

Infrastructure for common challenges

User needs report

Contents

Contents	1
About	2
Introduction	4
Exploring data access initiatives	5
How do we define data access initiatives?	5
Sketch	7
What frameworks can help design data access initiatives?	7
Systems thinking	7
Models of behaviour change and COM-B	9
Logic models	10
What did we learn from our exploratory research?	12
Data landscaping for initiatives	14
A note on terminology	15
How are data initiatives understanding their data landscape?	15
What tools and methodologies exist to support data landscaping?	16
Purpose of the tools	16
Multi-stakeholder approach	17
Ecosystem engagement	17
How do existing tools help initiatives with aspects of data landscaping?	18
Data ecosystem mapping	18
Developing logic models	18
Data Ethics Canvas and Consequence Scanning	19
Using the ‘Open standards for data guidebook’	20
How are the results of data landscaping exercises published and shared?	20
Main outputs and complementary tools	20
Ecosystem engagement	21
Impact assessment	21
How does data landscaping inform the design of initiatives?	21
What did we learn from our research?	22
Next Steps	24
Methodology	25
Research Methods	25
Known limitations and mitigations	25
Appendix	25
Table 1: Data landscaping tools	27
Table 2: Data landscaping outputs	29

About

This report has been researched and produced by the Open Data Institute, and published in January 2021. Its lead author was Leigh Dodds with contributions from Elea Himmelsbach, Emily Sinclair, Flor Serale, James Maddison, Jeni Tennison, Josh D'Addario and Olivier Thereaux. If you want to share feedback by email or would like to get in touch, contact the 'Infrastructure for common challenges' project team at research@theodi.org.

To share feedback in the comments, highlight the relevant piece of text and click the 'Add a comment' icon on the right-hand side of the page.



How can it be improved? We welcome suggestions from the community in the comments.

Introduction

Communities, societies and industries across the world face challenges. Some are at a global scale, such as the Covid -19 pandemic, while others are local such as municipal disaster response. Many of these challenges require, in part, better or more appropriate access to data.

[Data access initiatives](#) are joint activities and programmes of work that address a specific challenge by improving (which may mean increasing or decreasing) access to data, and by building or strengthening [data infrastructure](#).

As part of the Open Data Institute's '[Data infrastructure for common challenges](#)' project we have been exploring how these initiatives are developing the data infrastructure they need to tackle specific challenges. Our research so far has focused on the needs of those who are designing new data access initiatives.

People leading data access initiatives need to understand the current data landscape to help them prioritise and plan the initial activities for their initiatives. This data landscaping work often requires a range of activities including stakeholder mapping and engagement, creating data inventories, and understanding user needs.

Programmatic design tools such as data ecosystem mapping and logic models can help people leading data access initiatives to understand and address these challenges. Many initiatives use these tools, but do not fully exploit the opportunities they can offer.

This report is a summary of our initial user needs research on data access initiatives. It reviews design tools and outlines how different data access initiatives have embedded them into their design.

Our intention is to share insights into the challenges these initiatives encounter and how these tools can be useful to help them build the infrastructure they need to be successful. This report is not a set of recommendations, but represents our desire to 'work in the open': to share the approach and what we have learned, summarise our research to date and, more importantly, to collect external feedback on our insights and approach.

Exploring data access initiatives

The project began with an initial exploratory research phase which we used to develop a broader understanding of the scope and activities of a range of existing data access initiatives.

Our goal for this phase was to develop familiarity with a wide range of initiatives and identify a focus for further work.

Within this phase, we:

- **explored the concept of data access initiatives** to refine our understanding of the term and consider its relationships to data infrastructure, data ecosystems and data institutions. This helped to situate this project in the wider context of the kinds of interventions policymakers and others are making around data
- **identified and researched a range of data access initiatives** that conform to our definition, with a particular focus on those in the health sector and working around climate action. This helped us to understand the range of activities that initiatives were undertaking and how they described their work.
- **interviewed users from individual initiatives** to learn more about how they have scoped, planned and delivered their work. This helped us to understand the issues that initiatives face and where additional support and guidance might be useful
- **researched and explored a range of frameworks for designing programmes**, to identify how they might be used to frame the activities and work required to address common challenges.

We summarise some of these activities in the following sections.

How do we define data access initiatives?

As part of our work exploring data infrastructure for common challenges, we have been using the term 'data access initiatives'. We define a data access initiative as **a collaborative programme that focuses on collecting, using and sharing data to address a social, environmental or economic challenge**.

To differentiate these from other projects and initiatives, we characterise data access initiatives as:

- Initiatives or programmes that have **a clear challenge**, in the form of a specific social, environmental or economic problem that is the focus for the collaboration
- Involve **multiple stakeholders** that are actively working together to solve the problem

- Include a strong **focus on collecting, using and sharing data** as part of their work.

There are many types of alliances and partnerships that collect, use or share data but aren't tackling a specific problem: professional networks or data sharing platforms, for example.

There are also challenge-led initiatives in which data – or information flows – is not a strong element of tackling the challenge. All initiatives will need information to help understand a problem or to measure progress against its objectives. But using data for monitoring and understanding a challenge isn't the same as using data to tackle the challenge. Sometimes other interventions such as building skills, understanding or reallocating resources may be the best option.

We have chosen to focus on data access initiatives involving multiple stakeholders. The challenges they address tend to be more complex and have a higher impact, and sharing and publishing data is of particular importance when there is a need for cross-organisational collaboration.

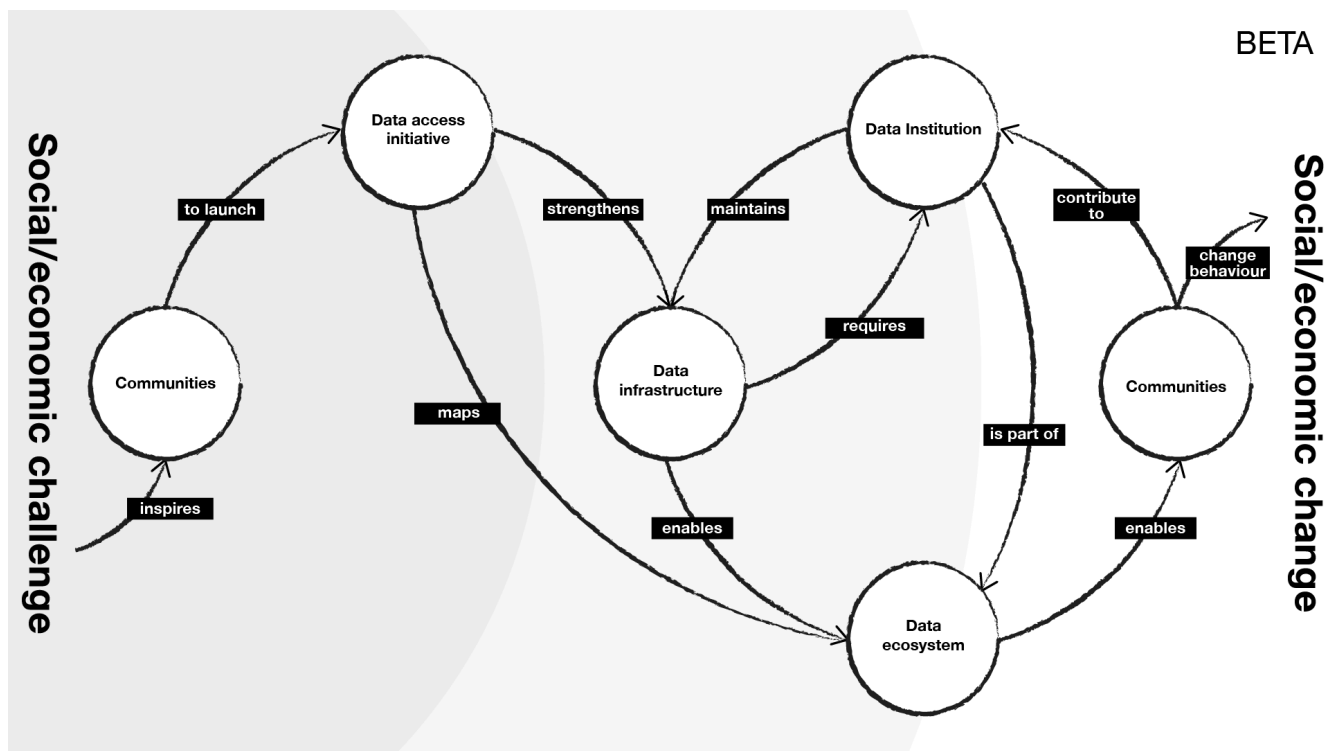
Our initial review highlighted that data access initiatives are engaged in a range of activities relating to build or maintaining data infrastructure, including:

- creating **new data assets** that will enable others to build understanding and tools to tackle the problem,
- developing **open standards** to support collecting, using and sharing data,
- creating **guidance and policies** that will help to inform how that data is used, in support of tackling a specific problem while minimising potential harms,
- building **new technologies**, including platforms, code and algorithms that will help support the management and use of the data.

Over time, to embed and secure the change necessary to solve a challenge, these initiatives might lead to new **data institutions** – organisations that steward data for a social, environmental or economic goal – or a strengthened role for existing institutions.

The sketch below maps how an initiative might evolve into a data institution and lead to wider change. The flow is illustrative only; in practice we have seen multiple approaches in this area.

Sketch



What frameworks can help design data access initiatives?

Data access initiatives engage in a range of activities with the goal of tackling a specific social, environmental or economic challenge.

To help understand how these activities might be planned, we looked at several different frameworks that are intended to support the design of impactful programmes of work. Our goals were to:

- understand how tackling a specific problem or challenge might be decomposed
- identify frameworks that might help to inform the design of new programmes
- understand whether these frameworks could help support analysis of different data access initiatives

We explored three frameworks: systems thinking, the COM-B behaviour change model and logic models.

Systems thinking

The '[systems thinking](#)' methodology encourages us to think about the world as a complex, interconnected system. By understanding the different parts of a system, and how they interact, we can better understand how the system as a whole might be changed. At their core, data access initiatives are about changing or intervening in

systems.

Researchers exploring systems thinking and its application to drive change have identified a range of interventions, many of which relate to building data infrastructure.

Donella H. Meadows identified [12 leverage points that can be used to change systems](#). One of those levers involves changing how data and information are shared, to provide better insights that can improve how systems work.

Les Robinson has developed Meadows' 12 points into a checklist of 18 ways to change a system (see the graphic 'How to change a system' below). Several of these are relevant to the design of data-access initiatives:

- **Infrastructure** – building or improving infrastructure, products and processes that make it simpler and easier to achieve the desired outcome
- **Creating feedback loops** and accountability to collect data on the consequences of behaviour and decisions and feed it back into the system
- **Building inclusion and community** to encourage more equitable access to data, share best practices, and encourage engagement with those who are affected by the collection and use of data
- Improving or building **data institutions** to help to build strong governance and rules
- **Encouraging innovation** to look for better alternatives and outcomes.

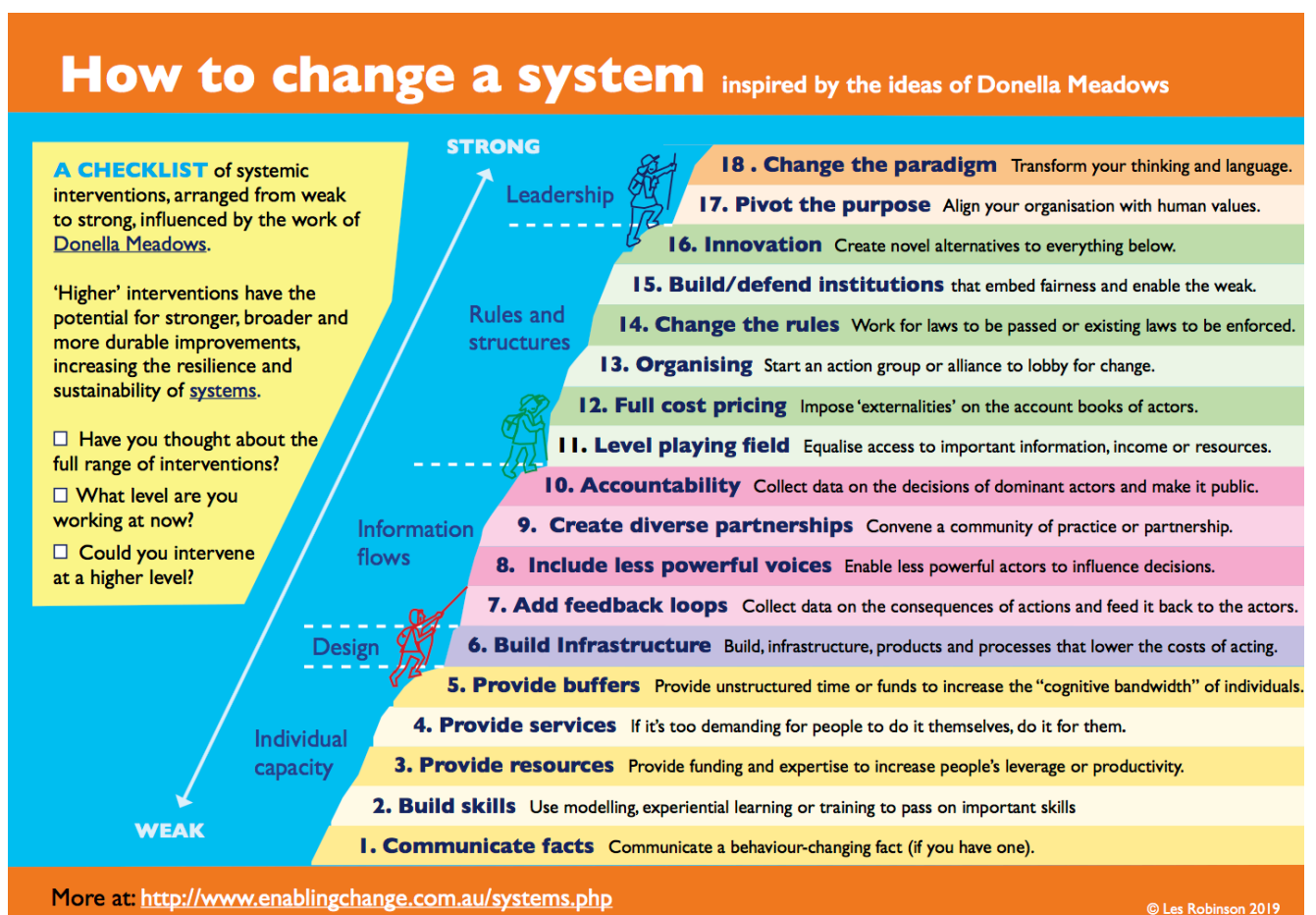


Image credit ©Les Robinson 2019

We found systems thinking a helpful way to think about the design of existing and new data access initiatives. It highlights the importance of a range of different types of intervention. Understanding the levers available to change systems can help to plan and prioritise impactful programmes.

Read more:

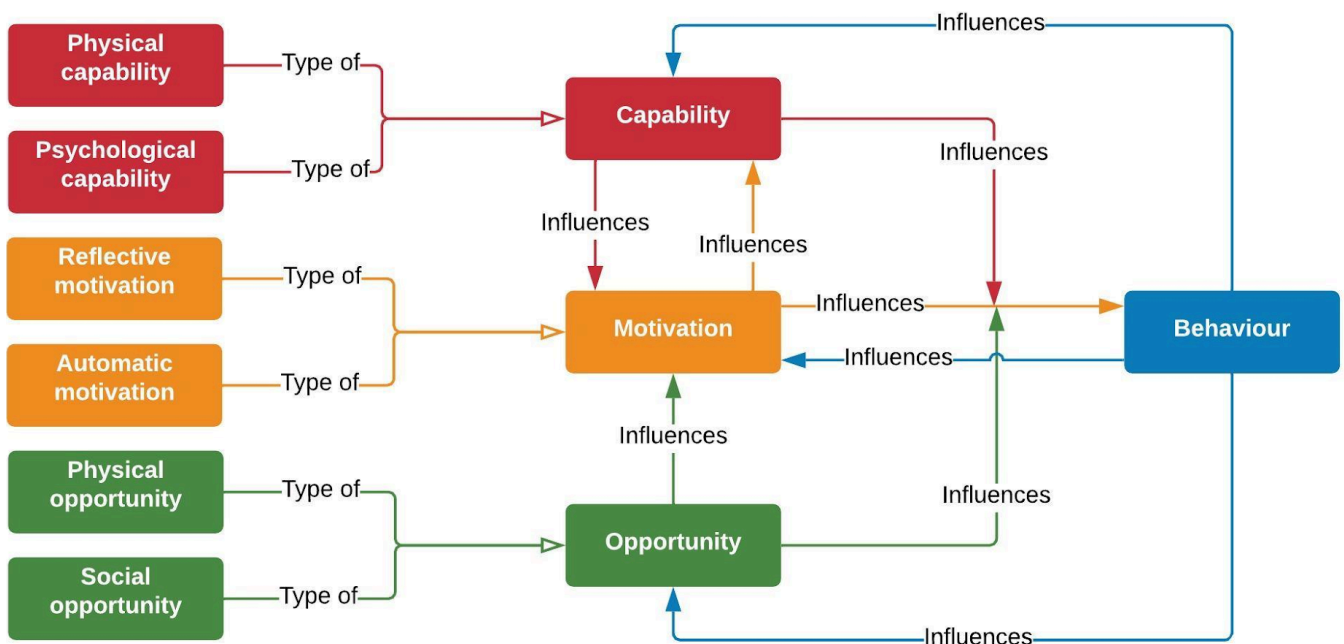
[Can thinking about the world as a complex, interconnected system help us solve difficult problems?](#)

Models of behaviour change and COM-B

Data access initiatives often require individuals and organisations to change their behaviour. This might involve sharing more data or changing business processes to be more data-enabled.

The [COM-B model](#) describes three major influences on behaviour:

- **Capability.** An individual or system's physical and psychological skills, knowledge and ability to act in a certain way.
- **Opportunity.** External factors which enable or inhibit certain behaviours. Influences include physical opportunities, such as time, location and resources, as well as social opportunities such as cultural norms and accepted behaviours.
- **Motivation.** Internal processes that influence behaviour. These consist of automatic processes, such as desire or aversion, and reflective processes, based on our evaluation of past events.



Susan Michie, Maartje M van Stralen, Robert West. (2011). *The behaviour change wheel: A new method for characterising and designing behaviour change interventions.*

A data access initiative can be designed to address one or more of them through delivering interventions. The '[Behaviour change wheel](#)' – a commonly-adopted framework which builds on the COM-B model – describes nine key interventions that an initiative might deliver to change influences on behaviour, which we can see applying to data access initiatives:

- **Education or training** – improving knowledge and skills around data can improve capacity in a sector or industry, and improve people's motivation
- **Persuasion, incentivisation or coercion** – providing incentives, such as prize money or work contracts, or introducing an element of punishment, are changes in motivation. The way that those incentives are communicated to ecosystem actors is key
- **Enablement** – providing resources, such as seed funding or skilled data professionals on fellowships, for organisations in an ecosystem can improve available opportunities, but also improve motivation and capability
- **Modelling** – providing relatable examples of other data access initiatives can improve motivation
- **Environmental restructuring** – delivering structural changes at an organisational or ecosystem level can create more opportunities for change
- **Restrictions** – introducing rules and guidance about the collection, dissemination and use of data into an industry or sector can motivate actors to engage with a desired behaviour or steer clear of undesirable actions

These interventions can be enabled through policy categories such as service provision, regulation or legislation, communications and marketing, fiscal measures, guidelines or social and environmental planning.

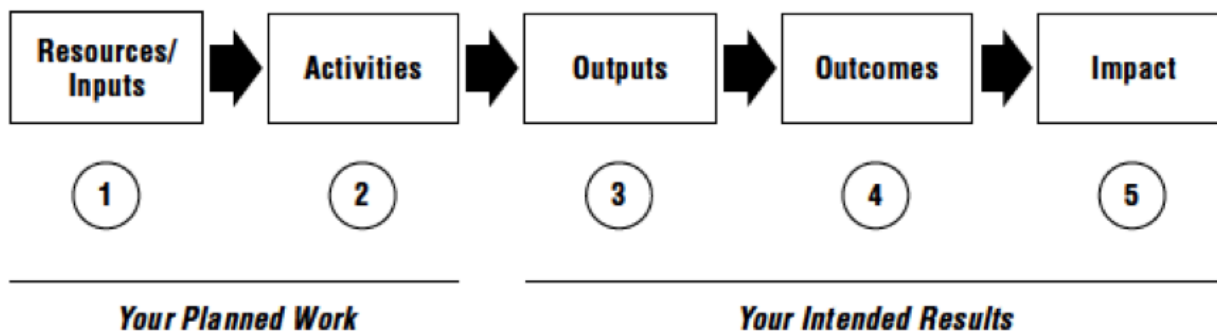
At the ODI we have previously applied thinking from the COM-B model to designing some of our own programmes of work. It can be helpful in planning out a range of activities that will ultimately influence the behaviour of the broader ecosystem in which the initiative is operating.

[Read more:](#)
[Changing behaviour to solve challenges](#)

Logic models

[Logic models](#) are tools designed to help people plan impactful projects and communicate those plans to others.

They provide a structured way of thinking about how to build a programme that will help to address a specific problem or challenge – essential for designing a good data access initiative. Logic modelling consists of mapping out the logical flow from the inputs and activities that an initiative requires, to the outputs needed in order to enable the outcomes and impact the initiative is set up to achieve.



W.K Kellogg Foundation Evaluation Handbook (1998)

Logic models are widely used to help to design programmes of work. However they can be applied in a variety of contexts. At the ODI, we have often applied logic models to our own programmes and projects, including [OpenActive](#) – a community-led initiative to help people get active using open data. Our activities have included things like developing open standards to support publication of open data. Those activities have produced outputs such as openly licensed standards and specifications, and lead to outcomes such as [up to £3 million per year in health costs avoided and up to a £20 million per year increase in productivity](#). Ultimately, the intended impact is to address the problem of inactivity in England by helping people get active.

Of the three frameworks, we found logic models to be particularly useful as they provided an overall design framework for describing how the activities of an initiative were intended to contribute towards its desired outcomes and impact. The other frameworks were helpful in supplying a menu of possible interventions that might be more or less appropriate for a given problem or context, that are then included in a logic model.

In our discovery phase we tested this tool further with some real world data access initiatives. This is described in [Developing logic models](#) below.

We also used logic models as an analytical tool. Based on our desk research we created simple logic models for a small number of data access initiatives, identifying their outputs and activities. This approach allowed us to identify a set of common activities relating to the building and maintenance of data infrastructure.

We used this list of activities to identify and compare a set of data access initiatives that are building data infrastructure based on adoption of common standards to support decentralised publishing of data. We have defined data access initiatives that follow this pattern as ‘[decentralised data publishing initiatives](#)’.

Read more:

[Using logic models to design data access initiatives](#)

[Comparing decentralised data publishing initiatives](#)

What did we learn from our exploratory research?

The research phase helped to develop our understanding of the field, built connections with a broad range of different initiatives that will be useful in our later work, and identified a range of challenges that initiatives are facing when planning and prioritising their work.

The interplay between initiatives, data infrastructure and data institutions is complex. Understanding the pathways for how initiatives develop into data institutions merits further research but is outside the scope of this project.

From our research and interviews with initiatives we learned that:

- Frequently, while initiatives have a broad vision, they focus on identifying a narrower set of use cases or problem statements to tackle within their early activities.
- Individuals driving initiatives typically have strong domain knowledge, but may lack experience scoping and designing the initiative, planning how to create or improve data infrastructure, and with how to select, develop and drive adoption of standards for data.
- Understanding the current data ecosystem and infrastructure is important for prioritising use cases, identifying the necessary activities required to build the desired future state, and demonstrating the value of their approach.
- Few initiatives seemed to be explicitly applying frameworks like COM-B, systems thinking and logic models to guide their design, although the core ideas were present in a number of plans and strategies.
- Often initiatives lack tools to understand and describe existing data infrastructure and choosing which interventions would be useful to help them tackle a challenge.
- Often initiatives lack guidance to ensure that data infrastructure is designed and built in ways that not only achieve their goals, but also maximise diversity, equity and inclusion, and minimise any potential harmful impacts from the initiative.
- Data access initiatives approach monitoring, evaluation and learning in different ways, often depending on their funding model. Those with rigid government funding have a rigorous evaluation framework, while startups and corporate programmes have a variety of key performance indicators.

These insights identify a number of different needs amongst those leading initiatives:

- **a need to better understand the role of open standards** in achieving their goals and how to go about creating them. The ODI's existing [guidebook on open standards](#) is designed to help with this
- **a need for external review and validation of plans and activities** to ensure they are well designed and executed. This might be met through developing further guidance and support for applying tools like logic models, or by developing templated models for certain types of outputs or activities, for

example 'creating a data inventory'

- **a need to undertake a review of the existing data ecosystem, datasets and data infrastructure** to help prioritise activities.

Of these three needs, we decided to explore the final one in more detail.

Data landscape review for initiatives

Different types of stakeholders involved in data access initiatives face different challenges during their design.

- Those **funding data access initiatives** may need more insight into what support initiatives need, help in identifying well-designed initiatives that are likely to create impact, or help supporting those who are designing and leading initiatives so they are more likely to create impact in practice.
- Those **designing new data access initiatives** may need help in designing an impactful initiative.
- Those **leading existing initiatives** may need to evaluate or adapt the way they are approaching their work

We have focused on those who are designing new data access initiatives. This group includes policy makers who are looking to scope new sector-wide initiatives and individuals who are trying to set up new initiatives.

We believe this group needs the most help and support. Identifying the needs of this group during their early planning and prioritisation will also lead to potentially greater impact.

We framed our problem statement as:

People leading data access initiatives need to understand the current data landscape to help them prioritise and plan the initial activities for their initiatives. These plans are frequently a subset of the wider challenge which they are seeking to address. This data landscaping work often requires a range of activities including stakeholder mapping and engagement, creating data inventories, understanding user needs, etc.

What help do they need to carry out these activities?

To help answer this question, we explored in more detail:

- how initiatives went about understanding their ‘data landscape’ or ‘ecosystem’, so we can understand the activities involved in that process
- which actors, skills and resources were involved in those processes
- what tools and methodologies already exist, had not been used by the initiatives we studied, and might support other initiatives in this process
- the types of outputs produced by those exercises and tools
- how these outputs could have usefully informed the plans and activities of the initiative

We used a variety of methods.

- **Desk research** where we looked at data ecosystem mapping and similar

approaches to understanding the wider context of initiatives, as well as the design of these approaches and terminology that is used to describe them.

- **User research** where we conducted interviews and a small survey to understand how individual organisations approach activities such as ecosystem mapping and data landscaping
- **Testing some existing ODI tools** with some existing initiatives to understand how those tools might support the design and planning of new data access initiatives.

A note on terminology

For the purposes of this report we are using the phrase “**data landscape review**” to refer to a broad range of activities that support a data access initiative in understanding the context in which they operate.

The context for a data landscape review might be a specific market, sector, or community, as defined by the goals of the initiative. As we note in the following section, the activities involved in “data landscaping” vary across initiatives.

In future outputs from this project we may refine this definition further and clarify the typical activities involved in the process.

How are data initiatives understanding their data landscape?

Our research showed that data landscape reviews involve a variety of different activities:

- **Stakeholder mapping**, to understand the people and organisations that an initiative might engage with, support, collaborate with, or influence.
- Creating **data inventories** to understand what data is currently collected, shared or managed that is relevant to the initiative.
- Doing **user research** to understand the needs of data users, data stewards and decision makers that might be supported by the initiative.
- Carrying out **needs assessments** with stakeholders to understand their skills and areas of support.
- **Mapping existing data ecosystems** to understand how data is currently being accessed, used and shared.
- **Mapping future data ecosystems** to help articulate the vision and potential areas of activity for the initiative and identify potential risks or unintended consequences.
- Gathering and developing **use cases and requirements** to inform plans for building and strengthening data infrastructure.
- Doing **market analysis** to understand the needs of data users, or opportunities for future products and services that will further the needs of the initiative.
- Performing **gap analyses** to identify where additional data may be needed to tackle a problem, or where existing data infrastructure might be improved.

- **Understanding the legal and regulatory context** within which the initiative operates.

These activities are carried out by data access initiatives using a variety of approaches, including conducting surveys, running structured interviews, holding informal meetings, carrying out desk research and facilitating workshops.

The activities are often done iteratively, with detail being refined based on learning or insights from other activities. For example, mapping out stakeholders might identify further potential sources of data to refine the contents of a data inventory.

However having a clear idea of the scope for a landscaping exercise is essential in order to narrow down the range of research, engagement and cataloguing work that might be required. Developing use cases and ensuring the initiative is properly scoped are important first steps.

In most cases we investigated, these activities were designed to engage with stakeholders and to work in the open. However the depth of stakeholder engagement varied from project to project.

We observed that the skills and experiences of people leading initiatives can have a direct impact on how they approach the data landscaping exercises. Many data access initiatives that we interviewed were led by individuals with a specialist or technical background. More technically minded leaders and teams focused more on mapping technical infrastructure and less on other areas, such as mapping stakeholders.

We also noted that data landscaping activities tend to focus on mapping existing relationships and assessing stakeholder demands but did not often extend to exploring and mapping the potential or desired future state of the ecosystem.

What tools and methodologies exist to support a data landscape review?

A number of existing tools, expressing a range of methodologies, have been published to help data access initiatives undertake data landscaping exercises. [Table 1 in the Appendix](#) summarises some of these tools, their intended use, the people involved in their use, and the types of outputs they produce.

We compared the listed tools according to their intended purpose, the scope of the activities they cover, the outputs produced and the intended audience.

Purpose of the tools

All these tools are designed to engage with and understand an ecosystem and increase access to sector specific data.

We also observed a set of shared activities between these tools related to data landscaping:

- Identification of problems and how data can help to address them

- Implementation risks and barriers for data publishing
- Identification of data governance and stakeholders
- Supply and demand gap assessments
- Recommendations and guidelines for data publication

Some of these tools were designed for important specific purposes that aim to address power imbalances, such as enabling data interoperability and increasing data quality in the Global South, or breaking barriers between data hoarders and publishers to increase the use of data and enhance development outcomes.

Depending on the main intended outcome of a tool (increase data quality, open public sector data, create an inventory, develop a standard, etc.) some outputs are designed to support data access initiatives in their landscaping exercises. These outputs include toolkits, step-by-step guides, canvases and assessment reports.

Multi-stakeholder approach

Many of the tools listed are community-oriented and often their users consist of self-selected groups, including public sector, academy and civil society. Their main audience and intended users predominantly includes stakeholders that are already involved in the problem space. Opportunities to involve other communities and actors, such as private sector players, to contribute to data landscaping activities are relatively limited.

Ecosystem engagement

We discovered that ecosystem engagement and stakeholder analysis are activities that are shared by many of the listed tools. However the language and purpose of the activity varies significantly from tool to tool. Data ecosystem terminology used to describe those activities can include:

- ecosystem mapping
- demand analysis
- stakeholder assessment
- data governance
- community building
- user analysis
- thematic expert engagement/networks
- stakeholder interviews

We also observed that activities to engage with stakeholders as part of the landscape assessment include a range of activities including:

- interviews with technical experts to build a diverse community around a specific problem
- desk research to identify actors
- workshops with the public sector and civil society
- canvas/questionnaire to identify stakeholder's demands and needs
- data asset inventory (mapping datasets and institutions in charge of publishing)
- ecosystem mapping and comparison (ideal versus actual).

How do existing tools help initiatives with aspects of a data landscape review?

We tested some existing tools to understand how they might support several early-stage data access initiatives in planning their activities. We chose data landscaping tools we were more familiar with and that fitted the user needs we identified for the various initiatives. This was done to be more relevant to the needs of the initiative. It also meant we were evaluating tools that we could then alter and adapt.

For each tool we ran workshops, provided additional guidance where needed and solicited feedback from the initiatives. This has helped us to better understand how people might, realistically, use them in the development of their initiatives.

Data ecosystem mapping

[Data ecosystem mapping](#) is designed to clarify how data creates value, illustrate the different actors in a data ecosystem, and show how value is exchanged across it.

We facilitated data ecosystem mapping workshops to help nine different organisations to scope their data access initiatives.

In the workshops we asked each organisation to:

- map the data they accessed, used and shared
- identify the data needed to build services
- identify the partners needed to deliver those services
- discuss findings with the group to find new opportunities.

User testing feedback identified ecosystem mapping as a productive tool to build a use case to engage stakeholders and build support. We learned that the most productive maps were those that had a clear focus, such as mapping a particular use case. We also learned that ecosystem mapping is a useful tool to find new opportunities, such as new partnerships, data sources or service provision, but that this is often overlooked and should be emphasised in workshops.

We found that, while not the same as in-person, an online workshop could still be a very effective way for people to engage with their data problem. Feedback from user testing explicitly highlighted that collaborating with other initiatives during the workshop was a valuable learning experience.

Overall, mapping participant data ecosystems helped:

- explore the variety of ways in which participants add and create value
- clarify the roles and responsibilities to deliver the use case
- identify ways to improve the initial idea of the initiative

Developing logic models

Logic models are designed as a framework to plan a path to impact.

Through workshops we tested how developing logic models can help early-stage initiatives structure their key activities, in order to achieve the desired impact for their data ecosystem.

User testing feedback indicated that the logic model process reflected the approach that organisations often have to present in order to gain funding from donors or investors. Users particularly liked the idea of having a clear, visual representation of their path to impact and felt that this would be useful to underpin conversations with internal and external stakeholders. In this way logic models can both structure an initiative internally, and be used to engage with funders and other stakeholders externally.

However, while most users felt that logic models could add value to their own projects, it was felt that in more exploratory projects, where the intended impact has not yet been established, the use of logic models might be a step ahead of the project.

Overall, developing logic models helped:

- structure the overall initiative clearly for use within the initiative
- provide a clear visualisation of the path from activities to impact
- support communication of the initiative with external stakeholders such as funders and partners.

Data Ethics Canvas and Consequence Scanning

The [Data Ethics Canvas](#) and [Consequence Scanning](#) are tools which help identify and manage ethical issues at any stage in a data project by asking questions about the use of data in the project.

We tested how these two tools can complement each. We used the Data Ethics Canvas to help an organisation identify the ethical implications of implementing their initiative and used Consequence Scanning to help them to think about ways to mitigate any potential harms.

User testing feedback indicated that the data ethics workshop helped illuminate difficulties in reducing harm while identifying and delivering positive interventions to people in vulnerable circumstances like fuel poverty. Users responded particularly positively to the way the workshop helped reveal blind spots around engagement and process such as how to engage with vulnerable consumers. They also liked its open and transparent process.

Overall using the Data Ethics Canvas and Consequence Scanning tools helped:

- illuminate issues and unintended consequences in a use case aimed at increasing access to data to provide services which could harm its target beneficiaries
- highlight the importance of, and potential for, engagement with the ecosystem to reduce harm
- inspire ways of having an open and transparent process in use-case development.

Using the ‘Open standards for data guidebook’

Open standards for data are reusable agreements that make it easier for people and organisations to publish, access, share and use better quality data.

To understand its contribution better, we interviewed an organisation who used the guidebook as their main reference to develop an open standard. Despite consulting the book extensively they did not make use of all the tools discussed in the book. We were able to advise them on the way that tools which are described in the guidebook, such as [Data ecosystem mapping for open standards](#), would help them develop their open standard.

This indicates that the guidebook, while useful, may not be sufficient in guiding an organisation to develop an open standard, and that additional tools and expertise can be useful additions.

Overall, using the ‘Open standards for data guidebook’ helped them:

- understand how to get started in the creation of an open data standard
- prioritise activities such as stakeholder mapping and engagement
- take action based on the insights gained through mapping their data ecosystem with tools described in the guidebook.

How are the results of data landscaping exercises published and shared?

As might be expected from the variety of approaches to data landscaping in our sample initiatives, a variety of different types of output were produced.

Some of these outputs were developed as a result of using the tools summarised in the previous section. Others are the results of data landscaping exercises undertaken by data access initiatives or organisations analysing a specific data ecosystem using a method that is not covered in the previous section.

[Table 2 in the Appendix](#) provides some examples of the results of data landscaping exercises, with a brief summary of their form and function.

Main outputs and complementary tools

We also compared the various output formats of these landscaping activities to understand the scope of the analysis, how results are published and what insights they offer.

Many of the listed outputs focus on data collection, availability and quality, and they publish the results of gap analysis and guidelines on how to strengthen data governance in specific sectors.

These outputs are often published as reports that synthesise the key findings from the landscaping exercise, with recommendations for the initiative and key

stakeholders.

The reports are often supported by additional structured information, such as diagrams of data flows, networks of stakeholders or other aspects of the data ecosystem. Data inventories are shared as spreadsheets or are used to populate simple data portals.

Ecosystem engagement

In most cases data landscaping activities are designed to engage with stakeholders and to work in the open. However, if there are privacy concerns or cultural specificity it can happen that research engages with a closed ecosystem, where data is not widely shared or accessed.

Some initiatives use ecosystem maps as a tool to help describe their environment to partners and funders. Others rely on reports, spreadsheets and canvases to communicate the data ecosystem to stakeholders.

Impact assessment

Overall we found that many of these data landscaping outputs are designed to support data access initiatives in understanding value streams in an ecosystem, assessing data gaps and increasing engagement with stakeholders and data publishers.

We also observed that most toolkits map stakeholder engagement but only a few use this exercise for impact assessment. Although some include use cases and lessons learned from implementation, they do not emphasise impacts in terms of increasing data use, solving the problems they addressed, strengthening data infrastructure or reshaping the data access initiative.

How does data landscaping inform the design of initiatives?

Data access initiatives identified a range of general benefits to the project from carrying out data landscaping activities, including:

- Improving and building stakeholder engagement and building community through involving them early on the planning and decision making processes.
- Improving their ability to communicate a project's outcomes, vision and value proposition to senior stakeholders and funders.
- Engaging with public policy makers to help to get support for the initiative and its goals.
- Furthering peer learning opportunities (if mapping is done collaboratively).
- Providing an opportunity to re-shape and re-evaluate the scope of a programme.
- Identifying new opportunities, structuring activities internally and prioritising next steps.

In terms of guiding how initiatives prioritise and plan work around data infrastructure, we identified that data landscaping helps with the following key activities for data access initiatives:

- Identifying policy recommendations and interventions that could lead to the collection, sharing and use of data in line with the goals of the initiative.
- Identifying data gaps that can be addressed through new data collection activities that the initiative can undertake or support.
- Clarifying barriers or issues with how existing data is currently being accessed, used and shared to help facilitate access and use.
- Identifying where development of new data standards, or the adoption of existing standards might help to strengthen data infrastructure.
- Documenting existing datasets that might be used to help to tackle the goals of the initiative. This can help to raise the profile of these resources and drive further uses.
- Describing quality issues, for example in coverage, representation or accuracy in existing datasets, that need to be tackled by improving existing data infrastructure.

What did we learn from our research?

Our research confirmed that organisations involved in early-stage data access initiatives need help with better understanding their data ecosystem early on in their work. This was true both for mapping out the data ecosystem of a particular use case, and for the data ecosystem underpinning an organisation's business model.

We learnt that the initiatives we looked at are led by subject matter experts who have deep content knowledge but not necessarily the right skill set to understand how to go about collecting and building the required data infrastructure to tackle a specific challenge.

We have seen that early on, people leading data access initiatives face similar challenges and have similar user needs to understand:

- **what data assets exist in their data ecosystem**, and relevant information about these assets such as accessibility, quality, reusability, etc
- **what other data infrastructure exists** to support the initiative, not just technologies and standards but also organisations and communities
- **what policies, regulations and ethical concerns** influence accessing, using and sharing data in the ecosystem
- **how to plan activities** in order to use and develop the necessary data infrastructure to underpin the data access initiative
- **how to engage with different stakeholders** in the data ecosystem, especially those outside of their sector, and how they can contribute to the use case or business model.

Our research and user testing has shown that existing tools such as data ecosystem mapping and the Data Ethics Canvas are good starting points to address some of these existing challenges. We identified additional tools such as Open Data Charter's 'Open up' guides, the Open Contracting Data Standard and the Data for SGDs toolbox that also address pieces of the puzzle.

However, the shortcomings we observed are that each of these tools are designed for specific activities within data landscaping, for instance to assess a particular part of

data infrastructure, to publish data openly, or to build a network. Based on the outcome of our user research we concluded that it would be useful to invest in developing a more versatile toolkit that can incorporate the various data landscaping needs identified.

Next Steps

Our next steps will be to develop prototypes to help address the user needs that we have identified.

In the next stage of the product development phase we plan to develop three prototypes in fast succession, which we will test with the following groups:

- Organisations we have worked with over the course of the project, in both interviews and applied user research such as workshops and working groups
- Organisations leading or working on data access initiatives that were not part of our earlier user research
- Organisations outside of our typical stakeholder network, specifically those representing more diverse views than we have so far, such as indigenous data governance groups, data initiatives in the Global South, and data initiatives led by and with a focus on various underrepresented groups.

The prototypes may involve:

- iterating or improving on existing tools and guidance developed by the ODI or other organisations, for example to highlight how they can be used as part of a data landscaping exercise
- identifying how to combine several tools to conduct a data landscaping process
- developing new tools, guidance or creative outputs that can help to support data landscaping activities

Based on the user feedback we will receive from these organisations, we will then make a decision to select one tool to take forward for further refinement in a final phase of the project.

Methodology

Research Methods

- **Desk research**

We did desk research to identify data ecosystem mapping activities and/or data gap analysis to strengthen data infrastructure and tackle a specific problem. For more information on these insights see 'Ecosystem mapping research' above.

- **Interviews & surveys**

We carried out additional user research through email, survey and meetings/interviews, with an explicit aim of getting information from perspectives that were not included in the desk research phase.

- **Applied user research**

We ran workshops and co-working sessions with a variety of different initiatives to explore the needs that emerged in earlier desk and user research in more detail. For more information on the tools used see 'Testing existing tools' above. We sought feedback on outputs such as our '[Comparing decentralised data publishing initiatives](#)' report as part of our commitment to work in the open. We contributed to the Open Energy Advisory Groups

Known limitations and mitigations

- **Capacity constraints for feedback**

The people and organisations that we reach out to to input into our research do so without financial compensation, often cutting into their own workdays. Although we have tried to reduce the effort it requires from them, research subjects are not always able to donate their time for interviews or to complete surveys.

In Alpha, we will be working with the winners of our stimulus fund who are being compensated for their participation, guaranteeing us a research subject cohort. We will also be reaching out to a broader set of stakeholders to reduce the reliance on each individual research subject.

- **Low stakeholder diversity**

The Open Data Institute has a broad and ever-growing network, however the most responsive people in our network tend to reflect the demographics of the ODI itself, and are predominantly from the Global North. This means that views from the Global South, and explicitly from various underrepresented groups, continue to be underrepresented in our research.

To mitigate this, we have started interviewing data access initiatives from the Global South, such as ILDA and Data4SDGs, or which address problems from marginalised communities, such as the British Columbia First Nations' Data Governance Initiative (BCFNDGI).

In general, we have recently been pushing to explicitly include diversity, inclusion and equity in the delivery of our projects, and have discovered tools

and frameworks that could help us be better in this space. Our project has aimed to use the [data feminism](#) framework and the '[Centering racial equity toolkit](#)' to identify issues around power and equity in data access initiatives. We will accelerate this work in the Alpha phase.

Appendix

Table 1: Data landscaping tools

Resource	Source	Scope	Actors	Outputs
City Data Sharing Toolkit	Future Cities Catapult	Data toolkit designed for governments and institutions with recommendations for sharing non-personal urban data	Local councils	Data sharing canvas, business case, data protection impact assessment
Civic Switchboard Guide	Institute of Museum and Library Services	A step by step guide to help libraries understand and become more engaged with their local civic data ecosystems.	Local libraries	Toolkit and use cases
Consequence Scanning	Doteveryone	Toolkit for product designers to understand consequences of planned services and products	Product design	Case studies and toolkit with prompts
Data Challenge Prizes for Health: a Playbook	Open Data Institute	Provides a step by step guide on how to scope out a health problem in detail and ensure it fits within the challenge prize model	Public health organisations, healthcare providers, policymakers, data scientists	Data challenge prize in the health sector
Data ecosystem mapping	Open Data Institute	Illustrates the different actors in a data ecosystem, and how value is exchanged across it	Data stewards, data users, data intermediaries and beneficiaries	Data ecosystem map
Data Ethics Canvas	Open Data Institute	Helps identify and manage ethical	Data stewards, data users, data	Completed data ethics canvas

		issues at the start of a project that uses data, and throughout.	intermediaries and beneficiaries	
Data Inventory Guide	GovEx Labs	A guide to support cities in designing data inventories.	Local governments	Steps, insights and use cases
Guide to civic tech and data ecosystem mapping	National Neighborhood Indicators Partnership (NNIP), Urban Institute and Living Cities	A guide on how to visualize and engage with civic tech networks at the local level	Local councils, housing agencies and civic tech community	Ecosystem assessment methodology, guides and resources for mapping.
How to create a data inventory	Open Data Institute, CABI	Provides a summary of the benefits of creating a data inventory, an outline of the steps to do so, and metadata recommendations	Data stewards, data managers, data users	Data inventory
Mapping Gender Data Availability and Mapping gender gaps	Data 2X/Open Data Watch	Methodology to assess sex-disaggregated data availability based on SDGs indicators	Governments and data practitioners	Data availability assessment and methodology to identify data gaps based on SDGs
Open Up Field Guide Methodology	Open Data Charter	A methodology to create a guide to increase data quality and use in a specific sector	Government, field experts and data practitioners	Field guide, data inventory, data governance framework
Open Standards for Data Guidebook	Open Data Institute	Helps people and organisations create, develop and adopt open standards for data	Organisations developing and adopting standards	Open standard for data
The 100 Questions	NYU Govlab	A methodology to build a community around a specific development issue and design the questions needed to tackle a problem, and the	Governments, academia and civil society	Key questions for each sector and partnerships for data collaboration.

		data needed to address it		
Toward an open data demand assessment methodology	IDB/NYU Govlab	Methodology to assess and segment data demand of a specific sector and strategies to match with supply to create impact.	Governments and civil society	Canvas and methodology to develop the assessment

Table 2: Data landscaping outputs

Output	Format	Methods Noted	Description
Bridging the Gap: Mapping Gender Data Availability in Latin America and the Caribbean	Report, Country Profiles	Data analysis	Detailed review of gender data in 5 countries, at the international, national and microdata levels. Assesses the availability of key indicators and presents recommendations for improvement.
Disaggregated demographic data collection in British Columbia: the grandmother perspective	Report	Community consultation, desk research and policy analysis	Framework for disaggregated data collection and recommendations on tools, processes and purpose to collect and access data from indigenous communities
Empowering communities in EITI implementing countries	Report	Desk research and interviews	Report with recommendations on how to strengthen communications and dissemination efforts to broaden and deepen local civil society engagement in natural resource governance in Colombia, Ghana and Indonesia. This study contains a mapping for each country, information needs and barriers to participation.
Femicide regional standardization lessons learned , key indicators and protocol	Report, Data Protocol and indicators	Desk research, workshops, expert interviews	Guidelines to enhance data collection and establish a data governance to protocolise and standardise data on femicides.
How can a sustainable open data system promote trust in our	Blog post, data ecosystem map	Desk research, interviews, workshops	Ecosystem mapping and analysis to understand the role that a third party data institution can play in the development of the emergent CCS industrial landscape

progress towards a net zero economy?			
How to prioritise data for global development	Report, Spreadsheet	Desk research, expert consultation, use case analysis	Identifies priority list of datasets to be released to tackle a range of development goals
Indigenous Data Network Roadmap Framework	Report	Community consultation, data analysis	A framework for auditing and assessing data services provided by potential partner organisations to the Indigenous Data Network
Mapping data ecosystems: Government Statistical Service (GSS) Alpha Project	Blog post, data ecosystem map	Desk research, interviews, workshops	Ecosystem mapping and analysis to understand how to bring data from different government departments together so that it is findable by related data collection instead of by organisation
Missing Numbers in Children's Services	Report	Literature Review, Expert Interviews	An assessment to identify how better data could improve outcomes for children and young people, by conducting a review of different types of data, their utility and potential impacts of use
Open Budget Data: Mapping the Landscape	Report	Literature Review Social Network Analysis Link Analysis	Empirical mapping and analysis of open budget data, drawing on digital methods and literature review. Intended to influence research direction, drive pilot projects, improve access to data, build alignment
Open Contracting Standard Development History	Report, 'Data maps', Comparison Tool	Data analysis and review Interviews	Mapping of supply and demand side of open contracting data, to understand user needs and alignment across existing datasets
Open Data Challenge Series Data Food Datasets , & Data Guide	Report, Spreadsheet	Desk research	List of suggested datasets useful for tackling a specific challenge in the area of improving food security. Includes assessment of data quality, openness, utility, etc.
Open Up guides on climate change , anti corruption and agriculture	Airtable with key datasets, data quality assessment and publication plan	Desk research, workshops with public officials and data exploration with civil society	Report with a set of data assets needed to tackle a specific problem and a quality assessment (identifying data assets, responsible institutions and level of openness)

360Giving Data strategy canvas	Canvas	Desk research and workshop with stakeholders	Guidelines on how to design a data strategy for data projects
--	--------	--	---