
2020

open data institute

Data 2020

This book provides a landscape review of hot topics in the world of data in 2020 – from digital competition, AI and trade to data rights, ethics and literacy.

You'll find an overview of the key issues and pointers to areas with potential for further exploration through research, development and practice.

You can find further resources relating to the topics set out in this book at theodi.org/data2020.

The Open Data Institute works with companies and governments to build an open, trustworthy data ecosystem

theodi.org | @ODIHQ

ODI Co-founders Sir Nigel Shadbolt and Sir Tim Berners-Lee at the ODI Summit 2019, theodi.org/summit

Data 2020

This book provides a landscape review of hot topics in the world of data in 2020 – from digital competition, AI and trade to data rights, ethics and literacy.

You'll find an overview of the key issues and pointers to areas with potential for further exploration through research, development and practice.

You can find further resources relating to the topics set out in this book at theodi.org/data2020.

ODI Summit 2019, theodi.org/summit

01

AI

and algorithmic accountability

AI and machine learning algorithms are increasingly being used to make decisions – including decisions about us. But how are these decisions made? How can algorithms be interrogated or understood, and how do we ensure that unfair bias isn't being built in – even unintentionally? More fundamentally, should some decisions be automated at all?

AI and algorithmic accountability

The UN Special Rapporteur on extreme poverty and human rights has highlighted the impact of AI and digital technologies on people's lives. They are being used in crucial decisions such as eligibility assessment, fraud prevention and detection, and risk scoring and needs classification.

AI Now has detailed the range of algorithmic systems used by the public sector. The UK's Centre for Data Ethics and Innovation is examining bias in algorithmic decision-making in financial services, crime and justice, recruitment and public services, and is creating recommendations about how any potential harms can be identified and minimised.

Transparency and an organisation's ability to explain the AI algorithms it uses builds trust. It also enables teams to monitor how decisions are made, and if necessary address failings, bias or problems in the system. Understanding and auditing automated decision-making is critical for society – not just in terms of ensuring decisions are accurate, non-discriminatory and fair, but to ensure people can maintain their autonomy.

Upturn and Omidyar Network have described how to design systems for accountability. The Alan Turing Institute and the UK's Information Commissioner's Office have also developed guidance for improving the explainability of AI.

Hot topics

- Regulating accountability and transparency mechanisms for private sector use of AI
- Engaging citizens around public service automation, including the use of facial recognition and biometrics
- Integrating AI into human decision-making
- Improving the explainability of algorithms created by machine learning
- Monitoring and responding to the impacts of algorithmic decision-making

Explore

- United Nations: Extreme poverty and human rights
- AI Now: Algorithmic account-ability policy toolkit
- Centre for Data Ethics and Innovation: Review on bias in algorithmic decision-making
- Upturn and Omidyar Network: Public scrutiny of automated decisions
- Information Commissioner's Office and The Alan Turing Institute: Consultation on explaining AI decisions guidance

For more tools and resources visit
theodi.org/ai2020

Collaboration

to solve societal problems

Some problems can only be tackled through collective action. For example, some major diseases can only be tackled if scientists, charities and pharmaceutical companies work together. Sharing data can be an essential part of these collaborations. But who manages that data? Who ensures it's shared and used in the right way? New data institutions could be an answer.

Collaboration to solve societal problems

Many sectors could benefit from institutions that steward and provide access to data. Take wildlife conservation for example. In 2019, we explored using a 'data trust' – one emerging type of data institution – to enable academics and conservationists to share data with app developers to help tackle international illegal wildlife trade.

We found that this approach – where trustees take on a fiduciary duty (a legal responsibility of impartiality, prudence, transparency and loyalty) on behalf of others for how data is shared – could help earn trust. We worked with the Greater London Authority to explore how data trusts could improve city services; and with WRAP to look at how data trusts could help with the mission to reduce global food waste.

But data trusts aren't the only way to manage access to data. Our own research found that there is huge demand from private, public and third sector organisations in countries around the world to increase sharing of data. Other types of data institution, such as data clubs or data cooperatives, might sometimes be a better fit.

Data institutions can bring together organisations or people to solve specific local or international problems. This would mean less 'data hugging' and more cross-sector collaboration.

Hot topics

- Designing data institutions to suit different challenges and local communities
- Empowering people through collective data institutions such as data cooperatives or data unions
- Making data institutions sustainable with appropriate revenue and business models
- Collaborative approaches to maintaining data as a shared asset
- Defining good data stewardship practices

Explore

- Open Data Institute: Are data trusts the answer?
- Nesta: The new ecosystem of trust
- Centre for International Governance Innovation: What is a data trust?
- StiftungNeueVerantwortung: Designing Data Trusts
- Aapti Institute: Data Stewardship - a taxonomy

For more tools and resources visit
theodi.org/collaboration2020

03

Competition

in digital markets

Digital services such as social media, e-commerce, peer-to-peer platforms, search, and online advertising pose a number of challenges to competition policy: they are international, evolve quickly, and it can sometimes be hard for new entrants to compete.

Competition in digital markets

Online marketplaces, social media, and digital services that provide recommendations are better if they have lots of data to use. Having more data means the companies are better able to offer personalised services to their customers.

Last year, Google accounted for more than 90% of all revenues earned from search advertising in the UK, with revenues of around £6bn. Knowing that there are big companies with large amounts of data may be causing small firms entering the market to position themselves for acquisition rather than expansion, and to harvest data as quickly as possible in order to be competitive.

The UK Digital Competition Expert Panel was set up in 2018 to consider the potential opportunities and challenges the emerging digital economy may pose for competition policy. Among other things, it recommended more standards for data portability and increased access to data to reduce barriers to entry.

The UK Competition and Markets Authority is carrying out a market study into online platforms and the digital advertising market. Its interim report includes potential measures such as providing access to click-and-query data and greater transparency around advertising.

The European Commission is also examining how to adapt competition law for the digital era.

For more tools and resources visit
theodi.org/competition2020

Hot topics

- Identifying alternative data business models
- Increasing data portability and interoperability
- Increasing both data protection and data access to promote competition
- Defining market dominance and consumer harm when many digital products are nominally free
- International discussion of competition policy, and its interaction with other geo-political issues

Explore

- UK Digital Competition Expert Panel: Documents relating to the Digital Competition Expert Panel's review of competition in digital markets
- The Competition and Markets Authority: Online platforms and digital advertising market study
- European Commission: Competition policy for the digital era

04

Ethics

and responsible technology

People and organisations rely on data to make better decisions or to innovate – from improving how we travel to improving cancer diagnosis. But recently, headlines have focused on data controversies – such as the Cambridge Analytica/Facebook scandal – leading to justified public concern around how personal data is being used. How can we ensure that data – both personal and non-personal – is collected, used and shared with minimal harm, and that technology is developed and deployed responsibly?

Data ethics and responsible technology

Handling data responsibly is a moral and legal imperative. In 2019, the UK's Information Commissioner Elizabeth Denham said: "Across the world people have woken up to the importance of personal data and how it's used. Individuals should be the ones in control and organisations must demonstrate their accountability to the public."

Retaining and building trust is also a practical necessity. If people stop trusting organisations and withdraw consent for data about them being collected or shared, then essential research and services – public and private – could be stifled.

A recent YouGov poll by the ODI indicates that, while 87% of the UK public think it's important for organisations to use personal data ethically, most are unconvinced that they will. Organisations need to find a way to retain or build trust.

The good news is that organisations are starting to recognise the need to be ethical and trustworthy in how they use data to hold on to their clients and customers. Organisations like the Co-op, Nationwide and the BBC are embedding ethical data practices into their day-to-day approaches. Doteveryone found that tech workers may leave their companies if they are worried about the negative consequences of their work. Organisations that fail to recognise this risk may get left behind.

Hot topics

- Turning ethical principles into practice
- Public participation and other community engagement
- Understanding international, regional and local variation in data ethics
- Demonstrating trust-worthiness through certification and accreditation
- Regulation and the role of civil society to hold organisations accountable

Explore

- Information Commissioner's Office: Organisations must continue to improve trans-parency and accountability
- Open Data Institute: Nearly 9 in 10 people think it's important that organisations use personal data ethically
- Doteveryone: people, power and technology
- Open Data Institute: Trust in data is 'new currency' for the Co-op
- Doteveryone: consequence scanning

For more tools and resources visit
theodi.org/ethics2020

DoxBox trustbot (2019), Alistair Gentry
theodi.org/culture

ODI Summit 2019, theodi.org/summit

Infra structure

to support our economies and societies

Data is a new form of infrastructure that underpins every sector of our society and economy, but it is often broken and neglected. If data cannot be accessed, used and shared in a trustworthy and consistent way, then many essential services could be under threat. Open data is the foundation of this data infrastructure.

Infrastructure to support our economies and societies

Having a strong data infrastructure will only become more vital as our populations grow and our economies and societies become ever-more reliant on getting value from data to meet people's needs.

For example, weather data is being used by everyone from farmers to the transport industry to individual citizens, and mapping data is created and shared by the public sector and then built on by diverse organisations, from Wheelmap to Google. Data is infrastructure for our communities and nations and across each and every sector globally.

In 2014, Henri Verdier, France's Chief Data Officer, said 'Data is currently at the centre of both public action and economic activity and it must be seen as an essential infrastructure to the functioning of the economy, just as the transport and telecommunications networks. The Government must be the catalyst, encouraging the rest of society.'

Good infrastructure should be simply there when we need it, but – at the moment – too much of our data infrastructure is unreliable, inaccessible or only available if you can pay for access. Data innovators struggle to get hold of data and to work out how they can best use it, while individuals do not feel that they are in control of data about them. Data infrastructure should be as easy to use as our road networks.

Hot topics

- Creating digital twins and data infrastructure that provide real-time pictures and predictive models
- Identifying high-value datasets that should be publicly funded open data
- Building stronger data infrastructure within specific sectors, for example engineering, physical activity and agriculture
- Incentivising the private sector to contribute to data infrastructure
- Designing data institutions to steward our data infrastructure

Explore

- Open Data Institute: Principals for strengthening our data infrastructure
- Etalab: Data as an essential infrastructure
- National Audit Office: Challenges in using data across government
- National Infrastructure Commission: Data for the public good

For more tools and resources visit
theodi.org/infrastructure2020

06

Mis inform ation

disinformation and fact checking

Fake news and deepfakes are the tip of the iceberg in how technology enables and aggravates misinformation at speed and scale. As regulators struggle to keep up, misinformation poses a clear and present risk to democracy and trust in politics.

Misinformation, disinformation and fact checking

In a report released in 2019, the UK's Department for Digital, Culture, Media and Sport (DCMS) select committee said Facebook had put democracy at risk by allowing voters to be targeted with disinformation, and called for Facebook to be regulated. Leading academic in the field of internet law, Lilian Edwards, also said last year that deepfakes are 'increasingly likely to be used to discredit politicians and business leaders.' The boss of Sky, Jeremy Darroch, has also called on more 'urgency' in regulation.

The UK's Centre for Data Ethics and Innovation points out that deepfakes still require specialist skills and professional software, but are likely to become more sophisticated over time. Meanwhile, fact checkers face staggering challenges in fighting disinformation spreading at increasing speed and scale on social platforms.

At the ODI, we think that Facebook and other ad platforms cannot fix all the problems they create alone. They need to collaborate with civil society, regulators and researchers to design solutions together.

Hot topics

- Agile and cross-border regulation of misinformation
- Open data to support more effective fact checking
- Improving media literacy and critical thinking
- Managing misinformation at scale through automation
- Access to data to support research into misinformation

Explore

- Centre for Data Ethics and Innovation: Deepfakes and audio-visual disinformation
- Lilian Edwards: Turing Lecture 'Regulating Unreality'
- DCMS: Disinformation and 'fake news'
- Open Data Institute: Facebook scandal: let's turn our attention from how data is collected to how it gets used
- Full Fact: automated fact checking

For more tools and resources visit
theodi.org/misinformation2020

07

Rights

and ownership

Can we own data about us? Can we sell access to data that isn't just about us but about others? Are enhanced rights and responsibilities the key to greater control over how data about us is used? The debate between ownership and rights is not as straightforward as it seems.

Data rights and ownership

While policymakers have long assumed there is a naivety or lack of interest from the public about data issues, we have found people are increasingly concerned with how data about them is used. People want to feel in control and not just be resigned to poor or harmful use of data about them. But they don't always have the language to explain their concerns or enthusiasm.

The idea that we can own data about us is gaining traction in some quarters – but it may not be that simple. Data can be about multiple people; for example DNA about you is also about your family. If we started to 'sell' data about ourselves, what impact could this have on others related to or like us?

We have a wealth of rights and responsibilities over data, which if enforced already provide controls over how data about us is used. How can these data rights, built around individuals, be adapted to recognise the rights of groups, communities and societies?

Data rights may need to vary as they reflect how societies view the relationship between individuals, communities, corporations and the state. Data regulations and policies need to change as technologies evolve, and as our expectations of data maintainers and publishers – and ourselves as data consumers and producers – change too.

Hot topics

- International differences in narratives and ways of thinking about personal data
- Public participation in decisions about uses of data
- Personal data stores and other technologies that provide controls over data
- Personal data representatives and data institutions for delegating decisions about data
- The role of data rights and regulations in international trade

Explore

- RSA, Open Data Institute and Luminate: About data about us
- Open Data Institute: How do we 'own' data?
- The Royal Society: Data ownership, rights and controls: Reaching a common understanding

For more tools and resources visit
theodi.org/rights2020

Skills

engagement and data literacy

"Data-driven companies that focus on continuous learning will be more productive and gain a competitive edge."

— Accenture's Data Business Group

Skills, engagement and data literacy

Employees and companies recognise the value in working with data, but recent research uncovered that 85% of such projects fail and only 8% of leaders report complete satisfaction with the outcome. There are many reasons for this failure: from complicated technical implementations and overly rapid uses of advanced analytics, to weak and undefined strategies for connecting people with technology as part of company culture and practice.

This imbalance in data literacy skills neglects capabilities in areas like data governance, data ethics, social science, and opportunities for innovation that take full advantage of data. It also leads to a lack of long-term impact and increases ethical risks.

Leaders need to nurture the right balance of data skills across their organisation and support their employees in becoming more confident and trustworthy in making better decisions with data. As recently reported, 48% of employees frequently make decisions based on gut feeling over data-driven insight. Only 21% report being confident in their data literacy skills, and 37% believe data literacy training would help them be more productive.

While organisations build data literacy across their workforce, they also need to increase the soft skills of data scientists and create multi-disciplinary teams. Schools need to incorporate data skills and critical thinking into existing curricula, and governments need to create new opportunities for workers displaced through increasing automation.

For more tools and resources visit
theodi.org/skills2020

Hot topics

- Building long-term data literacy programmes within organisations
- Building soft skills and multi-disciplinary teams
- Retraining opportunities in data
- Increasing basic data literacy for everyone
- Developing data literacy and critical thinking in school children

Explore

- Harvard Business Review: Companies are failing in their efforts to become data-driven
- Open Data Institute: Businesses need diverse skills to get the most from data
- The Data Literacy Project: The human impact of data literacy

Caroline Criado Pereze, speaker at the ODI Summit 2019, theodi.org/summit

ODI training, Tanzania, theodi.org/training

Trade

productivity and international innovation

Data flows in international trade support considerable economic activity across the world, but increasing trust in them could boost innovation and growth even further. To take part in complex data sharing for trade in services and the development of frontier technology, countries will be competing on the quality of their national data infrastructure – the data that they have access to, how they share it, and their ability to enforce privacy regulations.

Trade, productivity and international innovation

Access to data and the sharing of it across borders has become central to trade and economic competitiveness discussions across the world. Chile, New Zealand and Singapore recently signed the Digital Economy Partnership Agreement, which puts data flows at the heart of their trade relationships. The European Union is encouraging international adoption of GDPR, and Japan has started the Osaka Track at the G20 for multilateral discussion of data standards in the digital economy.

The effects of restricting data flows has become a leading area of research. The European Centre for International Political Economy has been pioneering in the field, discussing how constraints on data and other aspects of digital exchange can slow down innovation and growth.

Restricting data flows – such as through data localisation that requires data to be kept within national borders – is often a question of a lack of trust in trading partners, and debates on how to create institutions that increase trust are picking up momentum.

There are examples of international collaboration to learn from as data policy for trade and competitiveness develops. The Financial Conduct Authority has pioneered the use of regulatory 'sandboxes', and 'fintech bridges', to help regulators learn from each other and make it easier for small and young firms to export their services.

For more tools and resources visit
theodi.org/trade2020

Hot topics

- Designing national data infrastructure to support international trade competitiveness
- Building in data flows and sharing to digital trade deals
- Understanding the impacts of data localisation
- Creating greater international regulatory cooperation
- Avoiding data extraction and exploitation in countries with low data and institutional capacity

Explore

- New Zealand government: DEPA text and resources
- Open Data Institute: What are the links between data infrastructure and trade competitiveness?
- Open Data Institute: Data flows in international trade: what are the governance options?
- ECIPE: Digital Trade Restrictiveness Index
- Financial Conduct Authority: Regulatory sandbox

Value

estimation, prioritisation and distribution

Many people now recognise how important data is to our societies and economies, but how to value data is still an area of active research. Being able to put a value to data should help governments and organisations know where to invest.

Value estimation, prioritisation and distribution

Traditional ways of valuing data – based on how much it costs to create or what it can be sold for – don't reflect the full value of data to society. Approaches like Deloitte's 'Assessing the value of TfL's open data and digital partnerships' report for Transport for London show how much more value comes from the actions and decisions data supports, and the innovation and economic growth of data reusers.
This means it's contextual: data is more valuable to a community when it is used to help meet their goals.

Much of the value of data arises from combining it with other data, whether that's analysing billions of web searches each day, the medical histories of millions of diabetes patients, or combining land use and weather patterns to identify where to site new wind turbines.

Being able to value data helps governments work out how to prioritise increasing access to data. The European Commission will this year identify high value datasets that should be made open across six categories – geospatial, earth observation and environment, meteorological, statistics, companies, and transport.

For organisations sharing data, being able to estimate value helps to justify investment in improving data quality or interoperability. It can also help to work out how to fairly share value between those who collect or maintain data and those who use it.

Hot topics

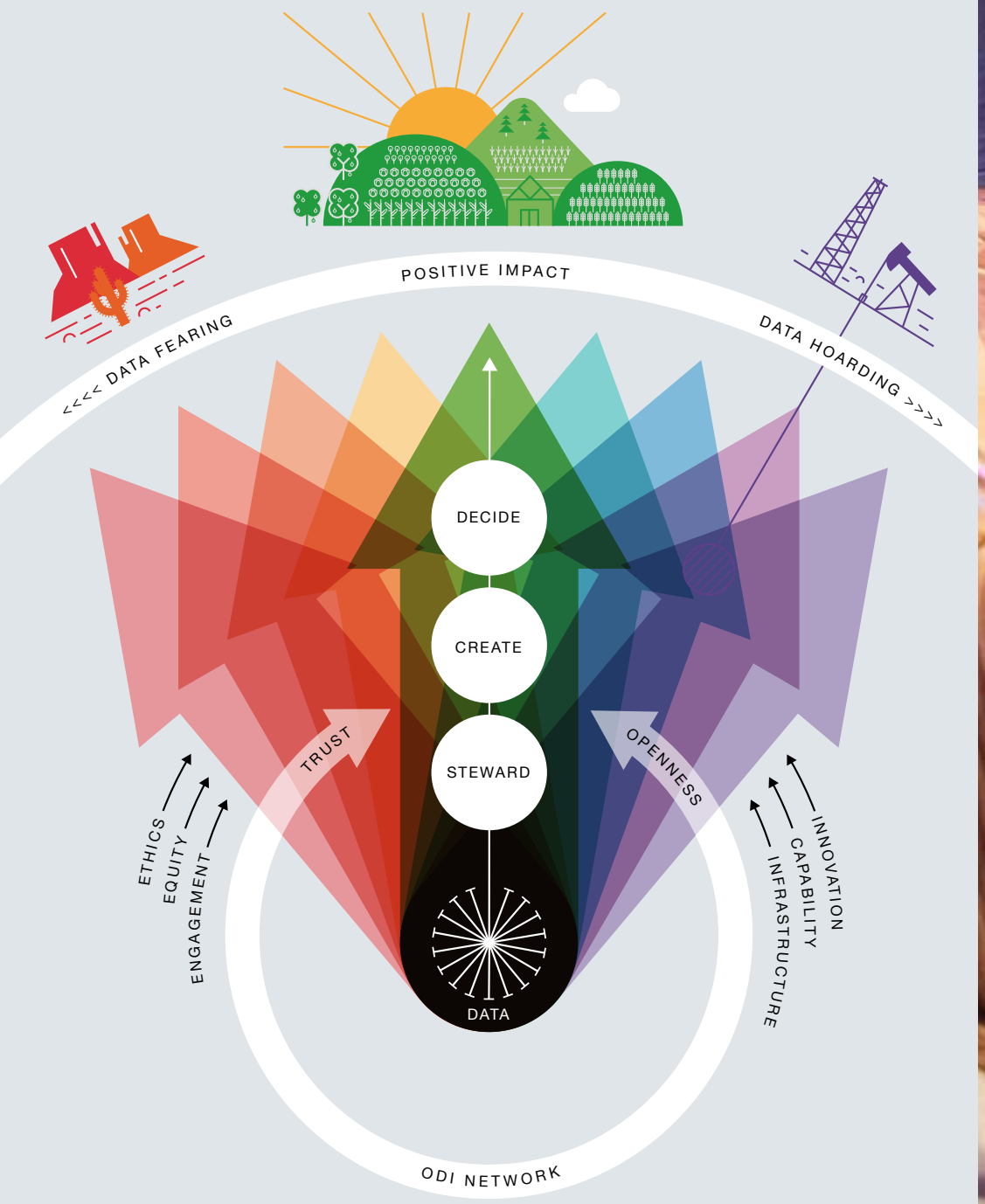
- Working out the current and potential value of a dataset
- Prioritising datasets for investment in particular geographies or sectors
- Classifying data to understand different types of value
- Incentivising investment in creating, using and sharing data
- Fairly distributing the value that arises from the use of data

Explore

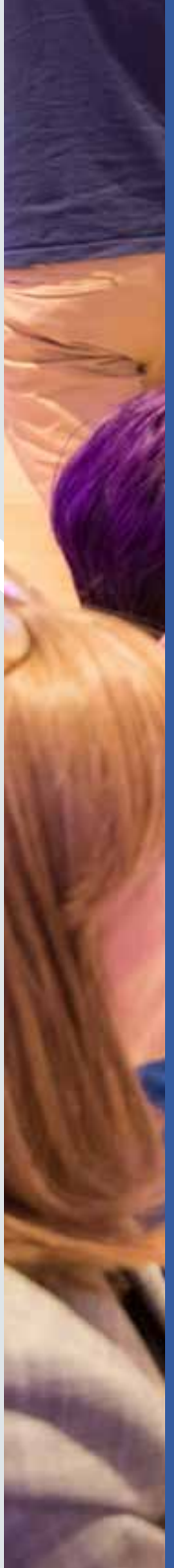
- Open Data Institute: Data's value and why we should measure it
- Open Knowledge Foundation: What data counts in Europe?
- Deloitte: Assessing the value of TfL's open data and digital partnerships report for Transport for London

For more tools and resources visit
theodi.org/value2020

ODI workshop, theodi.org/tools
Image source: CABI



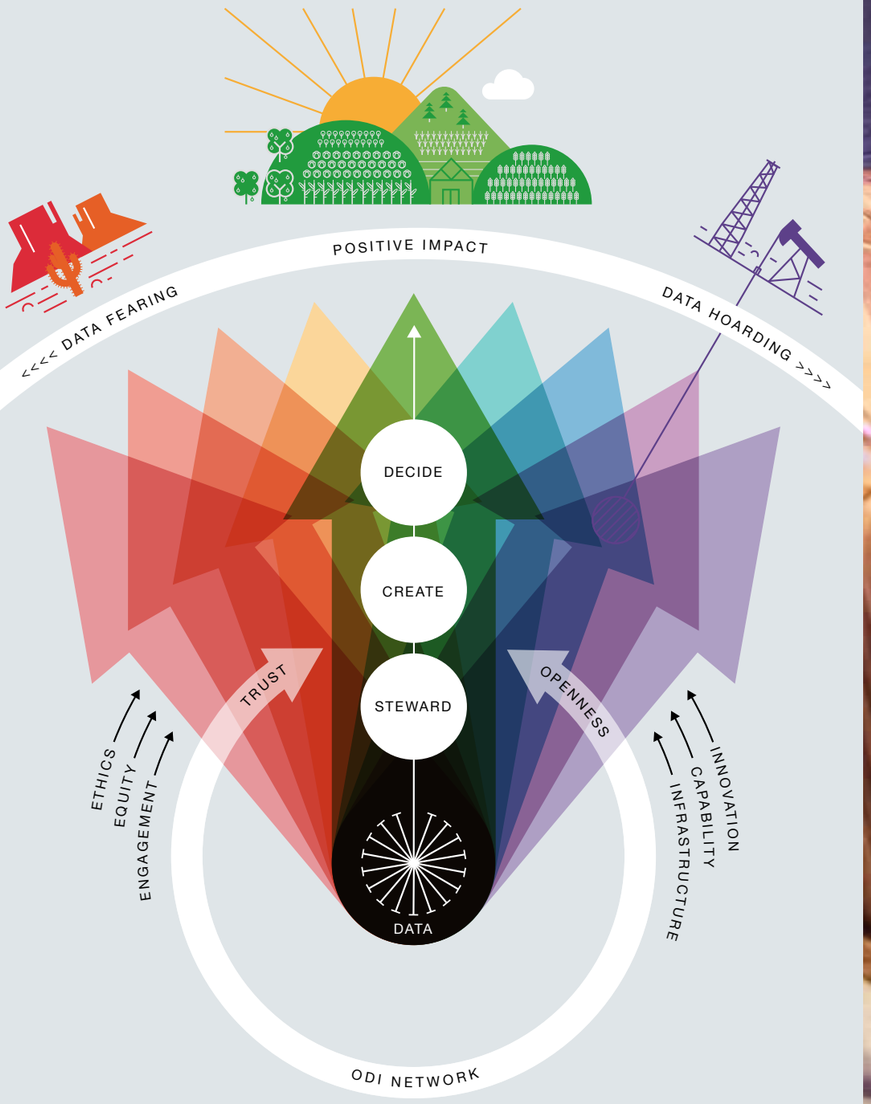
theodi.org/theory-of-change



You can find out more about our work and explore the resources relating to the topics set out in this book at theodi.org/data2020

Production and editing:
Hannah Foulds and Rebecca Ghani

Design: Adrian Philpott
@PHILPOTTdesign / 2zero3zero



theodi.org/theory-of-change

Datopolis at the ODI Summit 2019,
theodi.org/datopolis

AI

Collaboration

Competition

Ethics

Infrastructure

Misinformation

Rights

Skills

Trade

Value

theodi.org/data2020