The new ecosystem of trust

Here we attempt to open up part of the debate on data governance; suggesting how to address the twin goals of greater control for citizens, and greater value for the public as a whole. We argue that there are a variety of different solutions that need to be designed, and experimented with.

Overview

The world is struggling to govern data. The challenge is to reduce abuses of all kinds, enhance accountability and improve ethical standards, while also ensuring that the maximum public and private value can also be derived from data.

Despite many predictions to the contrary the world of commercial data is dominated by powerful organisations. By contrast, there are few institutions to protect the public interest and those that do exist remain relatively weak. This paper argues that new institutions—an ecosystem of trust—are needed to ensure that uses of data are trusted and trustworthy. It advocates the creation of different kinds of data trust to fill this gap. It argues:

- **That we need, but currently lack, institutions that are good at thinking through, discussing, and explaining** the often complex trade-offs that need to be made about data.
- **That the task of creating trust is different in different fields.** Overly generic solutions will be likely to fail.
- **That trusts need to be accountable**—in some cases to individual members where there is a direct relationship with individuals giving consent, in other cases to the broader public.
- **That we should expect a variety of types of data trust to form**—some sharing data; some managing synthetic data; some providing a research capability; some using commercial data and so on. The best analogy is finance which over time has developed a very wide range of types of institution and governance.

This paper builds on a series of Nesta think pieces on data and knowledge commons published over the last decade and current practical projects that explore how data can be mobilised to improve healthcare, policing, the jobs market and education. It aims to provide a framework for designing a new family of institutions under the umbrella title of data trusts, tailored to different conditions of consent, and different patterns of private and public value. It draws on the work of many others (including the work of GovLab and the Open Data Institute).

Introduction

The governance of personal data of all kinds has recently moved from being a very marginal specialist issue to one of general concern. Too much data has been misused, lost, shared, sold or combined with little involvement of the people most affected, and little ethical awareness on the part of the organisations in charge.

The most visible responses have been general ones—like the EU's GDPR. But these now need to be complemented by new institutions that can be generically described as 'data trusts'.

In current practice the term 'trust' is used to describe a very wide range of institutions. These include private trusts, a type of legal structure that holds and makes decisions about assets, such as property or investments, and involves trustors, trustees, and beneficiaries. There are also public trusts in fields like education with a duty to provide a public benefit. Examples include the Nesta Trust and the National Trust. There are trusts in business (e.g. to manage pension funds). And there are trusts in the public sector, such as the BBC Trust and NHS Foundation Trusts with remits to protect the public interest, at arms length from political decisions.

It's now over a decade since the first data trusts were set up as private initiatives in response to anxieties about abuse. These were important pioneers though none achieved much scale or traction.

Now a great deal of work is underway around the world to consider what other types of trust might be relevant to data, so as to fill the governance vacuum—handling everything from transport data to personalised health, the internet of things to school records, and recognising the very different uses of data—by the state for taxation or criminal justice etc.; by academia for research; by business for use and resale; and to guide individual choices. This paper aims to feed into that debate. Taking inspiration from Kieron O'Hara's recent paper, we use the term data trusts throughout to broadly denote institutions that work within the law to provide governance support for processing data and creating value in a trustworthy manner.

1. The twin problems: trust and value

Two main clusters of problem are coming to prominence. The first cluster of problems involve *misuse* and *overuse* of data*;* the second set of problems involves *underuse* of data.

1.1. Lack of control fuels distrust

The first problem is a lack of control and agency—individuals feel unable to control data about their own lives (from Facebook links and Google searches to retail behaviour and health) and communities are unable to control their own public data (as in Sidewalk labs and other smart city projects that attempted to privatise public data). Lack of control leads to the risk of abuses of privacy, and a wider problem of decreasing trust—which survey evidence from the Open Data Institute (ODI) shows is key in determining the likelihood consumers will share their personal data (although this varies across countries). The lack of transparency regarding how personal data is then used to train algorithms making decisions only adds to the mistrust.

1.2 Lack of trust leads to a deficit of public value

The second, mirror cluster of problems concern value. Flows of data promise a lot: better ways to assess problems, understand options, and make decisions. But current arrangements make it hard for individuals to realise the greatest value from their own data, and they make it even harder for communities to safely and effectively aggregate, analyse and link data to solve pressing problems, from health and crime to mobility. This is despite the fact that many consumers are prepared to make trade-offs: to share data if it benefits themselves and others—a 2018 Nesta poll found, for example, that 73 per cent of people said they would share their personal data in an effort to improve public services if there was a simple and secure way of doing it. A key reason for the failure to maximise public value is the lack of institutions that are sufficiently trusted to make judgements in the public interest.

Attempts to answer these problems sometimes point in opposite directions—the one towards less free flow, less linking of data, the other towards more linking and combination. But any credible policy responses have to address both simultaneously.

2. The current landscape

The governance field was largely empty earlier this decade. It is now full of activity, albeit at an early stage. Some is legislative—like GDPR and equivalents being considered around the world. Some is about standards—like Verify, IHAN and other standards intended to handle secure identity. Some is more entrepreneurial—like the many Personal Data Stores launched over the last decade, from Mydex to SOLID, Citizen-me to digi.me. Some are experiments like the newly launched Amsterdam Data Exchange (Amdex); the UK government’s recently announced efforts to fund data trust pilots to tackle wildlife conservation, working with the ODI; and the Information Commissioner's Office new regulatory beta sandbox to support the use personal data in innovative products for the public interest. Finally, we are now beginning to see new institutions within government to guide and shape activity, notably the new Centre for Data Ethics and Innovation.

Many organisations have done pioneering work, including the ODI in the UK and NYU GovLab with its work on data collaboratives. At Nesta, as part of the Europe-wide DECODE consortium, we are helping to develop new tools to give people control of their personal data while the Next Generation Internet (NGI) initiative is focused on creating a more inclusive, human-centric and resilient internet—with transparency and privacy as two of the guiding pillars.

The task of governing data better brings together many elements, from law and regulation to ethics and standards. We are just beginning to see more serious discussion about tax and data—from the proposals to tax digital platforms turnover to more targeted taxes of data harvesting in public places or infrastructures—and more serious debate around regulation. This paper deals with just one part of this broader picture: the role of institutions dedicated to curating data in the public interest.

3. Mapping the current state of data governance

Most of the debate about data governance has focused on the first problem we identified—designing new ways to give individuals more control over their data, protecting privacy and implementing privacy by design. Getting this right can generate a lot of private value—as we're beginning to see with open data in banking.

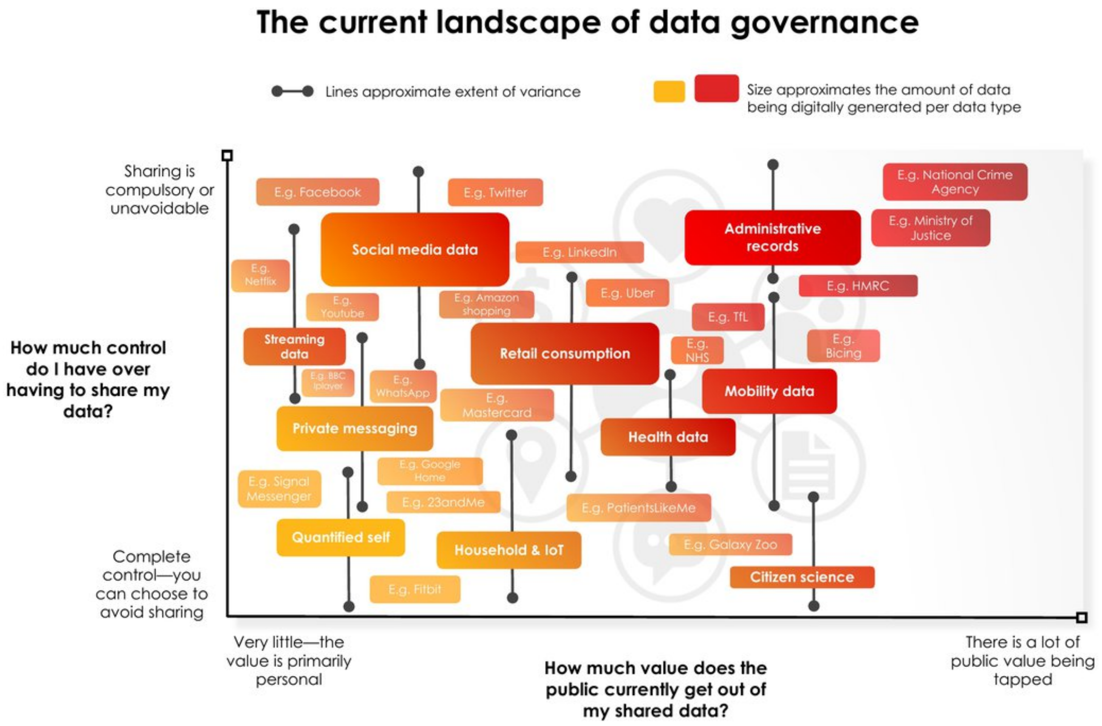
Much less work has been done on the second set of issues. Here, in fields where there is great potential public value to be reaped from linking data in new ways, very different solutions will be needed. To make sense of the options we suggest here a series of dimensions for thinking about the challenges of data, so that we can be clearer about which kinds of governance solution are most appropriate for which kinds of task.

3.1 Mapping the landscape

We first map the current state (or lack) of data governance on two main axes before introducing our framework for thinking about the future. To keep it as simple as possible, the examples of data we have mapped, e.g. health data, refer to very general types of database. Next to each type of database, we have included examples of the types of companies and organisations that are currently active in collecting and sharing data, e.g. NHS in the case of health data. Crucially, all data that we focus on here is personal data, that is primarily generated and collected digitally.

The y axis in our diagram, the control axis, concerns how much control or choice the individual can play in determining how personal data is shared and used. At one end of the axis there is near complete control: you have voluntary choice and can decide who can access or use personal data and for what purpose. At the other end you have little control; sharing is compulsory or unavoidable. That may be because an institution is legally empowered to collect data for particular uses, as in the case of governments collecting tax or criminal justice agencies tracking down and prosecuting criminals. But it may also be because data about you is harvested without your consent or because there are few realistic alternatives to using certain digital platforms and services, as they have become the dominant market incumbents (because of network effects). In theory you may have a lot of control and choice; but in practice you often have little—we have tried to show this divergence in the diagram by depicting variance for each category; for example, the extent to which you can avoid sharing private messenger data by using an encrypted communication app like Signal instead of alternatives that may be more popular but less secure.

The x axis, the net value axis, tries to map the balance of public and private value that is there to be realised from using the data. Again, this is also schematic and stylized rather than precise, but it's a crucial dimension when it comes to governance.



The current landscape of data governance

At one end of the axis the data is only really of value to you and only concerns your own behaviour and choices: for example the data from your own Fitbit, Netflix viewing history, or your own household data generated from an Internet of Things device, e.g. fuel consumption (of course all data can be useful for research purposes—the point is that the *main* value is private).

At the other end of the axis the primary potential value of data is public or shared: health data, mobility data, and what we are more broadly calling administrative data, e.g. data on crime or educational outcomes; which allow for large scale health analyses, or analyses of patterns of social mobility. Any data that is being collected for research purposes would also sit at the right side of the value axis.

Again, the patterns can be messy. Even the most public data has some private value (especially in health). But distinguishing the patterns of value is helpful in thinking through what kinds of governance are needed.

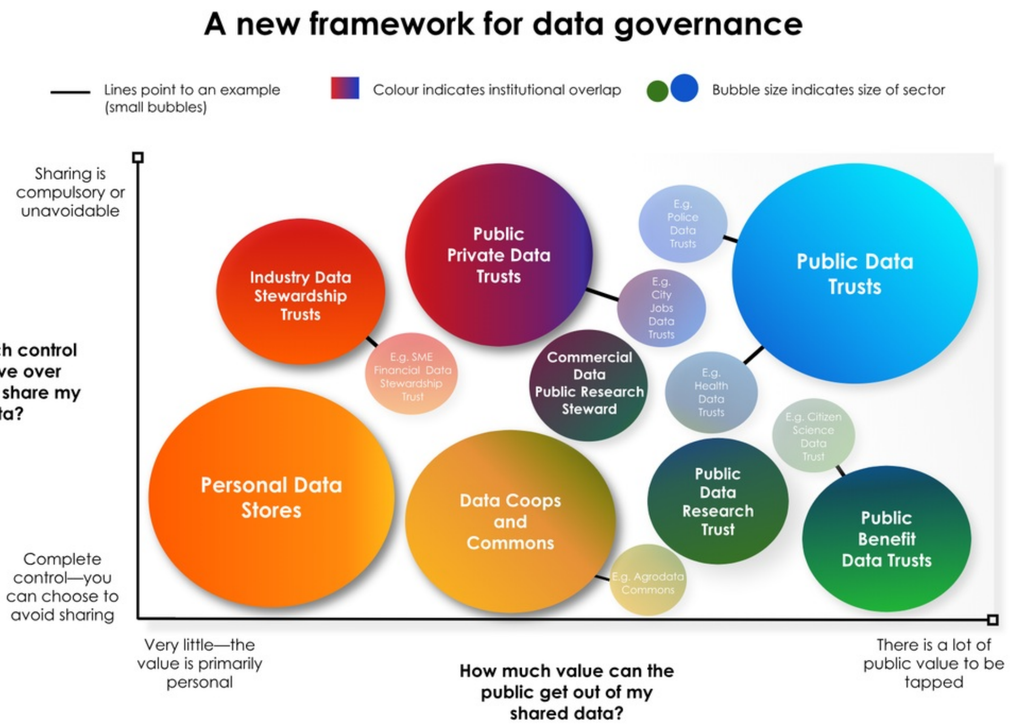
4. A new framework for data governance

Many would agree that the diagram above pictures a field where arrangements are problematic. Most retail consumption data currently sits in the top left corner but should increasingly be in the bottom left. Individuals should be able to decide if and when their data is used. Social media platforms have tended to harvest data with very little consent but are gradually, under public pressure or through new legal obligations, giving more rights to users to download or share their own data.

So what should sit where? Here we suggest a typology of forms of governance that could both enhance personal control and make possible more widespread sharing of data, with the right privacy protections, to serve the public interest.

Voluntary Data Associations/PDS/Data Coops

We start in the bottom left corner. Many of us want to control our own data and to be able to decide who it's shared with and on what terms. Intermediaries like Personal Data Stores (PDSs), well described here, offer a way to do this and a simpler route to accountability—since you can choose to leave a PDS that no longer works for you. The key point is that control is exercised through individual consent. The best models will be ones similar to other kinds of association or cooperative. There is lots of scope for developing these further, with some common standards to ensure data security, compliance and technical treatment.



There's also no shortage of possible applications. For example, farmers could share personal data on their crops (e.g. geospatial data) in exchange for a modest payment, or rights to access personalised analyses and forecasts. They could also pool data they are happy to share with a wider community for purposes such as measuring food security and form a type of agricultural data commons. Similarly, groups of patients with a rare condition could choose to pool data to accelerate the discovery of solutions. Some consumers might choose to share their data with retailers in exchange for payments (as we do in a limited way through loyalty cards).

Examples like midata.coop and saluscoop.org show what could be possible. These entities will be more like consumer cooperatives, or classic voluntary associations, with some kind of membership, and appropriate regulation by the Charity Commission or equivalent. Some may be offshoots of existing institutions. The BBC could potentially move in this direction too: highly trusted already, it is looking at creating reliable datastores of listener and viewer data to help improve services and accessibility.

Public Data Trusts

Next we move across to the top right hand corner: institutional arrangements for maximising public value when data sharing is compulsory or hard to avoid. Here we need new institutions which can be held to account for how well they manage data security as well as how well they maximise the value of this data. We expect that existing public services will not be able to generate trust through their existing machineries, but can benefit greatly from more data sharing. What will be needed are sector or field specific bodies, with their own personality.

Health is an obvious current example. We would expect a nested series of new entities (national and local **Health Data Trusts**) to be charged with maximising the public benefit from health data while respecting privacy and consent, for example around linking patient records, diagnoses, genomic and socio-economic and behavioural data. It will become increasingly important that there are visible roles here—people who can be held to account, and also explain on the media how and why choices are being made. The absence of these is becoming ever more of a problem, and is likely to inhibit the emergence of a smarter health field.

Another example would be crime data that combines citizen generated information, e.g. reports of robberies or burglar alarms going off, data from business (e.g. on patterns of crime in retail centres) and information from police themselves. Here much of the data is likely to be collected passively. There are clear potential benefits to be gained from better linking of data. But stronger public accountability will be needed to ensure that data use is genuinely in the public interest, with new intermediary institutions—**Police Data Trusts** (which in the UK would be linked to Police Commissioners)—charged with curating citizen-generated and administrative data, protecting privacy, and ensuring transparency.

Other variants of the Public Data Trust will be needed for other uses, including research. A **Public Data Research Trust** can act as the authorised guardian of a range of types of administrative and social data (from national and local governments etc.), providing this to authorised projects from authorised providers with built in audit of the data uses. Access could be provided via APIs and some commercial data could be added in to these data pools. Individuals could be given the option to opt in or out: it would be a matter for public policy to decide whether by default they would be opted in. As we know from organ donation very different behaviours result from different designs in this respect.

A good example might be the UK's LEO which links together school data with future earnings data from the tax authorities, opening up potentially very valuable insights on the effects of education on mobility. These would best be designed with governance that mirrors other kinds of research body, but with ethicists and non-academic representation. The UK's Office of National Statistics (ONS), for example, has a high reputation already and would be well-placed to branch out into a broader role of data curation.

Alternatively, another type of **Commercial Data Public Research Steward** would gather together large data sets (e.g. browsing behaviours) from commercial organisations and enable researchers to propose hypotheses for testing against the data, but without needing the researchers to access any of the data itself. These could be run by government research bodies (like the ONS) as part of the infrastructure for research or by new bodies.

Industry Data Stewardship Trusts

Our past work on banking data required the creation of new vehicles to look after commercial data, in this case the financial data of individual small businesses. This has to be guarded carefully; shared with authorised third parties under the authority of the consumer; and ideally with strong audit trails to ensure no misuse. We expect a range of vehicles of this kind to be needed for many industries, from energy to mobile phones, often using APIs. These would generally require governance that linked in industry regulators as well as big and small firms and consumers. For these, as for several of the other types of trust suggested here, it will be vital that business develops new internal roles for what Stefaan Verhulst calls ‘data stewards’ to navigate what is to be shared and on what terms.

Public Private Data Trusts

Another category for governance will be data around specific fields, linking public and commercial data. Mobility is a good example. There will be obvious gains for cities that can pool data around mobility to manage flows of all kinds. Ambitious projects like Shared Streets already show that a trusted intermediary which brings together city councils and local companies can be a catalyst for strengthening data standards (and improving the interoperability of data). There may also be a need to harvest some data on other issues. But it will be vital that all of this data is managed carefully to avoid abuses—with strong accountability and standards. These kinds of Data Trust would need more formal governance; a statutory footing; and multi-stakeholder representation. They will work best if clearly situated in particular sectors or fields rather than being too generic, as the choices and trade-offs will depend on context. Drones would be another case—Nesta’s Flying High programme is already well placed to design prototypes.

Our Open Jobs programme has shown how these could work in the labour market context. At a national level we propose a pooling of multiple data sources including administrative data; survey data; data drawn from scraping the web (for job advertisements and skills requirements); and ideally some commercial data. At a city level we also envisage **City Jobs Data Trusts** linking councils, business, universities and others, to ensure the maximum pooling of data to help people navigate careers and jobs choices.

Public Benefit Data Trusts

In the bottom right quadrant there are another group of examples that link together voluntarily provided data but primarily for public benefit, building out from the many great examples of citizen science. This might include households pooling air quality data; patients pooling their own carbon emissions data and so on. Here it is important that the arrangements suit the voluntary nature of data sharing; they should be open, convenient, and easy for citizens to use and accessible for researchers and policymakers.

We are sure that this list isn't comprehensive. But we hope that it already confirms the diversity of responses that is needed and will prompt others to propose their own solutions.

5. Achieving the right balance

Some types of data are particularly complex and straddle several quadrants of the diagrams. Genomic information arguably sits in the middle of the value axis—there is high individual value to tailor health treatments, but also a big public value in making sense of large scale patterns (linked to health outcomes and ancestry). Achieving the right balance between maximising public value and maintaining individual privacy is especially problematic, as DNA sequences are unique (with the exception of identical twins), meaning that a DNA sample can never be truly anonymized.

Schools data is another interesting example. The UK has been collecting a range of types of pupil data—on performance and many other indicators. This will have a direct value to pupils themselves, a value for research and for any providers of machine-learning based tools to help schools with assessment or curriculum. This diversity of potential kinds of value makes it even more vital that attention is paid to trust. The failed efforts of inBloom to improve education in the US, for example, were in part due to criticism the company faced for not requiring parental consent before moving personal student data from school district databases to its own ones.

As for highly-sensitive personal data sets that can never be fully open, it may become increasingly possible to anonymise it (through advances in cryptography) or substitute personal data with synthetic data—this has already been used in the Netherlands to create 'digital twins' of entire cities. Promising techniques include differential privacy, which has already been trialed for the US Census (albeit with limited success). The DECODE project is also experimenting with homomorphic encryption, as a method of deriving insights from encrypted data. But even the use of synthetic data should still be combined with strong legal protections and public accountability—as researchers at the University of Washington have previously argued.

5.1 Trust: who and for what

The ODI-YouGov survey cited earlier shows the range of attitudes to trust amongst the public. Most UK consumers (64%) trust the NHS and healthcare organisations with personal data about them, ranking top ahead of banks (57%), local government (41%) and online retailers (22%). Global comparisons of trust attitudes around the world today suggest an even greater variation between countries and over time (e.g. with trust in police often higher than trust in the rest of government).

These patterns reflect a mix of factors: perceived competence; moral integrity; direct experience, to name a few. Attitudes also vary across social groups. Age, for example, greatly influences the extent to which we are comfortable sharing personal information. In our 2018 Nesta poll, for instance, the figure for people who said they would share their personal data in an effort to improve public services jumped to 79 per cent among 18-24 year olds but fell to 68 per cent among 55-64 year olds.

All of the organisations and sectors on this list have been through periods of declining trust and periods when trust has been rebuilt. The key point is that trust has to be continually earned, and is not generic: it is trust to do particular things and at particular times. Any new institutions will need to show a track record of successful initiatives that provide demonstrable value; excellent communication and advocacy about what they do; abundant evidence of the care taken to get decisions right; and respected figures in leadership positions.

A healthy debate has now started on how best to design new institutions to grow public trust (as in this recent report from the Ada Lovelace Institute which echoes our view that data trusts need to be designed in ways that enable them to build trust, rather than these being technocratic or inward looking bodies like a previous generation of data and digital teams in government).

6. Finding a common language

An important challenge in any discussion of data is to use the right terms. Some widely used terms may impede rather than illuminate the choices. A key example is the use of a language of ownership and property. These terms seem intuitively clear and relevant, and lead some to want arrangements in which data is monetised—so if I agree to share personal data I should be paid for it.

But the language of ownership is not very meaningful upon closer inspection. In contrast to oil or other physical goods, data is everywhere, virtually infinite and non-rivalrous. It is more like an element than an object and just as factual information and abstract ideas can never be 'owned' by any single individual, neither can data—it exists conceptually separate from authorship. In a similar vein many have argued that data is really more like a public good than a private one (see, for example, this recent blog from Professor Diane Coyle).

The value of data tends to grow less through accumulation than through linking and analysis. Its value is determined not by the strength of boundaries around it but by the number of links it has. What matters is who is collecting the information, who is controlling it, and who is using it. The important analytical point becomes differentiating between 'holding ownership of data', and the 'rights to access and control data'.

But even 'rights over data' can be ambiguous, both from a legal and ethical standpoint (e.g. what rights do I have over personal health data that might help the world stop an epidemic?). As we have attempted to show in the above framework, the answers will always depend on the specific context and uses.

That diversity may however benefit from common standards, just as the diversity of the Internet rests on the TCP/IP protocol and URLs. Standards - potentially like Finland's IHAN, cited earlier, which aims to embed greater citizen control over data uses - could reduce the transactions costs associated with new designs and making it easier for trusted institutions (like a government or bank) to *hold* information about you, whilst you still have special access *rights* to that personal data—so that you can determine what happens with it. GDPR is a step in the right direction for this, but any future comprehensive 'Bill of Data Rights' will have to be a lot more detailed than current suggestions.

Conclusions

This brief paper has attempted to open up part of the debate on data governance suggesting how to address the twin goals of greater control for citizens, and greater value for the public as a whole. It suggests the variety of different solutions that need to be designed, and experimented with.

Growing the family of institutions set out here will take time. It will also require energetic action to grow the skills needed to run these well (a big challenge given pay and demand in the private sector); and a more fine-grained public debate about benefits, risks and trade-offs.

But this work is overdue. For a long period the prevailing ideological bias worked against the creation of new public interest institutions. That left us with a huge imbalance between strong, rich private organisations with global reach and a smattering of very weak public ones. That imbalance now needs to be put right.

We are grateful to other Nesta colleagues for inputs to this paper, including Tom Symons, Eddie Copeland, Laurie Smith, and Theo Bass.

Chief Executive Officer

Geoff Mulgan has been Chief Executive of Nesta since 2011. Nesta is the UK's innovation foundation and runs a wide range of activities in investment, practical innovation and research.

View profile
Executive Research Assistant

Vincent was the Executive Research Assistant Intern

View profile