Understanding the social and economic value of sharing data

March 2023

Open Data Institute

Contents

About

Introduction

There is consensus that the use and sharing of data can create social and economic value for people, organisations and for the wider economy.[1] However, there are multiple issues for estimating this potential value, and for designing mechanisms and incentives to make sure this value is realised and that ongoing sharing and 'fair value' exchange is supported.

In this report, we explore what levers organisations, and groups of organisations, can use to increase the value that is realised from sharing, or enhancing access, to the data they hold, or have access to. To do this we must understand how value is created from data. We must also understand the particular characteristics of data as an economic good and input for production, while also focusing on the infrastructure, assets and skills required to use and share data.

At the Open Data Institute (ODI), we interpret 'value' as ranging from economic value for people and companies, to the broad well-being of society. Data can generate value in multiple ways including giving monetary and non-monetary benefits. The ODI's theory of change groups activities that generate value from data into three categories: **stewarding data**, **creating insights** and **making decisions based on data**.

However, although data is widely regarded as valuable, organisations face multiple challenges when trying to understand how much value they can expect to generate from the data they hold or have access to, and how that value could vary when access to that data is granted to other actors. Therefore, we consider that there's a need to develop frameworks and tools that can support organisations, and groups of organisations, in assessing the potential value of the data they hold or have access to, and in identifying measures that can help 'unlock' it. This, in turn, requires developing ways to understand and navigate the different mechanisms and processes by which data, and particularly the sharing of data, create benefits for organisations and for society, and how these benefits are distributed. Additionally, we believe that developing these frameworks

---

[1] See The ODI (2021), 'Policy to unlock the economic value of data', and Bennett Institute for Public Policy & The ODI (2020), 'The Value of Data'.

and tools can help policymakers and industry groups explore the ways in which value-creating behaviours and practices can be sustainably incentivised and value appropriately shared.

Many of the issues associated with valuing and unlocking value from data arise from its economic characteristics. As laid out in the Value of Data report published by the Bennett Institute for Public Policy and the ODI,[2] data has particular characteristics that differentiate it from other goods or assets that organisations normally hold or have access to:

- Data can be used multiple times without being depleted, which makes it hard to know how it might be used in the future and thus to quantify its potential value over time.

- Data is 'non-rival', meaning that the same data can potentially be used by multiple actors at the same time without them needing to compete for access to it.

- Data may have different levels of excludability. Although it is non-rival in nature, in some cases some actors may have the capacity to restrict others from accessing or using certain data.

- Data often involves externalities. The use of data can produce positive or negative effects for actors beyond those who collect, process or use it.

- Data may have option value: its value may increase in the future if it can be put to productive uses that have not yet been thought of, or anticipated by, its current holders.

- Its value is strongly linked to access: all other things being equal, the more accessible data is, the greater the value that can be created from it.

It's extremely difficult to estimate or predict the uses to which data might be put and as a consequence, the total value or benefit that any individual actor or society as a whole could theoretically get from a specific dataset. Value can be created from data multiple times, for multiple stakeholders every time it gets used and reused, and its dimensions are highly dependent on the context in which the data is used and the infrastructure, assets (including, but not limited to, other data assets) and skills with which it can be combined. Clinical data that was originally collected to track a patient's progress, for example, can potentially become useful in the development of new health services or for health policy decisions once individual records are depersonalised, aggregated and combined with other sources of data.[3]

Therefore, when considering the value of data, rather than thinking

---

[2] Bennett Institute for Public Policy & The ODI (2020), 'The Value of Data'.

[3] The ODI and Roche (2021), 'Secondary use of health data in Europe'.

specifically about estimating an 'intrinsic' value of a given dataset or data point 'at rest', it can be more useful to focus on the value generated when a dataset is directed towards specific use cases. However, it's not possible to anticipate all the possible use cases that a given dataset could have, or the costs and benefits associated with all of them.

Instead of trying to estimate the value of a dataset beforehand, a productive way of approaching data valuation is by focusing on the levers that individuals and organisations can pull to increase value creation. This means focusing on the ways we can enable dataflows to ensure that data can be accessed by actors in a position to develop use cases that yield positive social and economic outcomes, and value flows so that the benefits are shared with those collecting and stewarding the data.

These characteristics also show that conceptualising data as an asset that organisations own and can trade may be counterproductive. Since the value of data is strongly linked to access, and it is a non-rival good that doesn't get depleted, it is more productive to think of data as an infrastructural good or resource, instead of as an asset that is 'owned' by private actors. Doing so helps emphasise the value that data can create for society rather than taking a narrow approach that focuses solely on the value that it can bring to those who collect or hold it.

To understand which levers organisations can use to enhance access and increase the value that can be generated for wider society from the data they hold, an organisation must understand the distinct economic characteristics of data listed briefly above. In this report, we explore these economic characteristics in more detail and how they influence decision-making when it comes to developing use cases, and to opening, sharing and accessing data, all of which are directly related to the value that can be generated with it. Then, we review some of the existing activities that organisations engage in to be able to realise the value from the data they hold. Finally, at the end of the report, we introduce a prototype of a Value of Data Canvas. The canvas is a framework designed to help organisations, and groups of organisations, structure and document discussions about the value that the data they hold can yield to them, and to society, and ways in which that value can be increased.

This report will be useful for:

- Organisations that steward data, and want to understand how they can increase the value they obtain from using it, sharing it or making it open.

- Groups of organisations interested in increasing the value they generate from the data within their ecosystems. For example:

  - Cross-sector alliances that want to leverage the data collected by their members, or by other organisations, to respond to social and environmental issues.

  - Groups of companies interested in tackling sector-wide challenges with data.

  - Trade unions and civil society organisations that want to use data to advance their common objectives.

- Researchers interested in the value of data and the economics of data from both a theoretical and a practical perspective.

- Policymakers interested in learning more about how data generates value, and how the contribution of data to local and national economies can be enhanced and quantified.

Data as an economic good

Data has particular economic characteristics that differ from those of traditional private goods or inputs for production. As a consequence, the levels of investment in data and the amount of data traded through markets may not be socially optimal. This section expands on data's economic characteristics and its consequences for value generation.

In recent years, data has become a highly sought after and highly-valued economic input. Organisations around the world are increasingly interested in developing ways to generate value from the data they hold, and to access new sources of data that can be useful to advance their economic goals. However, unlike most other inputs for production, such as physical capital, materials or labour, data is not a typical private good:

- Data is a non-rival good: it can be used multiple times and simultaneously by different actors without it being depleted.

- However, even though it is non-rival, it can be excludable: in certain cases, economic actors may have the power to prevent others from accessing or using certain data.

- Data has externalities, which means that it may create benefits or costs that are not easily assignable through market mechanisms.

- Data collection typically has high upfront costs and low marginal costs, meaning there are economies of scale in data production.

As a consequence, data markets are difficult to develop and maintain, and access to data – and the value that can be generated by enhancing access to data – is highly dependent on how governance frameworks incentivise or disincentivise data sharing and reuse.[4] In this section, we explore what implications the economic characteristics of data have for organisations when assessing the potential value they could create by increasing access to the data they hold.

---

[4]See SSRN (2021), Coyle, Diane & Diepeveen, Stephanie, 'Creating and governing social value from data', American Economic Association (2020), Jones, Charles I. & Christopher Tonetti (2020), 'Nonrivalry and the economics of data', and OECD (2022), 'Measuring the value of data and data flows'.

Non-rivalry and excludability

Data is a non-rival good. This means that a database or a data point can potentially be used and re-used by more than one actor at the same time, without one actor's use affecting or limiting the other's capacity to use it or depleting it. Non-rivalry means that there can be social gains of data being used broadly across organisations, potentially increasing its returns as its access and use increases.[5]

However, data can be excludable. This means that actors that hold data may have the capacity to decide who is able to access or use it. Being able to control access to certain data is one way in which organisations that hold data may be able to get value from it. For instance, they could charge others for access to the data or to services built using that data. The data holder's capacity to control access can lead to scenarios in which there's a gap between the value that private actors are able to create and capture from the data, and the value that society could get from it if access was increased. In cases in which society could benefit if access to data was increased, organisations holding the data may not have enough incentives to share or open it.

If there are no mechanisms to compensate holders for the value that sharing or increasing access to their data may create for other actors, the market power that actors can gain from controlling access may lead to data-hoarding scenarios.[6] This is especially true when the data that these actors hold can't be easily produced or collected by others. Certain kinds of data, such as mapping or environmental data, can be created or collected by multiple actors simultaneously. This makes those datasets less prone to monopolistic practices and creates incentives for those who collect it to develop value-added services that they can monetise. However, other kinds of data present characteristics that can make monopolistic practices more attractive for some. For example, data about consumer transactions in a specific store, data about internal administrative processes within a firm, or data about trips collected by ride-hailing platforms. These are all kinds of data that could potentially be valuable to multiple actors in an ecosystem, and create public good. Yet they are entirely excludable: those who collect it can perfectly exclude other actors from accessing it, as they won't be able to produce it themselves or get it elsewhere. Under these scenarios, without efficient

---

[5] American Economic Association (2020), Jones, Charles I. & Christopher Tonetti, 'Nonrivalry and the economics of data'.
[6] See The ODI (n.d.), 'Our theory of change'.

mechanisms to compensate data holders for increasing access to the data they hold, even when it's socially desirable, decisions made by data controllers to maximise their own private value can end up limiting the potential value to society.

This raises important questions around incentives and market or non-market mechanisms that can be designed to reward data collection and sharing for the value it brings to other actors and to society. But this also should be of interest to the data holders. Theoretically, as the value that society can realise from the data they control increases, they could potentially also increase the compensation they get from granting others access to it. This, however, requires the creation of legal, commercial or regulatory mechanisms that can redistribute part of that societal value towards actors that invest in collecting, maintaining, sharing and accessing data.

Open questions to explore in further research:

- How does the creation of value from data and its distribution vary when data is public or easy to collect by multiple actors versus when specific actors have the capacity to control who accesses it?

- What kind of mechanisms can be used to incentivise data controllers to increase access to the data they hold where doing so would be socially beneficial? How can data collectors and controllers be compensated for the value that opening up data would bring to an ecosystem? Are there any regulatory measures that can be taken to create the right incentives?

- How can we ensure a balance between private and public value is obtained from data collection, sharing and use?

Externalities

From an economic standpoint, an externality is a cost or benefit of an activity experienced by an actor that does not take part in the activity. A negative externality occurs when an economic activity creates costs that are not internalised by those taking part in it. On the other hand, a positive externality occurs when an economic activity creates benefits that are not entirely captured by those taking part in that activity. In both cases, externalities create market failures. When an economic activity causes negative externalities, then actors that engage in it may be incentivised to do more of it than what would be socially optimal, as they do not internalise the full costs. On the other hand, when an activity is associated with positive externalities, actors may produce less of it than is socially optimal, as they are not able to capture all the benefits it produces.

Data use commonly has population-level effects that are not entirely internalised by the actors engaged in collecting, sharing and using it. For example, increasing access to mobility data may have the consequence of making traffic in a city more efficient, reducing trip times even for those who don't access and use the data, such as commuters that don't regularly use mapping apps. Although these benefits can potentially be measured, they are not internalised entirely by the actors involved directly in collecting, sharing or using the data, nor in consuming data-enabled services. Because those who collect the data and those who use it are not able to internalise its benefits in full, when positive externalities exist the market will provide insufficient incentives to invest in data collection and sharing.

On the other hand, using data can also create negative externalities. This may come in the form of privacy costs: certain types of personal data can become valuable for society when shared and aggregated, yet doing so may affect people's privacy, creating harms (or costs) that are not internalised by those using the data. There's no easy way for the market to reallocate those costs to the people responsible for creating them. Neither the actors in charge of producing and sharing the data, nor those making decisions based on it, internalise the social costs created from its use.

Data externalities are closely linked to the relational character of data: data typically becomes valuable because of its power to establish relations or groups between different actors based on relevant shared features.[7] For example, a government may use data about people to group them by socio-economic status with the purpose of delivering more targeted public services, or a retail company may use data about people to group them according to their consumption patterns in order to improve their marketing efforts. Because of its relational character, from an economic standpoint it may be more appropriate to frame data as a collective

---

[7] The Yale Law Journal (2021),Viljoen, Salomé, 'A Relational Theory of Data Governance'.

resource and thus one for which market exchange will not be the best form of organisation.[8]

Open questions to explore in further research:

- How can data stewards or data institutions map and quantify the potential positive and negative externalities of increasing access to specific datasets?

- How can externalities be taken into account to design incentive models for organisations to increase access to the data they hold, favouring value maximisation at the ecosystem level rather than at an individual level?

---

[8] Daedalus (2022) Coyle, Diane, 'Socializing Data'.

Marginal costs of collection and sharing

When thinking about the value that can be created from data when access to it is increased, we need to consider the costs of collecting and sharing that data, and whether the value created from it can compensate for them.

As noted in The Value of Data, data collection often has high, fixed, upfront costs and low marginal costs.[9] Frequently, collecting data requires investing in assets, such as sensors or technical infrastructure, and once those assets are in place, the costs of collection tend to decrease as more data is collected. In addition, there are other costs associated with its processing and maintenance, such as data cleaning, storage, and the investment in the capabilities needed to use the data. Finally, data collection, use and sharing also creates financial and reputational costs associated with security breaches, which can push up costs through reduced revenue and legal fines.

However, for the purpose of making decisions about increasing access to data that has already been collected, the costs of collection can usually be considered 'sunk costs': costs already incurred by the organisation that won't be affected by future decisions. Instead, when evaluating whether to share data that has already been collected or is being collected, organisations need to pay attention to the marginal costs that they would incur in sharing it, and how those costs balance out with the potential value that can be obtained. Examples of costs associated with increasing access to data include:

- the need for further processing to remove any personal or commercially sensitive information,

- the development of standards and the technical infrastructure needed to enable access for external actors, or

- the investment needed for developing a data-sharing strategy.

A London Economics report into the cost considerations of data sharing suggests a taxonomy to classify data-sharing costs in four categories:

1. Strategy and planning costs,

2. Legal and regulatory costs,

3. Technical and organisational costs, and

4. Operational costs.[10]

---

9 Bennett Institute for Public Policy & The ODI (2020), 'The Value of Data'.
10 London Economics (2022), 'Research into the cost considerations of data sharing'.

It is important to add contingent costs to these direct costs, which can result in reduced revenue and legal fines. These include costs linked to the ways sharing data has the potential to result in financial and reputational harms from security breaches, and the ways sharing data may potentially affect the competitive position of an organisation.

When making decisions about which data to share, organisations should also consider the potential benefits that they could gain from it, alongside the benefits that could be created for society. As explained above, the benefits to be obtained by other actors gaining access to the data are not necessarily easily forecastable and may be hard to redistribute in ways that fairly cover the costs incurred by the organisation sharing the data. While it is socially desirable to create data flows that grant access to the actors most likely to generate the most value out of it, organisations collecting and sharing data also require mechanisms that ensure part of the social benefits of increased access to data can be used to compensate for their investment.

Open questions to explore in further research:

- How can organisations develop sustainable models for stewarding and sharing data that cover the fixed and variable costs of data collection and maintenance? What role do costs play in the design of incentive systems for data ecosystems?
- How can data ecosystems adequately compensate for the costs incurred by data collectors and controllers?

Enabling flows and creating value from data

This section outlines the journey that data follows from collection to value realisation. In doing so, it highlights the key issues that data providers and users face along that journey, and the processes and investments that needed to generate value from data.

The value of data is strongly linked to access. As data becomes more accessible, the value that is generated from it usually increases as well. Because of this, and because of the characteristics outlined in the previous section, markets are not the most efficient way to increase access to data. Treating data as an asset to be traded in markets leads to suboptimal social outcomes, as it may not create enough incentives for data holders to increase access to the data they hold. Instead of treating data as a tradeable good that is subject to market exchange, it may be more productive to conceptualise it as a collective resource and focus on developing processes and governance arrangements to regulate access and use.[11] Moreover, because of the economic characteristics of data, data markets don't naturally develop: according to the OECD, in 2019 only 2.1% of firms with more than 250 employees were selling data, and only 4.6% purchased data, and the percentages were lower for smaller firms.[12]

Increasing access to data in order to drive value creation remains a challenge that requires innovation and active participation from private and public actors.

By mapping the key steps in the data value journey, from data collection to management, sharing, use and reuse, this section aims to identify the key activities that data stewards need to engage in to increase access to their data and generate value from it.

---

[11] Daedalus (2022) Coyle, Diane, 'Socializing Data'.
[12] OECD (2022), 'Measuring the value of data and data flows'.

The stages that data goes through in order to generate value can be understood in terms of the ODI's theory of change (Figure 1). It identifies three groups of activities that create impact from data:[13]

- **Data stewardship:** collecting, maintaining and sharing data.

- **Insight creation:** creating valuable information from that data, in the form of products and services, analyses and insights, or stories and visualisations.

- **Decision-making:** making decisions based on information from multiple sources along with experience and understanding.

Figure 1. The ODI's theory of change

---
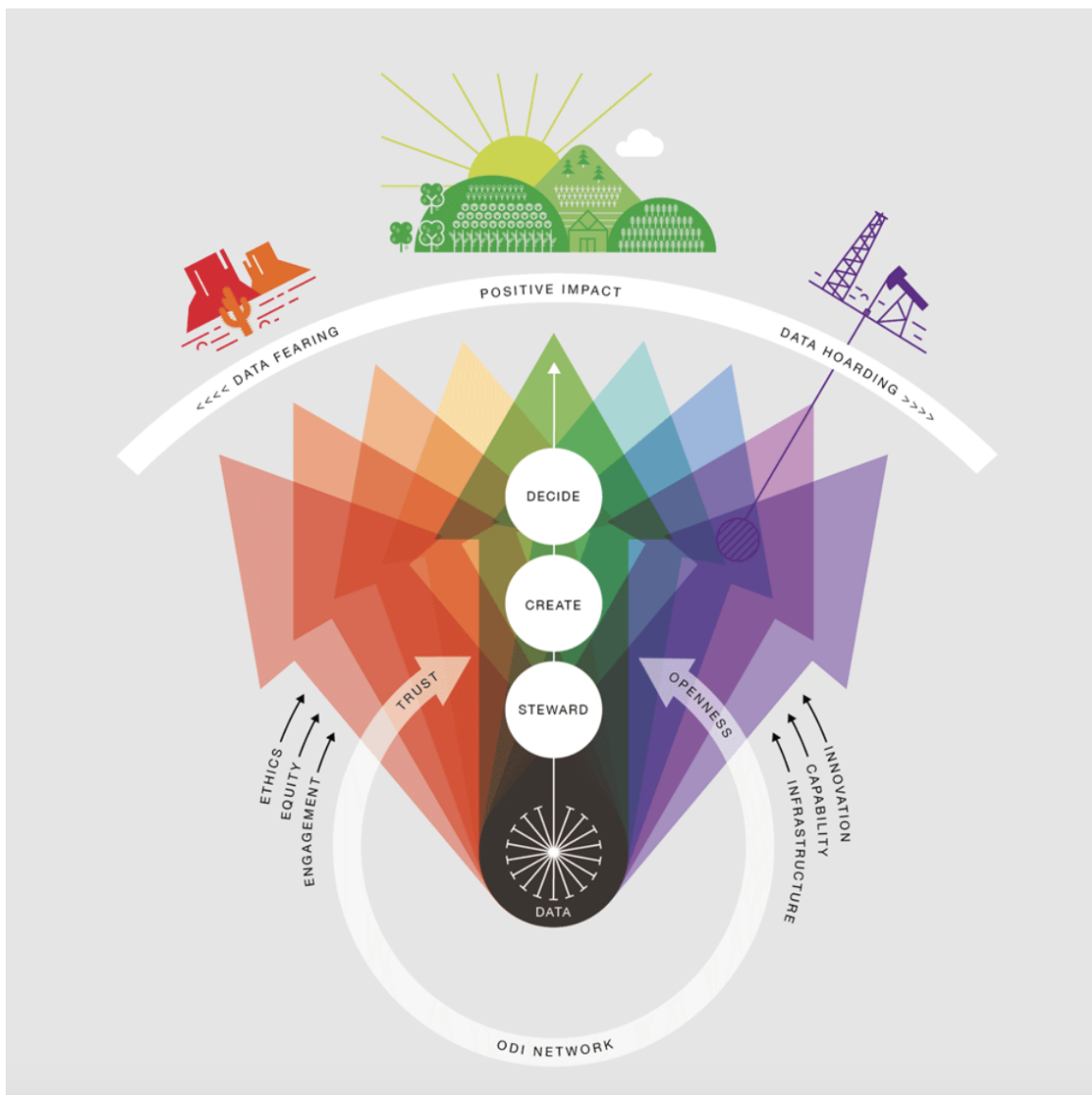
[13] The ODI (n.d.), 'Our theory of change'.

A more granular chain, focussed specifically on sequentially mapping the steps through which data value becomes incrementally tangible, was proposed by data scientist Chloe Mawer:

Figure 2. Data valuation chain

Source: Mawer (2015)[14]

Mawer shows that data valuation increases as it moves through this chain, and that the value of data in intermediate steps is estimated based on the potential value that could be realised at the end of the chain. However, the total value remains uncertain until the chain is completed.

While this chain is useful to understand how value becomes closer to being realised as data is processed, integrated and analysed, it doesn't fully take into account some steps that happen when data is shared or opened. That is, when some steps of the chain are conducted by different organisations. Combining Mawer's approach with the terms used by the ODI, we've mapped the key activities required in order to generate value from data when sharing or opening it. This journey will allow organisations to better understand what they need to do in order to maximise the value that can be generated by sharing or increasing access to the data they hold.

For data to become valuable it often has to be transformed by data holders or stewards first – actors who change it from its original form into a format better suited for use. If an organisation is interested in generating value for their wider ecosystem, either by granting access to other actors within their ecosystem or publishing its data openly, they must consider ensuring that the data can be trusted and easily reused, not just within their organisation but also by other organisations. One way to do this is to make it findable, accessible, interoperable, and reusable (FAIR).[15][16] Additionally, when sharing data, it is important for the data holder to understand the risks associated with it and take action to manage and mitigate these risks.

[14] Silicon Valley Data Science, (2015), Mawer, Chloe, 'Valuing Data is Hard'.
[15] The ODI (2023), 'Data assurance: what is it and why do we need it?'.
[16] GO FAIR (2016)), 'FAIR Principles'.

Finally, they need to determine a suitable model for safely granting other actors access to the data. It's also important to establish mechanisms that allow the value created by different actors at each stage of the chain to be redistributed, taking account of cost incurred. This requires business models to be developed that can sustain these data sharing activities, along with commercial agreements or terms that facilitate access to data and ensure those who benefit from access can adequately compensate those who publish, or provide access to, that data.

Assuring data and making it FAIR

At the ODI, we define data assurance as 'the process, or set of processes, that increase confidence that data will meet a specific need, and that organisations collecting, accessing, using and sharing data are doing so in trustworthy ways'.[17]

Trust is essential to realising the full value of data. Research carried out for the ODI by Frontier Economics in 2021 found that trust in datasets, in the institutions that steward them, and in the wider data ecosystems, results in increased data availability and therefore in increased economic performance.[18]

Data assurance can be used to improve confidence that data is of trustworthy provenance, that data meets ethical expectations or requirements, or that data is of appropriate quality. Additionally, the data practices of an organisation can also be assured to ensure that data is relevant and applicable for the intended use, that it is shared with the right people, or that there are no harmful impacts on people, organisations and communities arising from its collection, processing or use.

In addition to using data assurance to improve trust, there are technical considerations when organisations open or share data as that data needs to be reusable by other organisations. The FAIR data principles offer a good standard for doing so.[19] FAIR stands for: Findable, Accessible, Interoperable and Reusable. Making data FAIR requires preparing adequate metadata to ensure that data is easily processable by computational systems with little human intervention, and thus that it can be more easily reusable by different organisations without requiring much extra processing.

[17] The ODI (2023), 'Data assurance: what is it and why do we need it?'.
[18] Frontier Economics for the ODI (2021), 'The economic impact of trust in data ecosystems'.
[19] GO FAIR (n.d.), 'FAIR Principles'.

Assessing risks associated with data sharing

A common concern for organisations looking to share data is how to ensure that doing so will not generate negative impacts for themselves or for society. Assessing risks associated with sharing is an important step before making decisions about what data to share and how to do it.

Most risks associated with data sharing can be grouped into four categories:

- **Legal and regulatory:** Perceived or actual risks of breaching data protection law, intellectual property rights, other regulatory requirements or legal contracts

- **Ethical:** Perceived or actual risk of enabling unethical data collection or use, or of directly impacting people and communities

- **Reputational:** Perceived or actual risk of suffering reputational damage from sharing or using data that breaches trust, or that reveals limitations in processes or analyses

- **Commercial:** Perceived or actual risk of losing competitive advantage in the market

The ODI's guide on assessing risk when sharing data contains a more detailed explanation of each risk category, as well as guidance on how to identify, assess and manage risks.[20]

Developing data access models

From a technical standpoint, access to data can be increased in multiple ways, from APIs and blockchains to competitions and data institutions. Identifying which approach best suits the needs of a given organisation or group of organisations is not a trivial task. Deciding how to increase access to data depends on the kind of data that will be shared and the risks associated with it, the actors with whom data will be shared, and the purposes for which it may be used.[21]

---

[20] The ODI (2022), 'Assessing risk when sharing data: a guide'.
[21] See The ODI (n.d.), 'Mapping the wide world of data sharing.'

There are four key aspects of the data access models that data holders control through a variety of mechanisms. These are:[22]

1. **what** data can be accessed

2. **how** that data can be accessed

3. **who** can access the data

4. what that data is **used for**

Mechanisms to control *what* data can be accessed and how it can be accessed tend to be technical in nature. These can include things such as modifying data to make it less sensitive, aggregating it, or selecting specific portions of the data to be shared. In terms of controlling *how* data is accessed, organisations need to decide what technical features the access model will have: users can have access to the data through direct transfer, through data streams and APIs, through interfaces or platforms, through secure environments or others. Different levels of access to data and different mechanisms to access data may be more or less suited to different use cases, and convey different levels of risk

On the other hand, *who* can access the data and what that data is *used for* are normally controlled through legal, commercial and decision-making mechanisms and processes. These include things such as contracts, data-sharing agreements, access pricing mechanisms and policies that limit what kinds of users can access the data. Some of these mechanisms, such as contracts, data-sharing agreements or pricing mechanisms, are key to ensuring that part of the value generated from increasing access to the data can be directed towards those who collect and share the data. For organisations stewarding data, understanding the potential value that the data they hold could generate if opened or shared can help to establish what kind of compensation they should be able to expect from those who have been granted access to that data.

Figure 3. The Data Spectrum
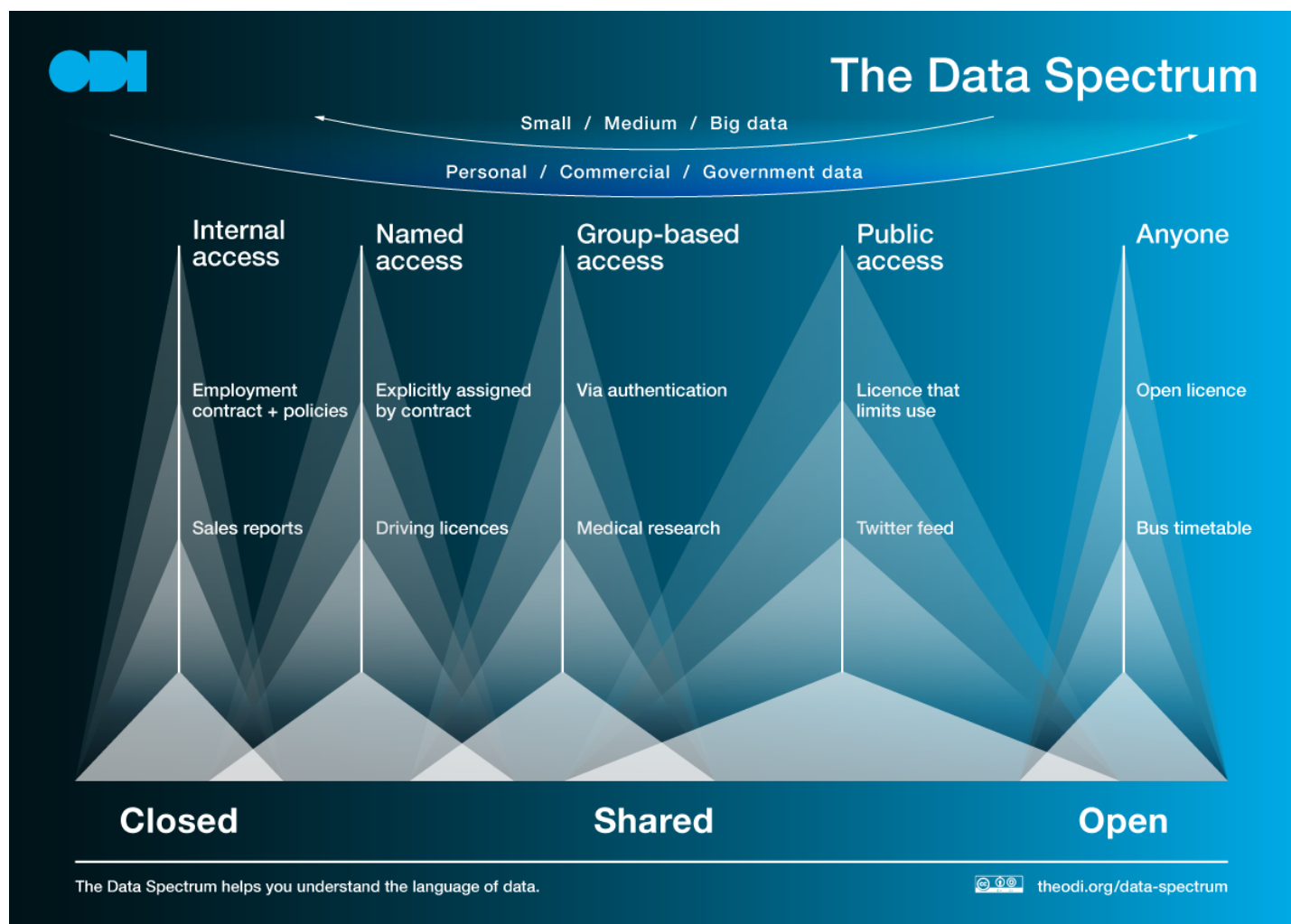
Source: theodi.org/dataspectrum

[The Data Spectrum](#) (Figure 3) can help people and organisations consider how widely data can or should be shared with their partners, stakeholders and ecosystem. As data moves along the spectrum – from closed, to shared, to open – an increasing number of people will have access to it, and the purposes for which the data can be used may also increase. As this happens, the potential to create value and impact for people, society, the economy and the environment rises. However, as data is shared more widely, the risks associated with it may also increase, especially in the case of sensitive data.

Assessing the potential value of datasets for an ecosystem

Thinking about the value of data from an ecosystem perspective allows us to understand and explain better where and how the use of data creates value, how data flows enable value creation, and how that value in turn flows across different providers and users of data.

The term data ecosystem refers to 'the people, communities, and organisations that are stewarding data, creating things from it, deciding what to do based on it, influencing any of those activities, or are affected by any of those activities.'[23]

As mentioned earlier, the value of data is highly dependent on access. Therefore, understanding the ways that value is generated requires us to understand how actors are able to access it, or not. Thinking about the value of data from an ecosystem perspective allows us to emphasise precisely the role of dataflows in the process of value generation. By taking this approach, we can visualise how the relationships between different actors are instrumental in value creation, and identify opportunities to increase value creation for the whole ecosystem or to improve value redistribution among different stakeholders.

The ODI has developed a methodology and tool to map data ecosystems[24] in order to understand how data creates value. A data ecosystem map identifies the data, data stewards and data users; the different roles they play; and the relationships between them.

Ecosystem mapping is an essential part of understanding value creation from data; in this section we want to propose two additional emerging frameworks that can help complement it. First, we present a two-by-two matrix framework, to help organisations and groups of organisations understand how different use cases for data may be categorised at the ecosystem level in order to prioritise decision-making. In this framework,

---

[23] The ODI (2019), 'Data Ecosystem Mapping tool'.
[24] ibid.

we propose that organisations or groups of organisations taking an ecosystemic approach need to balance value generation with value distribution. In the last part of this section, we introduce an early version of a Value of Data Canvas: a table designed to support organisations in understanding how the data they hold could generate value when shared or opened, and how the benefits can be distributed among different actors.

Value creation and value distribution

Generally speaking, use cases for data that require the participation of more than one stakeholder – such us use cases that require one actor to contribute data and another actor to process it, or use cases that require aggregating data held by multiple stakeholders – have the potential of increasing the total value that can be generated from data. However, these use cases also present a higher level of complexity as they require coordination and some level of alignment between different stakeholders that may or may not have competing interests.

Some use cases may generate enough value for each individual stakeholder to incentivise their participation, while other use cases may not. In the latter cases, organisations need to assess whether any measures can be put in place to ensure that the value generated from the use case is redistributed in a way that incentivises all relevant stakeholders to participate.

If it's possible to map who might benefit from a specific use case, and who needs to incur costs or contribute resources for that use case to be developed, then it should be possible to estimate whether the total benefits of a given use case would outweigh the costs for any party. However, the process of ensuring that the benefits are distributed in a way that creates the right incentives for actors that incur costs may sometimes in itself require incurring additional costs. For instance, when it means new strategies, governance mechanisms or infrastructure need to be developed to redistribute the benefits. Therefore, when assessing potential use cases at an ecosystem level, it is important that organisations not only assess the total net value that might be generated from the use cases in the abstract, but also the way in which that value is distributed among contributors.

Figure 4. Ecosystem use cases prioritisation matrix

	Low-value use cases	High-value use cases
Well-distributed benefits	Easier to implement, but lower value overall	High priority
Poorly-distributed benefits	Unlikely to be developed	Require coordination of efforts to ensure all participants are incentivised

We propose a simple two-by-two matrix (Figure 4) that can help organisations or groups of organisations prioritise use cases at an ecosystem level in order to guide decision making. The matrix divides the use cases into high-value use cases and low-value use cases, and into use cases with 'well-distributed benefits' and those with 'poorly-distributed benefits'. We consider use cases with well-distributed benefits to be those in which, without any additional effort, there would be sufficient benefits of participation for actors taking part in developing them to incentivise all or most of the actors involved to take part. Use cases with poorly-distributed benefits, on the other hand, are those in which some of the actors that are needed to participate in the use case implementation are not likely to be sufficiently incentivised by the benefits on offer, without specific mechanisms designed and implemented to do this.

The matrix helps map and group use cases in four groups:

1. **High-Value use cases in which the benefits are well distributed** should require relatively little intervention beyond just getting the right stakeholders around the table and agreeing on developing them. These cases should be prioritised.

2. **High-value use cases in which the benefits are not adequately distributed** may require some extra work including setting up governance frameworks, commercial agreements, or subsidies that create the right incentives. However, as a starting point it can be argued that society (or a given sector or industry) would benefit if they were developed. In addition to a total cost-benefit analysis of developing the use cases in the abstract, it is necessary to assess what coordination efforts are required to ensure all relevant stakeholders participate.

3. **Low-value use cases in which the benefits are well distributed** are use cases that may not be prioritised because of their low total contribution of value. However, because they require relatively little

additional coordination, they are more likely to be developed once the ecosystem is mature enough, or when most high-value use cases have already been developed.

4. **Low-value use cases in which the benefits are not well distributed** are unlikely to be developed, as the additional effort needed to make sure all participants contribute to the use cases may not justify the total value likely to be generated.

This subsection has proposed a way of classifying and ranking use cases at the ecosystem level by looking at their value contribution and at their possibilities for value redistribution among participants. However, organisations must also be able to assess, individually, how their data may contribute value to the wider ecosystem, and the ways they may be able to benefit from those contributions. The next subsection presents an early prototype of a canvas designed to support organisations in making those assessments.

The Value of Data Canvas

In order to help organisations and groups of organisations understand how the data they hold could generate value when shared or opened, we've developed an initial version of what we're calling the Value of Data Canvas (Figure 5). Similarly to other ODI guides and tools such as the Data Ecosystem Mapping tool and the Data Ethics Canvas, the purpose of the Value of Data Canvas is to structure and document conversations and guide decision-making. The canvas challenges organisations and teams to ask, discuss and document answers to questions about the data they hold, the value it contributes to their ecosystem, and how that value could be distributed across the organisation as well as the data ecosystems it is part of. This will be useful for organisations or groups of organisations interested in assessing the potential value that could be generated from the datasets in their ecosystems when opened or shared, and when discussing ways in which to enable and incentivise that value creation.

Figure 5. Value of Data Canvas

	Data collection	Data uses	Benefits	Complementary assets, skills and datasets	Costs and risks	Data access model
Organisation (Know your data)	• Name the key data sources and datasets you want to assess. • Are these datasets something that other organisations could also generate?	• What is your organisation currently using this data for? • What other uses could this data have for your organisation?	• What benefits is this data currently giving to your organisation? • What benefits or efficiencies could your organisation potentially get if access to this data is increased?	• What complementary assets or skills could your organisation invest in to increase the benefits it can obtain from this data? • What complementary datasets could increase the value of this data?	• What are the costs associated with the current and potential use cases of this data? • What risks does the collection, storage and use of this data create?	• What parts of the data should your organisation share? • How should your organisation share the data?
Ecosystem (Know your ecosystem)	• What actors could potentially be interested in collecting or accessing this data? • What actors, if any, could be in a position to produce this data by themselves?	• What uses could other actors give to this data if access to it was increased?	• What benefits or efficiencies could be gained within the ecosystem if more actors were granted access to this data?	• What complementary assets, skills and datasets are present within the ecosystem? Who has access to these assets, skills and datasets?	• What costs would your organisation or other organisations have to incur to increase access to this data? • What risks could arise as a consequence of increasing access to this data?	• Who should access the data? • What kind of access could be granted to other actors?
Incentives and value flows (Understand how value flows)	• What would other actors need to invest in to get this data or similar data elsewhere?	• What is keeping your organisation from increasing access to data that could potentially be useful for other actors?	• How can you quantify the benefits of using and sharing this data? • How are the benefits distributed across different actors of the ecosystem?	• How can the value of complementary assets, skills and datasets be quantified? • What are the factors that catalyse or limit their development?	• How do the costs of using and increasing access to the data balance out with potential benefits? • What actions can your organisation take to compensate for these costs and mitigate the risks?	• What mechanisms can be put in place so that the benefits of increasing access can compensate the organisation sharing the data for the cost of doing so?

The canvas consists of a matrix with three rows and six columns. Each row focuses on a different layer of the ecosystem. The first row invites teams to look into their own organisation and the data it holds and discuss things such as:

- How is the data being collected?

- How is that data useful for the organisation?

- How can value be generated from it?

The second layer invites teams to look at the ecosystem in which their organisation currently shares or could share data, and the actors within that ecosystem. It should support teams in thinking about which actors could be interested in accessing the data and what benefits could they get from it. Finally, the third layer moves the focus towards the incentives and value flows that exist or could be created within that ecosystem. It should support the discussion of questions such as:

- Who is capturing the added value that is created when access to data is increased?

- Who is incurring costs for collecting and sharing that data?

- How can value be fairly exchanged?

Conversely, each of the six columns of the canvas look at different aspects or parts of the journey of value creation and value distribution. The first column is useful to map and address key questions about the data itself and its process of collection – not just how data is currently being collected, but also what other similar data exists or could potentially be collected by someone else, as this is key when thinking about how value is distributed. The next column then challenges the teams to think about the current and potential use cases for the data:

- What is this data useful for or

- What could it be useful for?

As the canvas progresses further to the right, teams should also start mapping the economic benefits associated with using or sharing data, and thinking about how different use cases may yield different benefits.

After this, it invites teams to think about how the context plays a role in enabling the generation of value from data:

- What complementary assets or skills can make this data more valuable for organisations?

- Who has those skills?

- What complementary datasets (that is, datasets that become more valuable when combined) do we know of?

- What are the costs and risks associated with using and sharing this data?

Finally, after working through these questions, the last column invites organisations or groups of organisations to think creatively about how to implement processes to increase access to the data. This stage is designed to help prompt organisations to start documenting which key elements should be taken into account for future decisions. By the time teams reach this column, the questions they have discussed and documented in the previous columns should have given them valuable insights that inform the kind of data access models they could develop. In this column, teams can then start discussing the specifics of what a data access model that enables value generation could look like by asking questions such as:

- What parts of the data should be shared?

- How should your organisation's data be shared?

- What kind of access is valuable for actors who would use the data?

- What mechanisms can your organisation establish to ensure that data providers are compensated fairly?

The goal of this column is not to design a fully fleshed-out access and business model, but rather to help organisations start discussing what key features such models should have.

This is an early version of the Value of Data Canvas (Figure 5), and we expect to continue refining and iterating as we try it out in different projects, and with different organisations, across multiple sectors. We've also published it as a document open for comments along with a guide to help organisations in trying it out independently. The guide explains in more detail the recommended flow to use the canvas, noting what kind of discussions each section of the canvas should trigger, what elements are important to take into account in each section, and what other resources would be useful to help answer the questions in each section. Both resources can be accessed here.

Next steps

The ODI carried out this project to explore how we may leverage existing research to support organisations and groups of organisations in understanding and unlocking the social and economic value of the data they hold. We will continue to actively investigate these issues, and are interested in collaborating with industry, governments and civil society to continue working in this field.

As laid out in the report, there are still multiple aspects related to data valuation and the value of sharing data that require further research. In 2023, we will welcome a cohort of ODI Fellows that will focus on aspects related to the value of data to continue exploring this topic. Additionally, we've created an open and collaborative living repository of research resources and case studies, which can be accessed [here](#).

Finally, the version of Value of Data Canvas presented in this report is an early prototype which we expect to continue developing and iterating through collaboration with industry, civil society and governments. It's also been published as an open document, along with a guide, which can be accessed [here](#). If you or your organisation are interested in exploring how this research can support your goals, reach out to us by writing to research@theodi.org.