decode

Me, my data and I: The future of the personal data economy

September 2017

European Commission

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no.732546

Contract no. 732546

Authors

Tom Symons and **Theo Bass** (Nesta)

Reviewers/contributors

Pau Balcells Alegre, **Francesca Bria** and **Oleguer Sagarra** (Technology and Digital innovation Office, Barcelona City Hall - IMI); and **Gijs Boerwinkel**, **Tom Demeyer**, **Job Spierings** (Waag Society)

Project partners

BCMI Labs AB, City of Amsterdam, CNRS, Dyne.org, Eurecat, Technology and Digital innovation Office, Barcelona City Hall (IMI), Nesta, Open University of Catalonia, Politecnico di Torino/Nexa, Stichting Katholieke Universiteit Nijmegen Privacy & Identity Lab, Thingful, Thoughtworks Ltd., UCL, Waag Society.

Acknowledgements

This report was made possible thanks to the support of a number of people. We are particularly grateful to the participants of our workshop in May, which form the basis for the future scenarios that are outlined in Section 2 of the report.

The insightful comments of our peer reviewers on the draft of this report were invaluable in helping to shape the final outcome: **Pau Balcells Alegre**, **Gijs Boerwinkel**, **Francesca Bria**, **Tom Demeyer**, **Oleguer Sagarra** and **Job Spierings**.

We would also like to thank all of those who gave up their time to speak with us about the projects which they are involved with, both from within the DECODE consortium and further afield: **Jim Barritt** (ThoughtWorks), **Daniel Buchner** (Microsoft), **Bruno Carballa Smichowski** (Open University of Catalonia), **Marco Ciurcina** (Nexa Center), **Aik van Eemeren** (City of Amsterdam), **Mayo Fuster** (Open University of Catalonia), **Hamed Haddadi** (Queen Mary University of London), **Dr. Ernst Hafen** (Midata.coop), **William Heath**, **Fieke Jansen** (Tactical Technology Collective), **Stefano Lucarelli** (CNRS, Marcos Menendez (The Good Data), **Trent McCoghany** (BigchainDB), **Yves-Alexandre de Montjoye** (Imperial College London), **Annemarie Naylor** (Future Care Capital), **Antti Jogi Poikola** (MyData), **Giulia Rocchi** (CNRS), **Andrei Sambra** (MIT).

There are a number of people at Nesta who have helped steer this report and provide valuable contributions. Special thanks go to **Daniel Corredera** for his role putting together the case studies in Appendix 3, along with **Matt Stokes**, **Lydia Nichols**, **John Davies**, **Katja Bego**, **Eddie Copeland** and **Geoff Mulgan** for their comments.

Contents

Report purpose

This report is about **DECODE (DEecentralised Citizen Owned Data Ecosystems)**, a major EU Horizon 2020 project to give people control of their personal data. Its purpose is to:

1. Outline the problems that DECODE is trying to solve and why they are worth addressing.
2. Explore what the world is likely to look like 20 years from now if the status quo continues and present the alternative vision that DECODE offers.
3. Explain why the tools proposed by DECODE are a plausible solution to the problems identified.
4. Highlight and put in context the legal, economic, business model, technical and social challenges related to the project.
5. Investigate the domains/use cases where DECODE tools could bring real benefits to citizens, users and businesses and the key policy questions presented by each.

It is intended for a wide public audience of primarily non-technical readers, in addition to EU policy makers, cities and local governments, businesses, citizens, and innovators, entrepreneurs and developers within the open software and civic hacking movement.

Glossary

Personal data

The EU General Data Protection Regulation (GDPR) defines personal data as *"any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person."*

Internet

The internet is the network infrastructure which facilitates the exchange of digital information across the world. This includes a range of basic protocols, which are the rules that define how computers on the internet talk to each other.

World Wide Web

The World Wide Web is the main information sharing system that is built on top of the internet. It is made up of documents (text, videos, audio) which can be stored in physical servers and presented or shared in websites, all of which make up the majority of online services we use today.

APIs

Application Programme Interfaces are code which facilitate communication between different software. A company might create an API to define a (limited) set of rules which allow other developers to tap into its software's specific functions or data.

Internet of Things (IoT)

IoT refers to the growing network of physical devices that are connected to the internet. From 'smart' home devices to sensors in city environments, IoT devices are expected to rapidly increase the amount of information which is collected, stored and processed about our surrounding environments.

Anonymisation

The UK's Information Commissioner's Office define anonymisation as *"the process of turning data into a form which does not identify individuals and where identification is not likely to take place"*.

Pseudonymisation

Pseudonymisation is a procedure whereby any identifying characteristics of data are replaced with a value which prevents the individual from being directly identified. It provides weaker protection than anonymisation - it may still be possible to identify the individual by analysing other related data.

Attribute-Based Credentials

Attribute-Based Credentials offer a technology which would allow someone to reveal the minimal amount of information about themselves necessary for an online transaction to take place. The classic example is of someone wishing to purchase alcohol. All they need to prove is that they are 'over 18'; their specific data or birth, or any other information on their driving license or passport, need not be divulged. Similar techniques in Attribute-Based Encryption may allow a data provider to encrypt a secret message or piece of data, and then allow others to open it according to the access policy which the owner describes (e.g. 'only people who are over 18 may read this').

Distributed Ledger Technology (DLT)

DLTs enable a diverse set of untrusted actors to agree on a single record of events. One of its most influential implementations has been in the digital currency Bitcoin, a peer-to-peer method of exchanging digital money that removes reliance on a trusted intermediary like a payment processor or a bank. Bitcoin is an attempt to achieve a decentralised network, a system of exchange with no single locus of authority. This means that all responsibility over the ownership of Bitcoin is left to individual users.

Commons

The commons is a general term for shared resources in which each stakeholder has an equal interest.[1] Commons can be environmental (woodlands, rivers), cultural (literature, music) or digital (free and open source software, Wikipedia). In a commons the governance of the resource is carried out by a community that defines the rules of access and use.

Executive summary

This report is about the need to give people more control over their personal data. It is part of DECODE (DEecentralised Citizen Owned Data Ecosystems), a major EU project which is developing practical tools to give people control over how their data is used, and the ability to share it on their terms. The report identifies the problems current uses of personal data cause for people's privacy, the economy, and for society. It explores the vision of DECODE and how it would address these problems, before discussing the challenges which the project will have to overcome in order to be successful.

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. While some of that spirit remains, the reality today is very different.[2] We live in a world where Google now has over 90 per cent market share in internet searches. Facebook, having just hit two billion users, has a penetration of about 89 per cent of internet users.[3] For most people, these platforms provide a gateway to the web. But their dominance is growing far greater than their equivalents in previous industrial eras. So how did we get here?

The increase in global connectivity and the 'datafication' of everyday life means that people produce rich trails of information all the time - from location, to shopping and browsing habits, to 'likes' on social media. This information has become the central driving force for value creation on the web, as users have been content to sign away use of their data to advertisers, in return for a range of 'free' services.

For many people, this is simply part of the deal. It makes sense for companies to know more about their users in order to provide a more efficient and more personalised user-experience. But the side-effect has been a rapid consolidation of power among internet firms, largely because those that were able to use data to create the largest user-base in turn benefited from positively reinforcing network effects.

This matters because some of society's most important infrastructures - those of the internet economy - are now unaccountable. Their business model relies on asking users to trust platforms absolutely, yet companies benefit from opacity and lack of transparency about where they make their profits from data.

The use of personal data as a commodity has amplified in scale and complexity, leaving regulators struggling to catch up. People have surrendered their personal data and have limited control over how this is used. This has led to strong market concentration in the digital economy, with a handful of digital platforms being able to gather, aggregate and analyse large amount of data.

Five hundred million adblocker downloads is a symptom of a market which isn't working well for people. User experience of being on the web is increasingly one of being tracked, targeted, and hacked, and this is reflected in the numerous user surveys about feelings of disempowerment about how people's personal information is managed online.[4]

The untapped potential of personal data

We now talk about 'clouds' instead of mainframes, but the nature of data management hasn't changed that much since the 1960s.[5] Data is stored and maintained in vast warehouses. This means that data is more vulnerable, since large amounts of sensitive information in one place makes it a honeypot for hackers. It also reflects the desire among companies to hoard as much of the commercial value of data as possible.

As a result much of the potential value of the internet is not being realised. There is little incentive for companies to explore the untapped social value available in the data which they collect if it does not help them generate additional revenues. At the same time, the producers of personal data are unable to do this, disenfranchised through a lack of control of their say in how value could be created.

The argument of this report is that we urgently need to reconsider the way personal data is used in the digital economy. We need to flip the current model on its head, giving people back control of their personal data, being respectful of our data protection and fundamental rights framework. We argue that this will pave the way for more democratic alternatives to how value is generated from data while opening up new use-cases that are valuable to government, society and individuals themselves.

From siloed access to democratic control

For decades the open data movement has shown how information can provide massive economic and social benefits from the free sharing, use and re-use of organisational and government data. In a similar vein, many promising uses for personal data require the data to be blended, shared and compared as part of larger, population sized datasets to offer individual-level predictions, recommendations or insights.

The problem - and this is where we depart from the vision of open data - is that the risks of sharing and blending personal or sensitive data are far higher. These include risks of the data falling into wrong or incompetent hands, risks of revealing more about us than we're comfortable with, or even risks of data being used against us.

Part of the difficulty for making personal data open is the lack of legal, technical or economic norms that would allow people to control and share data on their own terms. If this were possible, then people might be able to share their data for the public good, or publish it as anonymised open data under specific conditions, or for specific use-cases. Currently there are no mainstream alternatives that provide users with this option.

Companies are waking up to some of the value of the personal data they collect, and have engaged in so-called data philanthropy, sharing useful information with governments or the third sector on an ad hoc basis. Facebook, for instance, has recently announced efforts to share users' anonymised location data to improve a local response to natural disasters. But more sustainable solutions are needed that prioritise accountability and user control from the start.

There need to be new models of governance that move beyond simply trusting major internet providers to share our data for public good when and how they see fit. What's more, there should be transparent models for managing the flows of data in a way that enables people who have contributed data to maintain democratic control, and where necessary, to decide collectively how that data is used or what third parties have access to it.

This report provides an alternative vision for greater ownership, democratic control and transparency over how personal data is used to generate social and economic value. Achieving this vision will require the uptake of new technologies for decentralisation, collaborative governance and alternative business models, building on use-cases which treat data as a common good. The DECODE project will pilot a number of digital platforms that promote this vision over the coming two years.

There are trends which suggest that this is a moment of opportunity for these types of radical alternatives. The introduction in May 2018 of tighter regulations in the form of the General Data Protection Regulation (GDPR) - the EU's ambitious new data protection regulation - together with rising public concern, and a wave of new technologies are paving the way to a future in which people have control of their data. This will help give them greater privacy, a more efficient and fairer market for innovation, and will enable society to benefit from the rich insights latent within data, considering data as a common good and protecting people's fundamental rights.

DECODE is part of this movement, and this report aims to summarise it, its aims, its vision and its challenges.

What is DECODE?

DECODE is about giving people ownership of their personal data so they can secure their privacy and reclaim their digital sovereignty.[6] It will create new technologies which put them in control of how their data is used so they can decide who has access, and for what purposes. In doing so, DECODE will create a new digital economy ecosystem, enabling in particular the rise of more localised, democratic models for pooling and sharing data. These new technologies will be piloted in Amsterdam and Barcelona. A key principle of this will be the pursuit of social value over purely economic return. It will also enable governments to be more responsive to citizen needs.

DECODE is a response to the problems created by people being disempowered of control over their own personal data which leaves them:

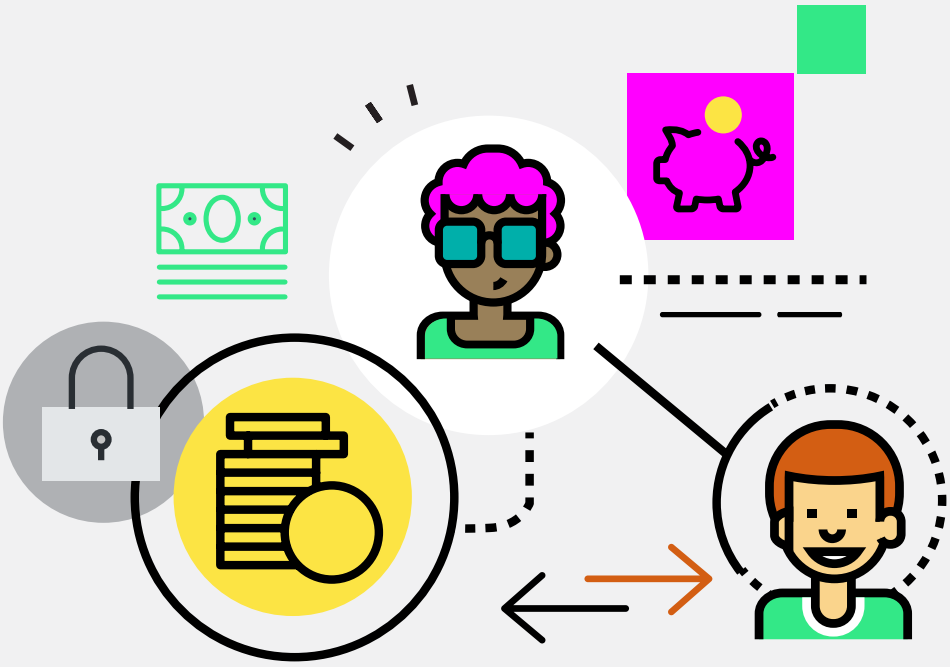
- Susceptible to exclusionary and discriminatory practices by digital platform operators
- Having their privacy and security eroded
- With limited voice or power in the digital economy as the benefits or value of digital economy are not evenly shared.

DECODE will create a number of practical tools. These will include personal data storage options, the ability to specify in granular detail how data is shared and with whom, privacy-preserving tools which enable authenticated digital participation without revealing private information, and a mechanism for data to be blended so that it can be used by governments, researchers, innovators and not-for-profit organisations.

There are a range of possibilities for how the technology created by DECODE could be used.

These break down broadly into three different categories

- **Giving people privacy-preserving mechanisms for interacting with online services**
  This could include enabling people to participate in digital democracy without having to reveal their identity but still being validated and accounted for.
- **Supporting the development of platform co-operativism - models of economic exchange which have social and ethical objectives**
  Platform co-operatives offer a feasible model to encourage the sharing of data, embedding co-ownership, transparency and democratic participation over how data is managed and used. Models could include the sharing of accommodation, transport, neighbourhood items like tools, or the exchange of labour.
- **Enabling data to be used for social good through the creation of a family of data commons**
  This could include a combination of personal data, city open data, and private data, as in the example of Smart Citizen Lab, which brings together citizen sensing data on air quality, noise pollution, water quality, or the creation of local co-operatives such as a neighbourhood energy scheme.



While elements of these possible use-cases already exist, DECODE's point of difference will be in enabling them in a privacy-preserving way which first and foremost puts people in control of how their data is used.

Achieving this vision will present a number of challenges, ranging from technical through to legal and ethical dilemmas. Resolving these challenges represents some of the most important work the consortium delivering DECODE will undertake during the project. Some of these challenges include:

- Creating a framework in which people want to share their data in a controlled way for the common good.
- Finding technological solutions that enable the enforcement of rules for data sharing, preventing the misuse of data.
- Testing whether there are viable alternative revenue generation models in an internet economy which finances itself predominantly through monetising personal data.
- Finding the right way to govern decentralised digital platforms so that contributors have a genuine stake and say in how they are run.

To respond to these and other challenges, the project has adopted a methodology which focuses on co-creation - with citizens, with technology experts and open internet advocates, and with our key stakeholders. The aim of this is to ensure that what is created responds to genuine problems and will have wide take-up. It will use a range of methods - research, citizen engagement, and exploring a range of new technologies - to reconcile some of the core challenges in the project. A commitment to an agile methodology and learning from testing will help the project iterate and improve throughout.

While this report has been written for a general audience, DECODE has a number of specific audiences for whom there are some key messages:

- **European Commission and national policymakers**: DECODE will be one of the first experiments in personal data control backed by major city governments. The learning from this process will in all likelihood involve lessons for national and supra-national policymakers about legislative and policy support that is required to enable this agenda. Nesta's next report, due in Summer 2018, will focus in particular on this issue.
- **Cities and local government**: the tools created by DECODE will be open source and free to use. In order for their use to scale, and for the highest impact to be achieved, the support and engagement of cities and local governments will be required. As the project progresses, we will be investigating how cities and local governments can best support DECODE initiatives, and how they can in turn can get the most value from them. Future reports will provide practical advice about how to adopt DECODE technologies across cities.
- **Citizens**: there is a chance to be involved in this work. Citizens of Barcelona and Amsterdam will have many opportunities to be involved, and we hope to reach other European cities through events and scaling activity over time. The DECODE website (decodeproject.eu) will have the latest information on how to get involved.
- **Innovators and entrepreneurs, civic hackers and data scientists**: there will be a range of events such as hackathons, challenges and summer schools designed to engage potential developer communities in the opportunities presented by DECODE. For the full social value of data to be unlocked, engaging with these communities will be essential.
- **Business**: DECODE will create an infrastructure which businesses can build on. DECODE's aggregated data could help lower some barriers to entry, and insights latent within the data could spur new products and services.

This an important time to be considering alternative approaches to the way we and our personal data interact with the internet. As technologies advance, and the amount of data we create expands, the ability to take a different path will reduce. DECODE aims to create a set of tools, legal and governance frameworks, which enable people to choose this different path. In doing so, we hope it will lead to a fairer kind of internet which can play a bigger role in solving the key social challenges of our time, while empowering citizens and protecting their digital sovereignty.

Introduction

Data is now the lifeblood of the internet economy. Without it, websites would stop functioning, services would go offline, tech giants would fold and the impact would ripple out through the economy as a whole.

Finding a suitable physical analogue for data is difficult. Data has been variously likened to oil, infrastructure, an asset class, a currency, property, digital labour, and even the toxic material asbestos. In reality, it shares common features with all of these, yet none describes its characteristics and trends perfectly.

One trend is of particular importance to this report. As the internet has grown, the data which enables the internet economy to function has departed from the control of the people who generate it. Large internet companies routinely collect personal data and a new breed of companies, known as data brokers, have led to a boom in the personal data market. They harvest, connect, analyse, exploit and sell this data, with the people this data relates to largely unaware of the process. The creators of this value - the internet users the world over - are left disempowered about how it is used.

The uses of personal data on the internet have developed and expanded faster than regulation can cope with. As is argued in *Platform Privacy: The Missing Piece of Data Protection Legislation*: *"across the world, privacy laws which govern the operational modus for companies providing services to consumers, may have been devised during a time when the Internet was predominantly used in research and academia."*[7]

In future, we may look back at the early decades of the internet and its use of personal data as akin to the Wild West - a loosely regulated space in which fortunes could be amassed through cavalier attitudes towards data ownership, ethics, and privacy. Regulators everywhere are playing catch up with the rapid growth of the internet economy, explosion of personal data and advance of the technologies which can make use of it.

As regulators grapple with a market growing rapidly in complexity and concentration, consumers are left with few options. Internet services have developed as free at the point of use, with sharing of personal data the de facto payment method. Some surveys suggest that this has led to a situation in which people are largely unconcerned about what happens to their data. Other surveys find that people will report that that they are concerned about what is happening to their data,[8] but the internet's 'take it or leave' terms and conditions leave them with little choice but to use services and platforms that hoard and sell their data.

The rapid advance of the internet has to some extent obscured how problematic much of this practice is. Technology's increasing sophistication means data is used and monetised in ways that regulators struggle to understand, let alone control. The potential uses - and abuses - of data will increase alongside advances in technologies such as the Internet of Things, machine learning and artificial intelligence. Without fully understanding these risks we could be undermining the foundations of the internet economy.

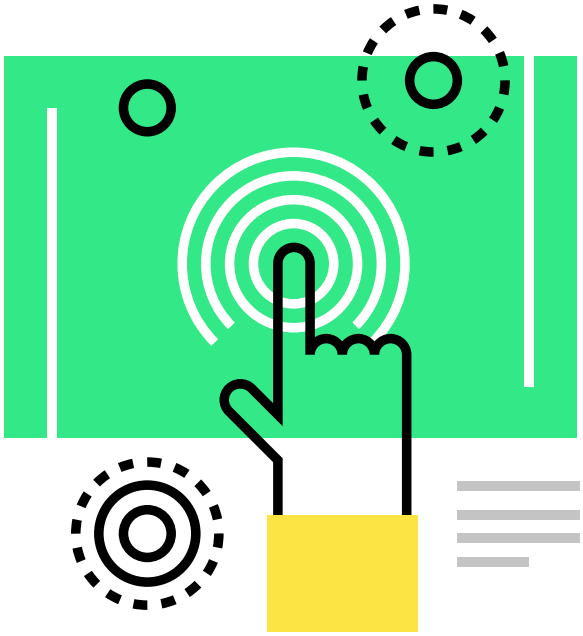
This is not to argue that everything is negative. The internet and web have given enormous value to its billions of users and accelerated technological innovation. Access to information, products and services has never been more free and more open. Our concern is that the centrality of personal data to the internet creates negative effects as well.

Centralisation and monopolisation of personal data enables the internet economy to function, but it does so in a way that produces inefficiencies and inequities for individuals, society and businesses, not to mention ethical concerns. In a world of ubiquitous connected devices, wearables and lives lived online, our data can paint incredibly accurate pictures of our identities. In today's personal data economy, these identities have limited privacy and autonomy.

The need for data to generate revenue and provide feedback creates barriers to new firms. Network effects - whereby a network such as the telephone becomes more valuable as the number of users increase - means that companies can exploit the role of data to become excessively dominant. Despite the infinite replicability of data, governments, researchers, not-for-profits and innovators are not able to freely access anonymised personal data because the people who generate it lack the means to share it on their own.

On the horizon, however, there are signs the tide could be turning. The arrival of the EU's General Data Protection Regulation (GDPR) in 2018, which imposes much stricter rules about the use of personal data, makes this model of the internet economy less attractive to business. The cost of holding personal data will become more expensive as regulations become more onerous. And the price of getting it wrong could disincentivise companies from holding personal data. The new GDPR brings a potential fine of up to €20 million or 4 per cent of global turnover (whichever is higher) for companies which misuse personal data. In such an environment, companies might start to feel that holding personal data is not worth their while, and explore alternative ways to use people's personal data which don't involve storing it.

While regulators mainly focus on creating new rules to protect people against the worst practices of the current dominant internet businesses, there is an emerging movement which believes that an entirely new approach, backed up by new technologies, may be required. DECODE is part of this, putting people in control of their personal data and changing their relationship with the internet economy.



Report structure and research methods

The aim of this report is to set out the rationale behind DECODE, a vision for the project and how it could provide an alternative to the current digital economy. Section 1 explores in detail the reasons that DECODE is needed. It is based on a literature review of the current personal data economy, and describes some of the consequences stemming from opaque secondary data markets and from the rapid consolidation of power among internet firms that has occurred in recent years.

Section 2 then explores the vision of DECODE by describing a world in 2035 in which people have greater control of their data. The stories and scenarios in this section section are based on a workshop which was held in May 2017 on the future of the personal data economy. Attendees included a range of activists, researchers, and policymakers, who were asked to to construct broad scenarios and individual personas about what the future might look like, based on a number of existing trends.

Section 3 presents a review of the policy areas and practical work which the project sits alongside, and the debates, dilemmas and challenges which DECODE touches on. It goes into further detail about how DECODE will respond to these challenges to deliver a novel alternative, including details about the pilots which will be carried out in collaboration with Barcelona and Amsterdam city councils over the coming months. The section is based on desk research and a series of research interviews conducted with members of the DECODE consortium, and a number of further interviews with other pioneering projects across the world who are providing solutions to enable greater control over personal data. For a full review of the case studies we gathered from this research, see the acknowledgements.

Section 1

What is DECODE and why do we need it?

DECODE (DEcentralised Citizen-owned Data Ecosystems) is an experimental project to enable practical alternatives to how we manage our personal data and interact on the internet. DECODE will develop technology that puts people in control of their personal data, giving them the ability to decide how it is shared. The technology will include an architecture for controlled and, if desired, anonymised data sharing, paving the way for the creation of a 'data commons'.[9]

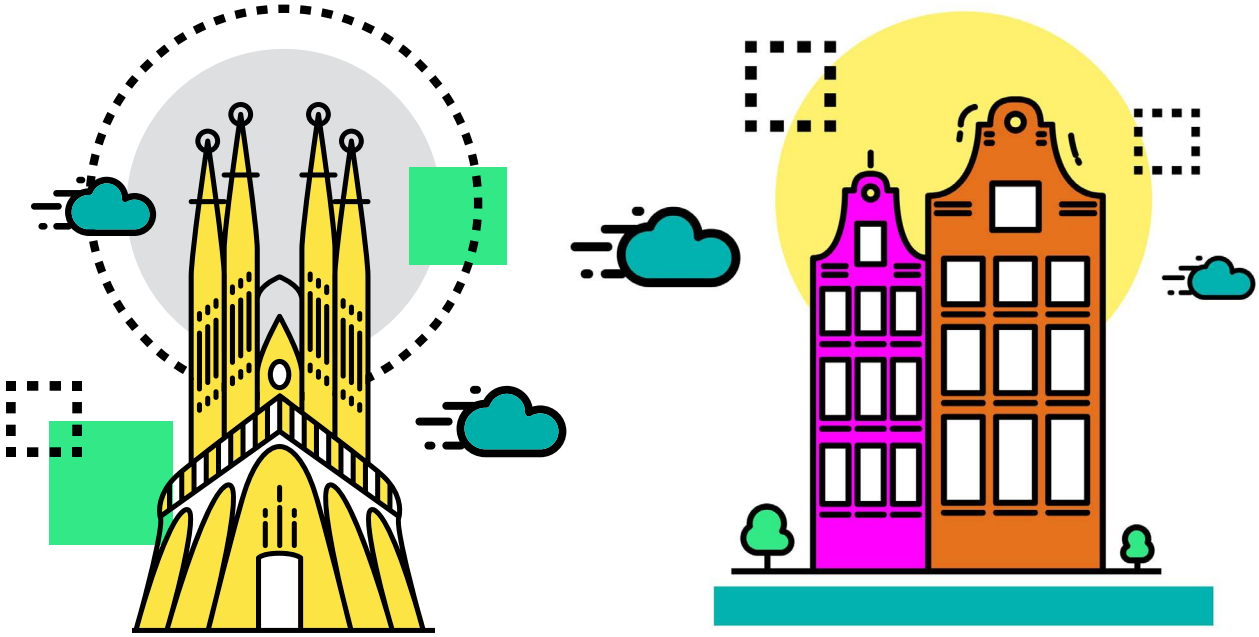
DECODE will test this technology in four pilots, to be held in Amsterdam and Barcelona, between 2017 and 2019. The pilots will trial the technology and demonstrate the wider social value that comes with individuals being given control of their personal data the means to share it differently. The four DECODE pilots have been chosen as follows:

1. **iDigital / BCNow Platform pilot**. This pilot partner with Barcelona City Council and the city's digital democracy platform Decidim.Barcelona, to allow citizen-generated data to be aggregated and blended from a range of different sources, including noise levels from individual sensors, healthcare data, and administrative open data. This will be displayed in a BCNow dashboard, and will give citizens the option to control the use of that information for specific purposes, including to inform policy proposals. It will also provide anonymous verification capabilities to minimise the sharing of sensitive of personally identifiable data with the city council.
2. **#CitizenSense Internet of Things pilot**. Residents will be given noise sensors that are placed in the neighbourhood. DECODE will provide sessions to train and support participants to help them setup and use the sensors to gather and analyse data to influence city-level decisions. The pilot tackles the technical challenges of collating and storing a stream of citizen-sensed data, while also enabling those citizens to control what information is shared. It builds on a local project, Making Sense, which was established in 2016 and has been co-funded by the European Commision.
3. **FairBnb pilot**. This responds to sharing economy platforms such as Airbnb which have caused disruption in cities such as Amsterdam by pushing up the the price of rents, while restricting access to occupancy data about hosts who break local legislation. The pilot involves collaboration with the Amsterdam Municipality and the FairBnB platform — which was started to provide a more sustainable solution for short-let accommodation, and reinvests profits back into local initiatives. DECODE will provide statistics and regulatory information to enable the community to govern the platform without compromising participants' privacy.
4. **Gebiedonline (Neighbourhood Online) pilot**. Gebiedonline is a pre-existing co-operative digital platform that enables local people, groups and organisations to view events taking place in their neighbourhood, share news, exchange and borrow products and services, and to meet people. Amsterdam City Council is keen to spread this to other locations across the city and leverage the platform to increase involvement with policy and decision-making. It provides an opportunity for DECODE to test a more privacy-preserving local social network, with granular controls so that residents can decide what information they share.

Through these pilots DECODE will explore how to build a data-centric digital economy where data that is generated and gathered by citizens, the Internet of Things (IoT), and sensor networks is available for broader communal use, with appropriate privacy protections. As a result, innovators, startups, NGOs, SMEs, co-operatives, and local communities can take advantage of that data to build apps and services that respond to their needs and those of the wider community. In this vision, cities can have a strong role, as custodians of these new digital rights of citizens.[10]

DECODE has five key objectives:

1. Providing alternatives to using dominant internet platforms, which offer a more democratic approach to creating and sharing economic resources.
2. Effectively using an extended range of data coming from people, sensors, devices and the city to enable collective, bottom up decision-making.
3. Ensuring that people are in full control of their data and identity, while maintaining privacy and trust in the systems they use.
4. Creating space for third parties to implement relevant innovative approaches and applications.
5. Preserving the digital sovereignty of citizens and preventing unauthorised use of their personal data on clouds, social networks and the Internet of Things.



Why DECODE?

The people creating the internet's most valuable resource have no say in how it's used. This has led to a profound disenfranchisement in the digital economy which creates problems for privacy, security, and economic and social fairness. DECODE is a response to this lack of control over personal data and the objectives listed above are manifestations of this disempowerment. But to fully understand this problem and its origins, it is useful to consider the context of the wider personal data economy.

Defining personal data

The EU's General Data Protection Regulation defines personal data *"as any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person."*

Such a definition broadens the scope of personal data to include any data created by an individual, even if it does not relate to that individual's identity. For instance, this definition of personal data includes data created by personal Internet of Things (IoT) devices, such as an energy meter or smart fridge. This is the definition we use for this report.

How people lost control of their data - a brief overview of the personal data economy

Across much of the web, services and websites are accessed free of charge without limitations. The original vision of the internet was the ultimate open network. Free and open for all to access, truly decentralised without barriers and without limits. It is perhaps because of this original vision that we have the internet economy that we do today. In the early days of the web, users became accustomed to accessing information, journalism, music, video and later social media for free. As these services grew, so too did the need to find a means of funding them which did not require the user to pay upfront.

In response, the internet economy has a set of business models which offer free to access services but generates revenue using the data of the people who use the service. This is primarily through the data of website users - their location, age, socio-demographic information, past browsing history, preferences and a range of other factors - to find monetisation opportunities.

However, to say that popular internet services are free to access is not strictly true. While no money changes hands when someone signs up for Facebook, they are in effect paying with a different currency - their personal data. As the means by which the wheels of the internet economy turn, personal data has significant value. For instance, Facebook's average revenue per user in the US and Canada was $62 in 2016.[11] This is not to say that each Facebook user could use their own data to make $62 a year - some of the value from data only comes when it is at huge scale, and Facebook has to spend money to create the infrastructure to gather data and show adverts - but it does illustrate how valuable this commodity is within the internet economy. The European Commission (2016)[12] estimates that by 2020 the value of European citizens' personal data is expected to reach €1 trillion solely in the European market, getting to almost 8 per cent of the total union GDP.

What is happening with our data?

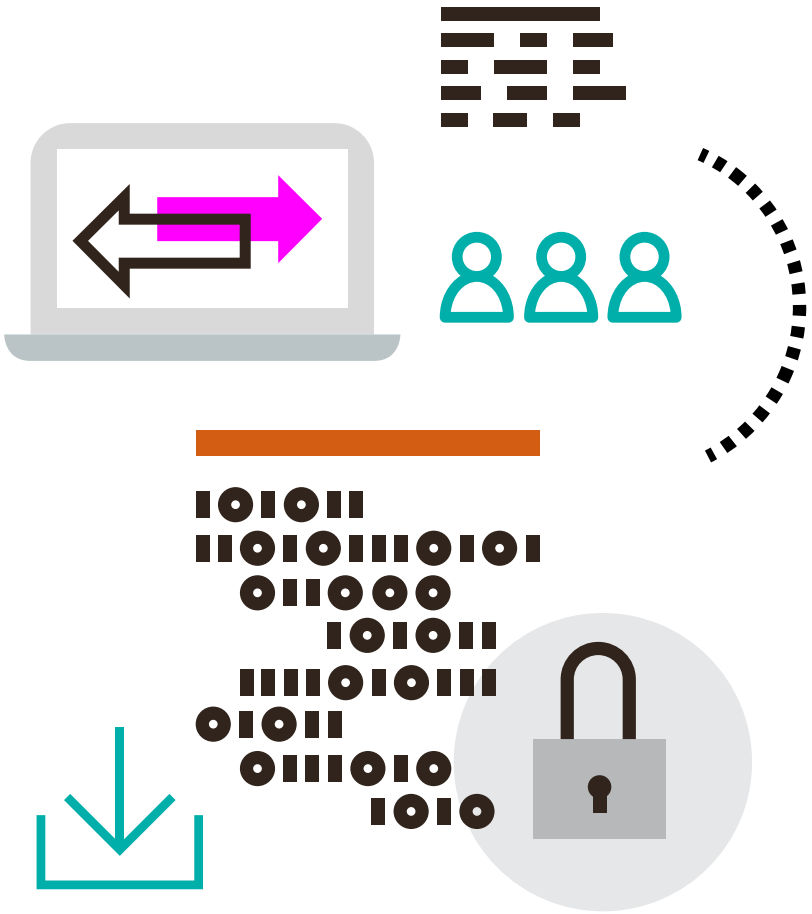
Users of some parts of the internet now find their every mouse movement and click electronically tracked, their data mined, aggregated, analysed and sold. Initially, internet companies found their access to personal data useful for improving services which helped to improve profits. Visitor behaviour could be used as a real-time feedback loop to experiment with new product and service innovations. This helped companies to grow more competitive and drive profit.

The potential for this data to better target advertising then became apparent. It is this which turns the side margins of websites and gaps in between content into valuable advertising real estate. Where a billboard can only carry one message at a time, advertising space on websites can show different products to different users; and even to the same user at different times of day and at different locations. Those differences are determined by the data of the people visiting the website.

In 2016, the revenue produced by online advertising was $72.5 billion in the US alone,[13] showing rapid growth from the $1.9 billion it generated almost 20 years ago.[14] In particular, social media and news websites that do not charge users money will collect personal data - usually via third parties - in order that they can charge more for targeted advertising on their sites. Some large internet platforms and advertising companies aim to collect as many pieces of information about people as they can with the goal of achieving the advertiser's holy grail: the ability to successfully predict buying intention.

The sophisticated insights that can be drawn from data have opened up another avenue of use for companies. Data technologies allow for more understanding and control of individual behaviour. This means that data can be used to inform attempts to 'nudge' people to behave in a certain way, such as to spend more money or to click on more links (which in turn generates greater advertising revenues). Mass data collection enables greater surveillance of activity and the ability to predict behaviour which stems from this. The virtual nature of the internet means this data informs live reshaping of environments and products around what is most profitable for companies.

This is in some ways an expansion of how companies have always collected and used personal data. Supermarket loyalty cards have been using people's shopping data to analyse, segment, predict and target promotions for decades. The internet has simply opened up the potential for magnitudes more data to be collected, and as a result, much more intensive analysis. But the scale at which this is happening, and potential for misuse, dwarfs any analogue or early electronic efforts.



The growing complexity of the personal data economy

Behind each personalised advert on the web there is a huge and complex personal data market. In some cases, such as Facebook, personal data is not shared with third parties but is used to make advertising space more valuable through profiling and segmentation techniques. In others, personal data can be sold on to a secondary market provided it is anonymised.

At the heart of this market are data brokers. These are firms which splice and dice personal data from a vast array of different sources to create individual profiles of internet users. These profiles are sometimes in the form of consumer segments, for instance 'new mothers living with an in-work husband', or 'affluent young professional without a mortgage'. These profiles are then sold onto other companies who use it for targeted marketing and advertising.

One such data broker, Acxiom, can combine data from multiple sources to build a personal profile of more than 1,000 pieces of information.[15] From this they can derive an even larger number of behavioural insights, primarily about an individual's buying intention across a range of different products. Just one US-based company, ID Analytics, holds information on more than 1.4 billion consumer transactions and 700 billion data elements. This enables them to build up detailed pictures of consumer browsing and purchasing habits, their interests, habits, hobbies, communities and opinions.[16]

As a result of this, the practice of mass personalisation has become common: websites and other platforms are shown differently to different people, depending on their internet profile and browsing history, to optimise user experience to different classes of people e.g. first time viewers or a particular type of shopper. This personalisation is very useful and convenient for the user, and people have become used and perhaps even reliant on it.

The Privacy Paradox - do we really have a choice about sharing our data?

This explosion of personal data has led to what is termed the 'privacy paradox'. This refers to the fact that people appear to be concerned that their personal data is not secure and that this erodes their privacy. But at the same time, people display an incredible willingness to surrender their personal data in exchange for free services and discounts, some of which may be relatively trivial.

For instance, the Attitudes to Data Protection, Eurobarometer Survey (2015) found that 67 per cent are concerned about not having complete control over the information they provide online or it being used for a different purpose from the one it was collected for. Yet, 71 per cent of the respondents said that providing personal information is an unavoidable part of modern life, necessary to obtain products of services.[17]

Only 20 per cent of respondents say they fully read privacy statements. This could be argued to be optimistic - in 2008 it was estimated by Stanford University that for the average person to fully read all the privacy statements they encountered it would take 7.6 days a year.[18] In practice, the 'take it or leave it' Terms and Conditions of most websites or online services mean that people have no choice but to grant access to their data to a large number of companies in any event. And once data has been shared, people have very little control over what happens to it, or even to know what has happened to it.

Disempowerment in the digital economy

Every person who uses the internet is creating a valuable economic and social resource in the form of personal data. These same people do not have control of how it used, even if their data is being used in ways which harm them. To compound this, there is a massive lack of transparency over how personal data is being used. This is a form of disempowerment which manifests itself in a number of ways. People are:

- Susceptible to exclusionary and discriminatory practices by digital platform operators.
- Having their privacy and security eroded.
- Left with limited voice or power in the digital economy as the benefits or value of digital economy are not evenly shared.

Susceptibility to exclusionary and discriminatory practices

The lack of control we have over our data comes with risks to the individual because of the increased use of algorithmic decision-making by companies and governments. This refer to the use of historic data points to make predictions about future behaviours. These predictions are then used to make decisions about how an organisation deals with an individual, and can have discriminatory and exclusionary impacts.

This is not to argue that algorithms are bad per se. There is huge potential for algorithms to support better decision-making in a wide variety of areas, not least in government. However, there are documented problems with algorithmic decision-making, evident across many sectors, from healthcare and insurance through to policing and criminal justice, which need to be addressed. In many cases, these problems are compounded by the lack of transparency over both how the algorithms work and whether or how people's data will be used by them.

The tyranny of the algorithm

Exclusionary Policies. The use of big data and algorithms to determine access - or the price paid for access - to a range of services, such as insurance, credit, education and housing, is growing. To some extent, this is just a more sophisticated version of what many companies have been doing for decades. Credit and loans decisions have always taken into account an individual's circumstances and history for instance.

The use of big data, however, opens up a far wider set of information to be used in decisions which could be considered much less benign. Big data enables us to find patterns and connections across a wide range of data, so that seemingly trivial information might be used in critical decisions. For instance a paper matching thousands of 58,000 volunteer's Facebook 'likes' to psychometric tests and demographic profiles tests found that liking curly fries was a strong indicator of a high IQ, and that religion, gender, sexuality, ethnicity and other characteristics could be strongly predicted from similarly unlikely clues.[19]

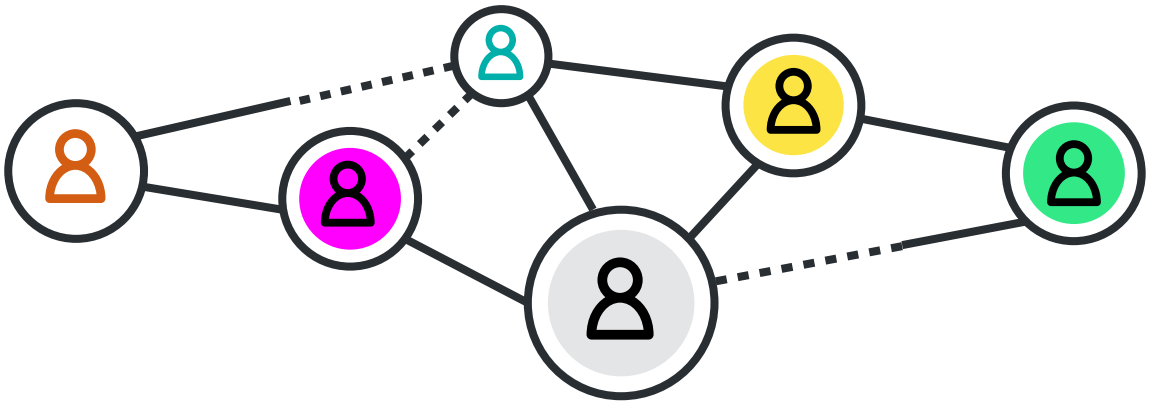
Other worrying uses include a credit card provider in the US that employed a behavioural scoring model which used information about expenditure on things such as marriage counselling, psychotherapy and billiards to determine how much credit to offer.[20] These examples suggest that data is being used beyond what could be fairly assumed as the scope of its original collection, to inform value-laden decisions which could exclude particular groups from vital products and services.

Social Discrimination. Some of the many organisations to seize on the predictive potential of big data and algorithms are police departments, especially in the US. 'Predictive policing' uses data collected by police departments, and in some cases integrates this with data scraped from social media sites, to assess which people are likely to commit offences, which areas are more likely to be affected by crime, and where to most effectively allocate resources. In theory, this can enable preventative interventions to be implemented which avert crimes. However, in practice, predictive policing algorithms have been found to disproportionately target people from ethnic minority groups, thereby perpetuating and strengthening the bias. The same patterns are present in algorithmic decision-making trials in courts and the justice system.[21]

There are a range of reasons why this can occur. Machine learning expert Moritz Hardt from the University of Berkeley argues that *"a learning algorithm is designed to pick up statistical patterns in training data. If the training data reflect existing social biases against a minority, the algorithm is likely to incorporate these biases."*[22]

There are additional reasons an algorithm could encode biases, such as being 'trained' on previous biased decisions, or that data is more available about some groups than others. Early adopters of technology or the technology developers - often from privileged groups - may generate more data and so therefore may be more present in the data sets. For instance, some facial recognition software fails to detect people with darker skin, because the people who coded it hadn't taught it to recognise a broad range of facial features and skin tones.[23]

Price Discrimination. Price discrimination refers to charging consumers different amounts for the same product or service, depending on factors such as time or the characteristics of the consumer. Algorithms and access to personal data enable internet businesses to exploit price sensitivity amongst their customer base to charge different amounts to different people. Online shops can, for instance, recognise a consumer based on cookies, IP addresses, past browsing history or a combination of these. In 2012, the stationery retailer Staples were found to be charging customers who lived further from a physical store more for products. This reflected the convenience of picking up the same item from a store, and therefore the decreased price sensitivity of those who lived further away.[24]



Uber has been experimenting with price discrimination, using its huge stores of personal data to make judgements about whether a passenger would be willing to pay more for a journey. Factors like whether they frequently travel from affluent area to affluent area, are frequently being picked up outside business addresses (indicating they may be charging the ride to an expense account), or how much battery life their phone has, have been used to influence the price Uber charges its customers.[25, 26]

It should be noted that price discrimination happens in the real-world economy too. Supermarkets charge different amounts for the same product, depending on the location of the store. Flight or train prices fluctuate significantly depending on a range of factors. It is also not clear that price discrimination is widespread on web[27] and data protection legislation does currently require firms to disclose in their terms and conditions or cookie policies that they may use this information as part of price discrimination.[28] However, the complexity and opacity of the personal data economy makes this kind of discrimination a continuing risk.

To compound these issues, there is a lack of transparency over how the algorithms in operation across the web are being used, and their original source code. This prevents scrutiny of algorithmic decision-making to see if coding has been done correctly and fairly. As a result there is an accountability gap between companies using algorithms and the decisions those algorithms are making. And, since the influence of algorithms is growing by the day, this accountability deficit creates more and more problems.

The risk become more severe the more technology improves

The value of data is increasing. Facebook and Google initially monetised the data they collected from users by using it to inform improvements to the service, and followed by creating profiles that it was hoped would increase advertising conversion rates. But in recent years they have discovered that data can be turned into any number of artificial-intelligence (AI) or 'cognitive' services, some of which will generate new sources of revenue.

These services include translation, visual recognition and assessing someone's mental health by analysing their use of social media — all of which can provide potential benefits for individuals and the community can also be sold to other firms to use in their own products. These new technologies, with the associated risks of powerful AI, will make us increasingly vulnerable if such technologies have extensive access to our personal data. With the emerging ubiquity of IoT, it will also be difficult for people to avoid being affected by these practices.

The erosion of privacy and security

Being able to maintain control over personal information is crucial. It allows us to create the necessary context for relationships of respect, friendship and trust. The abundance of our data and lack of control over it risks the erosion of our privacy due to the intimate profiles that can be created from our digital footprints. If left unchecked this could have profound implications for our ability to form relationships and participate economically and within society.

This threat to our privacy is happening despite national and supra-national level regulations created to protect our personal data from abuse by governments or companies. There are a number of specific reasons why there are fears over levels of privacy and security in spite of data protection regulations:

Anonymisation or pseudonymisation to data is inadequate. The EU Data Protection Directive (Directive 95/46/EC) on personal data enables companies that collect it to sell it on to third parties, provided it is stripped of the information which would enable someone to be personally identified. However, the data business has outstripped the EU's Directive. The large quantities of data we produce mean that pseudonymous data (datasets in which identifiable personal and sensitive data has been replaced with artificial values) and even anonymised data can sometimes be reverse engineered to re-identify people by commercial organisations using other data sets. This occurred, for instance, when Netflix released anonymised customer data as a competition for developers to build a better recommendation engine. Computer scientists managed to successfully de-anonymise two people in the dataset by blending it with some other easily attainable background information, uncovering their *"apparent political preferences and other potentially sensitive information"*.[29] Some argue that the concept of anonymisation is increasingly obsolete in a world characterised by rapid growth in networks and ubiquitous information sharing.

Consent to share is rarely true consent. Consent to sharing data does not always have to be explicit consent. Data handlers are entitled to assume consent from their terms and conditions (T&Cs). These are typically long and confusing. They will often request access to, and the ability to share, personal data which is not required to run the service. T&Cs can either be accepted or declined, but to decline is to be excluded from access to the service. In practice, this offers people no genuine choice about whether they grant access to their data or not.

Little transparency over how data will be used. There are well documented cases of T&Cs not adequately explaining how data will be used. For instance, the Global Privacy Enforcement Network has found serious shortfalls in the use of data by IoT providers.[30] Their study, which involved 25 data protection regulators around the world, looked at devices such as smart electricity meters, internet-connected thermostats and watches that monitor health. Their findings were that:

- Fifty-nine per cent of devices failed to adequately explain to customers how their personal information was collected, used and disclosed.
- Sixty-eight per cent failed to properly explain how information was stored.
- Seventy-two per cent failed to explain how customers could delete their information from the device.

Tracking data after it is shared is impossible. Once data has been shared with one service, individuals' rights over their own data become a lot murkier. Finnish academics Westerlund and Enkvist argue that EU legislation doesn't give the subject the right to review the data flows or the security practices of individual firms.[31] People are forced to fully trust platform providers to handle their sensitive data responsibly.

The Internet of Things creates new security and privacy risks. The rise of IoT brings new risks to personal data, offering more chances for data to be used to identify a person and other information about them, and more points of failure for security threats. In addition it is likely that IoT firms that collect this data will rely upon data processes to connect data generated from multiple devices (connected homes, cars, wearables etc.), and in doing so will be able to generate profiles of individual preferences and behaviours with a level of detail that was not possible before. That said, EU regulations do imply that data collected by IoT is personal data which should mean companies take additional measures to protect this information.

The EU's new General Data Protection Regulation is an ambitious and extensive piece of legislation designed to give additional protections to people. However, it is yet to be seen whether this will adequately address concerns that exist about the role of personal data in diminishing privacy. And even within the new regulatory framework, there is a need for high levels of trust on the part of citizens that their data is being handled legally.

Economic disempowerment

"We don't have better algorithms than anyone else; we just have more data."

Peter Norvig, Chief Scientist, Google

The internet economy is indisputably integral to the future prosperity of the European Union. Yet within this sector of the economy, the distribution of power is heavily skewed towards a small number of huge companies. This is despite the fact that one of the most valuable resources within this economy - personal data - is created by its users. Shouldn't the creators of this value have more say in how it used?

Many early users of the internet expressed their hope, and in some cases expectation, that it would develop into a radically open, utopian space in ambitious works such as the Cyborg Manifesto. Citizens and consumers would be able to find ideas, information, products and services unmediated by companies, governments or other organisations. Some elements of this original vision remain. It is still built on open technologies, and people can connect to other individuals, ideas, products and services based anywhere, in many instances for free.

But the internet is now characterised by a handful of monolithic companies which dominate different sectors. Facebook, Amazon, Alphabet (the parent company of Google), Microsoft and Apple are the five largest companies in the world.[32] Google accounts for 65 per cent of all search in the US, and 80 per cent in Europe. As of June 2017, Facebook had two billion monthly active users. Amazon now reaches almost half of all US households.[33] Facebook and Google together accounted for 99 per cent of the total year-on-year growth in advertising spend in 2016.[34] Significantly, they have achieved their dominance at least in part through the extensive capture and use of personal data. This is a long way from the original vision of the internet.

The internet economy is itself difficult to understand with conventional economic theory.

Personal data is important in this context for two reasons. Firstly, the control of data by companies rather than the individuals who create it contributes to market dynamics which mean those companies can become unassailable monoliths. Secondly, without control of their data, individuals are disempowered from having such an active say in how the digital economy functions. The result is that the economic power attached to data and value created from data is accrued by companies rather than individuals.

Personal data contributes to the excessive market dominance of some firms

Personal data as a barrier to entry

The distortionary effects that mass personal data gathering and use can have on the internet economy led the *Economist* newspaper to call this a barrier to entry for new firms and advocate changes to competition and anti-trust laws in response.[35] Access to personal data is important because it is the primary feedback loop through which companies can generate and test new innovations and service improvements. Crucially, monopolisation of data matters because the more data a firm has on performance, the more it can improve performance.

The economics community is divided on the extent of this problem. Economists Maurice Stucke and Allen Grunes argue that companies which control large amounts of data raise barriers to entry for potential rivals because they do not have the same ability to use the data to make competitive products.[36] A report commissioned by the French government points out that *"all digital economy firms use data to improve their supply, obtain productivity gains, diversify their activities or reinforce their position on the different faces of the business model"*.[37] In the UK, the Competition and Markets Authority found mixed evidence across a range of data markets, but did find that lack of access to consumer data, arising from the economies of scale and scope, disadvantaged small and potential new entrants in some markets.

In contrast, Anja Lambrecht and Catherine Tucker examined the use of data and found that access to data isn't enough to protect incumbents from superior products.[38] Legal academics Sokol and Comerford argue that internet firms can gain scale in ways which do not require access to user data, or can access data from other sources, such as data brokers.[39]

Other economists have made the argument that in the case of the sharing economy, data is an essential facility. This is an economic term referring to anything (material or immaterial) which a firm needs in order to compete on a level playing field. This is summarised by Bruno Carballa, an economist at the University of Catalunya:[40]

"The distinct features of every platform and their different performances in performing a task are the outcomes of the algorithms they are made of. These algorithms, in turn, work by processing large amounts of data and they improve due to it. Once an algorithm is correctly designed to fit its data, the more data that an algorithm can work on, the more likely it will be that it will improve over time."

This is particularly acute in the case of the sharing economy because there are few if any brokers that can sell relevant data to new market entrants. In some cases, this exchange of data is prohibited by law as platforms argue that the analysis that they perform on the data makes it their Intellectual Property, and therefore that they have no obligation to share.

Personal data as a factor enabling excessive market dominance

Personal data contributes to firms achieving the kind of network effects that reduce or prevent competition, known as network externalities by economists. This has led to some, such as Nick Srnicek, a lecturer in digital economy at King's College London, to argue that Facebook, Google and Amazon should be nationalised as network effects have enabled them to become so large that they are now unassailable.[41]

Network effects describe how platforms increase in value or attractiveness in line with increases in their user base. Services like communications, social networking, shopping, taxis services or apartment sharing rely on there being a large number of people who all use the same platform. As argued above, access to data is important for firms to grow and innovate, but if these firms become too big then competition and choice suffers.

This phenomenon is observable in today's sharing economy. When platforms reach a certain scale or critical mass, they are hard to compete with because any competitor's product would appear inferior in comparison to a larger network. People looking to rent a holiday apartment in a foreign city will typically look for the platform which has the largest number of apartments on it to maximise their chances of finding a suitable apartment. People renting their apartment to tourists will look for the platform with the highest number of potential customers to maximise their chances of a successful rent.

While a new taxi hailing app or accommodation sharing app could be created without much difficulty, it is hard to shift people onto these services once critical mass has been reached by a market incumbent. A range of factors create what is know as 'lock-in' which prevent customers leaving. Some of these affect any kind of company - moral and obligatory factors, personality factors - but others are linked to the inability to take their personal data from these systems. Someone may be deterred from switching to a different apartment sharing site for instance because their reputation score - which is an essential part of gaining future business - cannot move with them. Reputation passports are a possible solution to this, but existing sharing economy platforms have little incentive to develop or support these as there is no obvious benefit to platforms which are already dominant.

A key factor enabling these network effects to occur is whether data produced by dominant firms is available to potential competitors, or through meta-search (a tool which uses other search engines data to produce their own, usually aggregated, results). With some platforms, especially those in the sharing economy, data contributed by users of the platform is often not available to outside parties. This prevents the development of tools which aggregate data from multiple platforms which makes it easier for incumbent platforms to maintain positions of dominance.

If it was possible to have such data aggregation abilities in the sharing economy, for taxis, accommodation or gig-economy work, then the existing dominant platforms would be more vulnerable to competition. For buyers and sellers, using the biggest platform would no longer be so important as even small platforms would have high visibility. This underscores why who has control of data matters. When market incumbents have control of personal data, they can prevent competition by not making it available to the wider market.

Some economists argue that these barriers to entry, or excessive market dominance, are not necessarily a problem in competition law. These only become a problem if a company abuses that position, for instance by providing a poor quality product or service, or preventing competition. But if a dominant firm is providing a good service for free, and the product is popular and of a good standard, the fact they are dominant is less of a concern. It is also worth noting that network effects can have advantages, for instance in supporting the development of new products and services on the top of existing networks.

However, there is evidence that in the digital economy, and sharing economy in particular, market dominance and barriers to entry enabled by personal data are creating an unfair distribution of power and reward. This leaves people disempowered: they are creating a valuable resource which is intrinsically personal to them, yet they have little economic power to show for this. There are three key negative consequences:

Firms squeezing out competition and reducing choice - Firms with excessive market dominance can act to squeeze out new entrants who seek to compete with them, by reducing consumer and producer surplus, or simply by acquiring them financially. Facebook acquiring Whatsapp, Instagram and Oculus; Amazon acquiring GoodReads; and Google acquiring maps app Waze are just a few examples. This requires a financial advantage and is not limited to data nor the internet domain, but the incidence of this in the digital economy makes it highly relevant to this discussion.

In *Platform Capitalism*, Nick Srnicek argues that existing platforms are now so dominant that small competitors do not even attempt to take them on.[42] He cites the example of baby product retailer diapers.com, which found itself targeted and then purchased by Amazon as soon as it became moderately successful.[43] Facebook has developed an 'early bird' tool which warns them when startups which are potential competitors are finding success.[44] Uber has fought off competition from other e-hailing apps by reducing fares, making it unviable for other competitors to operate. This has resulted in drivers having to work more hours to make the same revenue.[45]

Data-driven platforms wield concerning levels of influence over their users - Platform users create the data that enable platforms to create value, yet they have little to no role in how that data is used. This becomes a problem where the company's profit motive outweighs the needs or interests of the people exchanging goods and services across the platform.

A large number of platforms and services can use their vast collections of data to experiment, nudge and innovate, often pushing the boundaries of what could be considered ethical research practices.[46] In January 2012, Facebook conducted a study on nearly 700,000 of its users to test the evidence of massive-scale emotional contagion, which tweaked the flow of positive and negative information arriving in people's newsfeeds to see if it affected their behaviour on the platform. Despite condemnation and concern from academics about the ethics manipulating users without their knowledge, these practices were legal within the terms of Facebook's Data Use Policy.

More recently Facebook's data has been used to build detailed psychometric profiles about their users wants, political preferences and insecurities for more intimately targeted advertising. The company offers one-to-one support to help high-paying customers to make the best use their vast database. This infamously included the Trump presidential campaign which spent millions of dollars on psychometric profiling and advertising through social media, and which was considered to have a profound influence on the final outcome of the election.[47] Cambridge Analytica, the firm that pioneered the method, justified this use of data as a new era of highly personalised advertising; others have labelled it an overly intrusive method of exploiting data to influence opinion on a massive scale.[48]

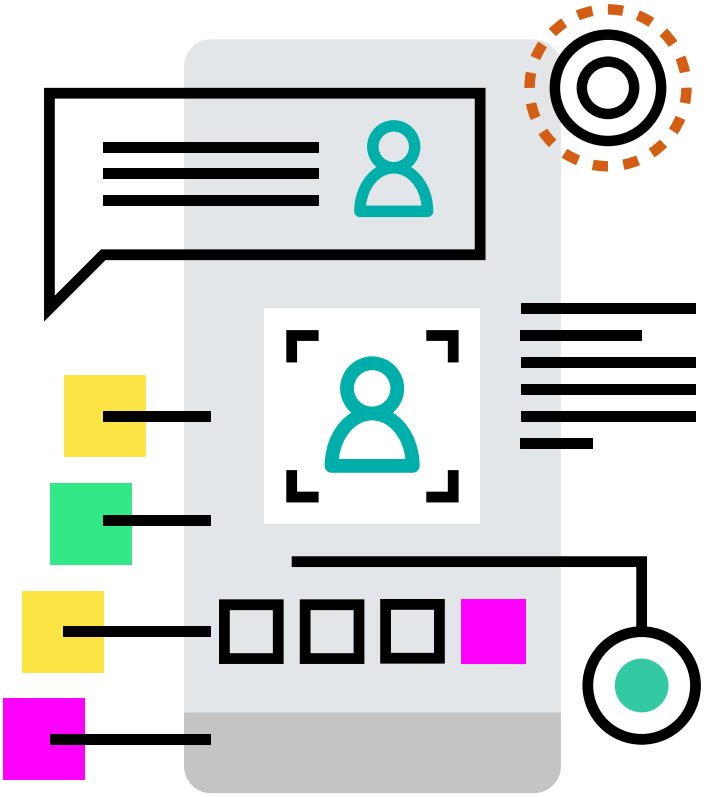
A reduced opportunity for innovation - while it can be argued that network effects are beneficial for innovation, the opportunities to innovate are dictated by existing platforms. At present, some platforms do make their data available through APIs in their websites. For instance, Facebook allow developers to build on top of their platform with access to data. Uber recently released anonymised data about travel patterns through its Movement application.

However, companies will set the rules about the sharing of their own data. Facebook use their API to control who gets access to customers' social graph, Facebook Connect and Graph API. They can use this to cut off any developer who poses a competitive threat. The result is that few developers invest seriously in creating alternatives.[49] Furthermore, platforms don't always reveal how the data has been treated which can diminish the usefulness of the data. This was observed during the Google flu trends, where the underlying algorithms (which turned out to overestimate flu prevalence) were not available to be inspected.[50]

Large internet platforms' market power can be leveraged into other parts of the economy

Looking further ahead, there is a risk that the number of platforms available will reduce as existing platforms start to consolidate other opportunities and businesses. Facebook began as a social networking site but has now added a marketplace. Recently it announced plans to offer free data plans to users in developing countries (though with restricted access to a small number of Facebook-approved websites).[51] Google began as a search engine but now offers email, document editing and cloud storage. It is not unusual for a consumer to access a Google service through a Google website using a Google browser on a Google device. Alongside this ubiquity they experiment in fields as diverse as quantum computing and piloting autonomous cars. Uber's long-term ambition is to become a global logistics and distribution firm, not just a taxi service.

As these firms grow and increasingly encroach on one another's space, they will seek to mine even greater quantities of personal data.[52] The risk is that with fewer but bigger platforms, possessing even greater amounts of our data, consumers and other companies will be captive in markets at the whim of a large and powerful company, able to extract greater consumer and producer surplus for their ends.[53]



Unlocking the social value of personal data

If this is the downside of not having control of our data, what would be the upside of gaining more control? What value to society would our data have if we had more control over how it was used?

The reason our data is so sought after by internet companies is because of the powerful insights it can provide. Data in and of itself has little value, but applied or analysed in the right way it can help us to find patterns, predict future behaviours, personalise service offerings, and even to shape people's behaviour and preferences. Governments and communities within the social economy are increasingly interested in data's ability, hoping it will help deliver positive social impact. Many individuals are also interested in the potential of their data to provide them insights, such as how to save money or improve their health.

Giving people control of their personal data and the ability to share it on their own terms would open a range of new possibilities for how people, communities and governments get value from their data. Crucially, giving people this control would give them the power to decide how their data is used, and for what purposes, democratising aspects of the digital economy. There are a number of existing practices which point the way to how DECODE would unlock social value from personal data.

Better run cities which respond intelligently to the needs of cities

The smart city agenda has long seen the promise of improvements in technology and big data as a means of optimising the running of a government or a place. Traditionally, smart city programmes have operated with data that governments generate and collect themselves. Sometimes this is expensive, requiring new infrastructure to be installed to collect the data. In other cases, they simply cannot collect the data they need as easily as a citizen could. Giving people control of their data would open up the possibility for them to share it with governments and public services in order to support both smart city agendas, but also more broadly to make government and public services more personalised and effective.

This concept overlaps with the idea of citizen sensing, which grew as a response to criticism of the Smart City movement as overly focused on technological solutions. Citizen sensing is enabled by the ability to use cheap devices (IoT and also smartphones) owned by citizens at scale to collect data relevant for research and other uses. Its premise is that if data represents a form of intelligence for a smart city to inform decisions, that data should come from citizens as well as other sensors and IoT devices. This will help the city to understand the city from a citizen perspective, and to collect greater quantities of data more cheaply and efficiently. The lack of citizen engagement has been a weakness of many smart city initiatives and the integration of citizen data is a key means of democratising the process of using data to inform the management of cities.[54] Barcelona is one of the cities that created a participatory and comprehensive digital strategy that starts from citizens 'needs,[55] with an explicit programme to develop and grow city data commons.[56]

Citizen sensing can take a variety of different forms. Bristol City Council has a citizen sensing project which aims to use a wireless mesh network across the city to source multiple types of data from citizens to be integrated with data collected by IoT devices. They have also been exploring ways to use citizen-generated data to understand people's needs and give them the mechanisms to address problems independently. For instance, one project involved giving residents damp sensors in the form of plastic frogs, which give people evidence with which they

can petition their landlords for higher quality housing.

In some cases, citizen sensing feeds directly into the management of the city's physical infrastructure, such as the Boston StreetBump app. This detects potential potholes from movements detected by people's smart phones as they drive around the city and sends them directly to the city management team. In Barcelona, the Making Sense project is helping citizens build their own sensing tools, make sense of their environments and address pressing environmental problems in air, water, soil and sound pollution. In other instances, citizen sensing can be a more passive process. In Jakarta, Twitter is used by citizens to map flooding in real-time on an online map.

Individuals providing transport data has led to creations such as OpenStreetMap which crowdsources data to create free maps and GPS services. Others examples include Geluidsnet, in which citizens living near Schiphol airport unhappy about noise pollution started a campaign by setting up a mesh network and installing sound sensors in and around their houses.[57]

So called data collaboratives involve partnerships between different sectors exchanging their data to create public value. Waze collects location data from millions of users to recommend routes which help users avoid congestion. The value of the dataset has attracted collaborations with local governments, like Boston, to help regulate the city's traffic. Popular proprietary apps like Strava and Uber are similarly sharing their data with city governments across the world to improve local public services.

These programmes highlight that data generated by citizens can have an important role in the management of cities. DECODE will take this further, offering citizens the ability to blend personal data from a variety of different sources, and share it on their own terms.

Data as a commons

The ultimate objective for DECODE is to be able to create a similar vision for personal data, whereby data can fulfill its social value as part of a commons. A commons is the sharing of a resource by a community according to rules set by themselves. A data commons describes the pooling of a community's data so value can be derived from it and benefit all members. Data is treated as a non-depletable resource and the more data is added to the commons, the more value can be extracted as new links can be drawn between datasets. As members reuse and integrate data, the commons becomes a desirable resource to researchers, businesses, public sector organisations, civil society and individual users who access it for data-driven insights and services.[58]

Data anonymised for academic research

Our public data, like taxes and medical records, are already available in an anonymised fashion so they can be studied. In fact, they have been used in the past in studies that investigate population trends and social biases. It is the openness of this data, how it's accessible and shareable, that has enabled this valuable research to be carried out. Jane Yakowitz describes in her paper *The Tragedy of the Data Commons*, how public datasets for academic research have made huge contributions to our understanding of public policies over the years. The examples given include:

- Public use data released by the Federal Financial Institutions Examination Council (FFIEC) provides a means of detecting housing discrimination and informs policy debates over the home mortgage crisis.
- Research performed by health economists and epidemiologists using Medicare and Medicaid data is now central to the debates about healthcare reform.
- Census microdata has been used to detect racial segregation trends in housing.
- Public-use birth data has led to great advances in our understanding on the effects of smoking on foetuses.
- Public crime data has been used to reveal the inequitable allocation of policy resources based on the socio-economic status of neighbourhoods.
- And the data commons is repeatedly used to expose fraud and discrimination that would not be discoverable or provable based on the experience of a single person.

The examples of research commons she provides often include strict conditionalities on which the data must be managed and shared. For example, the National Cancer Institute (NCI) Genomic Data Commons gives controlled access to the data by researchers, and a number of security measures are in place to reduce the risk of identification from patients' genomes.[59]

Although many of these examples refer to government data, it might be argued that a commons which includes highly personal data should provide greater involvement by those who the data refers to, both in terms of what data they provide and in setting the terms of access for that data. This is well demonstrated in the increasing prevalence of health knowledge commons and platform co-operatives.

Health knowledge commons

In recent years, attention has turned to a new form of data commons, created for and used by individuals as well as researchers. Some of these relate to health data, where there is huge potential value for people with health conditions from opening up their data to researchers in the hope they may find new ways of managing or curing their conditions.

Examples include PatientsLikeMe, which consists of groups of patients who share their symptoms, treatments and outcomes with each other; though data submitted is owned by the company and trials are done at the behest of pharmaceutical company funding. Midata.coop is a co-operative which allows users to manage and store medical data, from their medical records to their genomic profile, and keep it completely under their control. They can choose to share it with anyone (family members, doctors, and so on). The data producer needs to give explicit permission, especially with external bodies. This platform has already been used to conduct trials where patient information has been shared with pharmaceutical companies to aid in research studies.

DECODE will enable the creation of new data commons and co-operatives, formed and controlled by the people who contribute its data.

Shared personal data as a common resource for innovation

By making government data openly available to the public, greater value can be gained from it in the form of additional innovation, entrepreneurialism and scrutiny. This data can help innovators, entrepreneurs, community groups and civic enthusiasts to look for ways in which it can be used to respond to citizen needs. The most famous example is Citymapper, a global route planning app which was built on data provided first by the Greater London Authority's open data store.

With the right privacy-preserving mechanisms and data aggregation ability, there is the opportunity for personal data to become part of the resource made available by open data. Part of the difficulty for making personal data open is the lack of legal, technical or economic norms that would allow people to both collect, control and share their own data. If this were possible, then people might be able to share their data for specific projects or causes, or share it under specific conditions. Personal data also requires much higher levels of trust and transparency over data flows in order to facilitate effective sharing between individuals, and achieving these standards should be a prerequisite for third parties who wish to access or use that data.

This would be beneficial for both citizens and businesses. Making this data available under specific conditions to a selection of businesses, innovators and entrepreneurs would increase the chances that it is used to inform new products and services which respond better to citizen needs. As described above, the ability to do this currently is constrained and determined by a relatively small number of tech companies, who set the rules according to their own needs.

Section 2

An alternative vision for the personal data economy

The future risks of data monopolisation

The problems articulated in the previous section are pressing. It is inconceivable that people will produce less personal data as time goes on, and the technologies which use data to mediate our economic, social and personal relationships will continue to advance. Such developments, if left unchecked, would leave citizens disempowered within the digital economy would create an inefficient and unfair market for innovation, and would deprive society of the rich insights latent within personal data.

It is not the aim of the report to develop a detailed vision of this scenario. There are though a number of key risks which warrant a brief discussion.

The risks of not tackling the democratic deficit in data management

Imagine it is 2035 and over the past 18 years nothing has fundamentally changed about the internet economy: it still relies on personal data to create revenue streams and to provide tech giants with the intelligence they need for innovation and improvement.

The amount of data created by any person who uses connected devices would be magnitudes larger than people's current digital footprints. Location data of every place someone has ever been, everything they have ever bought, the time they spent hesitating over their choice to buy and details of their mood, levels of hunger, tiredness and frustration at the time, their full medical record and genomic information, a catalogue of photo and video data linked together by facial recognition technology larger than any existing library, granular details of their academic and employment history, every relationship, every public statement and and every online interaction.

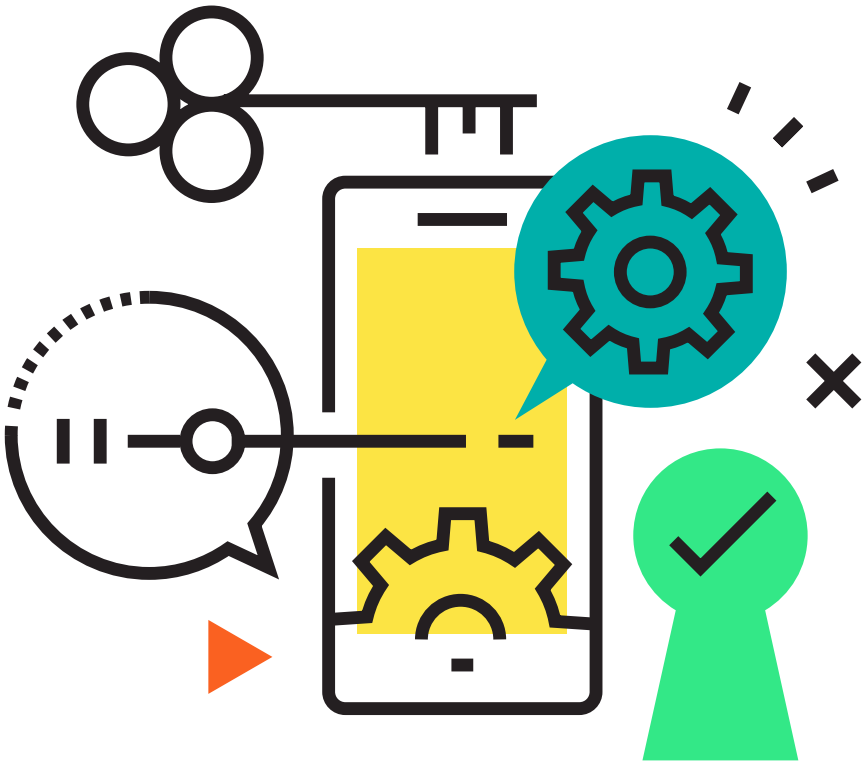
The services we have access to will be increasingly personalised, according to our interests but also, in the absence of competition, the interests of the platform owners and advertisers. The content we can access in education and news and social media will be tailored to elicit emotional responses making us more likely to buy, vote, or respond in certain ways. We may be encouraged into overspending, addiction, and dependency. Educators and employers could know intimate details of our physical and mental health, our inclination towards starting a family, our likely willingness to work more than our contracted hours without complaint.

Without changes to our approach to the internet economy and personal data management, this information would be collected and owned by private firms, housed in various data server farms around the world. It would have all been shared in exchange for access to the essential services provided by the internet. In doing so, control over how that data is used, shared and sold would have been surrendered.

With such a rich supply of data, the idea of digital - and non-digital - anonymity could become a historic concept. Data could be linked, sifted and manipulated to re-identify anyone from any anonymised set of data. Discriminatory practices could become rife. Discrimination and injustice could become baked into decision making processes across government and business. People of Colour could find themselves unfairly treated by the criminal justice system, or excluded from insurance or health services.

The increasing value of greater data supply would lead to a consolidation in the number of platforms people use to conduct their online lives. Network effects would enable internet giants to eliminate or acquire competition, making them even bigger, to a massively greater extent than we witness today. The handful of large platforms would begin to provide a comprehensive life management service, from shopping, banking, transport, work, health and social networking. All of these activities would be data-intensive, running on insights drawn from historic data all the while collecting evermore data.

The vastness of the personal data universe, and the sophistication of new technologies, would make regulation difficult. There would be considerable scope for personal data to be used to manipulate people for malign purposes. It could undermine democracy, a fair economy and social cohesion. People may be excluded from banking services, insurance, housing, jobs and even social activities on the basis of their data points. A sense of perpetual digital surveillance would cloak society, compromising any sense of freedom.



DECODE - an optimistic vision of the future personal data economy in 2035

DECODE is an attempt to give people, society and business a chance to take the different path. It is just one of a number of organisations, projects, collectives and communities who see that giving people control of their data would help avert the pessimistic scenario outlined above.

To illustrate why DECODE helps to address the problems we have outlined, we have drawn on our research to create an optimistic vision of a future in which people have been given greater control of their personal data.

Why 2035?

With acute problems facing the personal data economy today, focusing on the far-off future may seem like a distraction. But to fully understand the implications of such a major shift in how the internet economy functions, it is necessary to consider a longer-term horizon. It is not intended as a prediction of the future, nor which technological advancements are most likely to take hold. The purpose and value of this kind of futures exercise lies in the ability to think freely about the kinds of services, and society, that could emerge based on trends, technologies and movements visible today.

This report describes how decentralised control of personal data could change the economy and society in a European city in the year 2035. This scenario is explored through six imagined personas. We consider the opportunities and challenges of greater personal data ownership through imagining how a range of people would interact differently with the world as a result.

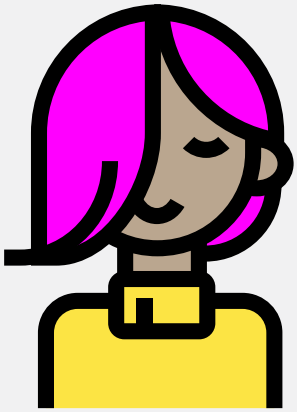
The personal data economy in 2035

In 2035, the majority of people now have their own personal data portals. These are in effect small servers, often located in their homes or a secure location of their choosing, which store all their personal data. This gives them control over how this data is used. It came about through a number of key trends:

- Governments and companies became increasingly worried about storing personal data. Beginning with the EU's General Data Protection Regulation, legislation for data protection became much stricter. It also gradually shifted from being reactive and instead began to anticipate advances in technology. This created a much more restrictive framework for how companies should use personal data, with harsh fines for malpractice. For most companies, the potential costs of getting it wrong started to outweigh the monetisation value of the data itself. Companies focused more on how they can use people's data only when required, rather than hoarding and monopolising it in the hope of future use.
- People realised that their personal data represented an increasingly complete picture of themselves – and the security of this worried them. Huge data leaks became increasingly common, including one from a high-end personalised genomics and biometrics health service, popular for the advanced personalised and pre-emptive insights it gave people to improve their health. Those affected were the victims of irreversible mass identity fraud. Alongside this, a series of successful cyber-attacks of large data silos suggested that data could never be completely secured when housed in server farms. As people became more wary, the climate became more amenable to regulatory change and businesses began to see the benefit of offering different tools and approaches
- A wave of new technologies offering people simple tools to store and manage their personal data differently emerged. At first these offerings were only taken up by small communities who were passionate about privacy or more socially-oriented forms of the digital economy. Over time, these technologies became more user-friendly and responsive to a range of citizen needs, increasing their take up. Following this, and as holding personal data became expensive for the private sector, large corporates saw the benefit in creating these tools, helping the idea to go mainstream.

Control of data and digital sovereignty

Polly



Age: 19

A student and political activist

Polly is part of a new generation of technology aware, hyper-cautious digital producers. Her generation have grown up with the ability to control their data and use private communication tools with encryption in preference to public social media activity that older generations use. Polly's mother was a victim of the genomic and biometric data leak and this had a profound impact on Polly's attitudes towards data and privacy.

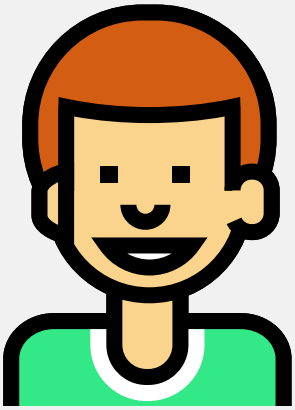
Polly keeps her data private and chooses not to share into her local neighbourhood data commons. She prefers a less personalised experience of using the internet if it means she can retain full control of her data.

Polly is a political activist who campaigns against many policies of the local city administration. She is concerned about leaving a digital trace of her campaigning and political opinions which the government could monitor. She uses a privacy-preserving identity authentication tool which enables her to participate on the city's digital deliberation platform about issues she is passionate about. This confirms she is a resident of the city, but does not reveal her identity to the city or to other users.

One of Polly's suggestions, for total transparency of all communications and meetings by elected officials, receives enough upvotes on the platform to be put to a local referendum. Polly is able to campaign for the motion without revealing her identity, but also campaigns offline in small, controlled events.

When the day of the vote arrives, Polly can vote from anywhere using her iris and fingerprint scanner to identify herself. She votes for the motion, though the vote is not recorded against her name for the council's audit purposes. The data collected through the process is collected in her personal data store so she has a record.

Ben



Age: 32

An online shopper and
self-employed graphic designer

Ben uses the sophisticated sharing settings on his personal data portal to select exactly which data he is sharing depending on the service. In effect, Ben is presenting a different digital version of himself to each service and individual he comes into contact with.

When shopping online for a Dream Writer, a tool which transcribes and illustrates dreams based on electronic frequencies in the brain, Ben is able to grant the website permission to past purchasing history of other electronics. He chooses for the website not to be able to see any other data. This gives him a personalised service, able to make use of the website's sophisticated analytics which help him find a Dream Writer that fits his needs. But it also means the website does not have access to any of the rest of his data as it would not be necessary to the transaction.

Ben has a disability which means he qualifies for some government benefits payments. He lives in public housing, and is a self-employed graphic designer. He has to deal with the government frequently - to claim his benefits, to manage his housing tenancy, and to pay his taxes. Using a combination of attribute-based cryptography and his sharing settings, Ben is able to manage these transactions in a way which keeps the data the government sees to a minimum. To claim benefits, Ben is able to authenticate his credentials and prove his eligibility without the government seeing who he is, or what specifically his disability is - the government just know that Ben is a genuine person who is eligible for the benefit. To pay his taxes, Ben can create a permission for the taxation office to have access to all his earning and spending data for a set time period. The taxation office's artificial intelligence enabled inspector automatically calculates Ben's taxation bill. Ben simply has to grant the permission and arrange the payment.

When he's travelling, or knows he will only be using a service once, he has a setting on his personal data portal which creates a 'burner profile' of himself. The service is able to verify that Ben is a real person, but Ben's data is blended and abstracted so that it is generalised and no longer personal to Ben. It represents a vague outline of Ben's preferences and demographic, enough for the service to provide a personalised offer, but not enough that it could identify Ben. This 'burner profile' is discarded after use.

An individual's personal data now represents a near complete picture of that person's life. A personal data store aggregates all of the data a person produces through internet-connected devices – phones, computers, most household utilities, health trackers, cars - their public service data, photos and videos, sentiment and preferences, their consumption data and their personal medical data. When compiled, there are few details of a person's life which cannot be reconstructed or inferred through their accumulated data.

The personal data stores give individuals genuine control of their data. Digital privacy is a right which can be exercised by all. They can see clearly what data they've created, and have the power to decide how this is used. The personal data stores people use have simple interfaces for sharing and using data to interact with digital services. While the scope of their data is huge, most people choose to use default options, based on culturally defined 'norm sets' which offer basic privacy, and then alter their preferences from an overview dashboard. This provides control via layers of granularity to keep the process manageable. From here people can grant permission to others to see their data, and call upon their data, as they choose when accessing online services.

Additional tools have given people a high level of trust in how their data is used. Each personal data store has a permanent and immutable record of every time an external party has used data from it. While the privacy and sharing settings mean that there should be no use which the individual hasn't authorised, the audit log gives them reassurance the system works and is secure from hacks.

Each data store can also utilise privacy-preserving technology. This is a means of authenticating credentials or other characteristics, without having to reveal any more information than is strictly necessary. To qualify for access to services with exclusion criteria, individuals can use this technology to gain access without having to reveal their personal information. The technology also enables people to be anonymous but authenticated. For instance, a city council can run a petition platform which people can sign without having to reveal their identity at any point.

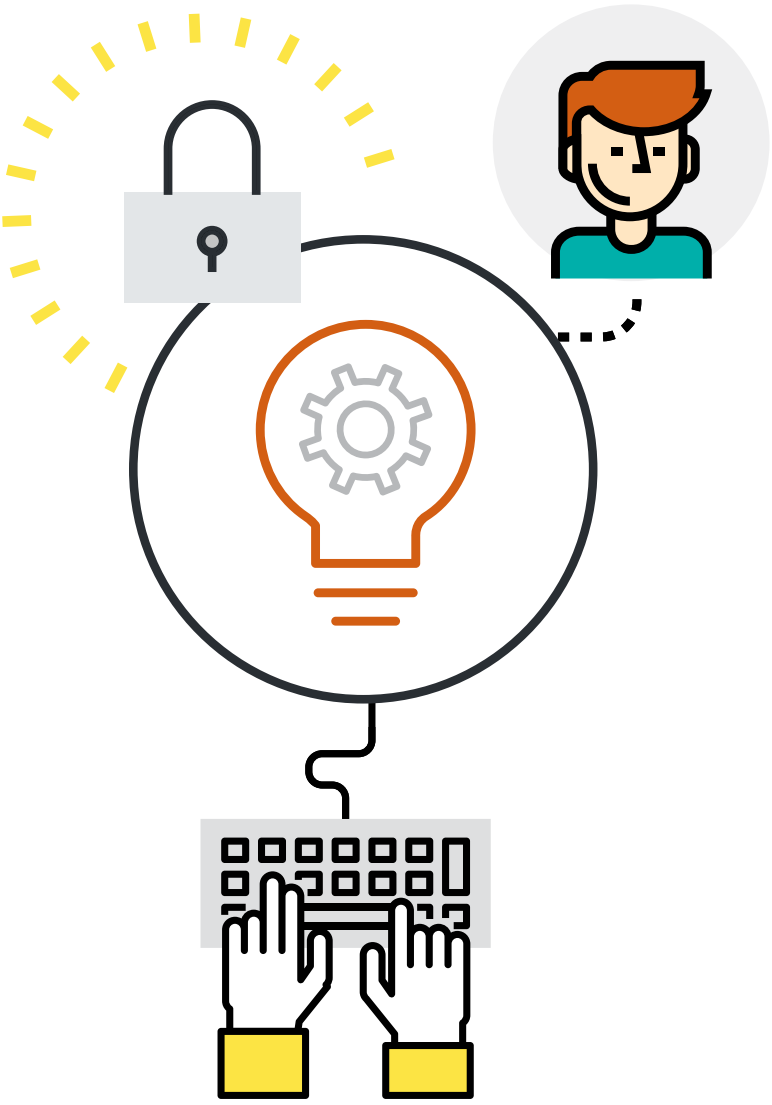
Having a personal data store makes interacting with services much quicker and easier. Permissions can be granted which give organisations all the information they need for a responsive service, but only the information they need and in a manner which can be revoked by the user.

This includes public services, which means that whenever people interact with an aspect of government they can grant permission for that interaction to all their relevant data. This is useful in health settings, where doctors can now get a complete picture of a patient - including family history, diet, exercise levels and data from wearables - without having to worry about information sharing or asking patients the same questions in every appointment. Another example of its benefits is for self-assessment taxpayers. As their data store contains all of their income and expenditure data, they can create a permission for the revenue service to see only the relevant data and only for the relevant time-period. This effectively automates the process of paying taxation when self-employed, an increasingly common status as the gig economy has continued to grow.

Users have started owning their own personal cloud servers. There is now little to differentiate cloud computing providers which have been reduced to simple, secure storage services of users' encrypted data. Initially known as 'Personal Information Management Systems' (PIMS), they have become as popular as services like Dropbox and Amazon Drive. Personal storage services never go offline, and can take actions or respond on behalf of users when they're not available.

Digital sovereignty has given people greater agency over how their data is used. People had different motivations for wanting control of their data. For some, the motivation to take control of their data was the opportunity to share it all for public benefit. They agree permissions for their data so that it is anonymised and shared with other people who have agreed to do similar with their data. Many people who see the benefits of data as primarily social use sharing settings to ensure that their data is licensed so that it can only be used for non-commercial, non-profit making uses.

For others, the appeal of greater control of their data was the ability to reclaim its economic value for themselves. Personal data remains essential to companies, even if it's prohibitively expensive for them to hold it themselves. This creates a new form of data economy in which companies can pay individuals directly for their data. It has also given rise to an industry of private data brokerage services offering to help people make money from their data. These firms offer the highest rewards for non-anonymised data, and for data sets from people with high purchasing power. The personal data economy is still a way of making money in the digital world, but individuals now have a greater insight into its mechanics and can claim some of the value for themselves.



Internet services can connect directly to an individual's personal data store using APIs. Most terms and conditions now come with a set of granular options which request access to people's data in return for free use of services. This means companies can still use people's personal data, and directly benefit from analysing it so they can perform personalisation and segmentation exercises. But some other activities, such as the on-sale of personal data into secondary markets, has dried up as a revenue stream. There are also now options to 'pay for' services which were previously free, if a person does not want to allow any access to their data.

The integration of personal data in one place, at the individual's fingertips, has also enabled most people to purchase 'data butler' services. These offer automated, personal assistance services which preemptively respond to people's needs based on analysis of their personal data. For instance, a data butler can look at a person's calendars, emails and current location and make a judgement that they won't be able to make it to the airport on public transport because their meeting is overrunning, so they need to order a taxi.

Data butlers can automate aspects of a person's life, and provide nudges and warnings when it detects problems emerging. If a person's physical activity levels drop off, and are accompanied by increasing spending on unhealthy food, the data butler can nudge the individual towards healthier behaviour. If physical signals indicate they are entering pre- stages of conditions such as diabetes or heart disease, the data butler alerts the individual and offers to arrange a doctor's appointment on their behalf

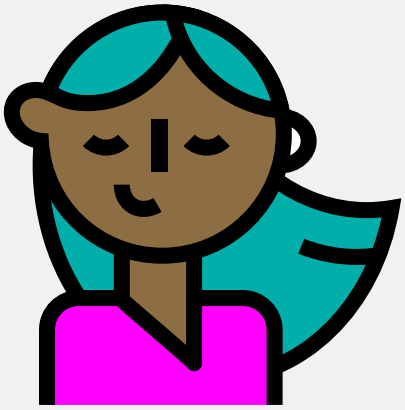
Before people had taken back ownership of their personal data, these kind of butler services left people feeling uneasy as it reminded them how much of their lives were being surveilled by large companies. By contrast, data butlers which link to personal data stores are given strict instructions by the user about what they can and can't see. The individual is able to set the level of assistance they want and which data the butler has access to. There are also no third party advertisements or recommendations offered. Users pay an upfront fee to use the service.

Some people are concerned about their personal data clouds malfunctioning. With everything stored locally, there is just one point of failure and a house fire or burglary can easily lead to a person losing all their data. Many people now pay for a data shadow service, whereby they pay for their data to be scattered securely across thousands of different data server farms. If a data back-up is ever required, for instance if their home store has malfunctioned, the person's biometric data key can blend all of their data back together.

There is an inter-generational imbalance in levels of digital sovereignty. For the early generations of internet users it was not possible to fully reclaim control of all of their personal data. It was too disparate and ownership had often been inadvertently transferred to private companies. These people, known as the 'lost digital generation', live with what is perceived as incomplete privacy. Younger generations who became digitally active after the introduction of tighter regulations and personal data stores enjoy a far greater sense of digital sovereignty. There is a resulting difference in how people interact with technology across generations. Younger generations are much more likely to value privacy above all else, while for the lost generation the benefits of sharing have fewer perceived risks and downsides.

A plurality of internet business models, platform operators and innovation ecosystem

Sarah



Age: 27

An entrepreneur and business owner

Sarah is an entrepreneur. She runs a small business which sells low-cost wearable skin sensors that detect sleeping patterns, heart rate and stress levels. She initially launched her businesses through a successful crowdfunding campaign. With over 1,000 products sold, Sarah is now steadily growing her business.

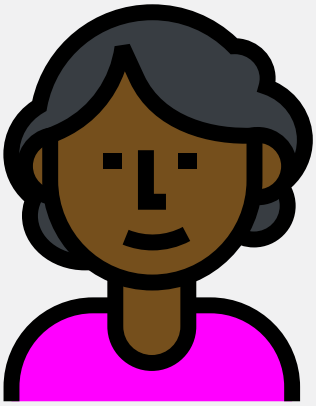
Sarah's intention was to make profit from her company primarily through the sale of hardware devices, but also by providing the user with deep, layered insights about their daily activities and behaviours through a monthly subscription app.

The company naturally minimises the personal data it holds. Users pull all their sensor data directly into their personal data stores, where the data is maintained either locally on users' hard disks or on their cloud storage. The app offers a range of different data-driven measurements and insights, blending sensor data with different sources in the user's data store.

There are several different personal data store providers, and Sarah's team had to work hard to create several different versions of the app that offer compatibility with all of them. Although it is not a legal requirement, is has become very common for companies to publish the source code of their software for security reasons, and Sarah has chosen to follow suit with this trend to try and build trust with her users.

Sarah has also decided that she wants to run some customer analytics to see how each of her services are performing. For this she uses a popular open source customer analytics app. The app sends requests down to each user's personal data store, and, after the user has granted permission, it only sends aggregated and anonymous insights back to Sarah's' marketing team. The company never has to see or store any fine-grained, detailed data about individual users, which protects the company from coming into conflict with data protection law.

Florence



Age: 56

Looking for work while suffering from a long-term health condition

Florence has lupus, an autoimmune condition which restricts her ability to work full-time, regular hours. The gig economy meets many of her needs, enabling her to work when she can but take time out when she needs to for health reasons. After a negative experience with a commercial gig work platform, which initially seemed promising because of the higher fees it paid, Florence joined a co-operatively-run gig work platform. Through this platform Florence can quickly find work, and all members have collectively agreed rules which prevent undercutting of payment rates. The platform has a Mutual Benefit Fund, which can provide financial assistance for people unable to work due to health reasons. While Florence tries not to call on its help, every now and then it provides a vital lifeline.

She travels regularly to her son's house in a neighbouring city to help look after his children. While she is away, she rents her apartment through her local short-term let platform co-operative. The platform co-op has become highly popular among local residents for ensuring that the benefits accrued from local tourism are fairly distributed to locals, renters, home-owners and holidaymakers. Each member commits to only renting their apartment for a maximum of 90 days. The community can inspect the occupancy data to check that these fair practices are obeyed, but the platform respects users' privacy and doesn't reveal any of the home-owners' identities. A share of the profits made by the platform are also invested back into a community housing scheme to build more local homes.

Florence is also active in a national data commons for people with lupus. Due to the rarity of the condition, a local scheme did not achieve the scale needed to create a dataset which could be used by researchers. Members of the co-operative can work with researchers and pharmaceutical companies to have their experiences and preferences reflected in the type of work that is carried out with their data. A small fee is paid by organisations accessing the data, which helps meet the running costs of commons. Members contribute as much data as they can in the hope that by considering a wide range of factors, there is more chance that positive interventions might be found.

The emergence of personal data stores changed the business model of many internet businesses. The use of targeted advertising became more difficult in an era in which people didn't want to give up control of their data. This made advertising space much less valuable. But this didn't fundamentally change how many internet businesses operate. They can still request to see someone's data, so that advertising can be targeted as they visit that site, they just do not harvest that data or have the ability to sell it on. The secondary data market shifted from brokers who would connect up data on behalf of third parties, to brokers who would connect companies directly to individual data. In practice, when people visit a website, brokers request permission (which can be agreed in blanket approvals) to view their data, in order to permit a real-time auction to advertisers for the space.

However, the wide take-up of personal data stores did enable new forms of digital production and platform services. The combination of personal data stores, regulations which levelled the playing field against large platforms, and technologies which enabled decentralised transactions unleashed a wave of platform co-operatives offering services such as taxis, cleaning and care. Social networks arose which did not all operate as competitive walled gardens selling advertising space in people's social lives; some charged for use, some, especially popular services which failed to turn a profit, became co-operatives owned by their core users. Others explored different models, funding themselves through opt-in surveys, through partnerships with local services, through access to marketplaces, and more.

Platforms operate with the seven founding co-operative principles, which distinguish them from profit-seeking platforms:

- Open membership.
- Democratic member control.
- Equitable member economic participation.
- Autonomy and independence.
- Education, training and information.
- Co-operation among co-operatives.
- Concern for the community.

The growth of these platforms was motivated by a desire for a fairer kind of internet. It was fundamentally about two things - *"shared ownership and shared governance of the internet's levers of power - its platforms and its protocols".*[60]

The internet and other technologies continues to disrupt the vast majority of industries and professions. This increased the precariousness of work and the unequal distribution of rewards, and led people to look for different ways of organising themselves economically. The leading voices behind this movement wanted people to be at the heart of business and consumption activity, rather than capital. Their vision was an economy which valued more than profit. Having control of personal data solved one of the problems which led to a small number of platform giants - the use of personal data to accelerate network effects. It helped create a more plural economy in which people could make a living while promoting socially-minded principles.

Platform co-operativism needed other factors to enable it to find a competitive advantage. Having begun as a counter-movement far from the mainstream, momentum started to build as people saw the exploitation and negative effects created by large commercial platforms. This momentum helped them campaign for governments to direct public investment towards platform co-ops to help them grow. This was vital as the traditional venture capital funding model for internet startups saw little value in investing in platform co-operatives. Over time, governments also recognised the need to change competition laws, focusing them instead on democratic ownership and not just breaking up large companies.

Platform co-operatives exist across all sectors of the economy, and reach into other more community-driven aims too. Local co-operatives operate in energy production and use, work/gig economy, sharing of tools and equipment, short-term rentals, car sharing, child care, elderly care, local currencies, finance and banking. In some cases, platform co-operativism has removed the need for some intermediaries. For instance, people can now buy and sell houses without the need for estate agents as the combination of data sharing and trust mechanisms remove the need for a middleman. This has shifted the balance of power in the economy towards individuals and communities, and away from companies. For instance cleaners can work together in a co-operative without a middleman to provide services and control the distribution of profits amongst themselves, supporting fair working conditions, while delivering accountability and quality to customers.

Some platforms have explicitly social objectives, such as to increase employment options for low-income or marginalised groups. Others have alternative pricing structures and business models such as a 'pay what you feel' which help them to be price competitive. The rules of the co-operatives, agreed democratically by its members, can promote and incentivise ethical consumption habits, and disincentivise harmful habits. For instance, a short-term rental co-operative has agreed rules that try to protect existing residents from being forced out through the increased prices which characterise commercial platforms. Instead they have a cap on what can be charged and commit to only allowing rentals for a set number of days each year.

Co-operatives reinvest any profits generated back into the platform, for the long-term benefit of users. In the case of the gig economy workers platform, some of this money is put into a Mutual Fund which can compensate workers when they are unable to work, such as through sickness or injury.

While in the past, 'democratising the internet' could be used simply to mean access to a service, this now implies that people have a genuine say in how the platforms they use are run. This makes them self-managing. Co-operatives also have specialist technical boards which handle complex decisions about changes to the technology which underpin the platforms. The members of the technical board are democratically elected and are accountable to the other members. Co-operatives promote stronger social ties which give people higher levels of trust - local, real-life connections make people feel more confident the technology and practices were ethical.

One inevitable consequence of a more decentralised world was the need for systems of trust to be introduced. Decentralising technologies could never completely remove the human and political elements of how goods and services were exchanged. This meant there still had to be mechanisms which people could use to confer trust. Open code and software was an important part of creating new mechanisms for trust. Transparency of algorithmic code, and data manipulation methods to strip datasets of bias were also valuable.

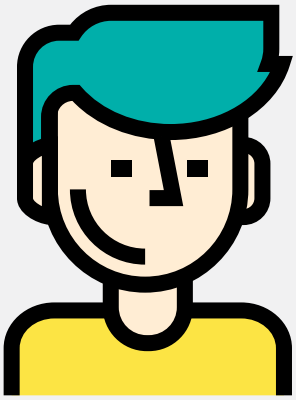
Many platforms used reputation scores as a means of incentivising good behaviour, and communicating reliability. In platform co-operatives, it became a good means of focusing members on the ethical priorities which had brought the co-operative together.

Over time, these reputational scores needed to be portable, and began to merge with one another. While helpful in some ways, the creation of a reputation passport also created challenges. People began to feel constant judgement, which some found oppressive. Others felt that reputation passports could entrench social inequalities. Efforts have tried to eliminate gender and racial bias from people's ratings to tackle fears that reputation scores may harm marginalised groups.

Platform co-operatives are also an integral part of a new technology development process. Artificial Intelligence (AI) is widespread and has found a role in nearly all professions. Platform co-operatives have contributed to the development of AI, bringing them closer to human needs and able to place a democratic and ethical framework around them through the governance of the platforms. With AI now responsible for large amounts of new computer coding, the ability to bring in safeguards and the ethical priorities of those who will use it is an essential feature preventing AI from becoming harmful.

Cities using citizen sensed data

Gerry



Age: 30

A gig-economy worker

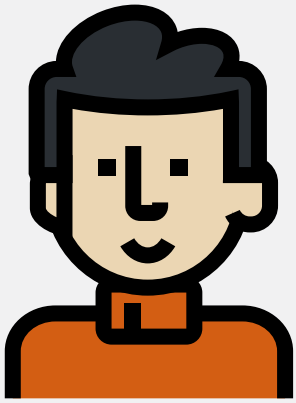
Gerry is in and out of work, though he usually freelances doing odd-jobs like gardening, cleaning, or handyman work. He uses a platform which collects information about his location, skills, and employment activity and quickly connects him with a range of companies looking for temporary workers. This access to rich personal data has become a widely-used method by which freelance workers and businesses are able match supply and demand for labour quickly and efficiently in the gig economy.

Gerry initially used a service which offered him jobs in exchange for exclusive use of his data, both from his connected devices and longer employment history. He agreed to the platform's terms and conditions, which also meant agreeing to be sent personalised product advertisements on a weekly basis. Over several months Gerry found work frequently, yet the jobs were often low quality and low pay. He was also uncomfortable with the advertisements he was being sent, which often included alarming insight into his activities and behaviours.

As a result, he followed a friend's recommendation to switch to an alternative platform co-operative that performs a similar function but practices higher standards about the companies which it partners with. If a company breaches the terms of its contract with a worker then the co-operative's ethics committee can decide to remove that company's access to the members' pooled data.

Instead of advertisements, the platform sends Gerry regular updates about local community events. At first he thought these were too frequent, but over time he liked being able to connect with his fellow peers. Gerry recommends that all of his friends join the platform to grow the co-operative and improve their influence in the local economy.

David



Age: 42

A part-time science teacher and stay-at-home dad

David is involved in a neighbourhood energy co-operative. The co-operative runs completely on renewable energy. It was formed after residents, discussing neighbourhood issues on an online deliberation platform, identified extortionate energy prices and a concern about a lack of focus on renewables from national government.

David had not previously been involved in community action but had been personally interested in climate change since his days as a student. He and his partner had been making environmentally conscious choices for their adult lives and saw this as a further chance to make an effort where they could.

The energy co-operative was formed and each member granted permission to access their energy consumption and production data to other members. This provided a picture of the total energy requirements of the neighbourhood, how much energy was produced by solar panels and mini wind turbines, and the profile of demand and supply. At first, not enough energy was produced locally to run the neighbourhood, so the co-operative used its bargaining power to negotiate a cheaper deal on energy from the national provider.

The co-operative members were conscious that the supply and demand of energy did not always match up - people wanted to use energy in the evening but solar panels typically produced most energy during the day. They spoke to another co-operative doing the same thing in a neighbouring town, who shared an app they have built that uses data from the co-operative and other open data to make predictions about periods of high demand and high supply. It then creates prompts for people to encourage them to use appliances outside of peak hours to reduce the strain on the local grid. Over time, the neighbourhood become more invested in the scheme and people committed to producing more renewable energy. Eventually the co-operative can run entirely using the local renewable energy production. A retribution scheme was put in place, so excess energy of some neighbors was used by others with a level of compensation offered.

Aspects of the city government are run through a combination of data and decentralised decision-making. The data shared by citizens enables the city government to be truly responsive. The city has access to real-time data generated by public services, IoT and citizens to create a truly intelligent local government body capable of undertaking analytics on demand to inform decision making.

With citizens now in control of their data, it is far easier for them to share it with the city for public benefit. An early pilot which used sensors in people's bins to create a more responsive refuse collection service demonstrated that sharing anonymised personal data improved the services people received in exchange. A strong political communication effort accompanied this, designed to change people's minds about the benefits of sharing their data in privacy-preserving ways. This created a better incentive for many people to share their data. While most of the use-cases for their data didn't change the amount of tax they paid they valued not having to wait as long for problems to be fixed and enhanced service quality. Gradually, a wider range of data sets were unlocked for the city to use, analyse and combine in order to understand how to improve the city for its residents. Citizen generated data sets include:

- Environmental data, such as air quality, noise levels, water quality.
- Experiential and sentiment data, such as how people feel about particular aspects of the city.
- Transport and movement data, for instance routes, footfall, congestion, public transport black spots and common points of failure.
- Issues and feedback with local public services, such as housing, health, education.

While some of these data-sets are possible for governments to collect themselves, having citizens contributing it makes the process cheaper and opens up massively more data, which is also more reflective the lives of residents. Using this data, the city government can be genuinely responsive. If there is repeated noise nuisance at night, the city can observe this in real time by connecting into the noise sensor data produced in people's homes, sending out a response team if necessary. By monitoring people's anonymised location and transport data the city can see where transport black spots are, where delays occur most commonly, where additional transport is required and where new cycle routes should go. Real-time feedback on public services alerts the council to problems stemming from poorly designed administrative processes.

Unlocking this data also supports increases in hyper-local decision-making. The city has been able to take advantage of personal data control and digital tools to enable neighbourhoods and communities to make decisions for themselves. For instance, communities are able to use data to understand where priority areas for funding are, then run participatory budgeting processes to allocate the money accordingly.

The Rise of the Data Commons

Personal data control has paved the way for the creation of a vast array of data commons, a new form of public good present in every city. These enable people to share their data, anonymously, for public good. The user interface on personal data stores offers both high-level and granular options for sharing, putting the individual in control of their data. Most people choose to share the data they produce as a byproduct of living with technology. The sensors on objects in their home, data from their wearables and chip implants about their location and movement, and even some experiential and sentiment data where there is a compelling benefit to the community from doing so.

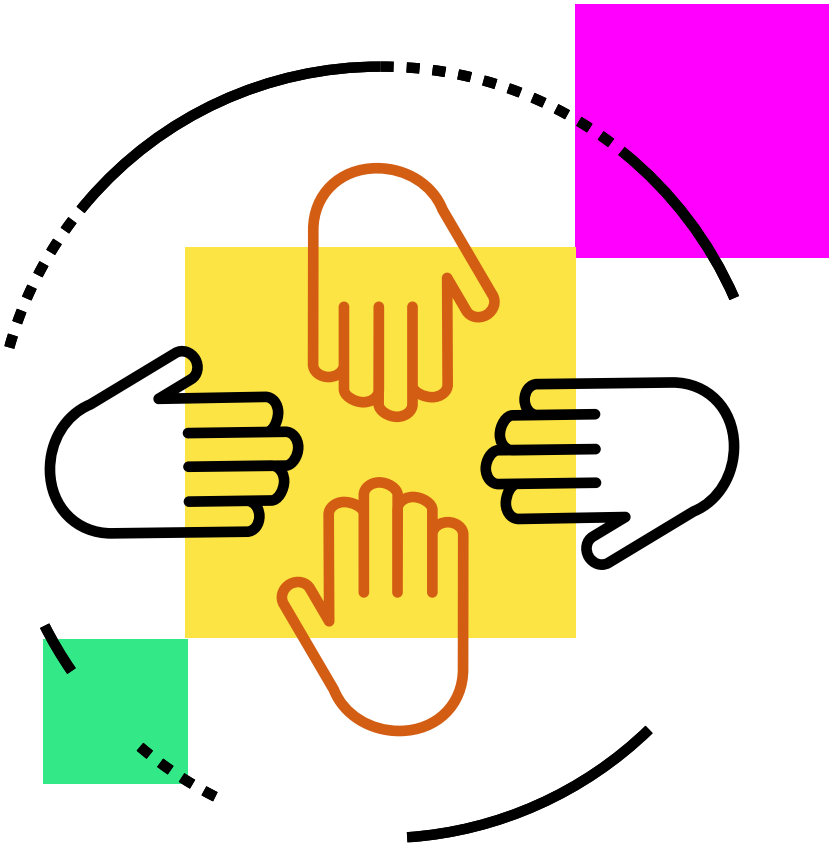
Some data commons specialise in particular types of data, such as health, and have a very specific purpose. These commons are not open to the public in the same way, as the data is far more sensitive and members of the commons use the restriction as leverage to give themselves a voice and stake in how it is used. For instance, health data commons offer a trade-off to pharmaceutical companies and medical researchers that access to the data is dependent on members of the commons being given a say in which research is conducted, and which pharmaceutical research and development is undertaken.

The creation of personal data commons has started to result in some powerful analytical insights. At first, data was limited and the use of it tended to be on small-scale responses and solutions. But as more people participated, and more personal data sets started flowing in, there are now significant insights which can be gained

In particular, the combination of traditional health data, genomic data, personally generated health data from wearables and self-monitoring devices, and consumption and lifestyle choice data has led to ground-breaking discoveries in the health field. Scientists and doctors can now predict with great accuracy which individuals are at highest risk of certain conditions. They also have access to vast amounts of rich data on all manner of health conditions, helping them to design better cures and interventions. Combining genomic data with lifestyle data, such as what people buy in the supermarket and the number of times they exercise each week offers unparalleled insights. Real-time risk detection and early warning systems can intervene with the aim of preventing people from tipping over the edge into chronic conditions, using data on their activity, spending and health data.

The data commons co-operatives are self-governing communities who agree rules collectively and democratically. This enables them to create rules about the use of data, and the digital and real world services connected to that, which prioritise social value over economic value. Data commons usually embed the principle of reciprocity into their governance. This requires that third parties who use the data have to give something back to the commons, to give it sustainability. This could come in a variety of forms - data storage space, donated worker hours, or simply finance for the commons. Reciprocity clauses can also be attached to the data, which require third parties interested in the data to give something back to the common in return - this ensures its survival.

The scientific and research community became extremely interested in access to these data commons. This was one of the first major philosophical challenges facing members of these co-operatives. There could be significant benefit for members, but the research community was not an active contributor to the commons and therefore not entitled to access the data. Various debates were held through various co-operatives and, in the end, the majority decided that data should be open for research purposes where there is a clear social benefit. This opened up the commons to outside eyes, reducing overall people's control but increasing the value of their data.



An immutable attribution technology enables people to be properly recognised when their data has contributed to social impact. The audit log which it provides of every time someone's data has been used means that it is possible when a new product, intervention or service becomes available that each individual can be alerted to the fact their data enabled it. In addition, where data from a data commons has been used to create commercial products or services, the same attribution system can ensure that a data royalty is paid back to the commons whenever that particular data contribution leads to a new sale or use.

In the past, there had been a failure of some early data co-operatives because decision-making and governance broke down – either through a lack of a truly democratic structure, or because people who did not know the technical side couldn't contribute to debate. There were instances of decisions being taken on behalf of other members which they didn't agree with, or which had negative distributional consequences for some members.

Most co-operatives now have three tiers of decision-making. General decisions are fully democratic, with each member having one vote per decision. For more technical decisions, co-operative members found they were unable to vote in an informed way, as the technical details were too complex. For these decisions, an elected committee take decisions on behalf of the rest of the co-operative, sometimes via a delegated democracy arrangement. There is an ethics committee, also elected, who specialise in making decisions about research requests on behalf of the co-operative.

In the past there had been an imbalance in the socio-economic profiles of people who share for good and those who share for money. Some of the private data brokerages insisted that in order to receive financial rewards for sharing data, people had to grant them exclusive access to it. This meant they were unable to participate in co-operatives and data commons. In practice this meant that poorer people tended to share for money, while richer people tended not to see the additional income (a much smaller percentage of their annual income) as a worthwhile sacrifice for loss of privacy and ability to help society. From this starting point, a bias in the commons data emerged.

Affluent people, already early and heavy adopters of new data-producing tech, began using the commons first as their data is more readily available. They continued to use the commons as relatively modest financial incentives to sell were not compelling. Companies wanted to attract these groups so deliberately orientated their services and products towards their needs. On the other hand, poorer people - who often lead more chaotic lives with fewer communications channels and technology - had less data to contribute and were more likely to be missed by data collectors. This led to a bias, with the commons data more likely to reflect the preferences, needs and status of more affluent neighbourhoods.

A class action lawsuit was taken on behalf of the people who had sold exclusive rights to their data. The case argued that they had in essence become digital slaves, unable to leave a financial relationship with the data broker. The Supreme Court ruled in favour of the collection of people who sold their data, giving them the right to use it for other purposes. National governments subsequently legislated to prohibit companies buying exclusive rights to data. This helped reduce socio-economic imbalances in the data held in the commons.

Supporting the digitally excluded

A minority of mainly older people do not participate in data commons, or store their personal data in one place. They rejected computer technology, either for practical or philosophical reasons, and live without it. This has posed a problem for the city government which has been attempting to integrate digital technology into all services. An outreach programme helped educate and train people who didn't use technology because they found it intimidating or confusing. For many services, such as payment of tax or applying for services, it is possible to retain analogue options for those who resolutely choose not to use technology.

However, by rejecting technology these people are absent from the picture created by data commons and smart citizen sensing labs. In response, the city government has to factor in that this section of society is not being represented by data. City data analysts have developed statistical techniques to factor in their absence. It also reinforces the need for the city government to continue offline consultation and engagement work, so that there is always a means for people to contribute which doesn't involve technology.

Section 3

Exploring current trends for the future of personal data

So far we have presented an optimistic vision for the personal data economy of the future. This section will bring us back to the present. It will build on these scenarios by outlining some of the current trends which support this vision, and how DECODE aims to navigate these trends to build a feasible and sustainable solution.

We begin by outlining DECODE's alternative model, which emphasises data's relational value, that is, as something to be freely circulated, used and re-used as a common good. Beyond simply redistributing the economic value captured by advertisers, a data commons has the potential to open up a whole range of integrated, ethical uses for data which provide benefits to individuals, society and the economy.

In the subsequent sections we describe how DECODE will create tools for individual ownership over data, which encourage the sharing of data and creation of new models for data use. A key challenge will be how new technology balances ethical practices and privacy of individuals, while realising the benefits from widespread sharing, use and re-use of data. These two objectives are seemingly difficult to reconcile. Data can lay bare intimate information about our personal behaviours and habits, but it fulfils most of its value as something which can be shared. For DECODE to achieve its aims, it will need to offer:

- **Flexible rules that give people ownership and control**. DECODE will need to devise technical tools to help give users full control over who can or cannot access their data. These must be designed around the user, offering options for anonymity, transparency, and privacy-preserving standards for sharing data while minimising leakage of personal or sensitive information.
- **A new generation of digital platforms**. Much of the opportunity will only be realised where individuals are able to pool their data together to leverage its potential economic and social value. Platform co-operatives offer a feasible model, highlighting the potential of digital technologies to help members to govern themselves. Effective data sharing has to be underpinned by high levels of user trust, and platform co-operatives achieve this by embedding openness, respect for individual users' privacy, and democratic participation over how decisions are made.
- **Creating incentives for people to join the DECODE system**. DECODE will need to offer compelling use-cases to individuals in order to be a sustainable alternative to the status quo. The project must be able to generate economic and social value without resorting to selling data to advertisers, and will need to make use of a number of strategies (including policymaker support) to build momentum around the pilots.

A simple solution - can't we just redistribute more of data's economic value?

For many people, the answer to the challenges set out in Section 1 is simple: give people the tools to realise and earn more of the economic value of their data. Imagine, for instance, if Facebook actually paid you money to see the adverts that it shows you. Dozens of new apps and personal data stores now lure people in by allowing them to become active players in the online advertising industry while reclaiming more of the value of their data for themselves. While this might be a step in the right direction, it doesn't grasp the full potential that data can have, for individuals, for society, or for the economy.

One response from the tech industry to problems in the personal data economy has been to win over customers by offering them tools to leverage more of the economic power of their data. Personal Information Management Services (PIMS) allow users to retain and store their information, in turn putting them in a better position to exchange that data for new services. Some PIMS also offer discounts, or even cash transfers in exchange for personal data, which hand more of the monetary value back to the individual.

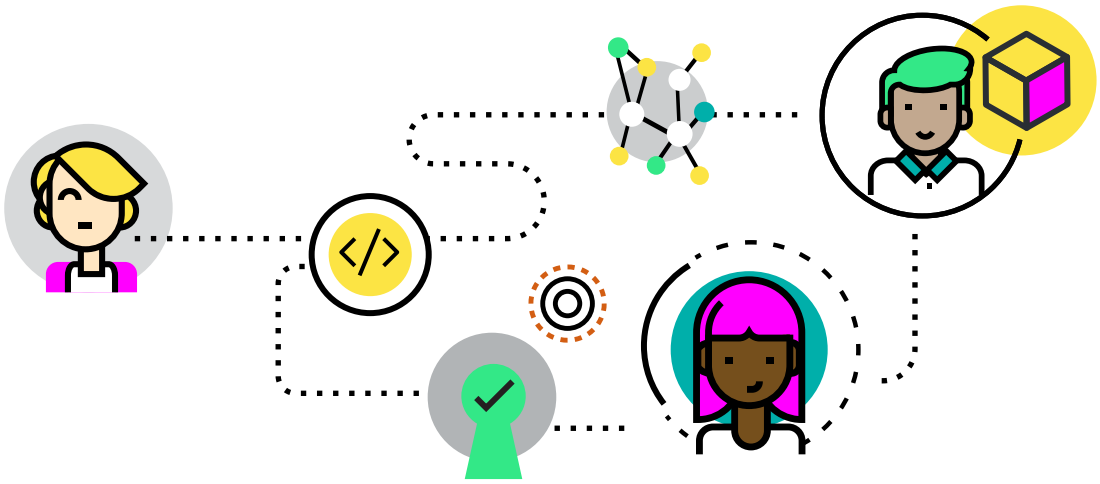
One example is Datacoup, a service which lets users pull their financial transaction data, wearables data and other information from third parties into a live dashboard. Users then have the choice to submit this anonymised data to companies in return for rewards. According to the co-founder Matt Hogan, an example might be a life insurance provider who might say that if someone is willing to share their spending data from their primary spending card, along with their Fitbit data, then that user will receive a discount on their insurance premium for as long as they're willing to share that information.[61]

CitizenMe is another service which encourages users to create personal profiles whereby all the data is stored locally on their mobile device. Josh Hedley-Dent of CitizenMe explains that CitizenMe never see or touch user information, the users themselves are the controllers for their own data. As in the case of Datacoup, users are offered deals by companies (mainly advertisers and market researchers) in exchange for insights from users' anonymised data:

If we're providing an insight to an individual that's based on their smart meter data, they can be informed that their fridge is about to pack in, and if they want we can show them appropriate offers to buy a replacement. That could replace the advertising industry (which needs to change) by moving to continuous marketing - where customer intent is matched with relevant market supply.[62]

By giving the user full choice in who has access to their data, customers may enter into more consensual and controlled relationships with companies who, in return, get access to fine-grained, high quality data. In other words, PIMS' central economic role is to allow the upward flow of rich, detailed information about the location, nature, and size of demand, in turn allowing for a much more efficient alignment of supply to that demand.[63]

Many PIMS operate on a similar assumption to the current personal data economy: that personal data indicates demand for a product and that indication of demand can be bought and sold at a price that advertisers are willing to pay for. PIMS simply make the user more active participants in transactions involving their data, cutting out intermediaries (like data brokers), and where possible even luring users by handing down some of the revenue from the sale of their data, via rewards or direct cash transfers.



While 'putting a price on data' has provided a useful way of alerting customers to the value that it has, many of these arguments rest on the problematic assumption that data should be seen primarily as digital asset. This is an argument echoed notably by the World Economic Forum's 2011 report *Personal Data: The Emergence of a New Asset Class*.[64] Some critics have argued that seeing data primarily as a new form of digital 'money' - akin to any other asset class such as bonds, equities or real estate - can be limiting to its potential use-cases.[65] Even with PIMS, focusing primarily on data's monetary value risks creating incentives which lead to speculation on its price, or even to new forms of 'data hoarding'.

What's more, since the value of data relies on its aggregation, it's likely that people would be paid little money for their data individually, and companies will be more likely to make profits on the aggregated value. One individual's data in isolation will give them very little bargaining power, making them more susceptible to draconian terms and conditions set by companies who wish to buy their data.

For many years it's been a cliche to claim that information or data are the new oil. Yet this analogy does not fit with the core properties of data. Oil is a scarce physical resource. Data and information by contrast can be replicated without limit and often become more valuable the more they are shared. This is why the ownership and business models outlined above may be ill-suited to making the most of data. Many of the people we interviewed had conceptions of data beyond data as a new 'asset class'. They instead view data as something closer to a common good that can facilitate co-operation, sharing, and collective use. As Annemarie Naylor from think tank Future Care Capital and Oleguer Sagarra from Barcelona Council suggest respectively:

I don't think that data, in isolation, is an asset and neither do I regard it as akin to a raw material - it is subject to individual 'controls' rather than property law. And, that logic implies that algorithms are more readily the 'assets' with which we should concern ourselves, to the extent that they contribute to the intellectual property needed to make productive use of data for everyone's benefit.[66]

Now the digital world is fully connected, much the same way that air is connected, we need to understand how it can benefit everyone, and the benefits from using it should be distributed. In the data commons, we need to make sure we build a common library of data from which to build projects on. People say data is the new oil, but I say it's the new soil.[67]

A theory of the commons

The commons is a general term for shared resources in which each stakeholder has an equal interest.[68] Commons can be environmental (woodlands, rivers), cultural (literature, music) or digital (free and open source software, Wikipedia). History is littered with examples of communities depleting or privatising these resources. In the 1960s Garret Hardin's famous theory popularised the term 'The Tragedy of the Commons' to describe this trend.[69] He argued that when access to a resource is free and is not defined by private or public property, individual self interest inevitably leads to overexploitation and the depletion of this resource. Only privatisation of it or resorting to make it the property of the state, would be able to safeguard the resource.[70]

In Governing the Commons, Nobel Prize-winning economist Elinor Ostrom provided an alternative analysis of historical case studies to show how common pool resources can and do generate norms and rules that encourage people to work collectively, as opposed to the logic of markets that are driven by competition over scarce resources. Ostrom suggests that Hardin's theory rests on an overly stark separation between the public and private spheres. In Hardin's view, property rights are thought of as something that are exclusive, making the idea of a non-absolute right of ownership inconceivable.

On the contrary, Ostrom thinks of rights of ownership of a good or resource as something that can be broken up into a multitude of rights and obligations - a 'bundle of rights' - distributed in more or less equal manner among the people who have access.[71] She argues that institutions are rarely characterised by a neat separation into either 'public' or 'private', but they are *"rich mixtures of 'private-like' and 'public-like' institutions defying classification in a simple dichotomy"*.[72] For example, commons may be accessible to all members of society - such as natural minerals like air, water and a habitable earth - but no individual has exclusive ownership over their use. A landowner may have a river passing through their property. This may give some rights to the landowner to use the river, but it does not allow exclusive control or the right to obstruct the water's navigation, and other users of the river may also have rights over its use as well.[73] Ostrom also states that in a commons the governance of the resource is carried on by a community that defines the rights of different people to access it.

In later work Ostrom moved her emphasis onto the application of commons for the management of intangible resources, including data and information. Unlike natural resources, knowledge and information are resources that are cumulative. In the case of data, the more that is shared, the higher the potential to link different sources and types of information to generate valuable new insights. Scarcity, in this case, can only be produced artificially by means of barriers to access such as siloing data or by intellectual property.

But in order for this to be sustainable, individual contributors need to be able define and enact fair principles over how that resource may be used. There are interesting parallels in the Free Software Movement, best known for creations such as Linux, Firefox, Apache and others. Access to the software is given to individuals entangled in a number different rights which the community decides upon. This can include access rights, or modifying rights to improve and re-share the software code. It might also involve restrictions over how the software is used, say, for non-commercial purposes only.[74]

In a data commons scenario (where monetisation of data is not excluded, but it is definitely not the goal), management by a community of users would exist by definition. For example, individuals that have consented to certain uses of data creating a valuable aggregated data pool could in turn leverage their collective bargaining power to sell anonymised data to a firm and only allow certain uses, for example, not allowing re-selling to third parties and making it available for free for researchers and government agencies.

In the dominant economic models that come from the 19th and 20th century, there is no place for ideas like the commons or the open source movement, since they are hard to define within the theoretical confines which define humans as fully rational and calculating agents. Ideas like Kate Raworth's *Doughnut Economics* offer useful recent contributions, seeing human well-being and the preservation of earth's critical systems as a starting point for a successful economy.

Flexible rules that give people control

A new conception of data ownership

DECODE aims to create a new legal and economic framework for citizens and public and private organisations to control their data, providing a means to manage who may access their data and for which purposes. DECODE aims to develop a set of intuitive, adaptive, context-dependent access rules, known as 'Smart Rules', which allow the use of data under specific circumstances.

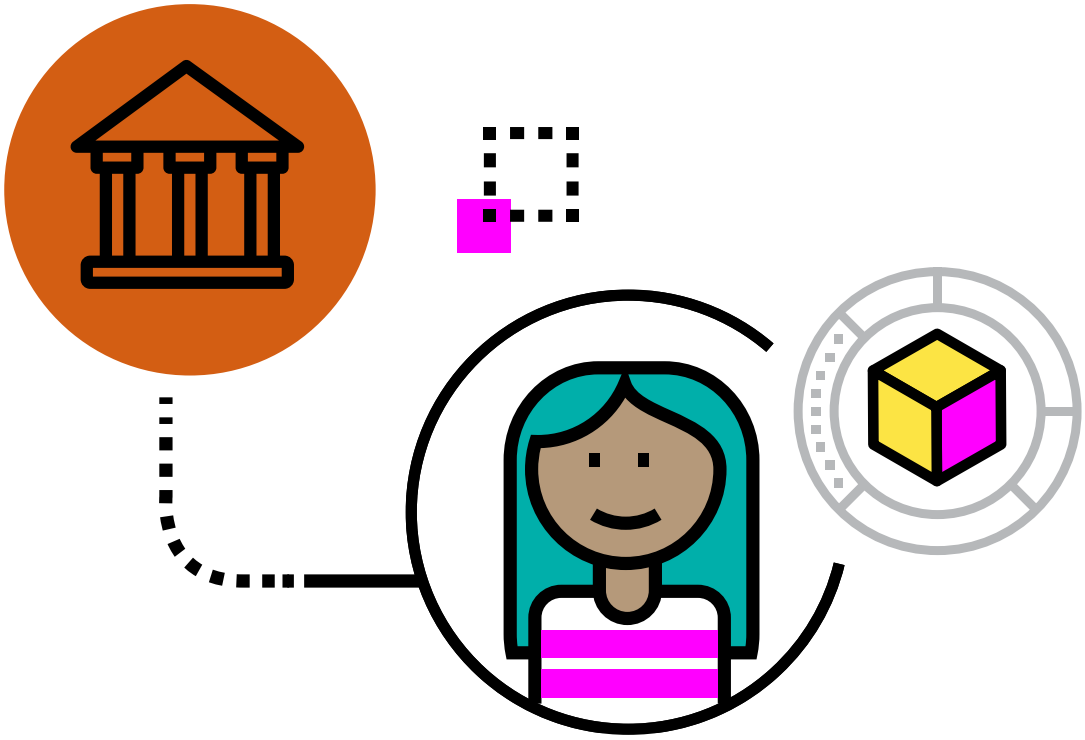
For instance, a citizen could enable access to location and other data to rescue services during a natural disaster, while in normal circumstances such data would not be granted. Or, access to personal health data would only be granted in the case of an emergency, or for a piece of research related to a specific health condition. In addition, citizens will be able to amend incorrect data and delete unwarranted data when legally possible.

A number of debates surround the issue of whether individuals should 'own' their data, and DECODE's 'Smart Rules' framework is a response to limitations with two contrasting sides to the argument.

Despite recent advances in data protection regulations in Europe, rules around collecting, managing and owning data in some ways tip the balance of power away from individuals. For many types of personal data (with the exception of personally identifiable information) copyrights and database rights are not naturally accrued to the person who it refers to. Instead they are owned by the entity that made the investment in infrastructure and effort of putting the database together. Furthermore, the new General Data Protection Regulation still gives database holders the freedom to use and release data, so long as it is properly de-identified. Irene Ng from the University of Warwick highlighted potential downsides resulting from this situation in a recent Nesta workshop:

Many data controllers strip data of all personally identifiable information. But using predictive analytics, they can combine data and then re-add identifiers with a number or ID. Once it's recombined, it's owned by the person who recombined it, not you.[75]

It is within this context that DECODE aims to give people the ability to express greater control over where there data is, how it is used, and who by. There is a long history of privacy advocates who have called for individuals to have the right to absolute control – akin to property rights – over data.[76] While this would seem to support DECODE's ambitions, there are some problems with this approach too. Specifically, some have argued that people may respond to having 'absolute property rights' by completely withdrawing all of their data from public use, even if this has damaging consequences for the rest of society.



This is well demonstrated in an article entitled *Tragedy of the Data Commons* by Jane Yakowitz, discussed in Section 1, who makes a strong case for anonymised research data as a common good. She gives examples of commons for academic research in the US, which have helped to uncover trends that would not be provable based on the experience of a single person. She cites examples such as census microdata being used to detect racial segregation trends in housing; public-use birth data leading to better understanding of the effects of smoking on foetuses; and public crime data being used to reveal the inequitable allocation of policy resources based on the socio-economic status of neighbourhoods.[77]

Yakowitz argues that if users were given the absolute right to control their data in these instances, the rational response would be for the individual to withdraw data completely to avoid the risk of identification. The analogy given is of a parent who has the incentive to withdraw their child receiving vaccination because there is a small risk of harmful side effects, even though only a few parents withdrawing will dramatically reduce the effectiveness of the overall process to deliver a public benefit.

This is similar to the view that data should be conceived of solely as a digital asset, to be exchanged or owned exclusively by those who create it. By encouraging exclusive property rights over data collected by individuals, we risk encouraging people to hoard and restrict its full potential.

An improved outcome may be achieved by allowing individuals to own their data in the first instance, but giving them user-friendly tools to share it, while specifying how it should or shouldn't be used. Currently there are a handful of legal tools which allow people to share certain types of data for public use, while retaining some control over its use (say, for non-commercial purposes). The Creative Commons has provided a range of legal templates allowing digital content producers to share their work for the common good, while the Open Knowledge Foundation's Open Data Commons project allows people to share datasets for unrestricted use. But according to Marco Ciurcina from the Nexa Centre for Internet & Society, these are not necessarily appropriate for the use of personal data:

These licenses do not deal with privacy issues. For example, you find in Wikipedia – which is a digital commons – there is a lot of personal data (articles that refer to people, contributions tracked by nicknames, IP addresses and email addresses). The use of free licenses does not imply that whoever reuses personal data in Wikipedia is free from the obligations provided by privacy laws and it's almost impossible to have consent and provide information to all the interested persons.[78]

In response to this challenge, DECODE aims to provide a much higher level of controlled access to certain types of data. Developers will be able to use these tools to create apps and services that offer users more choice, such as attaching specific terms to the use of data, or restricting access to it in some circumstances (where potentially sensitive data is involved).

According to DECODE consortium member Bruno Carballa from the Digital Commons Research Group, this can equate to something similar to property rights over data, but these rights can be modified, or reconceived as a 'bundle of rights', allowing data to be shared, pooled and re-used in certain circumstances.[79] In practice, this is a departure from notions of absolute ownership over data towards something akin to 'co-ownership' among those who have access. Co-ownership of data implies collective responsibility, transparency and permissioned delegation over the data's use and re-use, based on agreement by all stakeholders. Currently the tools allowing this kind of flexible permissioning over the use of data simply do not exist.

More concretely, Jeremiah Baarbé and other academics from the University of Ottawa have proposed a system of 'back-to-front' licenses for sharing data, which starts with an agreement between contributors and the collectors of that data. This is then accompanied by a distribution license – between the collectors and the consumers of the data – which would assure the standards of privacy, control and openness over the flows stated in the collection license are adequately met. These should be designed to be modular and easy to use, and clearly understandable.[80] In a similar vein, DECODE aims to show such a system of attaching specific permissions to data can be scalable given modern technologies. We go into more specific detail about the ideas and the tools that will lay the groundwork for this system below.

Technologies that flip data management on its head

DECODE will create technical architecture which allows users to decide where their data is stored, and who has access to it. Third parties will be built on top of this infrastructure, and they will not need to collect and store personal data in order to build useful applications.

One challenge that DECODE foresees is the unknown behavioural responses of everyday users, who will be given more choice over managing their personal data than they are used to. Therefore, DECODE's technological design will take place in close collaboration with end users, including local residents in each of the pilot cities, using agile and lean methodologies that learn by testing.

In DECODE's #MakingSense pilot in Barcelona, residents will be given noise sensors that are placed in the neighbourhood. DECODE will provide sessions to train and support participants to help them setup and use the sensors to gather and analyse data to influence city-level decisions. The pilot tackles the technical challenges of collating and storing a stream of citizen-sensed data, while also enabling those citizens to control what information is shared.

'Data portability' is a term which is often used to describe the ability for data to move freely between services. In particular, applications might be able to become much more lightweight, requesting access to data as and when they need to, and minimising the amount of personal information that needs to be stored or kept more permanently. In a recent Nesta blog, Reuben Binns stated the value that data portability could offer:

Wouldn't it be great if, regardless of which organisation I'm dealing with, there could be a standard way for me to express my preferences about how I want my data to be used? ... Similarly, what if I want to take my data out of one service and use it in another? Wouldn't it be great if there were a standard way to take my data - messages, media, contacts, behavioural trends - with me, and share or import them into another service? [81]

New European regulations may help to shift incentives towards greater portability. The GDPR - to be be implemented in May 2018 - requires organisations collecting personal data to provide a number of specific references, including who has access to the data, for how long, and for what purposes (Article 30). In response, many companies may begin to see that holding any personal data at all could be more costly than it's worth. Other provisions (Article 20) give individuals access to the data they have provided in a structured, commonly-used format, as well as the ability to extract and move that data to different services.

A number of personal data stores and PIMS are trying to push towards this goal. The Hub of All Things (HAT) project is building the infrastructure for a new digital exchange of personal data. It provides a database where users can congregate all their data, from browsing history to that produced by smart objects in their homes, but it also creates a platform where different bodies can buy and sell data. They envision a future where companies don't store personal data,[82] since

this is costly and risky for them. Instead, they sync their databases to each individual's HAT so they can access our data and pay individuals in return. The project is primarily focused on rebalancing the economic power between individuals and data-driven companies. Therefore, third parties may still have access to personal data - users only gain a copy of their data, and there is no specific emphasis on securing user privacy.

Another ambitious project called Solid (SOcial LInked Data[83]), developed by a team at the Massachusetts Institute for Technology (MIT), is attempting to encourage a whole new set of applications. The project has created a set of new technical standards that enable modular data-sharing agreements, giving users detailed control over the information they share and with whom. By linking user data across different applications, Solid aims to achieve a high degree of portability for user data:

With Solid individuals get the ability to use any number of services and decide what kind of data goes on each of these services. When you use an app, you basically bring all your history with you, in terms of what kind of brands you like, or what kind of products you like ... What Solid gives developers is this ability to not have to worry about user management, or not to worry about storage management.[84]

However, a key challenge will be whether people are ready to take control over their data, especially if this means implementing solutions with new (or more complex) user interfaces. It is also likely that giving people control over their data will increase the daily burden on people to make choices about how their data is stored and managed. As Andrei Sambra from Solid continues:

One of the challenging things of putting people in control is that they have to make all those decisions themselves. So you have to give them the right tools to make their lives easier, not more complicated. That will be also a very important thing.[85]

One of the problems is that many of these projects are still in their early stages and have very few live applications currently.

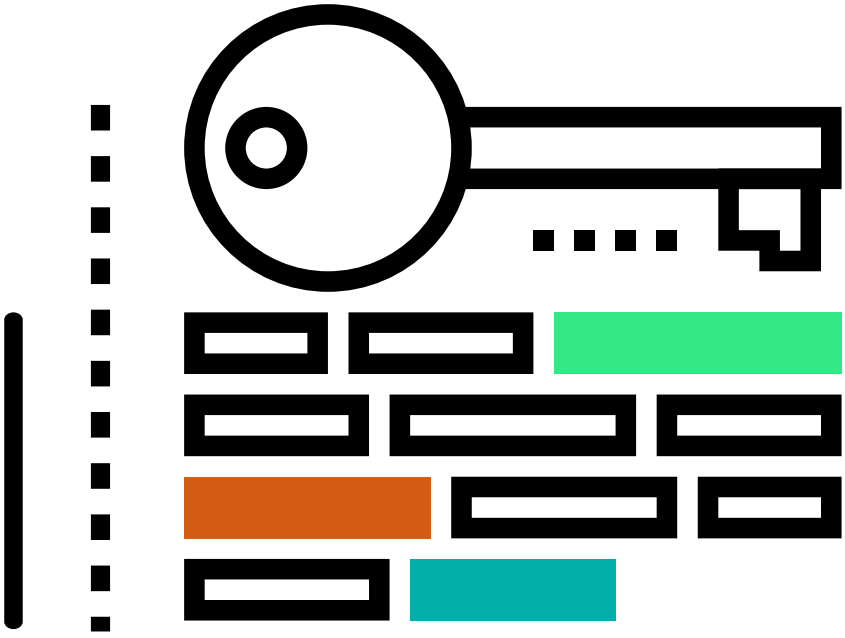
DECODE will contribute to the development of standards in this space by developing and designing solutions both for and with the end users. It will run pilots which test new technologies, bringing together developers, civil society actors (citizen organisations, digital rights advocates) and ordinary people to develop and pilot the tools following lean and agile methodologies that learn by testing. Finally, while DECODE shares some of the characteristics of Solid, it will have a much higher emphasis on user privacy and on transparency over data flows, as we describe in the following sections.

Greater privacy when sharing data

DECODE will define a number of machine-readable 'Smart Rules'. These rules will build on an emerging set of cryptographic techniques which allow users to selectively disclose information about themselves to service providers that are part of the DECODE ecosystem, in turn allowing them to prove that they meet certain access conditions without revealing unnecessary personal information. The first pilot for the use of this technology will be for anonymous user verification on Barcelona's digital democracy platform Decidim.Barcelona, based on a single anonymous verification process using the city council's census database.

Beyond anonymous verification, the pilot will also enable options for citizens to gather and share personal data from a range of sources. Citizen generated data will be aggregated and displayed in a 'BCNow' dashboard, and will include data that is donated to the public by individual citizens, such as, noise levels in a community, or healthcare data. Data from the Barcelona City Council Infrastructure (Sentilo, IRIS, CityOS) will be available on the dashboard as part of the Data Commons policy in Barcelona. Smart Rules will also enable users to express certain conditions under which their data may or may not be used. The data may be kept private, or used to inform a range of specific policy proposals which are created by citizens on the digital platform. This will allow citizens to actively involve themselves in local issues while remaining in full control over how their responses are shared.

The notion of privacy over data is increasingly difficult to maintain in a world characterised by rapid growth in networks, information sharing and processing power.[86] Some degree of trust in the data controllers is still a necessary precondition to sharing data with those actors. For instance, experience from Digital Rights Management (DRM) has shown that data can easily be copied after its output from the DRM system, bypassing any original restrictions on that data and quickly leading to the emergence of secondary markets which are difficult to audit or regulate.[87]



As a result, it is sometimes said that 'big data' and 'encryption' may be antithetical to one another in the modern information age. Big data involves large scale storing and processing of information to reveal patterns, whereas the cryptographic techniques have usually been employed to obscure or hide that data from prying eyes.[88] Yet there are promising new technologies which may enable users to achieve the best of both worlds, we describe some of those which are relevant to the DECODE project in this section.

One example comes from the MIT's OpenPDS project, which is a personal data store that allows algorithms to run on a user's device or trusted server or between devices, but without a central organisation ever having to collect any data. The project uses a mechanism called 'SafeAnswers' to asks questions and gets answers from individuals by *"sending code, not data"*[89] MIT's Yves-Alexandre de Montjoye describes it:

Our idea was that instead of sharing fine-grained behavioural data with 15 different services, each of whom record it and use it to provide you with a different service, the user holds the data and get these companies to ask questions. They send pieces of code that run specific questions ranging from music preferences, to how much time it's going to take you to get home tonight, to anything else, within the safe environment of your personal data store. The code is validated there, and then only the answer is sent back to the third party.[90]

The more recent Enigma project from MIT combines this SafeAnswers mechanism with new techniques in secure multiparty computation. This is a promising area of research which works by taking user data and splitting it up into unreadable encrypted chunks, which get processed by hundreds of other computers across the network. No raw data is ever revealed, only the final results of the computation are visible to participants. Imagine for instance that ten people want to find out what the highest earned salary is between them, without revealing to each other their individual salaries. Scale this up and it might be possible to allow thousands of people to submit personal health data for research by a pharmaceutical company, or allow a company to use purchase data in return for personalised shopping recommendations, without anyone other than the owner ever needing to see any unencrypted data.

Other promising cryptographic techniques in this area that are relevant to the DECODE project include technologies for the selective disclosure of data, such as Attribute Based Credentials. These techniques allow verification of a person to take place, without revealing any more information than is strictly necessary for the use of that service. Individuals can therefore be authenticated to access or use a service, without ever needing to be identified as anything other than 'over 18' or 'a resident in this city'. This may lead the way to much more effective data minimisation when sharing among different user groups

Functional cryptography is a broader field which encompasses similar techniques.[91] A system can be created which defines permissions or attributes which can be attached to data's use. Those who make requests will only be able to access a particular function of the encrypted data if they meet specific requirements. These may be chosen in a license created by the user (say, only for use by research organisations, or only for access at a specific time or place). Such permissions can be attached to specific pieces of data too. If the person or service asking for their data starts issuing requests for certain combinations of rows or columns that were not granted in the license, access can be denied. This area is of particular interest to the DECODE project, particularly for its ability to aid in the creation of cryptographically secure Smart Rules.

What many of these technologies show is that data-driven business models may be able to flourish, generating valuable insights from user data, yet users could have the assurance that companies or organisations never need to see or touch more data than is necessary, or even see any of the data at all. Giving this assurance to individuals might also reduce the incentive for people to 'lock down' their data entirely, which may address the unintended consequences of a giving people full control outlined earlier by Yakowitz.

Distributed ledgers to enforce decentralisation and individual control.

DECODE recognises the potential for distributed ledger technologies, like Bitcoin, to help enforce rules ensuring that the power over the exchange of data remains in the hands of individual users. At its core, distributed ledger technologies ensure that no single actor can manipulate the process of record-keeping.

In the case of DECODE, the ledger will be made up of the permissions which users attach to their personal data as Smart Rules. By storing these rules in a public distributed ledger, the Smart Rules will be highly transparent (in terms of showing where data is and who has had access to it) as well as tamper proof.

DECODE will use further cryptographic techniques to ensure that it is able to balance a high degree of transparency over data while preserving the privacy of those individuals that the Smart Rules refer to.

There have been a growing number of calls for a decentralising or 're-decentralising' of the web, including from the founder of the web itself.[92] Many advocates are placing a lot of faith in new technologies to enforce this. A number of these - including Tim Berners Lee's own Solid project - have already been mentioned. But a recent phenomenon, known as distributed ledger technology, is particularly interesting for its apparent ability to enforce new distributed power structures.

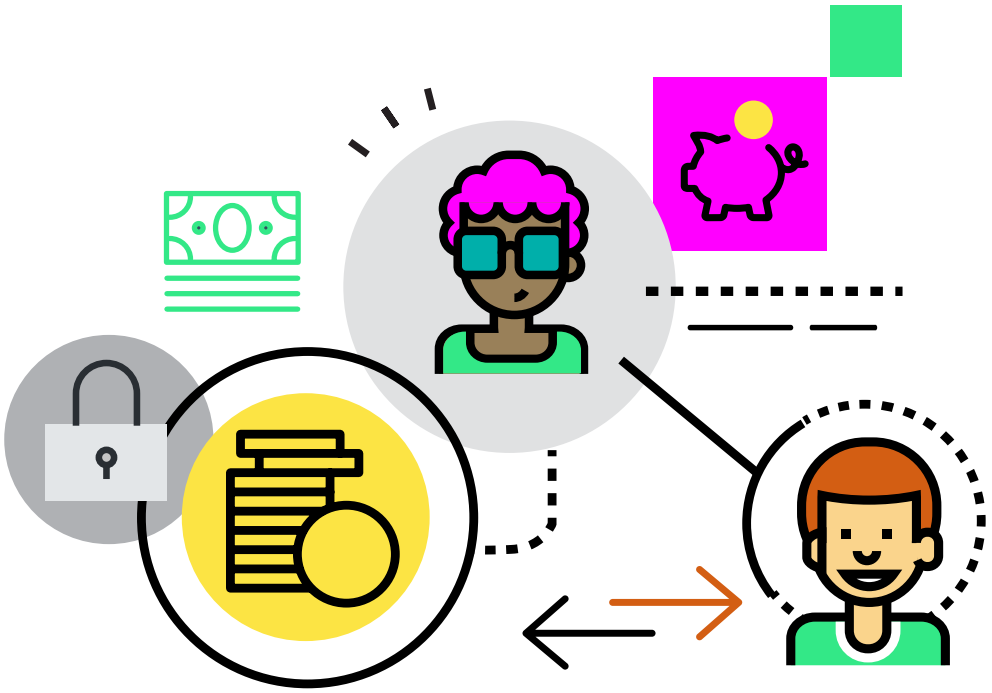
A distributed ledger is technology enabling a diverse set of untrusted actors to agree on a single record of events. One of its most influential implementations has been in the digital currency Bitcoin, a peer-to-peer method of exchanging digital money that removes reliance on a trusted intermediary like a payment processor or a bank. This particular implementation of distributed ledger technology has been referred to widely as the 'Bitcoin blockchain'.

No single entity owns or controls the maintenance of digital records which underpin Bitcoin. Thousands of computers across the network process and store the transaction data. Because of this, it's very difficult for anyone to erase, censor, or tamper with the data without everyone else in the network finding out.

In short, Bitcoin is an attempt to achieve a decentralised network, a system of exchange with no single locus of authority. This means that all responsibility over the ownership of Bitcoin is left to individual users. No customer service, no regulatory oversight - whoever holds Bitcoin is the sole bearer of all ownership and responsibility over that asset.

It's said that distributed ledgers' key characteristics could provide a foundational protocol for a fairer digital identity system on the web.[93] Beyond its application for digital currencies, distributed ledgers could provide a new set of technical standards for transparency, openness and user consent, on top of which a whole new generation of services might be built. More specifically they provide:

- An audit trail of ownership: For example, all Bitcoin transactions, including individual senders and receivers, appear on the distributed ledger as ever growing list of alphanumeric gobbledygook - lines of text known as cryptographic 'hashes'. With this, the technology balances a radical approach to transparency with user pseudonymity: anyone can check the record to verify that their money has arrived at the intended recipient's wallet, but the record provides just enough anonymity that people can use Bitcoin without being easily identified.
- Data integrity through decentralisation. Distributed ledgers achieve integrity by distributing the network's processing power and storage across tens, hundreds or thousands of computers. This is primarily about trust. By removing any single point of failure, distributed ledgers demonstrate to their users that no single authority can change or tamper with the underlying rules. In Bitcoin, if someone wants to alter or bring down the network, they need to acquire 51 per cent of the processing power of the entire network. Many distributed ledger applications are following a similar model, forming 'consortiums' which collectively maintain the network, and embed rules like majority voting over how data is recorded.
- Open innovation. One of the remarkable things about Bitcoin is the fact that it is totally open source. Bitcoin has inspired a new generation of open source enthusiasts, who have been able to freely build innovations on top of the network, and in some cases provide their own value added services for Bitcoin users. Most importantly, anyone can inspect or scrutinise the core Bitcoin code, offering suggestions, improvements and monitoring its operation.



A number of examples in the field of digital identity are already drawing inspiration from Bitcoin and distributed ledgers more generally. Blockstack, for instance, is creating a global system of digital IDs, which are written into the Bitcoin ledger. Nobody can touch them other than the owner of that ID. Blockstack is building a new generation of applications on top of this infrastructure which promises to provide a new decentralised internet where users own their own data, storing it wherever they like, and apps run locally.[94] The distributed ledger provides a more secure means to anchor things like identifiers and routing data than traditional systems that rely on trusted central parties. For instance, the traditional naming system for the web (which helps direct users find the right servers when they type in a URL) is maintained by a central organisation called ICANN, whereas the Blockstack name registry is run in a decentralised way, deriving its route of trust from the Bitcoin blockchain.[95]

Another new example is Sovrin, which attempts to provide users with 'self-sovereign identity'. The argument is that 'centralised' systems for storing personal data make it a 'treasure chest for attackers'. Sovrin argues that users should more easily be able to have 'ownership' over their data, and the exchange of data should be made possible through a decentralised, tamper-proof ledger of transactions between users.[96] DECODE takes inspiration from this model, but focuses less explicitly on identity management.

More broadly, DECODE will contribute to the development of standards in this space by developing a distributed data management architecture for the ownership and flow of data. Following a similar model to Blockstack, data permissions chosen by the user (rather than the data itself) will be stored in the distributed ledger. In DECODE these permissions will be referred to as 'Smart Rules', and they will act as a registry that tells authorised actors where to look in order to access someone's data, and under what conditions the data may be used.

As mentioned above, one of the most important innovations of Bitcoin's technology is that it balances both pseudonymity with full transparency over all transaction data. Although this provides a certain level of privacy for Bitcoin users, it has been shown that under certain conditions users can be re-identified.[97] This reflects a particular challenge for DECODE, which sees the benefits which distributed ledger technology offers as a transparent and decentralised database, but aims to create privacy enhancing tools for the exchange of personal data.

However, thanks to advances in modern cryptography - in particular a family of techniques known as zero knowledge - it is possible to add additional layers of privacy onto distributed ledgers. This is where DECODE's experimentation with additional anonymising features, such as Attribute-Based Credentials and Attribute-Based Cryptography, will help verifying that a transaction was correctly performed on a ledger, without revealing any sensitive information that could be used to identify a person.

A new kind of digital platform

Trialling new platforms for sharing data

DECODE will experiment with new models for promoting the management of common resources through digitally enabled co-operatives. These will be made compatible with the DECODE Smart Rules infrastructure. Where possible, effective data sharing should be further underpinned by high levels of user awareness and trust, and platform co-operatives achieve this by embedding openness, respect for individual users' privacy, and democratic participation over how decisions are made.

One of the initial pilots for the DECODE technology will be FairBnB, a platform co-operative in Amsterdam that will enable renters, guests, hosts and neighbours can collectively decide together with municipalities how to make the rental process fairer and more sustainable for local residents. The pilot aims to make it easier for accommodation providers to rent out their properties, while abiding by local laws and maintaining control over who can see their occupancy data. DECODE will provide statistics and regulatory information without compromising participants' privacy.

At a higher level, it is appropriate that DECODE strive to embed the values of the EU's Responsible Research and Innovation (RRI) agenda, which provide a valuable framework for big data applications with personal or sensitive data. The model focuses on education, governance, ethics and open access to results, but importantly it also tries to acknowledge the unpredictability of outcomes. In this framework, anticipation, foresight and privacy-by-design strategies should inform actions within a broader attempt to foster collective responsibility through public engagement with the participating communities.[98] DECODE's co-creation framework has been designed as such.[99]

One of the key objectives of DECODE will be to balance both high levels of user-privacy while also encouraging new, innovative applications of data. Indeed much of the opportunity of data will only be realised where individuals are able to pool it together to leverage its economic and social value. DECODE will need to experiment with broader strategies around data governance; particularly how it might be ethically pooled and managed to suit the requirements of all stakeholders.

In the information age, 'platform co-operatives' - which highlight the potential of digital technologies to help workers govern themselves - offer a promising vehicle for the management of personal data. The term was first coined by Trebor Sholz, to refer to the potential for digital technology to help foster new kinds of co-ownership and democratic management over common resources (whether they be labour, services, data or digital platforms themselves).[100] While traditional worker co-operatives are characterised by individuals contributing labour in return for co-ownership in the business, there are interesting parallels between labour and data which were expressed by some of our interviewees:

Ultimately we need to think about data as digital labour – we could then think about people contributing and appropriate remuneration. Data as digital labour would also imply that there's the potential to be exploited or discriminated against, such that appropriate safeguards are needed, but it would also point towards the potential for people to work together - whether for common cause on a philanthropic basis or for payment and profit.[101]

I see data as a factor of production in the new era.[102]

Platform co-operatives like the The Good Data collect, pool and sell users' internet browsing data, but they do so entirely on their members' terms. In order to build user trust the platform anonymises as much of the data it collects about its users as it can. This is done manually on a case-by-case basis, depending on who is requesting access to it and why. The co-op also refuses to collect any information that contains sensitive terms, say, from browsing explicit websites, health websites or political websites. The money made from the sale of the data is split between social lending in developing countries (chosen by members), and on improving technical development of the platform. Any co-op member can participate in deciding these rules, either through discussion on the site's collaborative platform, by attending general meetings, or by standing to be elected as a Company Director.

This offers a stark contrast to many new platforms which more often extract, store and use data at the expense of their users, or the democratic will of the communities in which they operate. Sharing economy platforms such as Uber have come under fire for using data they collect to manipulate user behaviour, from using psychological tricks to make drivers work longer, to claims that the company hikes the price of a ride when user's batteries are running low.[103]

Other platforms such as Airbnb have caused controversy in some cities by contributing to upwards pressure on rental prices and making it difficult to implement fairer regulation by concealing information about rule-breaking Airbnb hosts from local government.[104] Other controversial features include Airbnb's 'instant booking' process which removes the ability of hosts to vet prospective guests. Hosts do not have to use this function, but the platform strongly encourages them to do so. This is a benefit for the platform because it increases the number of bookings and therefore revenue, but it is not necessarily good for hosts. The difficulties created by Airbnb in Amsterdam are the starting point for one of DECODE's pilots. Using an online platform, DECODE will work Amsterdam's 'Fairbnb' community to build a democratic alternative that allows residency data to be shared with the municipality and the community without compromising users' privacy.

There will be new challenges in experimenting with new models like this. In co-operatives, the process of adding economic value in economic sectors such as industry or agriculture has been established over centuries. In the data economy the process of managing sensitive data, as well as finding new ways to innovate ethically with data is much less well understood. These are decisions which any platform will have to tackle head on, and a clear challenge will be how ordinary people are able to grasp the nature of these issues.

Revenue generation and incentives for participation

Economic models

Giving people full control over their data requires a different model to the current advertising based and surveillance revenue model. Gaining a significant number of users will mean that DECODE needs to reach out to a wide range of people, beyond those who are already active in digital social innovation communities. Assuming that only a small minority of people will ever care about the economic arguments or the technology, these commons and platforms will live or die by the extent to which they offer compelling benefits and use cases for individuals, civil society, government and business.

Unlike some PIMS, DECODE will not incentivise participation by offering cash rewards in return for the exchange of data. Instead, DECODE will provide applications which explore the collective benefits from data being pooled, shared and used by a range of different stakeholders, built on top of an ethically sound infrastructure that puts individual control at the centre. In the longer term, DECODE sees that this will unlock a wide range of revenue generation opportunities, where individuals can come together and agree to create new use-cases for their data, both for themselves and for others who agree to preserve and improve the data as a common resource.

Some of the most successful projects in fostering commons for knowledge and data, such as Wikipedia and OpenStreetMap have raised money through donations and annual membership fees respectively. OpenStreetMap has also sold products related to cartographical data collection with a number of retailers (such as for GPS services and books), demonstrating how commons-based crowdsourcing of data can create business opportunities for third parties.

For personal data, much of the opportunity will also be realised where individuals pool their data together to leverage its potential value, both for the individuals themselves and interested third parties. In this regard, co-operatives are a possible form that commons-friendly market entities could take. They allow people to both contribute their labour and, as co-owners and contributors to a common resource, generate livelihoods or useful services back in return for their participation. This also implies the possibility that members can collectively bargain with their data, both in terms of money and, more importantly, the conditions under which the data can be used and for what purposes. As entrepreneur William Heath puts it:

I do think setting the rules about how we should act collectively with our personal data could work. So as well as "I permission that data of mine to go into an open data commons, do what you will with it" we need "I can exercise a right, today and in future, to act with other people collectively, whether it's to express a democratic will, to support a cause, to switch from one supplier to another, or to share specific medical data for a specific research purpose in a very controlled way". I'm more focussed on enduring agency, control and responsibility, and not just throwing it over the wall. Above all, I believe that retaining provenance and control will create more value.[105]

A practical example are health knowledge commons, which facilitate the sharing of patients' knowledge about themselves, as well as personal data (whether sentiment, sensor data or genomic data), in order to generate new insights about diseases, diagnosis and treatments.[106] According to Dr. Ernst Hafen, creator of midata.coop, the individual members of the health knowledge commons should be given the tools to securely store and manage who has access to their data. By forming member-owned co-operatives, the members can leverage the collective power of their data, which researchers and pharmaceutical companies will pay for, while attaching specific conditions for access such as secure storage, ethical research practices, and so on.

Let's say a drugs company comes and says 'we want to recruit 1,000 patients with this specific profile', and the patients are then asked for their consent, and their data is transferred through a secure link to the clinical trial servers … We say to the company that this costs €10 million, on the condition that after the conclusion of the trial you get your data that was recorded during the trial back as a copy of the data bank. And we insist that whether the trial was positive or negative, that the data will be copied. With this, we now have the power, we're sort of a union in the 21st century, and this is also the way we can generate revenue that then the members of the co-operative have a say in how this revenue is spent.[107]

According to Dr. Hafen, who founded the platform as a non-profit, one of the principles of income generation is that any revenue made is reinvested back into the preservation of the platform, to avoid the capture by specific interest groups or shareholders. This follows the idea of 'reciprocity', used in Creative Commons. Anyone is free to tap into the common resources to create their own new services, but in return they must meet certain requirements that are conducive to, or provide direct contribution to, the preservation of the commons ('share-alike').

Another alternative was explored by Our Data Coop project, a research project co-funded by the UK government, to explore whether the 'intelligent sharing' of personal, organisational and public open data between third sector organisations could help improve their collective advantage for winning public procurement contracts. The ultimate aim of the project was to see whether non-profits could become 'ethical data-driven impact investment vehicles', helping to give them an equal footing with the private sector in the evidence-based policymaking and commissioning space.[108]

Projects like Our Data Coop try to demonstrate that a commons-based approach will generate increased value for a community as it grows. If the community is able to achieve a critical mass, then users, businesses, organisations and researchers will have a higher incentive to contribute to and make use of the commons because they will receive a higher value from that data for themselves. A similar proposal has been made for a data commons in New Zealand, which proposes that individual contributors submit a whole range of data from different services and connected devices to a trusted, democratically managed resource. In turn, everyone can benefit:

If Samsung stays off the commons, its heartbeat sensor will never be more than just a toy. But if a New Zealand entrepreneur's heartbeat sensor is hooked into the commons, it can be joined with people's medical records and what they eat from their supermarket shopping basket and directly shared with their general practitioner to become a complete personal health solution … These kinds of 'economies of scope' mean that you'll be able to design better services for your customers through the high-trust use of their integrated data rather than staying fragmented and in control of siloed data that is of lower value to everybody.[109]

A researcher or a citizen can add their data to a shared pool and get more back in return, whether through access to new data, or integrated data-driven services. This does not mean there is no opportunity to create financial rewards from using the commons. The New Zealand Data Commons Blueprint suggests, for instance, that re-users may make money from innovations (such as with 'apps' that rely on commons data), but that this may be taxed a portion of the sale, which would then be redistributed back to the participants of the commons.

In sum, the incentives are structured so that use of the commons also entails its preservation and improvement, which in turn will offer better insights and opportunity for personal, scientific, economic or social benefits. This opens up a world of opportunity to link wide-ranging information together from the wide array of different individuals and apps that submit data to the commons, all based around principles of user control, co-ownership and responsible management. This will eventually provide a much more powerful and attractive option than fragmented data, inferred from people's behaviours and kept in silos without their knowledge or consent.[110]

Measures for building momentum

DECODE will target local use-cases in the cities of Barcelona and Amsterdam, making it easier to build a critical mass of users. One of these will be a collaboration with the neighbourhood social networking site, Gebied Online. As a starting point, DECODE will provide granular privacy controls, so that residents can decide what information they share, including how they may choose to collectively put their data to good use in the community (this could be to help organising local events, pooling volunteers, sharing DIY tools, and so on).

The project will also partner with local government in the two cities, to create hackathons, challenge prizes and other public events to raise awareness. Finally, DECODE technology will aim to create a brand around ethical use of data by creating 'trust badges', which build an awareness and leverage consumer's desire for fairer data practices. In what follows we provide a more detailed overview of each of these strategies.

If the benefits of democratic control over data can only be realised when enough members are present to built a critical mass, then how will users be incentivised to join in the early stages? There are a number of measures which may be taken to help kick-start more ethical alternative platforms. These include keeping a local focus, policymaker support and even social marketing or branding to incentivise fair practices.

Targeting local and community-based use cases

The Good Data has tried to build a data commons by collecting its users' browsing data, and giving them co-operative ownership over who may access the anonymous dataset and for what purpose. Founded in 2014, their website states that they have so far earned a total of $1,760 from the sale of user data, and there are around 250 monthly active users. Founder Marcos Menendez describes some of the difficulties he has faced related to scale:

The problem is that we don't have a lot of data. Let's say that, for instance, somebody wants to use the data to do some research, and the first thing they are going to say is 'I need that segment of the population so I need men, living in UK, in cities larger than 500,000, that are this age'. And when you start cutting and cutting, you end up with a small group of people.[111]

Other use-cases which aim to collect and pool data at a local level have more easily achieved a critical mass of users. For instance, Gebied Online is a platform co-operative in The Netherlands, which began when founder Michael Vogler created a website for his Amsterdam neighbourhood of IJburg. Now the platform acts as a local social network for 4,000 active users in the neighbourhood (and has been rolled out in 15 communities in total), who share a wealth of information on the platform including local news, events and meetups, sale of products, and so on. Community members are co-owners in the platform, which is non-profit and does not sell customer data, creating local networks that *"together create value with (and in) their own online platform"*.[112] Part of the success of this platform may have also come from the platform owners using their good knowledge of the local community to get the initial value proposition right, and the benefits of word of mouth among local residents.

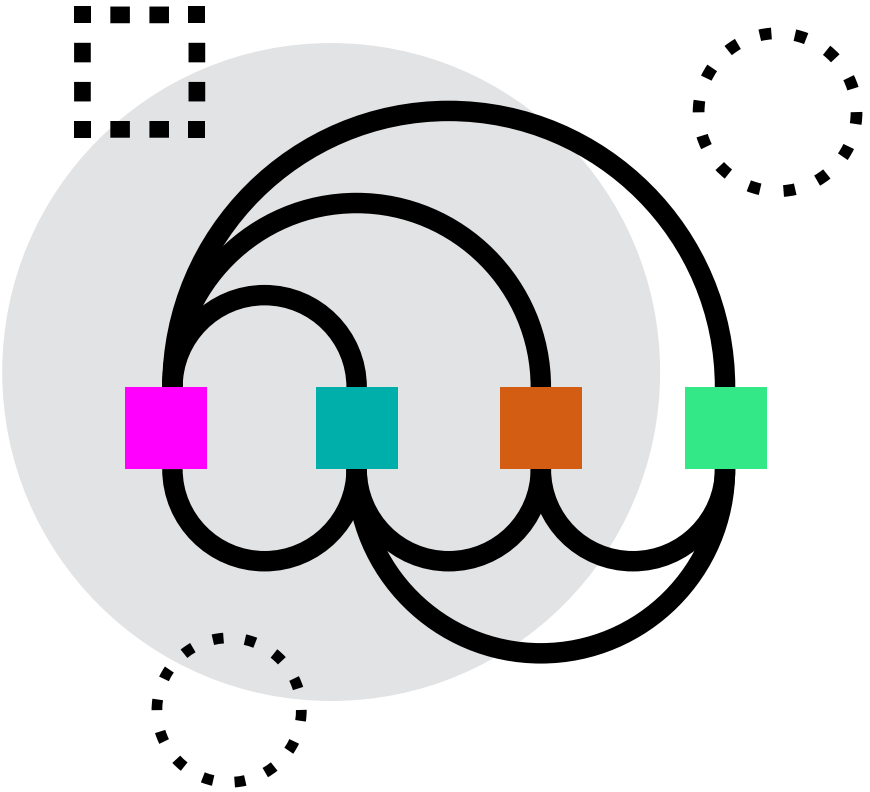
This platform has demonstrated its capabilities and been a success with residents in a number of cities and neighbourhoods, and DECODE will help to build on its successes so far by running a series of new pilots with residents. Now Amsterdam City Council is keen to spread this to other locations across the city and leverage the platform to increase involvement with policy and decision-making.

Policymaker support in two cities

DECODE is an experiment to test new, alternative models which begin at the local level. In this regard, the project will build on a partnership with Barcelona and Amsterdam city councils. The cities were chosen for their close links to active digital social innovation communities in makerspaces and innovation labs across the two cities, and the coordinator of the DECODE project is Francesca Bria, Barcelona CTIO. In terms of policy initiatives, the city councils will host hackathons, challenge prizes, regular meetups and large-scale events and conferences to engage a range of different stakeholders, and to raise awareness about the pilot use-cases.

Building an ethical brand

Once the technology itself has been built, a further incentive to attract support will be created by way of DECODE 'trust badges', which will be publicly viewable on any DECODE website. DECODE trust badges will be used to build a brand around the ethical management and use of data. Different service providers who wish to have access to the DECODE data commons will need to meet a set of standards around ethical data practices in order to attain their trust badge (such as open source, open and transparent standards, and so on). This builds on the success of social certification schemes such as Fairtrade and Fair Trade Music which place a premium on ethics and trust in the marketplace while leverage the consumer's desire for fairer practices.[113]



Conclusion

The problems with the personal data economy, and connected to that, the economics of internet are profound. Without a radical rethink about the future internet we want, we risk the continuing erosion of privacy, autonomy and digital rights, distorted markets dominated by a handful of monoliths, and a society deprived of the collective intelligence it needs to be democratic, sustainable and equitable.

DECODE will develop practical tools that address one aspect of this - the control of personal data and the ability to share them as commons. From these tools, a fundamental change in how people interact with the internet could occur. We hope this will be a democratising force, giving people a stake in how the internet economy functions. There will be other things required for the shift to occur - regulatory and governmental support, and a shift in public attitudes and economic incentives in particular - but creating scalable digital tools that work for people will be a major factor.

The path to this change is not straightforward. There are considerable technical, legal, ethical, economic, and practical challenges that DECODE will need to overcome. There are also external factors, such as the pace of technological development and social attitudes, which are beyond DECODE's control and are hard to predict.

DECODE is committed to tackling each of these challenges in a bottom-up and citizen-centric way. Our ambition is to build tools that respond to genuine problems experienced by citizens across Europe. This report sets out the vision we hope will come from this, and how we hope to get there.

This report coincides with the point in the project at which the projects are taking shape and early developments of the technology are emerging. At this juncture, there are some key messages for some of our audiences:

- **European Commission and national policymakers**: DECODE will be one of the first experiments in personal data control backed by major city governments. The learning from this process will in all likelihood involve lessons for national and supra-national policymakers about legislative and policy support that is required to enable this agenda. Nesta's next report, due in Summer 2018, will focus in particular on this issue.
- **Cities and local government**: the tools created by DECODE will be open source and free to use. In order for their use to scale, and for the highest impact to be achieved, the support and engagement of cities and local governments will be required. As the project progresses, we will be investigating how cities and local governments can best support DECODE initiatives, and how they can in turn can get the most value from them. Future reports will provide practical advice about how to adopt DECODE technologies.
- **Citizens**: there is a chance to be involved in this work. Citizens of Barcelona and Amsterdam will have many opportunities to be involved, and we hope to reach other European cities through events and scaling activity over time. Follow the DECODE website (decodeproject.eu) for the latest information on how to get involved.
- **Innovators and entrepreneurs, civic hackers and data scientists**: there will be a range of events such as hackathons, challenges and summer schools designed to engage potential developer communities in the opportunities presented by DECODE. For the full social value of data to be unlocked, engaging with these communities will be essential.
- **Business**: DECODE will create an infrastructure which businesses can build on. DECODE's aggregated data could help lower some barriers to entry, and insights latent within the data could spur new products and services with more sustainable, fair and privacy-preserving business models.

Appendix 1

Project methodology

DECODE is a multidisciplinary project which will use a range of methods to deliver on its objectives. DECODE's philosophy is to ensure that the DECODE architecture, requirements, and designs emerge from grounded bottom-up community requirements. The project will operate with a lean and agile methodology that builds iteratively and learns from feedback throughout. The methods we are using in the project to achieve this are:

- **Policy reviews** - academic research into some of the core policy areas related to DECODE, such as the collaborative economy. These will be used to create theoretical frameworks which guide both the design of the technology and the focus and delivery of the pilots.
- **Inception exercises** - These were used to gather the social requirements for DECODE pilots. Inception workshops were held in Barcelona and Amsterdam using a lean methodology, including user research with user stories from the identified lead users' communities, the formulation of a main hypothesis and active experiments. This led to the selection of the four pilots. There is more information about the inception exercises in DECODE deliverable D1.1.[114]
- **Co-creation** - DECODE will co-create its tools and pilots with end users. This is important to the project because giving people a role in creating tools helps to increase their usefulness. It also enables the development of a community of interested potential users, and networks through which the DECODE tools can be disseminated. DECODE deliverable 6.5[115] describes in more detail the approach to co-creation to be taken in DECODE.
- **Technical design and build** - DECODE's technical partners will begin by designing a distributed technological architecture to run applications developed by third parties. This will include a smart rule language that can be graphically represented, and a GNU/Linux based operating system that can execute signed smart rule applications. Further platform components are necessary to deliver the project and the underlying technical architecture will be determined through understanding the user requirements that emerge from the use cases. They will then begin to prototype technology, before testing it during the pilots. There is a weekly technical stand-up meeting in which the consortium comes together to review progress, solve problems and agree next steps. The project will aim to build on other open-source technologies, and will work in an iterative way.
- **Pilots** - four pilots will test the DECODE tools and assess the impact they can have. DECODE's four pilots are:
  - iDigital/BCNow Platform
  - IoT Pilot involving #CitizenSense
  - Holiday Rental Register/FairBnB
  - Gebiedonline (Neighbourhood Online)
- **Evaluation** - we will use evaluation to assess the economic and social impact of the DECODE pilots, integrating feedback from the users and from the fieldwork.
- **Dissemination** - we plan for DECODE to continue scaling after the project ends. To do this we have a dissemination strategy. This combines outreach, events, communications with open source tools. The project will employ challenges, hackathons, summer schools and training programmes to disseminate its results and tools. DECODE is creating a form of free and open source license for its tools, and for services and products which are compliant with DECODE's philosophy. This will create an ecosystem of organisations supporting DECODE, helping the tools to spread.

Appendix 2

A brief review of projects giving people more control of their data

Through this research we came across a number of organisations and projects with similar or complementary aims to DECODE. These projects have been summarised below. This is not intended to be an exhaustive list but is a starting point which DECODE or others could build on in future.

Name of project	Type of project	Distributed ledger?	Project aims
Bithouse	Platform co-operative	Yes	A decentralised contracting proposition for renting property. Will use Ethereum.
Blockstack	Self-sovereign identity	Yes	Decentralised naming and storage. Currently uses Bitcoin. Transactions in the blockchain contain simply a pointer to the data and its hash. Majority of activity happens off-chain.
Consent	Blockchain/ 'Consent tech'	Yes	A trust protocol to independently authenticate identity and selectively exchange personal information. Will be on Ethereum blockchain. Developing simple APIs and machine readable standards for consent.
Databox	Personal data store	No	An open-source personal networked device, augmented by cloud-hosted services, that collates, curates, and mediates access to an individual's personal data by verified and audited third party applications and services.
Datacoup/ Give with data	Personal data store	No	A proprietary platform to capture data from third party providers to generate better insights. Uses third party APIs to collect data from various sources, anonymises them, then sells on to brands/ marketers/charities and hands value back to user.
The Good Data	Platform co-operative	No	Allows people to sell their browsing data and donate the profit to good causes. Members are co-owners of the platform and the data they submit. The service is a browser plug-in which blocks third party trackers, pools and anonymises the data, before selling it

Name of project	Type of project	Distributed ledger?	Project aims
Hub of all things	Personal data store/exchange	No	A 'private data account' that let you store your personal data. They're also trying to create a personal data 'exchange'/marketplace. Have created a dashboard - 'Rumpel' - to help users visualise and understand their data. And 'DataBuyer', which allows third party providers to offer services on top of this infrastructure.
Maidsafe	Blockchain	Yes	A 'crowdsourced internet', which breaks up and stores data across multiple computers, and pays people who have spare hard disk space in a virtual currency: safecoin. Consensus without a blockchain' https://blog.maidsafe.net/2015/01/29/consensus-without-a-blockchain/
Meeco	Personal data store	No	Meeco is a tool which allows people to manage their digital relationships. Through Meeco they can add, organise, edit and securely share all their information.
Midatacoop	Platform co-operative	No	A co-operative for owning and sharing health data for social good.
MyData	Personal data store	No	An alternative vision for personal data is managed and stored, and a network and annual conference bringing together these conversations with a range of different global communities. http://hiit.github.io/mydata-stack/ https://github.com/HIIT/mydata-sdk
MyDex	Personal data store	No	Provides a hyper-secure storage area and service so you can manage your personal data your way, from any aspect of your life.
Open PDS	Personal data store	No	An MIT project that allows users to collect, store, and give fine-grained access to their data all while protecting their privacy. Uses technology called 'SafeAnswers' which *"allows services to ask questions whose answers are calculated against the metadata instead of trying to anonymize individuals' metadata"*.

Name of project	Type of project	Distributed ledger?	Project aims
Our Data Coop	Platform co-operative	No	A project to test whether data co-ops could combine personal, organisational and public open data and function as asset-locked vehicles so that the third sector is better placed to understand the costs/ impact associated with different approaches to service delivery.
Our Data Mutual	Platform co-operative	No	We are creating a platform for the management, monetisation, control, protection and 'use for good' of our data. Fundamentally we'll be presented with a tool that lets us define how comfortable we are with the trading of our data - and with which types of organisation.
Resolution	Platform co-operative	Yes	A music co-op that provides a globally accessible database to store artists, songs, licensing terms and royalty payout rules. BigchainDB: 'blockchainified' big data service. Permissioned ledger https://www.bigchaindb.com/
Schluss	'Data vault'	No	A safe place on the internet where you can store images, text, video, and share things granularity with other people. Aims to create a 'global standard'
Sovrin/Project Indy	Self-sovereign identity	Yes	Distributed ledger and consortium for attribute-based, self-sovereign identity system (in partnership with Hyperledger). Public permissioned blockchain'
Swarm City	Platform co-operative	Yes	A decentralised ridesharing alternative to Uber. Will use Ethereum.
Synero	Decentralised social network	Yes	A privacy-aware social network, and decentralised attention economy.
Uport	Self-sovereign identity	Yes	uPort is an open source software project project to establish a global, unified, sovereign identity system. Create a self-sovereign identity on your mobile device. Uses credentials and reputation systems. Will use Ethereum.

Appendix 3

Empirical case studies

While researching this paper, we compiled a number of case studies of projects and organisations which have particular relevance to DECODE.

MIDATA.coop

MIDATA.coop is a Swiss co-operative that aims to give people control over their their medical data. Users can collect a variety of health-related information, including their hospital records, data produced by fitness trackers, even their genomic profile, and safely encrypt and store it in a local cloud. The user can then track their progress or share the data with whomever they want: their doctor, their family members, or relevant clinical trials where their data could prove useful.[116]

One of the things that makes MIDATA.coop different from other data storage platforms is that it does not use monetary rewards to encourage people to share their data. As Dr. Ernst Hafen, co-founder and president of the co-operative, said *"We do not want to introduce financial incentives for data sharing because that's exactly the wrong incentives. That's what everyone does and we want to change."*[117] The primary motivation for people to share their data is to help with medical research they care about. Patients gain collective influence by pooling data together, creating a valuable resource which pharmaceutical companies want to access. Any money made from this data is invested back into the community as decided by members of the co-operative, rather than provided as dividends to shareholders. This is an important aspect of MIDATA.coop's model, which Dr. Hafen strongly believes should remain a co-operative. His argument is simple: everyone has similar amounts of personal data, whether it be number of heartbeats or genome data, so everyone should have an equal say in how it is used.[118]

Though a fairly young initiative (founded in 2015), it has already seen some successes. The first pilot sees post-bariatric surgery patients recording health data like their weight loss and sharing it with doctors investigating the postoperative recovery period. The latest study examines a drug's effect on Multiple Sclerosis patients by analysing the data they input about motoric and cognitive capabilities on an app.[119] From late 2017 onwards anyone will be able to join and become a member without an access fee.

openPDS

A PDS (Personal Data Store) is a platform through which people can securely store, manage and share their data. There are quite a few organisations that have developed their own PDSs, though openPDS stands out due to its innovative SafeAnswers tool and focus on large-scale behavioural data.

Personal data shared with corporations is usually anonymised (or de-identified) by having all identifying labels removed. However, data science researcher and designer of OpenPDS Dr. Yves-Alexandre de Montjoye and his team at MIT proved that with only four geolocation data points a person could be uniquely identified within large scale mobile phone dataset of 1.5 million people.[120] To solve this issue and maintain people's privacy, they created the Q&A system SafeAnswers. When a third party like an app needs some information from you, it sends a couple of lines of code to openPDS. If approved and validated, the personal data store runs this algorithm on your data accordingly and sends back a validated 'answer' containing only the indispensable information. Through this strict need-to-know basis, apps can't actually see your data and it never leaves the safety of your PDS.[121]

What differentiated them when the enterprise was started back in 2014 is that they focused on big behavioural datasets; *"the breadcrumbs, anything generated as a side effect of you using technology"* as opposed to personal identity information which was the standard at the time.[122] Trials to test its performance in the field have already taken place. The largest is now taking place in Senegal, where openPDS has partnered with Orange to implement a large scale experiment in how the company can use people's mobile phone data whilst preserving their privacy.[123]

Hub of All Things (HAT)

HAT is building the infrastructure for a new digital exchange of personal data. It provides a database where users can congregate all their data, from browsing history to that produced by smart objects in their homes, but it also creates a platform where different bodies can buy and sell data.[124]

Though it cannot stop Internet Service Providers and third parties from collecting personal data, it uses apps and browser extensions to take a copy. A person can then combine data from different sources, which allows novel links between data sets to be drawn and uncover relationships that corporations might be interested in. Users can then sell them this data and therefore monetise their online activity, shifting the profits of the personal data economy to the users rather than unknown third parties. This way, HAT aims to help both users in giving them ownership and revenue from their data, as well as companies who'll have access to richer data sources.[125]

They believe the infrastructure through which data is managed and shared should not be privately owned, but rather a public and transparent platform on the internet. By providing access to all types of users, like companies, governments and average citizens, it results in a common marketplace for data transactions, but always with individuals having full control over their data's usage. In the words of Project Lead Irene Ng, *"data exchanges must be actioned upon by users themselves"*.[126] They envision a future where companies don't store personal data, since this is costly and risky for them. Instead, they sync their databases to multiple users' HATs, accessing personal data while paying individuals in return.[127]

The program's launch in 2013 was motivated by the unfair one-sided control of personal data by third parties. Originally a university research project funded by the Research Council's UK Digital Economy Programme, it received £1.2 million to create the platform,[128] and afterwards raised an extra £50,000 in funding through an Indiegogo campaign to launch HAT globally. Since then, over 250 people have already acquired their own HATs.

GebiedOnline

GebiedOnline is based on the belief that connecting people, both within a community and between different neighborhoods, will encourage active discussion of the issues and develop ideas that have positive social and economic impact.

It began in 2012, when Michael Vogler created a website for his neighbourhood of IJburg, Amsterdam, to bring citizens together to discuss how to improve the area. The project grew popular enough that other neighbourhoods became interested in adapting the tool for themselves. As a result the website template, providing a number of modular and customizable features, was made available to other neighbourhoods in 2016. It allows users to share events; start projects; propose, debate and rank ideas for the community democratically; and even create marketplaces for exchanging goods.[129]

This network of neighbourhoods is organised as a co-operative, allowing the community to decide how the platform should be managed and customised. So far the initiative has been moderately well received in the Netherlands with five neighbourhoods deploying their own sites, the largest of which now has 4,000 citizens enrolled.[130]

The platform's business model and governance structure stands in contrast to other emerging digital platforms for local communities. The Silicon Valley startup NextDoor, which describes itself as a *"private social network for your neighbourhood"*, already boasts of having over 147,000 neighborhoods enrolled in the US.[131] Recently, they have expanded to European markets in both the UK and the Netherlands.[132] Since it is funded by private equity investment, it will likely focus on satisfying its stakeholders by increasing revenue through advertising and the selling of personal data to third parties. On the other hand, co-operatives like GebiedOnline tend to be more driven by satisfying the community and reinvesting profits back into improving the area.

CitizenMe

The mobile app CitizenMe is based on the premise that people should understand and obtain value from their data, using integrated artificial intelligence (AI) to draw connections between different data sets and extract valuable information the user can benefit from. For example, linking your social media accounts to the app allows you to see what your posts say about your personality, and users can also share subsets of their data or complete short surveys to directly receive money through PayPal. There is also the option to donate data to charities.[133]

These features allow people to see what third parties can do with the data they extract (sometimes in unethical ways), and raise awareness of the importance of people taking control of their personal data. This was the sentiment the founder, StJohn Deakins, had in mind when he created the company back in 2014. In his words: *"Every individual should have visibility and control of their digital soul".*[134]

In accordance with this view, all data shared with the app is kept in a person's smartphone, and it only sends encrypted information to brands the user has consented to. It even provides a detailed guide to other sites' terms and conditions, and alerts you when they change so you can stay informed of issues affecting your privacy rights.[135] According to CitizenMe's Josh Hedley-Dent, in the future they plan to introduce *"smart AI"* that *"safeguards the privacy of the data further as [CitizenMe] touch less and less data to do the number crunching required to provide insights."*[136] The platform now has over 10,000 downloads on Google Play.

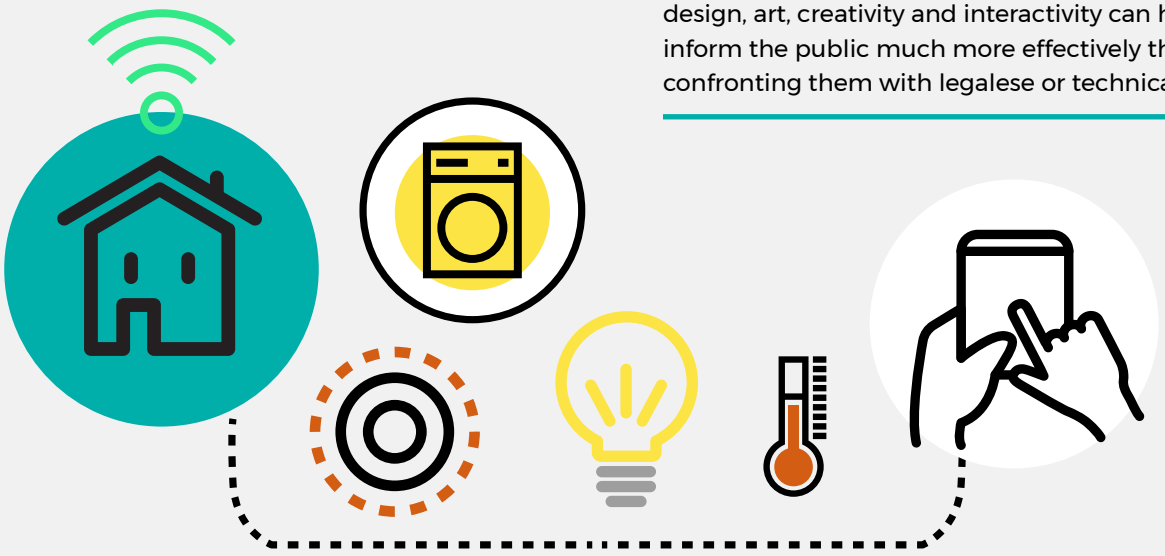
ProjectsbyIF

ProjectsbyIF is a London-based design studio started in 2015 which wants to make data science and technology more understandable for all. They aim to raise awareness about technological issues, like data rights and privacy, in an engaging and coherent way.

Their work focuses on creating tools that facilitate the average layperson's understanding of complex issues, both for the benefit of the user and for companies to improve their design. A clear case of this is their collection of common data permissions: design patterns found in many websites such as cookie banners, terms of service, opt-in for consent prompts, and so on. ProjectsbyIF fashion their own versions of these making them as clear and easy-to-use as possible.[137] Their interactive exhibition at Somerset House in 2015, which allowed visitors to create their own fictional data licenses, saw almost 16,000 people using the tool,[138] showing that *"when you give people the opportunity to understand, people choose to take control."*

Thanks to this approach they have collaborated to make prototype services with some well-known brands. Paperfree, in partnership with Co-op, explored how the co-operative could help its members store and manage financially sensitive data, like receipts, in a trustworthy manner which respected users' privacy. Rights and Networks, as part of Google Open Research, investigated how common technologies like public Wi-Fi could impact some of our basic rights to privacy.[139]

Sarah Gold, who founded and now directs ProjectsbyIF believes that *"together we need to build patterns that help people understand what happens with their data. This will put them in greater positions of control and power."*[140] Using design, art, creativity and interactivity can help to inform the public much more effectively than by confronting them with legalese or technical jargon.



Solid

Solid is a completely new framework for the Internet which aims to give users complete ownership of their data, increase their privacy and allow the development of new applications focused on reusing data. Users can store different types of data in distinct storage spaces: for example, sensitive financial information could be saved in the home, whereas social media data would be in an online cloud. It also includes the protocols for applications to access that data, so only authorised apps can do so.[141]

Tim Berners Lee, creator of the World Wide Web, and his team at MIT, have been developing this technology since 2016 in an effort to re-decentralise the web. Internet giants like Google or Facebook control most internet traffic, generating large personal data silos that give them an unfair advantage over competitors. With Solid, every user owns their data, so if an individual becomes dissatisfied with a service, they can just take their data to another site. There is no penalty for changing, and vendor lock-in is avoided. Taking it one step further, they also want to make data expression standard, so that diverse data types can be shared by completely different services.[142] It is this interoperability that gives Solid its name: SOcial LInked Data.

The project is still young and no major use cases have been developed yet. The team envisions third parties creating their own apps will drive the expansion of the platform. As developer Andrei Sambra says: *"[The adoption] is going to come from the private sector"*.[143] In the meantime, the team itself has built a series of prototype applications as examples, including a blogging platform, an address book and a profile editor, which are available on their website.[144]

Databox

Described as a 'privacy aware platform', Databox is currently a physical device you can install in your home that manages access to your data. You can login with your social media accounts and it can even connect to IoT devices in your home, so all data can be accessed within the hub. There are apps for each data set that control the flow of data, sitting between personal data and the cloud. By doing so, it wants to stop data like your location from continuously and indiscriminately going up into the cloud without your permission.

They believe your personal data should be shared with organisations, called data processors, that can analyse it and provide useful services in return, like insights into your expenditures or personalised offers. However, the priority is always user privacy. Rather than send the data to the cloud where its access cannot be controlled, apps are sent down from the cloud into the Databox to analyse data there, so information never leaves the box. Users can then allow the app to send the analytics back to the data processor, which only receives the minimum amount of data, or it can be sent directly to the user.[145] According to Hamed Haddadi, a researcher at Databox, this approach is perfect for businesses that want market research information from their users, but *"don't necessarily want to have access to fine-grained, detailed data"*.[146] With this, the aim is not to completely change the data market, but direct it further towards user privacy by minimising the amount of sensitive information which needs to be shared.

Databox is a collaboration between many U.K. universities with several prominent industry partners which will run from 2016 to 2019, using the £1.5 million in funding provided by the EPSRC.[147] In the six months they've been operating, they have developed several standard apps for the service, like sentiment analysis based on Twitter data and IoT device controllers. In the future, they expect third parties to develop their own apps. The team is working with the BBC, for instance, who is already planning to create a user behaviour analytics Databox app.

TheGoodData

The Good Data is a co-operative that collects, pools and sells members' internet browsing data, but it does so entirely on their members' terms. In order to build user trust the platform anonymises as much of the data it collects about its users as it can. This is done manually on a case-by-case basis, depending on who is requesting access to it and why. The co-op also refuses to collect any information that contains sensitive terms, say, from browsing explicit websites, health websites or political websites.[148]

Revenue is generated from the sale of anonymised data to data brokers and other advertising platforms, and the profits are split between social lending in developing countries, and on improving technical development of the platform. Any co-op member can participate in deciding these rules, either through discussion on the site's collaborative platform, by attending general meetings, or by standing to be elected as a Company Director. The investments can be tracked on their website: so far they have funded 34 projects in developing countries.[149]

On one hand, the coders are incentivised to build better services for the site by setting aside a portion of the company's revenue for them. On the other, the users (those whose data is actually being sold) retain the power to make decisions, making sure the company's interests are always in line with those of its members. Altogether, this approach allows for transparency of data flows, participation and fair distribution of profits.[150]

New Zealand Data Commons

A data commons describes the pooling and sharing of data as a resource by a community to derive value from it that will benefit all members of said community. Whilst the concept itself has been around for a while, it was first suggested as a New Zealand-wide project in April 2015 by James Mansell, as part of a government report on increasing productivity. It led to the creation of the New Zealand Data Commons project a year later: a project with the goal of forming a network of interested partners to research the topic and create a blueprint for how this alternative model of data sharing would work.[151]

Over the six following months, three design teams formed by experts worked independently on separate issues like market or governance, and at the end pooled their conclusions in a report. It covers the theory of data commons, from how it should be set up, the people that should participate, and how its establishment can be promoted.[152]

According to their vision, a successful data commons will have some key principles, the biggest of which is trust between participants. Members of the commons will have to provide potentially sensitive information to a pool of data, so they'll need to trust that it will be used for the greater good of the community. To this end, the commons' rules should be designed by the participants in a transparent and inclusive manner that accommodates everyone's interests. Of course, data providers should always be in control of any interactions, with the ability to terminate any data sharing agreement with another individual at any time.

This report is now publicly available, welcoming anyone to contribute their thoughts. Whilst the website describes it as *"a safer, lower-cost and higher-value alternative to the current approaches to the challenge of data integration and reuse"*, NZ Data Commons' does not presume to have found the perfect model for a data commons. It is a complex issue which needs to be explored in more depth, and with this project they want to fuel the discussion so as to progress the field.

Endnotes

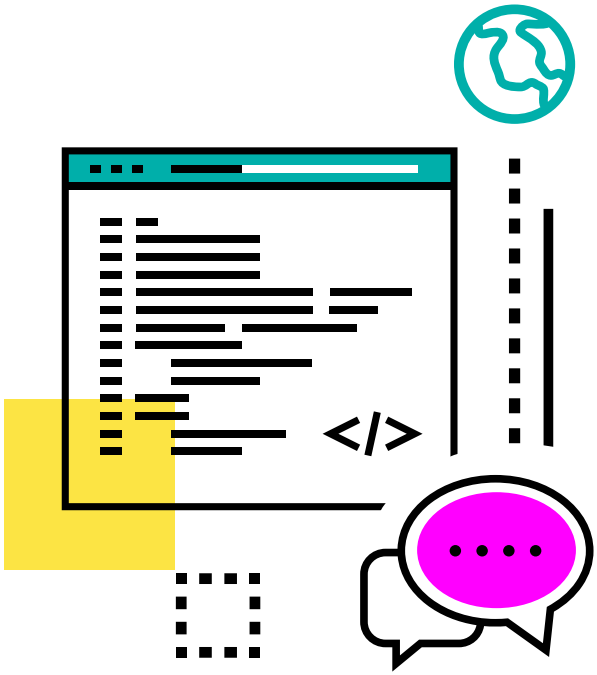
1. See: http://dlc.dlib.indiana.edu/dlc/contentguidelines [accessed 31 August].
2. Bria, F and Primosig, F. (2015) Internet as common or capture of collective intelligence. D-CENT project. Available from: https://dcentproject.eu/wp-content/uploads/2015/08/D3.3-Annex-Internet-Identity-Seminar_annex.pdf
3. Zingales, L. and Rolnik, G. (2017) A Way to Own Your Social-Media Data [online]. 'New York Times'. June 30. Available from: https://www.nytimes.com/2017/06/30/opinion/social-data-google-facebook-europe.html [accessed 24 August 2017].
4. Jourová, V. (2015) 'Data protection Eurobarometer Factsheet' [online]. European Commission. Available from: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_eurobarometer_240615_en.pdf [accessed 24 August 2017].
5. Deakins, S. (2017) 'Understanding how people interact with their data' [online]. Nesta Blog. Available from: http://www.nesta.org.uk/blog/understanding-how-people-interact-their-data [accessed 24 August 2017].
6. Bria, F. (2015) Public policies for digital sovereignty. OR Books, NYC. Available from: https://www.academia.edu/19102224/Public_policies_for_digital_sovereignty
7. Westerlund, M. and Enkvist, J (2016) Platform Privacy: The Missing Piece of Data Protection Legislation [online].. 'Journal of Intellectual Property, Information Technology and E-Commerce Law'. Vol. 7(1). Available from: https://www.jipitec.eu/issues/jipitec-7-1-2016/4390/#ftn.N103FA [accessed 24 August 2017].
8. European Commission (2015) Special Eurobarometer 431 - Data Protection Report. Available from: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf [accessed 30 August 2017].
9. Bria, F. (2017) People should control their digital identity. CityMetric. Available from: http://www.citymetric.com/horizons/people-should-control-their-digital-identity-barcelona-s-chief-technology-officer-decode
10. Interview with Francesca Bria, 19 June 2017. How Barcelona's citizens will control the use of their data. GovInsider. Available from: https://govinsider.asia/inclusive-gov/barcelona-city-council-citizen-data-sharing-francesca-bria/
11. See: https://s21.q4cdn.com/399680738/files/doc_financials/2017/FB-Q1'17-Earnings-Slides.pdf [accessed 25 August, 2017].
12. European Commission (2016) 'The EU Data Protection Reform and Big Data Factsheet' [online]. Available from http://ec.europa.eu/justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf [accessed 25 August 2017]
13. Business Insider (2017) Online ad revenues are surging, but 2 companies are getting most of the spoils [online]. Available from: http://www.businessinsider.com/online-ads-revenues-going-to-google-and-facebook-the-most-2017-4?IR=T [Accessed 30 August 2017].
14. Pricewaterhouse Coopers (2000) IAB Internet Advertising Revenue Report [online]. Available at: https://www.iab.com/wp-content/uploads/2015/05/IAB_PWC_1999Q4.pdf [accessed 30th August 2017]
15. See: https://www.acxiom.co.uk/what-we-do1/data-quality/ [accessed 31 August 2017].
16. Federal Trade Commission (2014) 'Data Brokers: A Call for Transparency and Accountability.' Washington DC: Federal Trade Commission. p.iv.
17. European Commission (2015) Special Eurobarometer 431 - Data Protection Report. Available from: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf [accessed 30 August 2017].
18. McDonald, A. M. and Cranor, L. F (2008) The Cost of Reading Privacy Policies. 'I/S: A Journal of Law and Policy for the Information Society'. 2008 Privacy Year in Review Issue.
19. BBC news (2013) Facebook 'likes' predict personality [online]. 'BBC News,' 11 March, 2013. Available from: http://www.bbc.co.uk/news/technology-21699305 [accessed 25 August 2017].
20. Hirsch, D. (2015) That's Unfair! Or Is It? Big Data, Discrimination and the FTC's Unfairness Authority. 'Kentucky Law Journal.' Vol.103.
21. Larson, J., Mattu, S., Kirchner, L. and Angwin, J. (2016) How We Analyzed the COMPAS Recidivism Algorithm [online]. 'ProPublica'. Available from: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [accessed 25 August 2017].
22. Hardt, M. (2014) How big data is unfair [online]. 'Medium.' Available from: https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de [accessed 31 August 2017].
23. Evans, M. (2017) Media Lab student wins national award for fighting bias in machine learning [online]. 'MIT media lab post'. Avaialable from: https://www.media.mit.edu/posts/media-lab-student-recognized-for-fighting-bias-in-machine-learning/ [accessed 25 August 2017].
24. Valentino-Devries, J., Singer-Vine, J. and Soltani, A. (2012) Websites Vary Prices, Deals Based on Users' Information [online]. 'The Wall Street Journal'. 24 December, 2012. Available from: https://www.wsj.com/article_email/SB10001424127887323777204578189391813881534-lMyQjAxMTAyMDIwMzEyNDMyWj.html [accessed 25 August 2017].

25. Kominers, S. D. (2017) Uber's New Pricing Idea is Good Theory, RIsky Business [online]. 'Bloomberg'.' 13 June, 2017. Available from: https://www.bloomberg.com/view/articles/2017-06-13/uber-s-new-pricing-idea-is-good-theory-risky-business [accessed 25 August 2017].
26. Chowdhry, A. (2016) Uber: Users Are More Likely To Pay Surge Pricing If Their Phone Battery Is Low [online]. 'Forbes.' 25 May 2016. Available from: https://www.forbes.com/sites/amitchowdhry/2016/05/25/uber-low-battery/#44a5827b74b3 [accessed 25 August 2017].
27. Office of Fair Trading (2013) 'Personalised pricing: Increasing transparency to improve trust in the market [online]'. Available from: http://webarchive.nationalarchives.gov.uk/20140402165101/ http://oft.gov.uk/shared_oft/markets-work/personalised-pricing/oft1489.pdf [accessed 25 August 2017].
28. Borgesius, F. and Poort, J. (2017) Online Price Discrimination and EU Data Privacy Law. 'Journal of Consumer Privacy.' pp.1-20.
29. Information Commissioner's Office (2016) Privacy regulators study finds Internet of Things shortfalls [online]. 'Information Commissioner's Office News and Blogs'. 22 September 2016. Available from: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2016/09/privacy-regulators-study-finds-internet-of-things-shortfalls/ [accessed 31 August 2017].
30. Singer, N. (2015) With a Few Bits of Data, Researchers Identify 'Anonymous' People [online]. 'New York Times'. 29 January 2015. Available from: https://bits.blogs.nytimes.com/2015/01/29/with-a-few-bits-of-data-researchers-identify-anonymous-people/ [accessed 25 August 2017].
31. Westerlund, M and Enkvist, J. (2016) Platform Privacy: The Missing Piece of Data Protection Legislation JIPITEC, Vol. 7. (urn:nbn:de:0009-29-43904). Available at: https://www.jipitec.eu/issues/jipitec-7-1-2016/4390 [Accessed 30th August 2017].
32. The Economist (2017) The world's most valuable resource is no longer oil, but data [online]. 6 May 2017. Available from: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [accessed 25 August 2017].
33. Isidore, C. (2016) Amazon Prime now reaches nearly half of U.S. households [online]. 'CNN Tech.' 26 January 2016. Availble from: http://money.cnn.com/2016/01/26/technology/amazon-prime-memberships/ [accessed 25 August 2017].
34. Richter, W. and Street, W. (2017) Online ad revenues are surging, but 2 companies are getting most of the spoils [online]. 'Business Insider'. 27 April 2017. Available from: http://www.businessinsider.com/online-ads-revenues-going-to-google-and-facebook-the-most-2017-4?IR=T [accessed 25 August 2017].
35. The Economist (2017) The world's most valuable resource is no longer oil, but data [online]. 6 May 2017. Available from: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [accessed 25 August 2017].
36. Stucke, M and Grunes, A. (2016) 'Big Data and Competition Policy'. Oxford: Oxford University Press.
37. Collin, P. and Collin, N., cited in Smichowski, B. C. (2016) Data as a common in the sharing economy: a general policy proposal. 'Document de travail du CEPN.' Paris: Centre d'économie de L'Universitié Paris Nord.
38. Lambrecht A. and Tucker, C. E. (2015) Can Big Data Protect a Firm from Competition? [online]. 'SSRN'. Avaialable from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2705530 [accessed 29 August 2017].
39. Sokol, D. and Comerford, R. (2016) Does Antitrust Have a Role to Play in Regulating Big Data? In: Blair, R. and Sokol, D. (eds) 'Cambridge Handbook of Antitrust, Intellectual Property and High Tech.' Cambridge: Cambridge University Press.
40. Smichowski, B. C. (2016) Data as a common in the sharing economy: a general policy proposal. 'Document de travail du CEPN.' Paris: Centre d'économie de L'Universitié Paris Nord.
41. Srnicek, N. writing in 'The Guardian' (2017) We need to nationalise Google, Facebook and Amazon. Here's why Available at: https://www.theguardian.com/commentisfree/2017/aug/30/nationalise-google-facebook-amazon-data-monopoly-platform-public-interest [accessed 30 August 2017].
42. Srnicek, N. (2016) Platform Capitalism. Polity Press.
43. Pasquale, F. (2017) Will Amazon Take Over the World? [online] 'Boston Review.' 20 July 2017. Available from https://bostonreview.net/class-inequality/frank-pasquale-will-amazon-take-over-world [accessed 29 August 2017].
44. Clark, B. (2017) Facebook's new 'early bird' spy tool is just the tip of the iceberg [online]. Thenextweb.com. 10 August 2017. https://thenextweb.com/insider/2017/08/10/facebooks-new-early-bird-spy-tool-is-just-the-tip-of-the-iceberg/#.tnw_CcBjYpcY [accessed 30 August 2017].
45. Smichowski, B. C. (2016) Data as a common in the sharing economy: a general policy proposal. 'Document de travail du CEPN.' Paris: Centre d'économie de L'Universitié Paris Nord.
46. Hogg, C. (2014) How the internet makes us all part of the social experiment [online] 'The Long+Short. Available from: https://thelongandshort.org/machines/ab-testing-facebook-social-experiments
47. https://www.wired.com/2016/11/facebook-won-trump-election-not-just-fake-news/

48. Doward, J., Cadwalladr, C. and Gibbs, A. (2017) Watchdog to launch inquiry into misuse of data in politics [online]. 4 March 2017. 'The Guardian'. Avaialable from: https://www.theguardian.com/technology/2017/mar/04/cambridge-analytics-data-brexit-trump [accessed 31 August 2017].
49. Zingales, L. and Rolnik, G. (2017) A Way to Own Your Social-Media Data [online]. 'New York Times'. 30 June. Available from: https://www.nytimes.com/2017/06/30/opinion/social-data-google-facebook-europe.html [accessed 24 August 2017].
50. Lohr, S. (2014) Google Flu Trends: The Limits of Big Data [online]. 'New York Times.' 28 March, 2014. Available from https://bits.blogs.nytimes.com/2014/03/28/google-flu-trends-the-limits-of-big-data/ [accessed 29 August, 2017].
51. See: https://info.internet.org/en/story/free-basics-from-internet-org/ [accessd 29 August 2017].
52. Tarnoff, B. (2017) Silicon Valley siphons our data like oil. But the deepest drilling has just begun [online]. 'Guardian'. 23 August 2017. Available from https://www.theguardian.com/world/2017/aug/23/silicon-valley-big-data-extraction-amazon-whole-foods-facebook [accessed 29 August 2017].
53. Pasquale, F. (2017) Will Amazon Take Over the World? [online] 'Boston Review.' 20 July 2017. Available from https://bostonreview.net/class-inequality/frank-pasquale-will-amazon-take-over-world [accessed 29 August 2017].
54. Saunders, T. and Baeck, P. (2015) Rethinking Smart Cities from the Ground Up. London: Nesta.
55. Interview with Francesca Bria, 20 June 2017: http://magazine.ouishare.net/2017/06/building-the-networked-city-from-the-ground-up-with-citizens-interview-with-francesca-bria/
56. Barcelona City data commons programme: http://ajuntament.barcelona.cat/digital/en/digital-transformation/city-data-commons
57. Lange, M. and De Waal, M. (2016) Owning the City: New Media and Citizen Engagement in Urban Design. In: Etingoff, K. (ed) Urban Land Use: Community-Based Planning. Boca Raton: CRC Press.
58. See: http://datacommons.org.nz/ [accessed 29 August 2017].
59. Yakowitz, J. (2011) Tragedy of the Data Commons. 'Harvard Journal of Law and Technology.' Vol.25(1).
60. Scholz, T. and Schneider, N. (2017) What this is and isn't about. In: Scholz, T. and Schneider, N. (eds) 'Ours to Hack and Own'. New York: OR Books.
61. Interview with Matt Hogan, 30 May 2017.
62. Bass, T. (2017) The monetary and social value in our data - An interview with Citizenme [online]. 'Nesta Blogs'. Available from http://www.nesta.org.uk/blog/monetary-and-social-value-our-data-interview-citizenme [accessed 29 August 2017].
63. Ctrl-Shift (2014) 'Personal Information Management Services: An analysis of an emerging market'. London: Ctrl-Shift.
64. World Economic Forum (2011) 'Personal Data: The Emergence of a New Asset Class' [online]. Available from http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf [accessed 29 August 2017].
65. Morozov, E. (2014) Silicon Valley is turning our lives into an asset class [online]. 'Financial Times.' Available from https://www.ft.com/content/9d2a73fe-a54a-11e3-8070-00144feab7de?mhq5j=e1 [accessed 29 August 2017].
66. Interview with Annemarie Naylor, 22 June 2017.
67. Interview with Oleguer Sagarra Phd, 21 June 2017.
68. See: http://dlc.dlib.indiana.edu/dlc/contentguidelines [accessed 31 August 2017].
69. Hardin, G. (1968) The Tragedy of the Commons. 'Science'. Vol.162(3859). pp.1243-1248.
70. For more extensive description of the debate see Vercellone, C., Bria, F., Fumagalli, A., Gentilucci, E., Giuliani, A., Griziotti, G. and Pierluigi, V. (2013)'Managing the commons in the knowledge economy. 'DEcentralised Citizen ENgagement Technologies D3.2'. Available from http://www.nesta.org.uk/sites/default/files/d-cent_managing_the_commons_in_the_knowledge_economy.pdf [accessed 31 August 2017].
71. Ibid., p.17.
72. Ostrom, E. (1990) 'Governing the Commons.' Cambridge: Cambridge University Press.
73. http://www.nesta.org.uk/sites/default/files/research_on_digital_identity_ecosystems.pdf p.106.
74. Interview with Bruno Carballa Smichowski, 3 July 2017.
75. Ng, I. (2017) Personal data innovation for whom? [online]. 'Nesta Blogs'. Available from: http://www.nesta.org.uk/blog/personal-data-innovation-whom [accessed 31 August 2017].
76. Rossi, M. (2001) Is Computer Data 'Tangible Property' or Subject to 'Physical Loss or Damage'? - Part 1 [online]. 'IRMI'. Available from: https://www.irmi.com/articles/expert-commentary/is-computer-data-tangible-property-or-subject-to-physical-loss-or-damage-part-1 [accessed 31 August].
77. Yakowitz, J. (2011) Tragedy of the Data Commons. 'Harvard Journal of Law and Technology.' Vol.25(1).
78. Interview with Marco Ciurcina, 19 June 2017.
79. Smichowski, B. C. (2016) Data as a common in the sharing economy: a general policy proposal. 'Document de travail du CEPN.' Paris: Centre d'économie de L'Universitié Paris Nord.
80. Baarbé, J., Blom, M. and de Beer, J. (2017) A Data Commons for Food Security. 'Working Paper 7: IASC 2017 Conference Paper, June 20 2017'. openAIR.
81. Binns, R. (2017) The ongoing struggle for personal data standards [online]. 'Nesta Blogs'. Available from http://www.nesta.org.uk/blog/ongoing-struggle-personal-data-standards [accessed 31 August].
82. See: https://hubofallthings.com/ [Accessed 19 July 2017].
83. See: https://solid.mit.edu. [accessed 31 August 2017].
84. Interview with Andrei Sambra, 7 April 2017.

85. Interview with Andrei Sambra, 7 April 2017.
86. Weitzner, D., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J. and Sussman, G. J. (2008) Information Accountability. 'Communications of the ACM'. Vol.51(6).
87. Lazaro, C. and Le Metayer, D. (2015) The control over personal data: True remedy of fairy tale? 'CRIDS: Research Report No 8681'.
88. Greenemeier, L (2014) How to reconcile big data and privacy. 'Scientific American'. 6 March 2014.
89. See: http://openpds.media.mit.edu/ [accessed 31 August 2017].
90. Interview with Yves-Alexandre Montjoye.
91. https://arxiv.org/ftp/arxiv/papers/1512/1512.06000.pdf
92. Clark, L. (2014) Tim Berners-Lee: we need to re-decentralise the web. 'Wired'. 6 February 2014.
93. See: https://blockstack.org/whitepaper.pdf [accessed 31 August 2017].
94. See: https://blockstack.org/ [accessed 31 August 2017].
95. See: https://blockstack.org/docs/blockstack-vs-dns [accessed 31 August 2017].
96. See: https://sovrin.org/ [accessed 31 August 2017].
97. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., Mccoy, D., Voelker, G. M. and Savage, S. (2013) A Fistful of Bitcoins: Characterizing Payments Among Men with No Names.'; login:'. Vol.38(6).
98. See: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation [accessed 31 August 2017].
99. Dyne.org (2017) Co-creation framework, methodologies and templates. 'DEcentralised Citizen Owned Data Ecosystem, Deliverable 6.5'.
100. Scholz, T. (2014) Platform Co-operativism vs. the Sharing Economy [online]. 'Medium'. Available from https://medium.com/@trebors/platform-cooperativism-vs-the-sharing-economy-2ea737f1b5ad [accessed 31 August, 2017].
101. Interview with Annemarie Naylor, 22 June 2017.
102. Interview with Marcos Menendez, 12 May 2017.
103. Scheiber, N. (2017) How Uber Uses Psychological Tricks to Push Its Drivers' Buttons. 'New York Times'. 2 April 2017; Dakers, M. (2016) Uber knows customers with dying batteries are more likely to accept surge pricing. 'The Daily Telegraph'. 22 May 2016.
104. Van der Zee, R. (2016) The 'Airbnb effect': is it real, and what is it doing to a city like Amsterdam? 'The Guardian'. 6 October 2016.
105. Interview with William Heath, 13 July 2017.
106. Loder, J., Bunt, L. and Wyatt, J. (2013) 'Doctor Know: A Knowledge Commons in Health'. London: Nesta.
107. Interview with Dr Ernst Hafen, 31 May 2017.
108. See: http://commonfutures.eu/developing-data-coops-for-community-benefit/ [accessed 31 August 2017].
109. Mansell, J., Laking, R., Matheson, B. and Light, R. (2016) 'Data Commons Blueprint' [online]. Available from: http://datacommons.org.nz/ [accessed 31 August 2017].
110. Ibid.
111. Interview with Marcos Menendez, 12 May 2017.
112. Gebiedonline.nl (2016) 'A co-operatively owned online platform for neighbouhood communities' [online]. Available from: https://gebiedonline.nl/engine/download/blob/gebiedsplatform/69870/2016/40/2016-10-02_Gebiedonline_IntroEng.pdf?app=gebiedsplatform&class=9096&id=242&field=69870 [accessed 31 August 2017].
113. Baarbé, J., Blom, M. and de Beer, J. (2017) A Data Commons for Food Security. 'Working Paper 7: IASC 2017 Conference Paper, June 20 2017'. openAIR.
114. Submitted to the European Commission in July 2017.
115. Submitted to the European Commission in July 2017.
116. See: https://midata.coop/ [Accessed 19 July 2017].
117. Interview with Dr. Ernst Hafen, 31 May 2017.
118. See: https://platform.coop/featured/midata-coop [Accessed 19 July 2017].
119. See: https://midata.coop/#projects [Accessed 19 July 2017].
120. De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., Blondel, V. D. (2013) Unique in the Crowd: The privacy bounds of human mobility. 'Nature', Vol.25(3).
121. See: http://openpds.media.mit.edu/ [Accessed 18 July 2017].
122. Interview with Yves-Alexandre de Montjoye, 2 July 2017.
123. De Montjoye, Y.-A., Shmueli, E., Wang, S., Pentland, A. S. (2014) openPDS: Protecting the Privacy of Metadata through SafeAnswers. 'PLoS ONE'. Vol. 9(7).
124. See: https://hubofallthings.com/ [Accessed 19 July 2017].
125. See: https://hatresearch.org/what-is-the-hat/the-research-and-its-implications/ [Accessed 19 July 2017].
126. Ng, I. (2017) Personal data innovation for whom? [online] 'Nesta Blogs'. Available from: http://www.nesta.org.uk/blog/personal-data-innovation-whom [Accessed 19 July 2017].
127. See: https://www.indiegogo.com/projects/hat-claim-your-data-organise-visualise-control-technology#/ [Accessed 19 July 2017].
128. See: http://www2.warwick.ac.uk/fac/sci/wmg/research/business_transformation/ssg/research/value/hat/ [Accessed 19 Jun 2017].
129. Bouman, K. (2017) Hallo Ijburg: Gebiedonline in de praktijk #1: Ijburg als eerste gebruiker. 'Nieuw Amsterdam'. 6 March, 2017. Available from: https://stedenintransitie.nl/stadbericht/hallo-ijburg [Accessed 18 July 2017].

130. Gebiedonline.nl (2016) 'A co-operatively owned online platform for neighbouhood communities' [online]. Available from: https://gebiedonline.nl/engine/download/blob/gebiedsplatform/69870/2016/40/2016-10-02_Gebiedonline_IntroEng.pdf?app=gebiedsplatform&class=9096&id=242&field=69870 [accessed 31 August 2017].
131. See: https://nextdoor.com/ [Accessed 18 July 2017].
132. Lev-Ram, M. (2016) Exclusive: Nextdoor Expands to First Market Outside U.S [online]. 'Fortune Tech' Available from http://fortune.com/2016/02/16/nextdoor-expands-netherlands/ [Accessed 19 July 2017].
133. See: https://www.citizenme.com/ [Accessed 19 July 2017].
134. See: https://www.citizenme.com/public/wp/welcome/ [Accessed 19 July 2017].
135. Hahn, G. (2014) Managing terms of service when the only constant is change. 'Citizenme Blog'. Available from https://www.citizenme.com/public/wp/managing-terms-of-service-when-the-only-constant-is-change/ [Accessed 19 July 2017].
136. Bass, T. (2017) The monetary and social value in our data - An interview with Citizenme [online]. 'Nesta Blogs'. Available from http://www.nesta.org.uk/blog/monetary-and-social-value-our-data-interview-citizenme [accessed 29 August 2017].
137. See: https://catalogue.projectsbyif.com/ [Accessed 19 July 2017].
138. See: https://projectsbyif.com/projects/data-licences [Accessed 19 July 2017].
139. See: https://projectsbyif.com/projects [Accessed 19 July 2017].
140. Gold, S. (2017) Designing tools for people [online]. 'Nesta Blogs'. Available from http://www.nesta.org.uk/blog/designing-tools-people [Accessed 19 July 2017].
141. Finley, K. (2017) Tim Berners-Lee, inventor of the web, plots a radical overhaul of his creation [online]. 'Wired'. 4 April, 2017. Available from https://www.wired.com/2017/04/tim-berners-lee-inventor-web-plots-radical-overhaul-creation/ [Accessed 25 July 2017].
142. Weinberger, D. (2016) How the father of the World Wide Web plans to reclaim it from Facebook and Google [online]. 'Digital Trends'. 27 December 2016. Available from https://www.digitaltrends.com/web/ways-to-decentralize-the-web/ [Accessed 25 July 2017].
143. Interview with Andrei Sambra, 6 April 2017.
144. See: https://solid.mit.edu/ [Accessed 25 July 2017].
145. See: http://www.databoxproject.uk/2016/12/27/so-whats-the-databox/ [Accessed 25 July 2017].
146. Interview with Hamed Haddadi, 13 April 2017.
147. See: https://www.databoxproject.uk/about/ [Accessed 25 June 2017].
148. See: https://www.forbes.com/sites/northwesternmutual/2017/07/11/why-my-family-talks-about-peaches-and-pits-every-night-at-dinner/#10d3fe1b7f6d [Accessed 25 July 2017].
149. See: https://www.thegooddata.org/good-data [Accessed 25 July 2017].
150. Interview with Marcos Menendez, 12 May 2017.
151. See: http://datacommons.org.nz/ [Accessed 26 July 2017].
152. Mansell, J., Laking, R., Matheson, B. and Light, R. (2016) 'Data Commons Blueprint' [online]. Available from: http://datacommons.org.nz/ [accessed 31 August 2017].



decode

58 Victoria Embankment
London EC4Y 0DS

info@decodeproject.eu

@decodeproject

www.decodeproject.eu