



A NEW FAMILY OF DATA COMMONS?

Geoff Mulgan

Over the next two years Nesta will be involved in an interesting European project on data commons, with a partnership from across Europe that includes the cities of Barcelona and Amsterdam. The aim will be to design and test out new ways of combining data that give citizens greater control, while also making it easier to generate public value through linking data together.

What kind of Internet?

The project is a response to a remarkable historical irony. The Internet, and later the World Wide Web, emerged out of hope that it would be a democratising force, open to anyone and free to use, without walls or ownership: the sort of world John Lennon's Imagine tried to conjure up.

Some of that spirit survives. But much of today's reality is very different. Power has been consolidated in a very small number of organisations, run as traditional commercial firms. They've grown huge thanks to the combination of business models that rely substantially on how they use and sell your data, and network effects which mean that the ones that first became big then tended to become enormous, because the marginal cost of another user was so much lower for them than for their competitors. That's why we live in a world where the dominance of firms like Google, Amazon and Facebook is so much greater than their equivalents in previous industrial eras.

As consumers we've done well out of the deal that offered us free services in exchange for handing over our data. But such consolidation of power is problematic in all sorts of ways. Our most important infrastructures are now unaccountable, driven by incentives that are often against our interests (we're not their customers, but rather a commodity which they sell). We have very little control over our own data. Meanwhile much of the potential value of the Internet is not being realised because it doesn't fit the dominant business models. The major players may not intend to do harm. But harm often results as a by-product of their core activity.



What's the answer?

Part of the answer is simply to promote more pluralism. Other fields have thrived by fostering pluralism – combining commercial business with public services, non-profits and small firms. TV and radio combine commercial, new digital channels, alongside public services such as the BBC. Health and education the world over are similarly pluralistic. The Internet by contrast has become much more of a monoculture – a huge irony given the high hopes with which it was born.

Changing how we think about data

So how could a better balance be achieved? A good starting point is to rethink how data is organised, since the sale of data underpins much of the current oligopoly Internet model. If instead of asking what data could be sold by one business to another, we instead ask what people really need from data, very different answers come to the fore, that are about rather more than better marketing of the latest black Friday deals.

Transport has shown some of what can be done. Thanks to open data it's now easy to navigate complex routes, and find out how long a journey will take us through road, buses, trains or on foot. In the UK, for example, transportapi has aggregated 80% of all UK transport data, and now has some 1500 developers building new services on top of it. The data is much more valuable to us because it's pooled together. Some of the services are still advertising financed – but the underlying data is treated as a common good.

This example is a reminder of the many types of data that need to be considered very differently, data that's small or big; public, commercial or private. There's highly personal data; data that is necessarily held within organisations; data that is covered by contractual arrangements between organisations; data that is governed by specific authorisations (eg medical data); or fully open data. The ODI's description of a ['data spectrum'](#) provides one useful framework for avoiding confusions.

Here I summarise a few of potential options for more pooling of data and their potential implications. I then look at some of the competing technologies, including blockchain and other distributed ledgers and tools for more 'general transparency'.

Fields where data commons have the greatest potential



A rough list of important new uses of data that have some of the characteristics of commons includes:

Help with everyday choices for individuals and families – there are fantastic tools for helping us to shop or go on holiday, but far fewer good tools for some of the decisions that really matter, mainly because these are hard to finance through advertising. A good example of what’s needed is the [Skillsroute app](#) that helps teenagers choose what courses to do. The [health knowledge commons](#) which we have advocated is another example, bringing together what’s known about diseases, diagnoses and treatments, making it easy to use, and then, in time, making it feasible to link this to very personal data, for example about genetic dispositions. In the long run we should expect a range of types of service of this kind – some purely commercial, some philanthropic or voluntary and some provided by public organisations.

Help with everyday choices for organisations, from small businesses to charities. The new prize Nesta is [doing with the banks](#) to open up data for small firms so that they can better plan their finances is a good example, and one that I hope will be copied by other regulators in other fields. It complements the move to use standards and regulation to require banks to open up their data with APIs. Opening up law to make it easier for small organisations to find relevant legislation and judgements, and then to use AI to predict how the law might be applied, is another good example of a very important public good that is currently missing because it doesn’t fit orthodox business models.

Prediction and prevention of harms. Predictive algorithms have long been used in criminal justice and health. New York and London have used predictive algorithms to better target fire services and many public agencies are beginning to gather and pool data so that they can use resources more smartly. Deepmind’s new project with London hospitals to map blood tests and better predict health risks is a good example (and is creating a kind of data commons to hold the data, even while Deepmind will own the algorithms).

Better services. In some cases open data makes it possible radically to reimagine services, for example showing where buildings are underused, or linking up datasets when someone has a health emergency. Some of these new services may be based on the blockchain, for example in energy, housing, transport and finance. Others will build on open APIs, like the ones already used in transport. These could already be used to regulate, and tax, parts of the sharing economy like Airbnb and Uber. Others still will help with targeting public action to where it’s most needed,



and Nesta is currently working with many of the big cities in England to set up [offices of data analytics](#) that will help them make the most of existing public data.

Better coordination of big systems. Next generation infrastructures will depend much more on data and information commons. New forms of transport – including driverless cars, and drones or VTOL craft carrying people – will require far more real-time pooling of data to ensure that routes are optimised (as is the case with air travel, and the Internet itself). Traditional models where this data is proprietary are beginning to look very anachronistic.

Better policy. Another growing field of activity is more systematic use of data to guide policy. [Nesta's work](#) using open, commercial and web data to better track how the economy is changing is one of many examples of how data can support better decisions, for example showing the makeup of local economies, their complexity and growth prospects. This is clearly a public good – and if analyses of this kind are only open to big, rich firms that's likely to be bad for growth.

More accountable power. [DCENT](#) in Helsinki, Barcelona and Madrid has shown how new kinds of platform can make public power more accountable, opening up to the public to monitor, propose and comment on policies and programmes, including through the use of open APIs.

What links all of these very different examples? All of them require new kinds of commons. They don't work if data is privately owned, or kept in separate silos. They also don't work if the public don't trust the holders and handlers of data. Instead they generate value by linking data up and treating it as a public good, while also protecting privacy.

For many years it's been a cliché to claim that information or data are the new oil. Yet this gets it very wrong. Oil is an inherently scarce physical resource. Data and information by contrast can be replicated without limit and often become more valuable the more they are shared. But also, unlike oil, data bears an aura of our most intimate secrets.

This is why the ownership and business models suitable for industries like oil are so ill-suited to making the most of data, and why other models may be much more relevant, including commons.

Some cities such as Barcelona are developing explicit strategies using data commons to handle services like water, transport and energy. Seoul has done the same, combining commercial and public data to generate useful new insights, and setting it free from the public and private silos it gets stuck in.



Others are experimenting with bottom up commons that allow people to control their own personal data, and decide who should access it and on what terms. As I've [written before](#) there are many technological and legislative ways of making that happen.

But the best forms for these remain unclear. Some argue that distributed ledgers like the blockchain are ideally suited for helping people to manage their own identity and determine who gets what personal information without the need to rely on intermediaries or governments. Others [warn that blockchains](#) – whose essential nature is that information is publicly verifiable – are particularly *unsuitable* for storing personal information of any kind, and problematic in other ways, including their openness to capture.

Tim Berners-Lee, founder of the WorldWideWeb is working on one alternative, under the name [Solid](#). Several of the big tech companies are developing their own versions of distributed ledgers, and trying to avoid some of the design peculiarities of blockchain. And [many researchers](#) are thinking about data collaboratives and commons could work.

The best of these new commons may achieve the ideal of a more comprehensive general transparency – where data can be independently verified; where it's impossible to tamper with past data inputs; and where the rules governing how data can be used are built into the system, so that any breach of privacy is automatically flagged up. In the long-run this may prove a better solution to the challenge of privacy than attempts to anonymise or de-identify data – which is hard now, and likely to become ever harder. In any event common standards will become increasingly important.

Types of commons

For some time to come we may see a coexistence of multiple repositories, some commercial, some run by governments and some resulting from citizen initiative. At least four different types of commons are likely to emerge:

- **Personal data commons** that are repositories of personal data, and that allow the individual control over who uses their data and how, including control over authentication, and the right to remove any past data. These could evolve in a purely voluntary way, bottom up. But much of their value comes from scale and my guess is that the successful versions will be hybrids, with some involvement of the public sector and the big digital companies.



- **Public data commons** are repositories of anonymised statistical data, for example on economic activity or mobile phone usage, that can be opened up for anyone to use. The current [city data stores](#) are a good example, as are many of the open data initiatives.
- **Knowledge commons** aggregate and orchestrate useful knowledge, for example on health, and in some cases can be linked to personal data, for example on genetic makeup. The websites of some of the 'what works' centres [also point in this direction](#).
- **City Data Commons** will allow pooling of public and private data in real time to allow the most efficient coordination of big infrastructures like transport and energy. These will need new structures of governance – PPPs – to balance what will often be conflicting interests. They will have very strict privacy provisions – since they will be tracking who goes where all the time. But they will promise big gains in efficiency.

Most of these are only likely to work if they have a hybrid character – combining some of the authority of the state with a civic quality, while also creating space for entrepreneurs, and big firms, to come up with creative ways of using data.

The challenge of designing them parallels that which faced the cities that emerged from the 19th century. They turned out to need a series of new commons from public health to roads, sewers to public parks and galleries, which underpinned an expansion of individual freedom and wealth. The postal service, for example, provided a public good, but also guaranteed that individual communications would be private.

The challenge is that this will require very different ways of thinking. Digital business is doing fantastically well. But profit-maximising, privately-owned companies are only one way of using technologies, and it's myopic to assume that there are no alternatives. The commercial models of today's Internet could soon look very anachronistic, with their dependence on naïve users who don't ask too many questions about what happens to their data.

There are [good reasons](#) to want greater pluralism and prevent a dominant monoculture. But for that to happen we need much more imagination to show how these many commons can best be run, best be financed and best be held to account.

We'll get the most from data if we can simultaneously share it more easily, because



its value rises through sharing and combination, and protect it more securely, because confidence depends critically on how well the intimate details of our lives are kept secret. The answers won't be simple - which is why we need the very best brains to help find them.

[This paper is based on a talk given by Geoff Mulgan at ESADE in Barcelona in November 2016]

info@nesta.org.uk [@nesta_uk](https://twitter.com/nesta_uk) www.facebook.com/nesta.uk www.nesta.org.uk

Nesta is a registered charity in England and Wales with company number 7706036 and charity number 1144091. Registered as a charity in Scotland number SC042833. Registered office: 1 Plough Place, London, EC4A 1DE.

December 2016