, consulting, involving, collaborating and empowering people, whereby people go from being recipients of information about the data to be stewarded, to being active, empowered actors with agency and the power to influence how that data is to be stewarded.

---

[15] Arnstein, S (1969), <u>A Ladder of Citizen Participation</u>. *Journal of the American Institute of Planners*, 35(4), pp.216-224

2. Government interest in bottom-up data institutions

Over the past decade, how data is stewarded has become a significant issue for policymakers. Scandals such as Cambridge Analytica's use of data from Facebook users in political advertising have raised awareness of the impact that irresponsible data stewardship can have on individuals and society.

From an economic perspective, access to data plays a major role in developing and delivering products and services: the five largest private companies in the world are technology companies that offer services predicated on the collection, use and sharing of data.[16] By supporting better decision-making, data – and technologies which use data such as Artificial Intelligence (AI) – can contribute to the creation of value across public, private and third sectors, and in industries from health to transport.[17]

There are a number of market failures associated with data availability,[18] meaning the right data isn't always available to create the necessary information and support decision-making. In response, we've seen governments and other policymakers begin to think about how to steer the stewardship of data towards their goals, priorities and values.

In the UK, where the government has set out an ambition to become a 'global science and technology superpower',[19] the Innovation Strategy[20] and AI Strategy[21] both highlight the role of data in the development of new technologies, products and services. Mission 1 of the government's National Data Strategy – 'Unlocking the value of data across the economy' – sets out measures that aim to boost data availability across sectors.[22]

Research carried out in 2021 for the Department for Digital, Culture, Media and Sport (DCMS) by Frontier Economics identified six barriers to optimal data sharing:[23]

- **Lack of incentive to share data:** data providers may not be sufficiently incentivised to share or grant access to their data - for example, sharing may incur costs that they are unable to recoup from the beneficiaries of expanded access.

---

[16] PWC (2021), 'Global Top 100 public companies by market capitalisation'

[17] Open Data Institute (2018), 'Our theory of change'

[18] Open Data Institute (2019), The Value of Data

[19] Gov.uk (2021), "We're restoring Britain's place as a scientific superpower"

[20] Gov.uk (2021), UK Innovation Strategy: leading the future by creating it

[21] Gov.uk (2021), National AI Strategy

[22] Gov.uk (2021), National Data Strategy Mission 1 Policy Framework: Unlocking the value of data across the economy

[23] Gov.uk (2021), Increasing access to data held across the economy

- **Lack of knowledge:** data providers may lack enough knowledge of the potential uses of their data, while data users lack enough knowledge of what data could be made available. Both parties may have limited understanding of how data can be shared even where a potential use has been identified.

- **Commercial, reputational and ethical risks:** the perceived or actual risk of losing competitive advantage, of suffering reputational damage from data uses that may breach others' trust, or from data sharing exposing controversial business practices, may deter providers from sharing data.

- **Regulatory and legal risks:** the perceived or actual risk of breaching data protection, intellectual property rights or regulatory requirements may be a deterrent to sharing.

- **Costs of data access/sharing:** costs may be prohibitive because of a lack of coordination, common foundations, infrastructure and technologies that are needed for data sharing to be cost effective.

- **Missed opportunities to use data in the public interest:** cases where sharing data may be particularly likely to lead to economic and social gains beyond the benefits that accrue to the organisations providing and obtaining access to data.

Data institutions can help overcome some of the market failures associated with data availability. The UK government's Centre for Data Ethics and Innovation has argued that data intermediaries can provide partial or complete solutions to each of these barriers,[24] and has set out measures to 'support the development of a thriving [data] intermediary ecosystem'.[25]

Recently published ODI research demonstrates the ways in which different types of data institutions create value within their ecosystems.[26] Focussing on illustrative case studies of five UK-based data institutions, this research found that existing data institutions are creating significant value in the form of cost savings, efficiency gains and positive social impacts. For example, the Marine Environmental Data and Information Network (MEDIN) – which promotes the sharing and reuse of marine data – was found to produce cost savings of £1.32 million per year from making data reusable, while OpenActive – which develops open data standards to publish and use data about sport and leisure activities – was estimated to save activity providers £267,200 a month in time saved on promotional activities.

The field of bottom-up data institutions is relatively nascent, so claims about their impact are necessarily speculative. However, we think it is likely that bottom-up approaches to data stewardship can:

---

[24] Gov.uk (2021), Unlocking the value of data: Exploring the role of data intermediaries

[25] Gov.uk (2021), National Data Strategy Mission 1 Policy Framework: Unlocking the value of data across the economy

[26] Open Data Institute (2022), 'Measuring the impact of data institutions'

- provide new incentives for sharing data

- build trust in data use

- increase competition and reduce costs

- unlock data for public causes

- mitigate harms associated with data use.

2.1 Providing new incentives for sharing data

Bottom-up data stewardship can create new incentives for individuals and communities to share data.

Some bottom-up data institutions enable people to donate data for **altruistic reasons**. An example of this is Open Humans, which allows people to collect data about their health, and make it available for new research and citizen science projects. It provides opportunities for people to come together to address health conditions they are passionate about, and has enabled high-quality, peer-reviewed health research. The Open Humans model is intrinsically bottom-up in nature: users collect the data, choose to pool it, and make decisions about how access should be governed; and this leads to the creation of data sources that would not otherwise exist.

Other examples include ZOE Covid Study and Corona-Datenspende, data donation apps for Covid-19 in the UK and Germany respectively. Participants donate data on Covid symptoms, either through an app-based interface (in the case of ZOE) or through linking their fitness wristband or smartwatch to measure values such as sleep patterns, heart rate and number of steps (in the case of Corona-Datenspende). This data is then transmitted to academic partners to inform research into the spread of Covid-19 and its health impacts.

These data donation practices are increasingly termed 'data altruism' due to their similarity with more traditional forms of altruism such as philanthropic donation. Data altruism is attracting increasing interest from policymakers who want to encourage its spread - for example, the European Commission will establish a legal register of data altruism organisations with its forthcoming Data Governance Act.

Other data institutions depend on **volunteer contributors** who contribute to collective stores of knowledge. Examples of this include OpenStreetMap, iNaturalist and SafeCast. Each of these projects allows users to collaborate on the creation and governance of a comprehensive store of data which is made publicly available:

- OpenStreetMap is a free, editable and user-created geographic database of the world, which has grown to become one of the few viable alternatives to Google Maps.

- iNaturalist allows users to share observations of biodiversity across the globe, with 66 million observations of plants, animals, fungi, and other organisms currently listed.

- Safecast stewards the world's largest open dataset of background radiation and air-quality measurements, contributed by users gathering data from physical sensors.

Unlike Open Humans, the data contributed to these projects is not personal data - what they have in common is the use of participation to create valuable new data sources.

Organisations like Swash, which describes itself as a 'data union', take the form of a browser extension installed by users, who can then consent for their browsing data to be automatically captured, pooled and sold on their behalf. Unlike the examples above, Swash is not intended to facilitate altruistic practices or civic engagement, but seeks to generate **financial return** for its users.

At the ODI, we're sceptical about financial returns as an incentive for bottom-up data stewardship for different reasons. On a practical level, data is hard to value: its value largely depends on how it is used and in what context, but it is difficult to predict how it might be used by others or in the future.[27] Moreover, most of the value from data arises when it is aggregated or combined with other datasets - this means that unless such institutions can achieve significant scale, the returns to individuals are likely to be minimal.

The use of financial incentives as a means of encouraging bottom-up data stewardship has also been challenged on political and ethical grounds.[28] Critics of these models argue that incentivising people to participate by compensating them for data could lead to those with fewer resources being more likely to be subjected to exploitative data gathering practices.[29] The eventual result of this could be that privacy and control become luxuries, creating a tiered and unequal data economy.[30]

Lastly, some bottom-up data institutions enable people to contribute data **in exchange for a service**. For example, Open Humans also provides users with access to tools to enable them to generate their own insights into their health.

In recent work with the Global Partnership for AI, the Aapti Institute and the Data Trusts Initiative[31], we have produced a design for a 'London Cycling Data Trust'. In return for sharing data about their cycling journeys around the city, users would get access to improved route planning, statistics and insights about their journeys, and an opportunity to directly feed into the development of London's cycling infrastructure through assessments on the quality of different routes.

---

[27] Open Data Institute (2020), The Value of Data

[28] Centre for International Governance Innovation (2019), Should Tech Firms Pay People for Their Data?

[29] Electronic Frontier Foundation (2020), Why Getting Paid for Your Data Is a Bad Deal

[30] World Economic Forum (2017), What If: Privacy becomes a luxury good?

[31] The Open Data Institute and Aapti Institute (2022), Enabling Data Sharing for Social Benefit Through Data Trusts: Data Trusts in Climate.

2.2 Building trust in data use

Bottom-up data stewardship can help to mitigate the risks associated with some types of data use, and thereby increase trust in the data ecosystem.

Research carried out for the ODI by Frontier Economics has shown that **trust in data ecosystems is a key determinant of data sharing**:[32] greater trust – both in individual institutions, and in the data ecosystem as a whole – results in higher levels of data sharing, and increased data availability.

There is evidence to suggest that **empowerment and control in data systems can help to build trust**. Research carried out by the ODI in conjunction with the RSA found support for more granular control over how data about individuals is collected, shared and used,[33] and surveys carried out by the Centre for Data Ethics and Innovation (CDEI)[34] and the University of Sheffield[35] have also indicated demand for greater individual control over how personal data is managed.

In the course of this research, we carried out interviews with policymakers and bottom-up data institutions (see Methodology) and both indicated that **participatory approaches could help to meet this demand for greater control**, particularly in the case of sensitive data. Similar arguments have been made by the Ada Lovelace Institute, which has argued that participatory data stewardship could increase the legitimacy of – and therefore public trust in – the use of data,[36] and by the Information Technology Federation of Japan, which has advocated the use of 'personal information banks', a type of bottom-up data institution, as a means of supporting trust in the data ecosystem.[37]

2.3 Reducing costs and increasing competition

Another important policy goal to which bottom-up data institutions can contribute is reducing the cost of data access and sharing by increasing competition in the digital economy.

The Furman Review, 'Unlocking digital competition', found that **a lack of competition in digital markets leads to a number of harms**,[38] including reduced innovation in services, reduced consumer choice, and higher prices for businesses using platforms – which are then passed on to consumers. The review notes that this lack of effective competition is amplified by 'network-based and data-driven platform

---

[32] Frontier Economics for the Open Data Institute (2021), The economic impact of trust in data ecosystems

[33] Open Data Institute (2019), Data About Us: 'the people' know and care more than they are given credit for

[34] Gov.uk (2020), Online targeting: Final report and recommendations

[35] Hartman, T, Kennedy, H, Steedman, R, Jones, R (2020) Public perceptions of good data management: Findings from a UK-based survey

[36] Ada Lovelace Institute (2021), Participatory data stewardship

[37] Nikkei Asia (2019), Japan's 'information banks' to let users cash in on personal data

[38] Gov.uk (2019), Unlocking digital competition: Report of the Digital Competition Expert Panel2019)

business models [that] tend to tip markets towards a single winner', and that data openness must be a key tool for the new Digital Markets Unit for promoting competition.[39]

ODI research[40] has identified similar patterns in the artificial intelligence (AI) sector. Organisations with access to large silos of high-quality data are better able to train high-performing AI systems with it; these high-performing AI systems are more likely to be more widely adopted; and the wide usage of these high-performing AI systems generates more data for the data silos, in a feedback loop that tends towards monopoly.

Bottom-up data institutions can help to tackle this problem by **creating new, alternative stores of data in specific sectors**. This, in turn, creates greater competition in data providers, reduces the competitive advantage enjoyed by incumbents and creates greater choice for consumers.[41] One example of this is OpenStreetMap, which has emerged as a viable competitor in a concentrated market dominated by top-down mapping projects such as Google and Apple Maps.

2.4 Unlocking data for public causes

Bottom-up data institutions can **unlock data for public causes,** including the UK government's net zero carbon ambitions and United Nations' sustainable development goals.[42]

Recent ODI research has set out the potential for data trusts – themselves a form of bottom-up data institution – to help **tackle the climate crisis**, providing practitioners and policymakers with usable models.[43] Other types of bottom-up data institutions are already providing new sources of data that could help to support climate change mitigation efforts. For example:

- Globe at Night is an international citizen science campaign to raise public awareness of the impact of light pollution by inviting citizen-scientists to measure and share their observations of the brightness of the night sky.

- Sensor.Community is a network of individuals with DIY air quality sensors. The community pools data about air quality from all over the world to create an open environmental dataset.

- The Energy Data Co-op is being developed by Open Data Manchester and Carbon Co-op, where people can pool and share data to improve the efficiency of their homes by changing how energy is used.

---

[39] Ibid

[40] Open Data Institute (2018), The role of data in AI business models

[41] Autonomy (2020), Platforming Equality: policy challenges for the digital economy

[42] United Nations (no date), The 17 goals

[43] GPAI (2022), Enabling data sharing for social benefit through data trusts

Other examples can be found in **the field of labour rights**, where an increasing number of tools and initiatives rely upon data about workers being made widely available, accessible, usable or open:[44]

- Worker Info Exchange supports trade unions and grassroots worker organisations to collect data gathered on them by employers (through subject access and data portability requests), aggregate this data and analyse it in support of collective bargaining.

- Driver's Seat allows gig economy drivers to combine their driving data with data from other drivers to gain insights about their gig work. Driver's Seat then sells this data to municipal agencies on behalf of drivers and distributes the profits back to members.

- WeClock's worker collectives allow workers to collect, pool and analyse data on their own activities, supporting collective bargaining and allowing instances of employer malpractice to be recorded and investigated.

Some of the above initiatives, as well as efforts by trade unions to steward member data in creative ways were highlighted in a recent roundtable run by the ODI as part of its Experimentalism and the Fourth Industrial Revolution project.[45]

2.5 Mitigating harms through participation

Finally, participation in data stewardship can help to mitigate harms associated with data use.

Data is the result of a process of collection, standardisation and categorisation of social behaviour. It is not necessarily neutral or representative of an objective version of reality; it is instead situated knowledge, which results from a particular experience of the world. Many commentators have noted how data materialises human prejudice and reflects the knowledge, bias, and experience of those who collect, manage and analyse it.[46]

As a consequence, technologies which use data tend to reflect the existing social structures and power relations they are developed in. This, in turn, risks further exacerbating and intensifying the exclusion of already marginalised communities and social groups.[47] For example, a study from the Massachusetts Institute of Technology

---

[44] Open Data Institute (2021), Data institutions for labour rights and supply chains

[45] Open Data Institute (2021), Le Guin and data subversions

[46] O' Neil, C. (2016), Weapons of math destruction. How big data increases inequality and threatens democracy. New York: Penguin Random House; Benjamin, R (2019), Race after technology: The new Jim Crow. Durham: Duke University Press; Eubanks, V (2018), Automating inequality. New York: St. Martin's Press; Noble, S (2018), Algorithms of oppression: How search engines reinforce racism. New York: New York University.

[47] Broussard, M (2018), Automated unintelligence: How computers misunderstand the world. Cambridge, MA: MIT Press;Taylor, L (2017), What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2), pp.1-14; Benjamin, R (2019), Race after technology: The new Jim Crow. Durham: Duke University Press.

(MIT) has shown that voice and face recognition systems do not recognise the voices and faces of those who are not considered or participants in the programming of these models. In particular, black women are recognised far less often than both white men and women.

Empowering people who will be directly affected by the use of data to participate in the data stewardship process can help to mitigate some of these harms. Bottom-up approaches can help to ensure that the design of data institutions takes into account as many experiences and viewpoints as possible, improving their representativeness. This can enable a higher degree of confidence that the systems used to steward data are fair and just.

3. Mechanisms for government support

The UK's National Data Strategy, published in 2020, pledged to 'create an environment that supports existing data institutions in the public, private and third sectors'. Support for data institutions falls under Mission 1 of the National Data Strategy, 'Unlocking the value of data across the economy'.

'Data: a new direction', the 2021 DCMS consultation on proposed changes to the UK's data protection regime, referenced data intermediaries, asking whether, and how, the government should have a role in enabling the activity of responsible data intermediaries. The ODI's response highlighted that data institutions can help to improve data availability across the economy, and argued that the government needs both to actively support them and manage associated risks.[48]

At a European Union (EU) level, the European Strategy for Data[49] and accompanying legislation such as the Data Governance Act[50] and Data Act[51] aim to stimulate economic growth through industrial data spaces and promote a human-centric data economy via new data altruism organisations.

Over the course of our research for this report, we identified four areas that need to improve if bottom-up data institutions are to further grow:

- **Knowledge** – The existing evidence base on bottom-up data institutions and the size of the potential market opportunity they represent is limited.

- **Trust** – Bottom-up data institutions need to be trusted by the individuals and communities that use them, but there is some evidence that users are likely to be wary of new or unproven approaches to data stewardship.

- **Rights** – Bottom-up data institutions often rely on the contribution of user data through the exercise of data portability rights, but these are limited in UK and European law.

- **Infrastructure** – The technological infrastructures required for bottom-up data stewardship aren't always available.

While preparing this report, we have drawn inspiration from policies that have been implemented and proposed in the UK, EU and elsewhere in the world. The mechanisms suggested below are intended to be considered in the context of the UK

---

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

[49] European Commission (no date), Strategy for Data

[50] European Parliament (no date), Data Governance Act

[51] European Parliament (no date), Data Act

– although we hope they will be useful to policymakers in other countries – and any references to 'government' should be understood to refer to the UK government, except where otherwise indicated.

3.2 Knowledge

One of the factors preventing the further emergence and growth of bottom-up data institutions is the limited evidence base that currently exists on their effectiveness and potential.

Bottom-up data institutions are a new type of organisation, and few currently exist at scale. While there is some evidence (explored in Chapter 2) that participatory approaches to data stewardship can unlock new stores of data and enhance trust in existing institutions, there is a need to better define these opportunities and assess their potential to grow. Doing so will help to inform and align the activities of different actors operating in this ecosystem, including bottom-up data institutions themselves, government, and potential funders or investors.

There are a number of actions that could help to improve knowledge of bottom-up data institutions and their potential:

Join up and rationalise existing government support

The government needs to work across departments, arm's-length bodies, regulators and relevant third and private sector actors to join up and rationalise the provision of support for data institutions.

Currently, numerous departments and arm's-length bodies – including the DCMS, the Department for Business, Energy and Industrial Strategy (BEIS), the Department of Health and Social Care, and UK Research and Innovation – provide direct or indirect support for data institutions. Rationalising and joining up government support for data institutions – perhaps by designating a central team with an enhanced remit to coordinate support for data institutions across government – could help to:

- ensure value for money on government investments in this area

- improve the government's own knowledge-gathering and horizon-scanning capabilities by enabling more detailed ongoing monitoring of the ecosystem and its own role as a funder

- simplify the landscape for data institutions, including bottom-up data institutions, currently receiving support, and those seeking to receive it in future.

Getting this right and improving the government's understanding of its own data investments would support Mission 3 of the National Data Strategy: 'Transforming government's use of data to drive efficiency and improve public services'. This is a public good in itself: the government has a duty to get its own use of data right, to

improve how it makes policies, delivers services and otherwise effectively and efficiently runs its operations.

But it is also essential to, and cannot be separated from, achieving other missions of the National Data Strategy, including Mission 1, 'Unlocking the value of data across the economy'. The government can influence what others do through setting expectations and sending signals.[52] By understanding its own data needs, and identifying areas where participatory data stewardship could help to meet these needs, it can help to establish this as best practice across the economy.

Commission research and trials to assess the potential market for bottom-up data institutions across sectors

The government can continue to commission and support research and trials that will help improve understanding of the potential market opportunity for bottom-up data institutions.

This opportunity will no doubt vary depending on the specific use case or sector under consideration: some sectors may have data needs that could be met by new bottom-up data-gathering approaches, while others will not. Equally, in some cases existing silos of data might be improved by more participatory approaches to stewardship, while in other cases these approaches will not be appropriate.

In addition to this, the relative merits of different business models for bottom-up data institutions need investigation. As discussed in Chapter 1, bottom-up data institutions can rely on a variety of different incentives – including data altruism, access to services and financial reward – to drive user interest and engagement, but little research has been done on the effectiveness of different approaches.

By funding horizon-scanning research into new bottom-up approaches, the government will be able to provide private and third sector actors with the knowledge they need to build and engage with these new institutions.

Maintain the UK's rich institutional landscape and knowledge on bottom-up data stewardship

The government can continue to provide support for the UK's rich institutional landscape and knowledge on bottom-up data stewardship.

In recent years, the UK government has funded pioneering research into potential models for data intermediation, including DCMS funding for the ODI's own work on data institutions, Office for AI research into data trusts,[53] and in-house CDEI work on data intermediaries.

These organisations, including the ODI, can drive the research agenda, as well as provide benefits to UK-based data institutions in other ways, such as through

---

[52] Open Data Institute (2021), Mapping data in the UK government
[53] Open Data Institute (2019), Data trusts: lessons from three pilots

mentoring, advocacy and the provision of practical guidance. Maintaining the UK's hard-won thought leadership and practical knowledge in this area, will contribute to the information-gathering goals set out above and ensure that guidance is available to those bottom-up data institutions that need it.

3.3 Trust

Research carried out for the ODI by Frontier Economics has shown that trust in data ecosystems is a key determinant of data sharing.[54] There is some evidence – discussed in the previous chapter – that data institutions can enhance trust in data use, and that the empowerment and control provided by bottom-up institutions might be particularly conducive to this goal.

However, data institutions themselves require trust from the individuals and communities that use them. People need to be able to trust that they can safely share data with or through data institutions; data users need to know that they can reliably use data from data institutions; and people, organisations and communities affected by data institutions need to trust that they are having a positive impact.

For bottom-up data institutions, trust is a particularly acute issue: first, their explicit reliance on user willingness to engage means that they are more vulnerable to fluctuations in public trust than others data institutions; and second, users are likely to be wary of new approaches to data stewardship without first seeing proof of their effectiveness and independence.

There are a number of actions that could help to improve trust in bottom-up data institutions:

Provide a legal definition for (bottom-up) data institutions

The government can provide legal definitions for data institutions and specific data-gathering practices such as data altruism.

This could help bottom-up data institutions work towards a particular legal standard, and enhance public trust by providing clarity on what they can expect data institutions to do with data about them or in which they have a vested interest. It wouldn't necessarily require new primary legislation – it could be done through secondary legislation, or through guidance from the government or a regulator through the Information Commissioner's Office (ICO).

The European Commission's Data Governance Act seeks to do this by defining data intermediaries as organisations which provide 'data sharing services' and requiring that they comply with strict rules on how they handle data. For example, the act requires data intermediaries to 'act only as intermediaries in the transactions' and not use the data exchanged for any other purpose. It expects structural safeguards to be

---

[54] Frontier Economics for the Open Data Institute (2021), <u>The economic impact of trust in data ecosystems</u>

put in place to ensure this, with data sharing services provided through a legal entity separate from that which carries out any of the organisation's other activities. By clearly defining what activities data intermediaries can legally carry out, the Commission hopes that public trust in them will be strengthened, allowing the sector to grow.

Introduce a register for bottom-up data institutions

Another option would be to introduce a register for bottom-up data institutions.

The European Commission's Data Governance Act will require that individual member states, and the Commission itself, establish a 'register of recognised data altruistic organisations'. Once the act is passed, only controllers listed in this register will be able to call themselves a 'data altruism organisation recognised in the Union'. To register as a data altruism organisation, a data controller will need to meet the following criteria:

- be a legal entity constituted to meet objectives of general interest

- operate on a not-for-profit basis and be independent from any entity that operates on a for-profit basis

- performs the activities related to data altruism through a legally independent structure, separate from other activities it has undertaken.

By regulating which organisations can and cannot describe themselves as data altruism organisations, a register could provide greater clarity about the purpose of the data institutions with which they are sharing data. This would reduce instances of incorrect or fraudulent claims to data altruism, and help to build trust in data altruism as a practice. It could also aid the development of such organisations, supporting growth and innovation.

In the UK context, there are a number of different routes – both legislative and non-legislative – by which a register could be created. Its maintenance could be taken on by an existing body such as the ICO, Ofcom or the CMA.

Support the development of an assurance sector for bottom-up data institutions

Government can support the development of a mature assurance sector for bottom-up data institutions, including audit and certification services.

Audits and certifications are assessments that can be conducted internally by a data institution to improve its own trustworthy operation, or externally by a third party. Previous ODI research has explored the role of these mechanisms in providing assurance of data trusts[55] and of data institutions more broadly.[56]

---

[55] Open Data Institute (2019), Data trusts: will certification work?
[56] Open Data Institute (2020), Designing trustworthy data institutions

A study carried out in 2021 by Frontier Economics on behalf of the ODI identified a nascent but buoyant market of firms and services focused on trust in data, and revealed a potentially significant market opportunity for these organisations in the coming years.[57]

An example of this is the Japanese government, which has worked with the Information Technology Federation of Japan to create guidelines and a certification scheme for 'personal information banks', a form of bottom-up data institution. The first of these organisations was announced in 2019, and a number of established companies now offer this service.[58] The operation of personal information banks is dependent on a robust certification system, which ensures that businesses operating an information bank meet certain standards of reliability and trustworthiness. The certification scheme is government-backed, but managed by the Information Technology Federation of Japan – illustrating how governments can act as stewards for private sector-led schemes, building trust in data sharing.

However, the wide variety of different models that exist for bottom-up data institutions – as discussed in Chapter 1 – may act as a barrier to the rapid establishment of certification schemes. Researchers from Harvard University exploring the possibility of a certification scheme for platform cooperatives found 'an environment that was too new and diverse for a formal certification process or shared standards [in need of] greater collaboration among cooperatives on best practices'.[59]

3.4 Rights

A third challenge to the further emergence and growth of bottom-up data institutions is the relative weakness of data portability rights in the UK and the EU. The ODI has previously defined data portability as:

> 'the ability to share data between people, groups and organisations. A company, for example, might 'port' data – which could involve the transfer of data, or the provision of access to it – to a third party in order to deliver a particular service.'

The GDPR introduced a right to data portability, allowing users to demand the transfer of their data in a machine-readable format. This is the primary legal route which many existing bottom-up data institutions use to receive data. An example of this is WorkerInfoExchange, which makes use of data portability requests from individual users to receive data initially collected by gig economy platforms, such as Deliveroo and Uber.

---

[57] Open Data Institute (2021), New research for ODI reveals a nascent but buoyant market of firms and services focused on trust in data

[58] Nikkei Asia (2019), Japan's 'information banks' to let users cash in on personal data

[59] Berkman Klein Centre (2022), What we learned from our research sprint on cooperative data governance

As it currently stands, however, the data portability right isn't suited to supporting bottom-up data stewardship. Requests are often frustrated or denied outright on data protection grounds, and the right is subject to further limits that make its exercise cumbersome:

- it only supports one-off requests rather than continuous access

- data holders can wait up to a month to fulfil them, and often longer

- it is only a right for individuals rather than all account holders (which may be organisations or households).

Strengthening the right to data portability, and introducing new related rights, could help improve access to data for bottom-up data institutions and allow them to grow. There are a number of actions that could achieve this:

Enhance rights to data portability

Government can enhance rights to data portability, empowering individuals to share and transfer data concerning them to new types of organisations. Revisions to regulation could include:

- supporting continuous access to data through application programming interfaces (APIs), rather than one-off requests

- requiring data holders to fulfil requests within a month

- requiring data holders to cater for data portability requests from organisations, households and groups (for example shared accounts).

Further changes to data portability regulation should be considered on a sector-by-sector basis. It may be that this functionality develops naturally, without regulation: providing access to data is a strategy that organisations that hold data adopt if they think they can become the centre of an app ecosystem, as it makes their service more useful for consumers. It may be that regulatory intervention is only needed in sectors where no access is provided, or where companies form a club and only support data portability between each other.

The BEIS Smart Data agenda is an example of how greater data portability can be achieved in specific sectors without recourse to primary legislation, although new government powers may be needed to mandate participation in Smart Data initiatives.[60] The government should consider how Smart Data initiatives can be designed to facilitate the growth of bottom-up data institutions in sectors where greater user, consumer or citizen participation would be valuable.

---

[60] GOV.UK (2019), Smart data: putting consumers in control of their data and enabling innovation

Government and regulators should also be aware of the potential adverse consequences of enhanced data portability rights. While facilitating the flow of personal data between organisations could be hugely beneficial to bottom-up data institutions, in some circumstances having some friction around moving personal data to new services will be a good thing and prevent abuses.[61]

Provide a right to data reciprocity

A related reform could include the introduction of a right to data reciprocity, as advocated in a report by employment rights lawyers Robin Allen QC and Dee Masters for the Trades Union Congress.[62]

At present, data collection in most sectors is one-sided: employees, workers and consumers provide data, and employers or vendors use it, providing only generalised insights to data subjects in return. It is this asymmetry that motivates the use of data portability requests by bottom-up data institutions.

Allen and Masters argue that to redress this imbalance 'there should be a requirement that an employer can obtain consent to use data only if a reciprocal right is given to the worker'.[63] If extended to other groups such as consumers, such a reform could stimulate further growth of bottom-up data stewardship by allowing organisations to circumvent the friction created by limited data portability rights.

It could also help to enable consumer organisations, cooperatives, trade unions and membership organisations to take on data stewardship responsibilities on behalf of their memberships and become bottom-up data institutions themselves.

Loosen data processing regulations under GDPR for certified bottom-up data institutions

In addition to data portability, the GDPR introduced further legal rights for data subjects and requirements on data controllers, including bottom-up data institutions and their users.

These rights and responsibilities are important safeguards: they help to support trust in the data ecosystem. As we argued in our response to DCMS's 2021 consultation on data protection, weakening protections could potentially carry long-term and indirect costs in the form of negative externalities from the increased collection and use of sensitive personal data.[64]

---

[61] Open Data Institute (2018), Data portability: the role governments should play

[62] TUC (2021), Work and the AI revolution

[63] Ibid

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

However, it's also clear that harms and benefits from data policies and data practices don't remain static: they change over time. The growth of new, bottom-up forms of data stewardship – and new incentives to share data, such as data altruism – presents new opportunities for data to be used in innovative and socially beneficial ways. In some cases, where existing legislation serves as a barrier to this, there could be an argument for loosening or refining the obligations placed upon bottom-up data institutions.

German non-profit group AlgorithmWatch have highlighted a number of ways in which this could be done, including:[65]

- excluding the processing of personal data for altruistic purposes from the scope of the GDPR by creating an altruism exemption, similar to the existing household exemption

- providing data controllers deemed altruistic with an advantage when balancing interests required under Article 6 of the GDPR

- privileging further processing for altruistic purposes, either in all circumstances or in relation to personal data which is already publicly available.

These ideas are worthy of further consideration, but they would likely need to be accompanied by some of the measures discussed in the previous section on trust to ensure that greater rights to personal data are not abused by organisations claiming altruism on spurious grounds.

3.5 Infrastructure

The functions of a data institution – for example holding data on behalf of an organisation or person, or combining data from different sources – require well-functioning technical tools. This need is particularly acute for bottom-up data institutions, which will usually require some sort of front-end software to enable user participation. There are a number of ways in which this infrastructure could be improved:

Support the development of shared technological infrastructure

The government can support – through funding, piloting and testing – the development of shared infrastructure that can be used by bottom-up data institutions.

---

[65] AlgorithmWatch (2022), Data altruism: how the EU is screwing up a good idea

One example of this is DECODE (DEcentralised Citizen-owned Data Ecosystems), an EU-funded project designed to develop practical solutions that would put individuals in control of personal data. DECODE ran four pilots in Amsterdam and Barcelona to test this technology between 2017 and 2019. DECODE provided free, open source, interoperable tools to enable bottom-up, community-driven projects using data, including:

- DECODE OS, a private, anonymous, peer-to-peer network

- DECODE app, providing anonymous authentication for digital democracy applications

- BarcelonaNow, allowing the creation of interactive dashboards for the exploration, interpretation and sharing of data.

By supporting the development of freely accessible, open source tools such as this, the government could facilitate the growth of bottom-up data institutions in different sectors that depend on this infrastructure.

Experiment with data institutions as technology providers

The government can support the development of infrastructure for bottom-up data institutions by experimenting and procuring services that already exist.

In 2020, the government of Flanders – the most populous sub-national unit of Belgium – partnered with Inrupt and Solid to pilot a service called My Citizen Profile. The aim of this was to provide every citizen with a Secure Data Store or 'Solid Pod' to store data, with control over who would be able to access it and when.

Other public and quasi-public bodies, notably the NHS and the BBC in the UK, have taken part in similar trials of Solid Pods.[66]

The ODI's Experimentalism and the Fourth Industrial Revolution project is exploring how governments can adapt through experimentation to the unprecedented and fast-changing circumstances created by greater data availability.[67] Participants in project roundtables have highlighted the need to work across sectors when engaging in public sector experimentation around data as a way of building public trust and mitigating resource constraints.[68]

By involving bottom-up data institutions in this process, government can support them to develop and improve their services, as well as shielding itself from some of the potential risks involved in developing bespoke digital infrastructure.[69]

---

[66] Financial Times (2022), NHS signs up for Tim Berners-Lee pilot to reinvent web

[67] ODI (2021), Project: Experimentalism and the Fourth Industrial Revolution

[68] ODI (2021), Asimov's Data Revolution

[69] House of Commions Public Accounts Committee (2019), Government flagship digital identification system failing its users

Mandate interoperability and the development of open standards

The previous section highlighted the need for improved rights to data portability and control to make it easier for users to engage with bottom-up data institutions. These rights will only be executable in practice if organisations adopt common standards, allowing data to flow between them.

Standards for data are reusable agreements that make it easier for people and organisations to publish, access, share and use better quality data. Open standards are standards which are "made available to the general public and are developed (or approved) and maintained via a collaborative and consensus-driven process."[70]

---

[70] ITU (2022), Definition of "Open Standards"

By driving the greater adoption of open standards, the government can support the flow of data within and across sectors. One way of doing this would be to mandate interoperability: that is, requiring organisations in a specific sector to adopt a shared set of open standards which would enable users to freely transfer data between different service providers.

Recent work from the Ada Lovelace Institute has highlighted the use of interoperability mandates under existing competition law in the European Union, including rulings that have prevented Microsoft and Google from privileging use of their own products and services on hardware using their operating systems.[71]

In the UK context, interoperability mandates are a feature of existing Smart Data initiatives: Open Banking makes use of a mandate from the Competition and Markets Authority (CMA) requiring current account providers to interoperate with authorised payments providers.[72]

Building on these initiatives, a new Digital Markets Unit (DMU) has now been established within the CMA, with responsibility for 'enabling greater personal data mobility and systems with open standards'.[73] The government has proposed that the DMU will need the power to mandate interoperability and third party data access, with forthcoming legislation expected to enable this.[74] Considering the impact of open standards on bottom-up data institutions could form an important part of this work.

---

[71] Ada Lovelace Institute (2021), Making interoperability work in practice: forms, business models and safeguards

[72] Open Banking (2021), Variable Recurring Payments required for Sweeping

[73] GOV.UK (2019), Unlocking digital competition, Report of the Digital Competition Expert Panel

[74] GOV.UK (2021), A new pro-competition regime for digital markets

Methodology

This work set out to better understand 'How can the UK government create an enabling environment for bottom-up data institutions?'.

Our research questions were:

- What are the main barriers that bottom-up data institutions encounter?
- Why should policymakers care about bottom-up data institutions flourishing?
- What is already being done to support bottom-up data institutions?

We undertook a mixed methods approach, consisting of desk research and interviews with expert stakeholders.

Our desk research focused on identifying bottom-up data institutions and relevant policies in the UK and around the world. We sampled for bottom-up data institutions from the data institutions register. We examined:

- general information about the organisation – core goals or mission, governance paradigm, and sector

- form of data institution – the functions it fulfils or services it provides

- revenue models, considering organisations that are both for-profit and not-for-profit

- main activities and projects.

For existing policies, we assessed:

- existing governmental support for bottom-up data institutions in the UK

- existing governmental support for bottom-up data institutions in other countries

- broader data priorities and goals.

We conducted 13 interviews with 17 people from bottom-up data institutions, government bodies, and relevant experts and stakeholders.

We aimed to cover a range of bottom-up data institutions, of different maturity levels, stewardship roles and sectors, prioritising socially oriented institutions. We asked the organisations about their governance model, their needs, the barriers and challenges encountered, the support received from governments and the support desired. We interviewed ErnieApp, Schluss, Salus.coop, OpenStreetMap and WorkerInfoExchange .

Through our interviews with government bodies and other stakeholders, we sought to understand the support they currently provide to bottom-up data institutions, the barriers they encounter in doing so and the outcomes desired from that support. We

spoke to the Centre for Data Ethics and Innovation, Department for Business, Energy and Industrial Strategy, Department for Digital, Culture, Media and Sport, Treasury, European Commission, the Ada Lovelace Institute, the Aapti Institute and the International Centre of Expertise in Montréal on Artificial Intelligence.

We're conscious of the following limitations of this work:

- A limited number of interviews with bottom-up data institutions and policymakers, which restricts our interview findings to a limited amount of lived experiences, needs and perspectives.

- A focus on emerging and successful bottom-up data institutions without covering instances of failed bottom-up data institutions, which has not illustrated their lessons learned or barriers that could not be overcome.

- Sampling from the ODI's ecosystem, which skews towards European organisations and to views that may be similar to our own.

This study is part of our ongoing effort to increase awareness and understanding of data institutions among policymakers, funders and others. We hope that our findings indicate some potential routes for future policy development that the UK government and other policymakers can draw on to create an enabling environment for data institutions to thrive. Future work in the ODI's data institutions programme will aim to build on this research, further developing our understanding of the mechanisms discussed in this report and how they might be implemented.