



DATA SHARING TOOLKIT

MARCH 2020

Approaches, guidance and resources
to unlock the value of data



About Smart Dubai

His Highness Sheikh Mohammed bin Rashid Al Maktoum, Vice President and Prime Minister of the UAE, and Ruler of Dubai, launched the Smart Dubai Initiative in 2013 with a vision of making Dubai the happiest and smartest city on Earth.

The Smart Dubai office was formed in 2015 to oversee Dubai's smart transformation and accomplish the leadership's vision. Collaborating with government and private sector partners, Smart Dubai is consistently adopting the latest technological innovations to provide efficient, seamless, safe and personalised city experiences for residents and visitors.

As of January 2020, the Smart Dubai office was officially renamed as the Smart Dubai Department.

www.smartdubai.ae

About Nesta

Nesta is an innovation foundation. For us, innovation means turning bold ideas into reality and changing lives for the better.

We use our expertise, skills and funding in areas where there are big challenges facing society.

Nesta is based in the UK and supported by a financial endowment. We work with partners around the globe to bring bold ideas to life to change the world for good.

www.nesta.org.uk

Acknowledgements

This toolkit was created by Tom Symons and Camilla Bertocin from Nesta in partnership with the Smart Dubai team.

As part of the research process for this toolkit, we held two workshops and presented our work to external experts in Dubai and London.

We'd like to thank all those who have attended the workshops and contributed along the way. In particular, we would like to thank Thea Snow and Eddie Copeland for their insight and support.

This toolkit is a copyrighted work of Smart Dubai.

About this toolkit**04**

Who is it for?

05

When should it be used?

05

How should it be used?

06

Data sharing**07**

Public problems addressed through data and analysis

08

Barriers to data sharing

10

Examples of existing data sharing initiatives

11

Labelling data sharing initiatives

12

Data sharing decision tool

13

A: The decision matrix**14**

Canvasses

16

Why share data?

17

What data to share?

19

Who to involve?

22

What is the overall governance structure?

24

What is the appropriate data infrastructure?

26

How is data accessed?

27

B: Project foundations**29**

Project foundations

30

Checklist

31

Requirements

32

ABOUT THIS TOOLKIT

Globally, innovation in data governance is fairly embryonic. Recent years have seen a flurry of new activity, but much of this remains poorly defined or nascent. Few, if any, countries have 'cracked the code' of responsible and effective data sharing initiatives and governance.

Data's value can be unlocked by creating trusted and ethical mechanisms for individuals, private and public sector organisations to share data.

Our research has focused on how trusted data sharing arrangements can be formed that ensure data is in the right hands and value can be extracted from it.

We approached this work by analysing a range of models for data sharing and boiling them down to the essential components that are common to all of them.

Who is it for?

- Innovators around the world interested in exploring data sharing initiatives and looking for a tool to spark the right discussions, anticipate issues and find new collaborative approaches to data sharing.
- International organisations, public institutions, businesses and non-profits.
- Those familiar with data and/or previously involved in traditional data sharing.
- Those grappling with slow or failed progress in activating data sharing collaborations on a bigger scale or dealing with complex initiatives.

This toolkit is designed to be used both by individuals and by teams and groups working through the activities and canvasses together.

When should it be used?

We designed a flexible decision tool that:

- Provides useful guidance and resources for private and public organisations to prepare for and design data sharing initiatives.
- Helps them identify the right combination of options for the specific and unique context for the given circumstances.

We have created six canvasses for this toolkit. While this guide presents the stages in sequential order, in reality users may find that they begin in the middle and end at the start. This is fine. It may also be necessary to walk through the steps more than once.

Additional material can be found in the 'Useful tools' sections. The tools are all referenced in the toolkit, accessible online and easy to use if you wish to dive deeper.



How should it be used?

You are new to data sharing and would like to know more about what is theoretically out there.

- The first part of this toolkit provides the context, an overview of a range of different data sharing initiatives and plenty of case studies. [07 ▶](#)

You are considering proposing a data sharing partnership to your team, but want to have an idea of what it would take to put one in place first.

- Start with the decision matrix. [14 ▶](#)
- You can go through the material of the canvasses and dive deeper using the additional resources in the 'Useful tools' sections.
- Move to the 'Project foundations' section. Do you tick all the boxes of the checklist? [29 ▶](#)
- If not, is there something you can do to facilitate the process of change that leads to all boxes being ticked?

You already have a partnership in place and want to kickstart aligning expectations and incentives, and need a tool to spark conversations across different stakeholders.

- Start with the 'Project foundations' section. [29 ▶](#)
- Discuss the checklist items and then move to the more practical exercise of the decision matrix. [14 ▶](#)
- Print each canvas on A3 (or a large piece of paper), divide into multiple groups and repeat the process if necessary with other stakeholders to guarantee maximum representation.



DATA SHARING

It is now widely recognised that the value of data held by individuals and organisations alike can increase exponentially if it is shared and combined with other sources of data.

Bringing together data sources breaks down traditional silos and unleashes the potential for data to generate important and meaningful insights. Public services in many parts of the world are exploring the potential of data analytics to address public problems, using their own data to become more efficient and effective.

The open data movement has led to governments worldwide sharing their data.

Yet many datasets that could help solve public problems remain closed, proprietary or difficult to find or share.

For this reason, governments, public and private sector organisations around the world are experimenting with different ways of accelerating data sharing and collaboration between those who hold valuable data and those able to deliver solutions to unlock the value from data.

Public problems addressed through data and analysis



Labour markets

- Future of work
- Skills
- Jobs

The Open Skills Project is a public–private partnership focused on providing a dynamic, locally relevant, up-to-date and normalised taxonomy of skills and jobs. Its aim is to improve our understanding of the labour market and reduce frictions in the workforce data ecosystem by enabling a more granular common language of skills among industry, academia, government, and non-profit organisations.

Consumer data and retail

- Consumer sentiment and experience
- Consumption patterns
- Business operations

Linking consumer confidence index and social media sentiment analysis is an analysis of the correlation between the official consumer confidence index obtained from the MIER (Malaysian Institute of Economic Research) and social media big data (via sentiment analysis, from Twitter) on consumer purchasing behaviour for two types of products over the course of two years.

Smart cities and city data

- Transport, mobility and urban planning
- Congestion reduction
- Energy sector

City Brain, a partnership of the Chinese government with the commercial platform Alibaba, provides real-time data from 750 sources to tackle problems of traffic congestion, analyse energy and water consumption patterns and identify vulnerable residents in need of additional support.



► Public Problems

Education

- Tailoring learning materials based on a student's needs
- Diagnosing strengths, weaknesses or gaps in a student's learning experience
- Providing automated feedback

The Baltimore Early Childhood Data Collaborative is a partnership of Baltimore City agencies serving young children and their families, sharing data to understand the experiences of young children in Baltimore and how those experiences relate to later educational outcomes.

Health research

- Rare diseases
- Genomic data mapping
- Population growth forecasts

Personal health datasets shared by voluntary patients and research centres are increasingly used to support research and preventive health care services, such as in the cases of MIDATA, the NCI Genomic Data Commons or Healx.

Environment

- Air and water pollution reduction
- Flood risk modelling
- Forest change monitoring

Fluxnet is a repository of eddy covariance measurements of carbon dioxide and water vapour exchange from more than 800 active and historic flux measurement sites, dispersed across most of the world's climate space and representative biomes.



Barriers to data sharing

Data breaches and privacy missteps regularly make headlines and have recently caused a profound and widening lack of trust among individuals, institutions and governments in the notion of safe data sharing.

In addition, a culture of risk aversion in public sector agencies can mean that the privacy risks are seen to outweigh the potential benefits.

While data sharing offers many opportunities, there are also significant challenges to be addressed and barriers to be overcome.

Other barriers to data sharing include:

- Risks associated with sharing commercially sensitive information
- The complexity of facilitating cross-border data flows
- Reputational concerns
- Regulatory or legal uncertainty
- A lack of dedicated personnel to drive and steward such initiatives
- Mixed levels of data maturity across organisations
- Unclear incentives (especially when engaging private companies)

Given the challenges, but in light of the significant opportunities outlined, there is a growing demand for trusted mechanisms for sharing data and meaningful privacy and data protection regulations.

Institutional frameworks needed to support the safe and trusted sharing and use of data between multiple different organisations are not yet well-established – there are not yet clearly codified processes that facilitate responsible data sharing.

The challenge facing public and private sector entities globally is how to strengthen trust and implement effective public-private data collaboration.

This toolkit aims to offer some insights and answers to tackle this challenge.



Examples of existing data sharing initiatives

At the outset of our research, we surveyed a range of different data sharing initiatives and grouped them by the following types.

While these may seem very different, they are all united in their attempt to release greater value from data through sharing.

Data commons

A spectrum of initiatives in which data is shared as a common resource among individuals or organisations, who collectively decide on the rules that govern access to it.

- [Dataverse](#)
- [DECODE](#)

Data exchanges and markets

Usually this is a data platform where data is treated as an economic good, and access is regulated through price mechanisms.

- [Copenhagen–Hitachi City Data Exchange](#)

Data trusts

Legal structures that provide independent stewardship of data.

- [Open Data Institute \(ODI\) data trusts pilots](#)

Open data platforms and open APIs

Curated sets of open datasets and APIs (application programming interfaces).

- [data.gov.uk](#)
- [Transport for London Unified APIs](#)

Data collaboratives

Includes all forms of collaboration in which participants from different sectors – including private companies, research institutions and government agencies – exchange data to create value for the public good.

- [Global Forest Watch](#)
- [California Data Collaborative](#)

Data co-operatives

Mutual organisations owned and controlled by their members, formed to collect and share data in the interests of their members.

- [MIDATA](#)

Offices of data analytics

Multiple organisations share data sourced from public sector bodies to improve services and make better decisions.

- [Essex Centre for Data Analytics](#)

Research partnerships and data hackathons

Projects focused on the identification of specific problems, collaborative work sessions and events where people work on data-related projects.

- [Consumer Data Research Centre](#)
- [Analytics Vidhya hackathon contests](#)

Labelling data sharing initiatives

As a very nascent field, there has been an emergence of organisations specialising in researching and promoting specific data sharing models.

To get a sense of the great variation in names and approaches, below we show the ODI Data Access Map. This is a clear example of how many terms and definitions can be attached to data sharing initiatives, and it is an attempt to try and navigate them, by crowdsourcing definitions and case studies. These labels can be used interchangeably depending on the organisation in question, leading to confusion and a lack of rigour in their application.

Some of the resistance to exploring data sharing opportunities is due to the fact that this definitional confusion means that decision-makers have no firm framework to help them understand their options for sharing data in practice. In addition, each collaborative project involving data sharing is unique, set in a specific dynamic composed of different actors, laws and rules, expectations, levels of expertise, incentives and relations. In this

context, predefined models and labels, such as those described above, are not useful as tools for translating theory into practice.

Rather than considering whether a 'data trust' or a 'data collaborative' or a 'data commons' is the right approach, it is much more important

to think about the specific problem that sharing data is set to solve. From this problem and the specific context in which the partnership is being developed, other questions around governance, power and access will flow.



Data sharing decision tool

Our work has involved an in-depth review of a range of different data sharing collaboratives and partnerships to understand the essential components that are common to all of them.

This analysis was then translated into a flexible decision tool that covers two stages of development, which can happen almost in parallel, as the process of designing a data sharing initiative will involve some degree of iteration.

A: The decision matrix

Identifies six key decision points, prompting and guiding discussions about all key elements of a data sharing arrangement.

14

B: Project foundations

Identifies the conditions required to move forward with the data sharing project and provides the overarching legal, technical and relationship considerations.

29



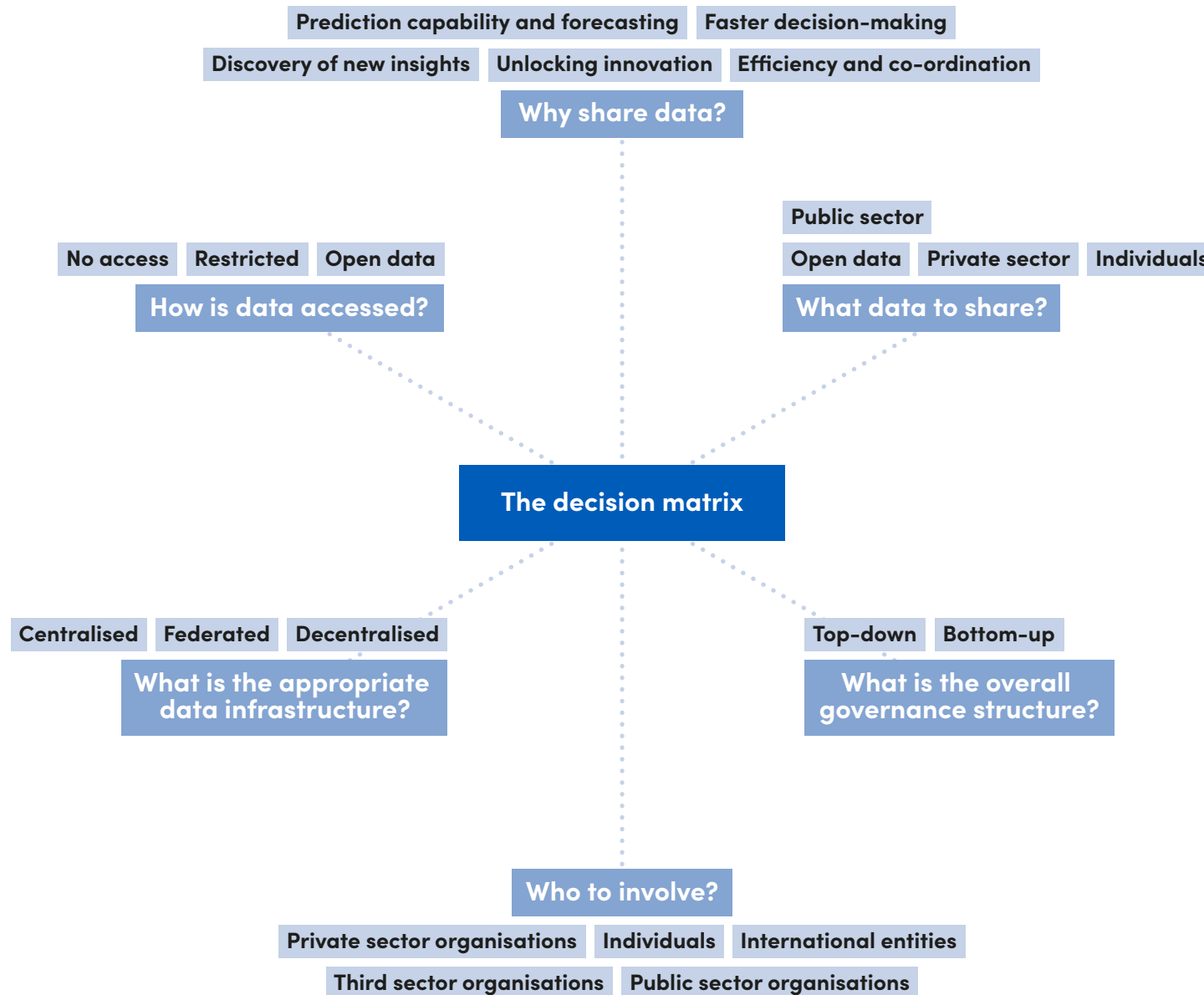
A: THE DECISION MATRIX

Analysis of a range of models (such as trusts, collaboratives and co-operatives) indicates that there are six essential components of a data sharing collaborative. These are framed in the decision matrix as the key decision points that anyone developing a data trust should consider when designing a data sharing approach.

These are ordered in what is in theory an ideal sequence, but often reality is more complex.

There is a degree of necessary circularity to the process, meaning that the tool may be used multiple times, switching from one section to another, refining the level of detail and the participation of different groups of stakeholders.

Each decision point is described in more detail in the following section, with guidance around which option is best suited for the particular set of circumstances.



Canvasses

Why share data?

This step encourages you to define the purpose of sharing data. Defining this upfront, and having shared agreement of the purpose and what success looks like is a precondition to a project's success.

— Review Canvas 1 [17 ▶](#)

What data to share?

This step of the decision matrix supports consideration of datasets needed for the partnership and prompts stakeholders to discuss the types of data that need to be shared, their form and the ethical considerations that need to follow.

— Review Canvas 2 [19 ▶](#)

Who to involve?

This step supports consideration of which parties need to be involved in the arrangement to make it successful.

— Review Canvas 3 [22 ▶](#)

What is the overall governance structure?

This step of the matrix prompts parties to consider where power lies, and how the structure and roles of all parties interact in the data sharing arrangement.

— Review Canvas 4 [24 ▶](#)

What is the appropriate data infrastructure?

This step sets out the options for data storage infrastructure, which enable the data exchange across stakeholders.

— Review Canvas 5 [26 ▶](#)

How is data accessed?

This step helps you choose which form of data access best fits your initiative.

— Review Canvas 6 [27 ▶](#)



Why share data?

How to use this canvas

Review the questions, identify the reasons why you might use data to achieve the purpose and define your problem statement/ value proposition in the box below.

Questions to ask

- What type of problem are you trying to solve?
- What other things could you do instead of sharing data?
- Are there examples of similar projects that could help identify benefits and make the case for the partnership to be formed?
- Is the output of the initiative going to be a one-off analysis or does it require continuous data sharing?
- Is the data you want to share going to be used for strategic reasons or operational?
- Are you starting with a specific problem in mind or with the data?
- How is the project being communicated?

Useful tools

- GovLab [Data collaboratives explorer](#)
- Nesta DIY Toolkit [Problem definition](#)

Select the reasons you might use data to achieve your purpose



DISCOVERY OF NEW INSIGHTS

Creating new knowledge, sharing it with multiple partners and identifying the key questions that need to be addressed.



UNLOCKING INNOVATION

Identifying new sources of value by opening up data to third parties that can turn them into new, innovative data products with shared value.



FASTER DECISION-MAKING

Providing stakeholders with a more complete and accurate picture of complex issues for rapid decision-making. This can include improving public service design and delivery, and emergency response.



INCREASED PREDICTION CAPABILITY AND FORECASTING

Identifying new drivers of more accurate forecasts from disparate, interrelated and interconnected data sources from the use of advanced data analytics.



OPTIMISED PROCESS EFFICIENCY AND CO-ORDINATION

Providing additional insights from new data sources to augment co-ordination and reduce inefficiencies in day-to-day operations.

Define a specific problem statement or value proposition below

► Why share data? – Case studies

Discover new insights

AMDEX (AMSTERDAM DATA EXCHANGE) A data exchange initiative by the Amsterdam Economic Board, backed by Amsterdam Science Park and Amsterdam Data Science, and supported by the City of Amsterdam. The project is still at concept phase, and aims to collect city data held by government agencies, companies and others to provide broad access to data for researchers, businesses, governments and individuals in a secure marketplace. AMdEX explores possible use cases where data exchanges might be useful, including:

- **DATA LOGISTICS FOR LOGISTICS DATA (DL4LD)** An innovation project of the Dutch national technical institute TNO and the University of Amsterdam on sharing logistic data at a large scale.
- **CHIEF E-MOBILITY** A data-driven optimisation project aimed at the creation of an electric car charging infrastructure in Amsterdam.
- **KNOWLEDGE MILE** One of Amsterdam's long streets, made into the smartest city street by the Amsterdam Creative Industries Network.

Unlocking innovation

OPEN BANKING Since 2018, a regulation from the Competition and Markets Authority mandates that UK-regulated banks allow authorised providers (such as licensed startups offering budgeting apps, or other banks) direct access to customer account information and data at transaction level through APIs. The idea behind this initiative is that it will bring more innovation to financial services thanks to third-party developers, who will create new tools that will positively impact on vulnerable communities' financial inclusion.

Faster decision-making

GOOGLE WAZE A platform that provides real-time anonymised crowdsourced traffic data collected from participating drivers. In its Connected Cities programme, it shares its large amount of traffic data with government agencies, which can use this data to better inform policy or quickly deploy traffic assistance if needed. Some use cases of Waze data include:

- Reducing traffic
- Reducing incident response times

- Fixing potholes and identifying high-priority streets for maintenance services
- Sharing garbage trucks' real-time location

Increased prediction capability and forecasting

FLOWMINDER During the 2010 Haiti earthquake response, Flowminder researchers pioneered the use of de-identified data from mobile operators to follow population displacement. As a result, mobile phone data is increasingly used both in emergency contexts, providing operationally useful insight to humanitarian staff, and in discrete pieces of research looking, for instance, at the intersection of migration and the climate crisis. To support partners, Flowminder has created FlowKit, a suite of software tools designed to enable access and analysis of mobile data for humanitarian and development use cases.

Optimised process efficiency and co-ordination

SEOUL OWL BUS In Seoul, South Korea, where the metro system shuts down from midnight to 5 am, the municipal government used its citizens' late-night calls and texts to plan routes for a new night bus service. A telecom company (KT) provided the government with anonymous phone data, which officials used to colour-code regions of the city by call volume. They then analysed the number of passengers who were getting on and off at each bus stop in the heavy-call volume regions and, based on this information, implemented the Seoul's Owl Bus service along the nine most heavily trafficked late night routes. This partnership was able to create a service that not only saved late night commuters \$1.2 million in taxi fares from 2012 to 2014, reducing car trips by 2.3 million annually (city buses emit 80 per cent less carbon monoxide than private cars); it was also beneficial to low-income communities, providing them with a viable solution to commute home to the outer boroughs after working night shifts in the city.

What data to share?

How to use this canvas

Identify the datasets that would help you solve the problem or achieve the value proposition and place them in the diagram below (group by data owner).

In addition, you could also use Nesta's [Dataset Catalogue](#), which we created to help you list the datasets identified, rank them from most to least essential, and include useful information about the data.

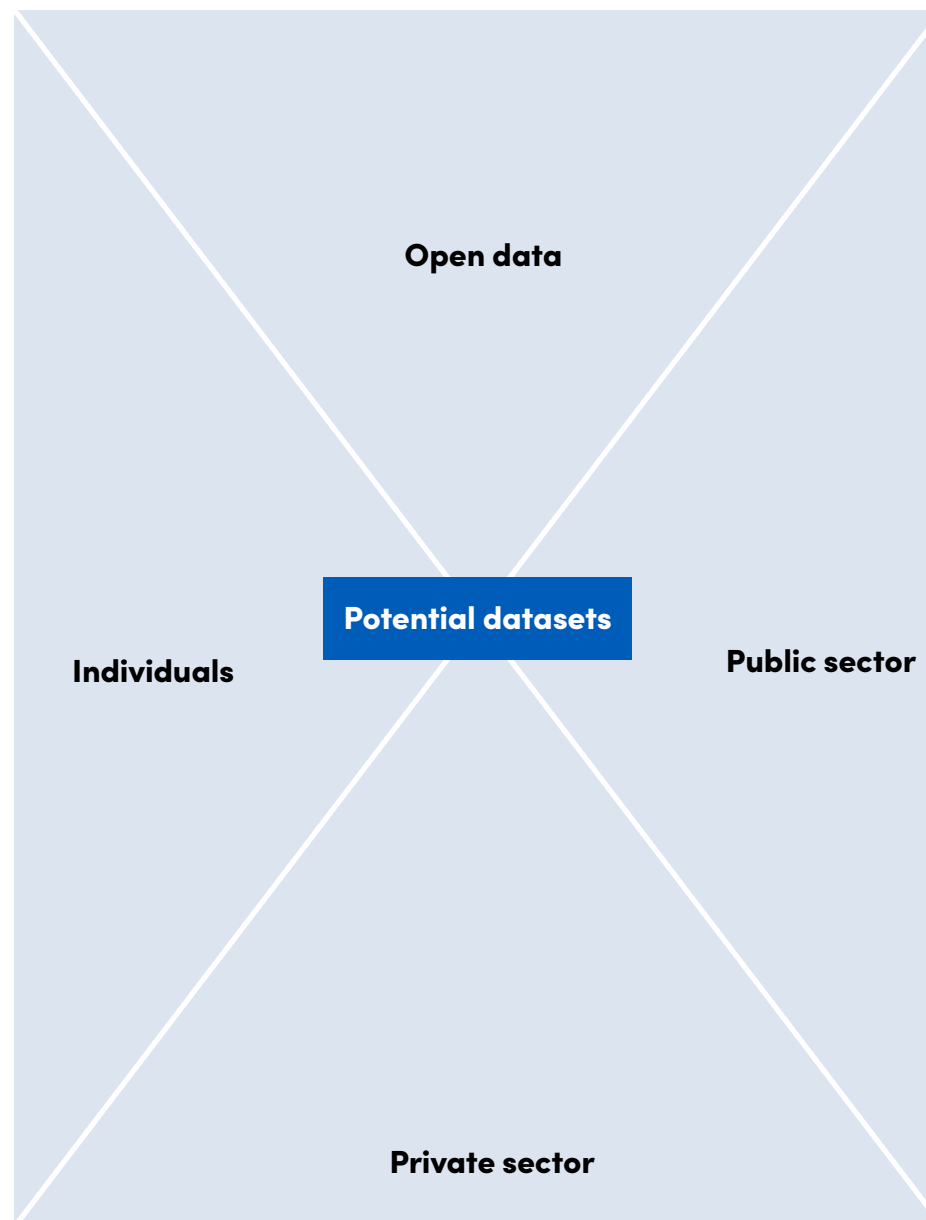
Questions to ask

- What data would you need, and how much of it is already available?
- How are you going to incorporate data if it becomes available in the future?
- What are the data gaps and can you mitigate the effect of inequality in data availability?

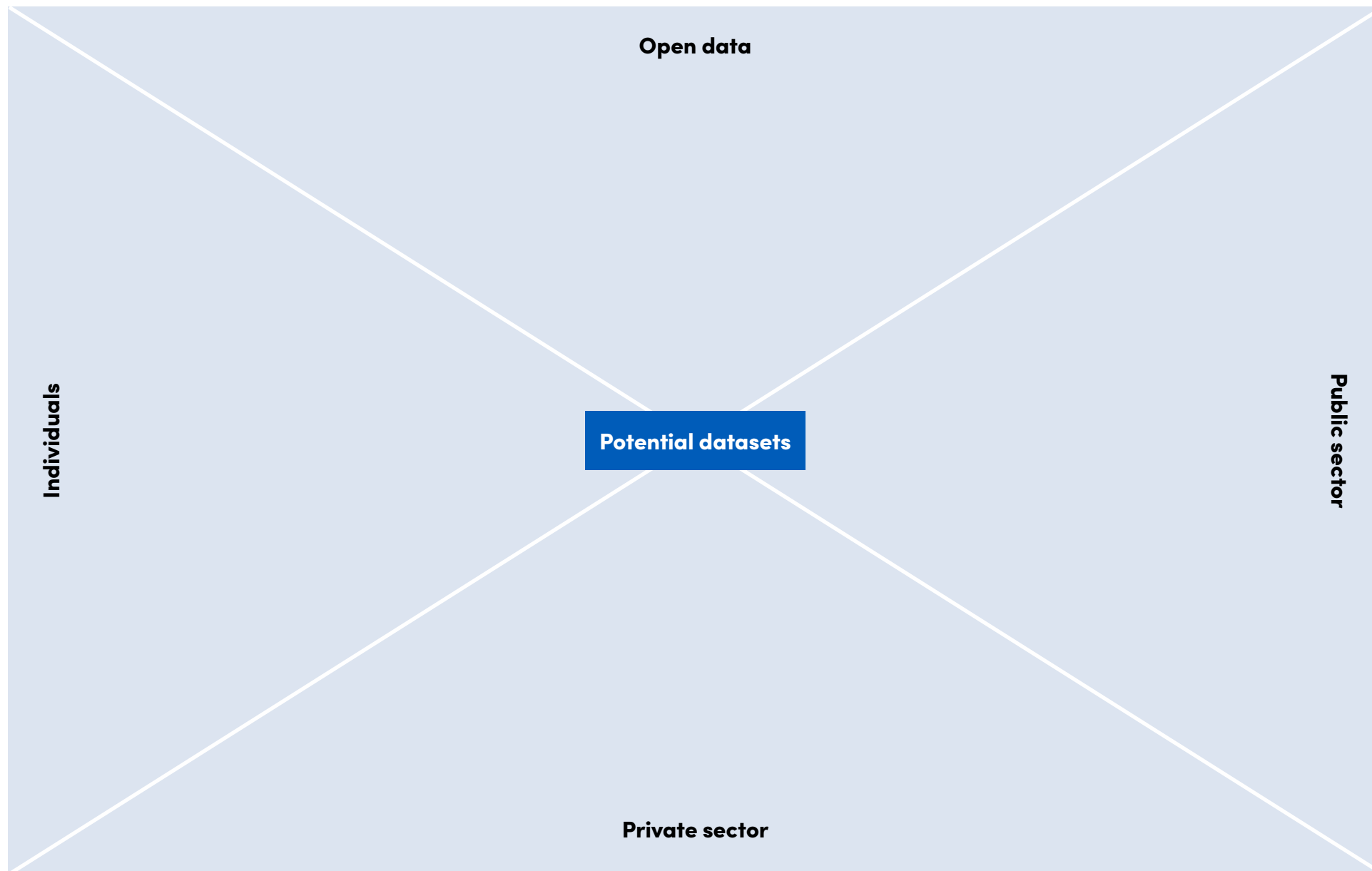
- Where do these datasets fit on the spectrum of closed–open?
- How do we understand public norms around this type of data (e.g. mandatory portability)?
- What form does each of the datasets take?
- How mature is the data?
- How shareable and easy to link is it?
- What is the level of anonymisation of the data?

Useful tools

- Nesta [Dataset Catalogue](#)
- ODI [Data Spectrum](#)
- ODI [Data Access Map](#)
- ODI [Data About Us](#)
- Wellcome Trust, [Understanding patient data](#)
- Local Government Association [Data Maturity self-assessment tool](#)



► What data to share?



► What data to share? – Case studies

Open data

Open data can be published by different actors, in different formats or through APIs. For a list of open data hubs see [27 ►](#)

Public sector

There are many examples of initiatives using data held by the public sector.

Interesting examples are the [pilots](#) from the ODAs, which, combining data from multiple sources, tackle issues of domestic abuse, school readiness, business and building inspections and gang violence.

A growing number of cities are now interested in unlocking the value of urban data, and considering to make it freely available to the public, such as in the case of [Transport for West Midlands](#), using Chordant's oneTRANSPORT Data Marketplace to help deliver improved transport services for residents and travellers across the region.

Another example is environmental and air quality data (e.g. [AirNow](#)).

Private companies

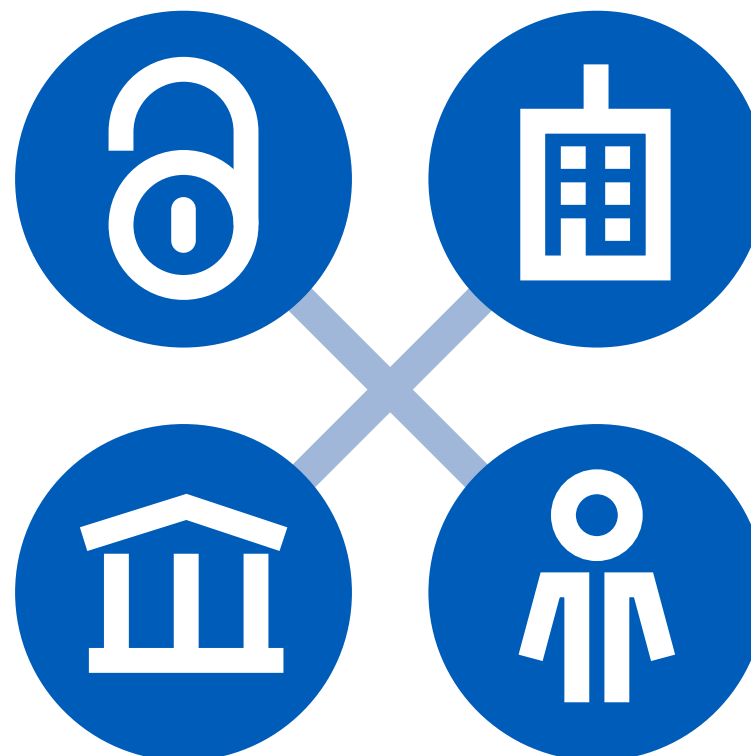
Data held by private companies are a very rich resource. Examples include:

- Bank account information and customer data at transaction level ([Open Banking](#)).
- De-identified data types produced by mobile network operators, increasingly used to inform public health, urban planning and crisis response (e.g. [Flowminder](#), [Seoul Owl Bus](#)).
- Energy consumption data (e.g. the [Ontario Smart Metering Initiative](#)).

Individuals

Examples of initiatives involving citizen data include:

- Crowdsourced traffic data and road conditions from drivers and public transport ([Google Waze](#)).
- Record data such as noise levels, pollution, temperature and humidity ([DECODE's citizen sensing pilot in Barcelona](#)).
- Health data ([MIDATA](#)).



Who to involve?

How to use this canvas

Review the questions and, using the stakeholder map below, identify the people you will need to involve (place at the centre the most essential ones).

Questions to ask

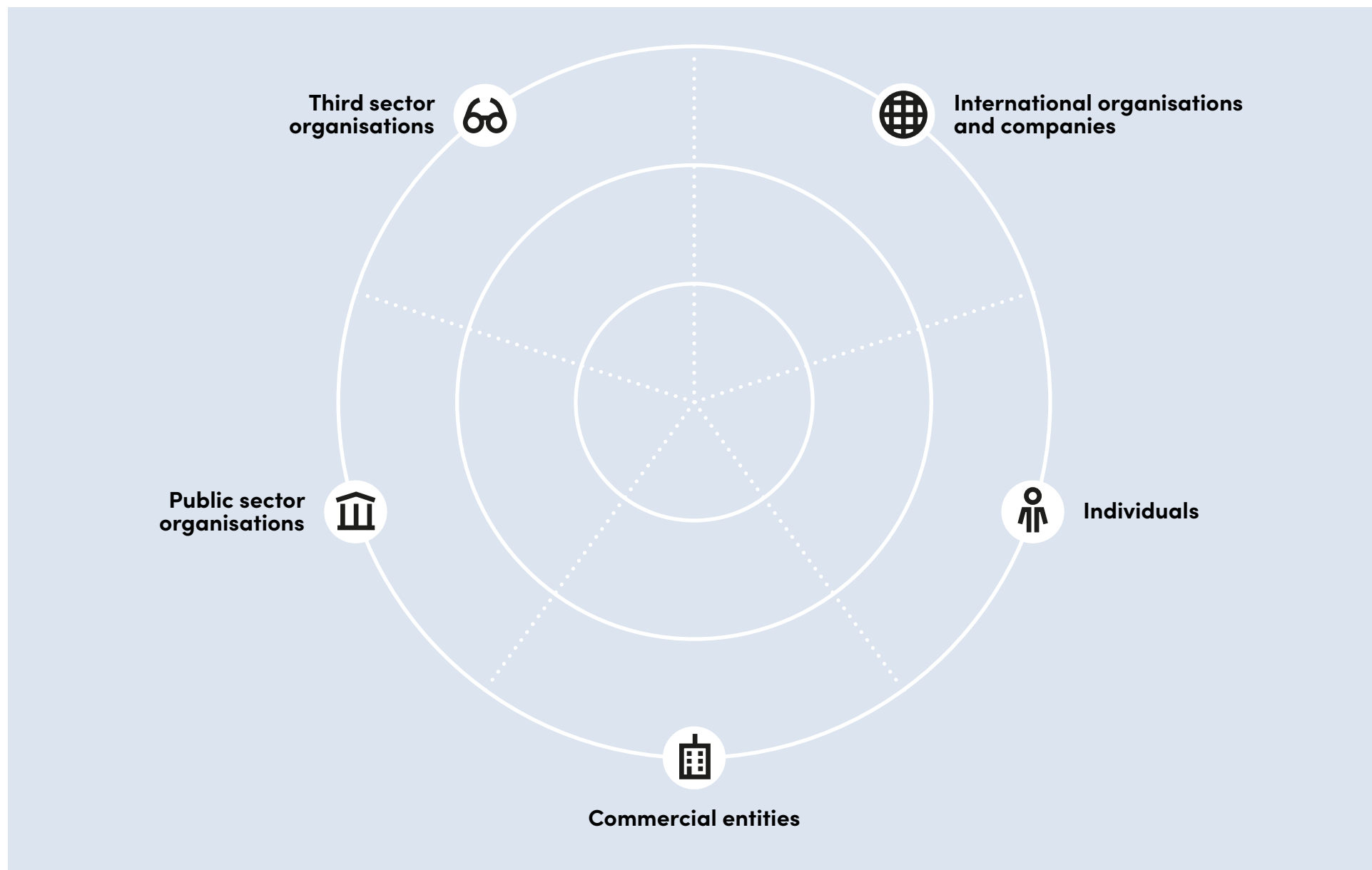
- Who is initiating the partnership?
- What are the incentives of stakeholders to take part in the partnership?
- Assess what the value distribution is in the partnership (i.e. is there an equity of value among all stakeholders? Who will not benefit?)
- Where is funding coming from?
- Who holds data that relates to this use case?
- Is there anyone else besides data providers who are needed to make this project work? For example, do you have the expertise in terms of data science?
- When data subjects' involvement is required, how are they represented in the decisions?
- Is there an option for opting in or out?

Useful tools

- ODI [Mapping Data Ecosystems](#)
- Nesta [Partnership Toolkit](#)



► Who to involve?



What is the overall governance structure?

How to use this canvas

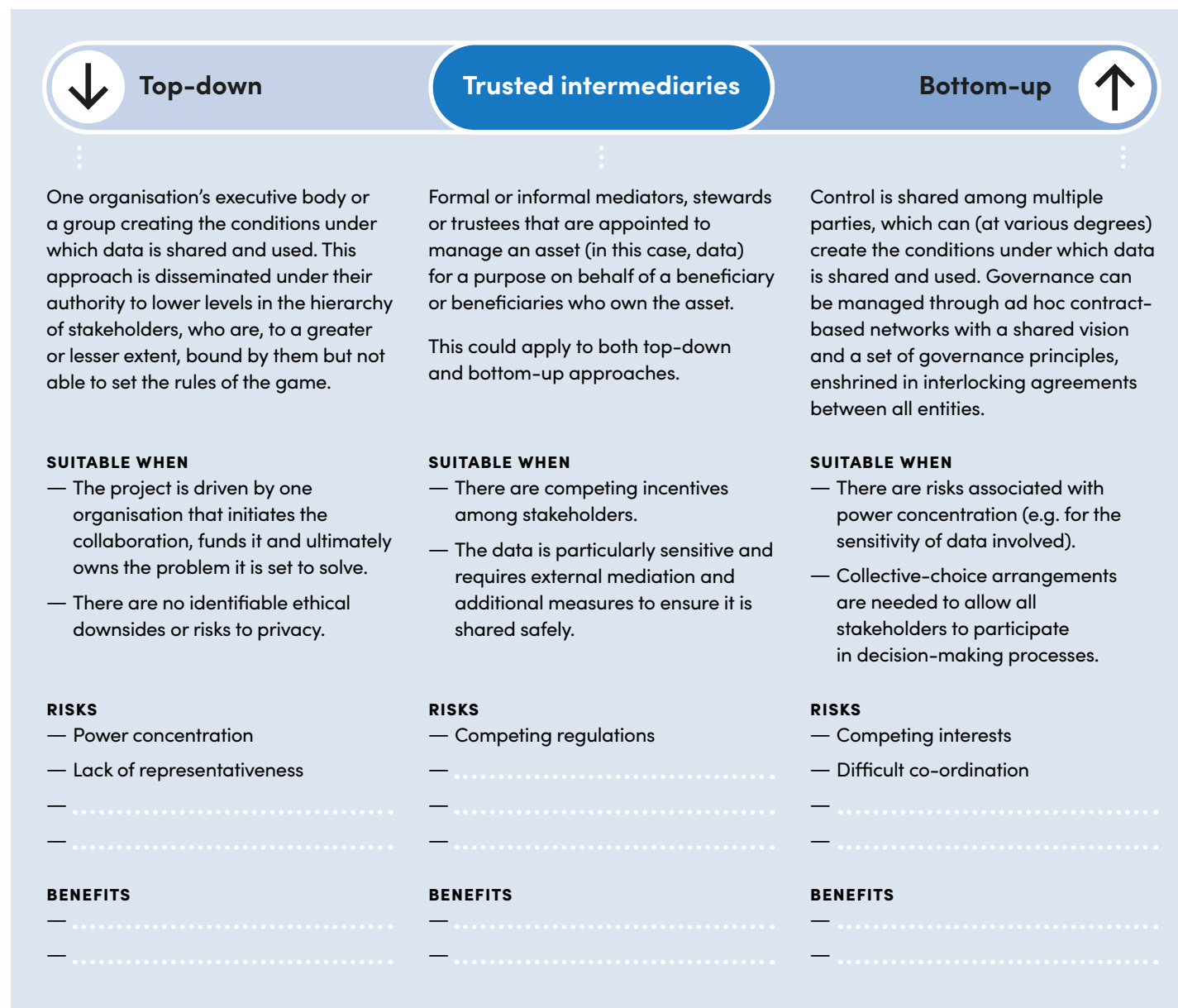
Discuss the questions below and identify the approach your data sharing initiative should take, considering risks and benefits summarised below.

Questions to ask

- How will each party report progress?
- How will decisions be made and in which forum?
- How will conflicts between the parties be resolved?
- Would anyone external be brought in?
- Who is accountable for what?
- How is risk being managed and, if needed, mitigated?

Useful tools

- Royal Society and British Academy, [Data management and use](#)
- ODI [Lessons from pilots](#)



► What is the overall governance structure? – Case studies

There are various options for governance of data sharing initiatives, which here are presented as a set of options on a spectrum of top-down to bottom-up approaches, depending on the power dynamic among those who make the decisions on the structure of the partnership, those who will run the data sharing initiative, those who will provide the data and the outputs derived from it.

In reality, many options will take the form of a hybrid, with some top-down involvement (either from the public or private sector), combined with an element of self-governance by other stakeholders (e.g. a city, citizen group, commercial entity).

- Examples at the top-down end of the scale include the Sidewalk Labs' smart city project in Toronto and the partnership between DeepMind and NHS in the UK, both beset by controversy and criticism, mainly around the lack of meaningful public engagement, the choice of providers and issues of data governance. There are, however, very successful top-down initiatives that involve data sharing such as:

- The Cancer Genome Atlas (note that this, like many other health research initiatives, involves highly sensitive data).

- The AirNow partnership on air quality data (therefore using less sensitive data).

- Top-down approaches where an intermediary is introduced include initiatives such as:

- The Ontario Smart Metering Entity in Canada

- The ODI data trust pilots in the UK

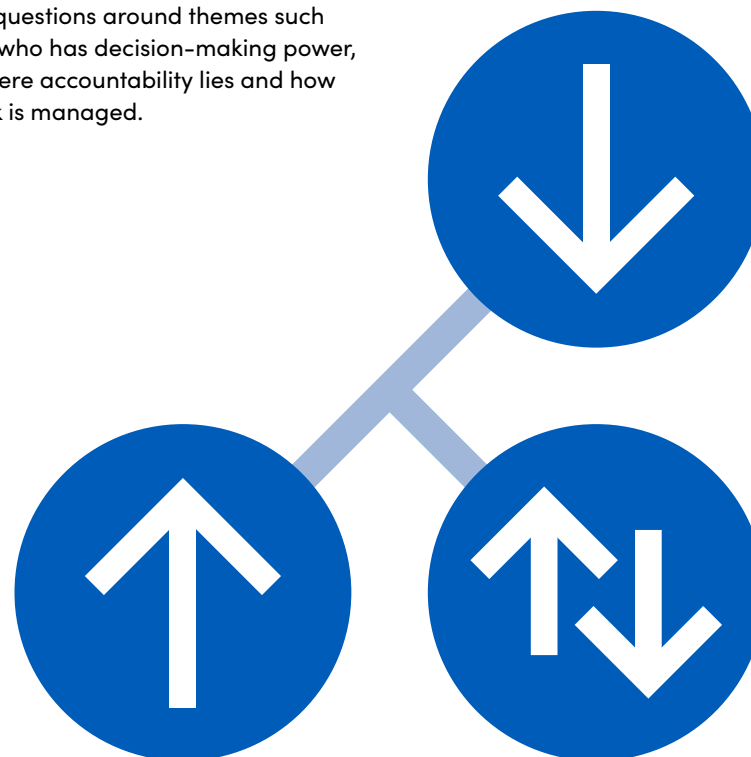
- At the opposite end of the spectrum, it is difficult to imagine a purely bottom-up model, as infrastructure and technologies often include decisions made by those outside of the stakeholders' group, and can end up unrepresentative of the population. Examples of initiative that get closer to bottom-up approaches are:

- Membership models, often defined as 'data co-ops', which give people shared ownership and decision-making power over a platform that gathers

and shares data. These are early-stage, but examples like MIDATA or Saluus.coop have tried to show how this could work in practice.

- DECODE project pilots, which demonstrated how bottom-up approaches supported by enabling city authorities can operate as an effective hybrid model.

The best-fit governance model will depend upon the answers to a set of questions around themes such as who has decision-making power, where accountability lies and how risk is managed.



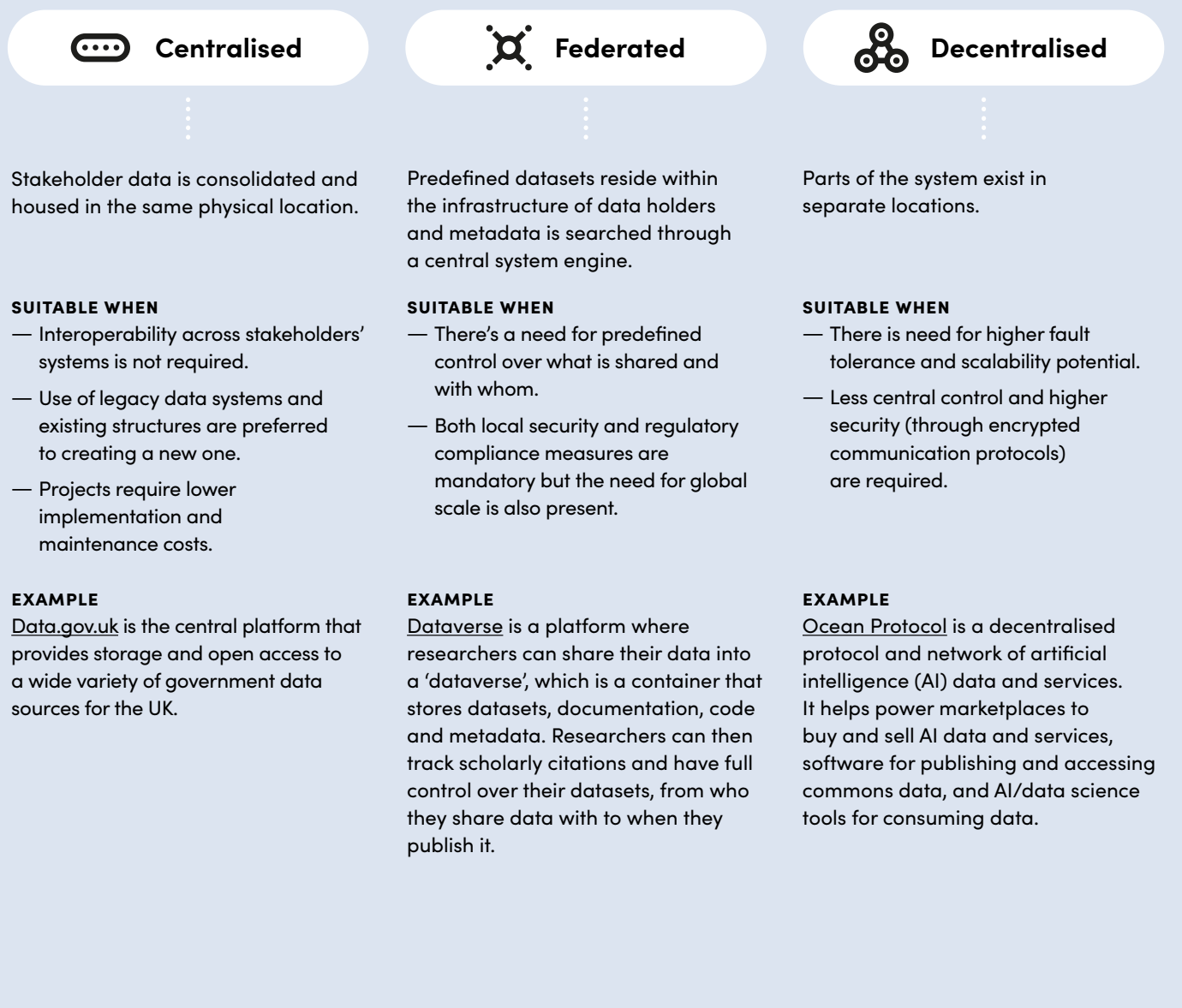
What is the appropriate data infrastructure?

How to use this canvas

On the spectrum, identify the architecture that is most suitable to your needs by reviewing the options provided below.

Questions to ask

- What is the structure that best fits the purpose and why?
- Will the data need curation? If so, who is responsible for it?
- Future proofing: what measures have been considered if circumstances change?



How is data accessed?

How to use this canvas

Think about the levels of security needed for your data. Within your own data sharing initiative, indicate what kind of data needs to have no access, restricted access and open access.

Questions to ask

- Will data be accessible?
- If so, what is the access model to the data?
- Is there the need to have multiple access models?
- If restricted, what kind of restriction does it require and why?



No access

The partnership requires sharing data, but this will not be made available to stakeholders.



Open access

Nationally and internationally, there is increasing commitment to the principle that data which are publicly funded should be publicly available. Open data is becoming increasingly available throughout the world, released by governments as part of the transparency agenda, in the forms of open APIs or open data hubs.

Below is a selection of links to open data hubs. This is not intended to be a complete list, but an indicator of what is openly available.

- **INTERNATIONAL LEVEL** Global development [World Bank Open Data](#), [HDX](#) for humanitarian data.
- **REGIONAL/FEDERAL/NATIONAL** [US City-Data](#), the [EU Open Data portal](#), the [UK data.gov.uk](#), the [ODI certified datasets](#).
- **CITY-LEVEL** [NYC Open Data](#) for the city of New York, [Amsterdam Open Data](#).

Some cities, like London, are now exploring how their open data platforms could support the sharing of data that isn't necessarily suitable to be published for public consumption. Access [here](#) more information on the future of the London Datastore.

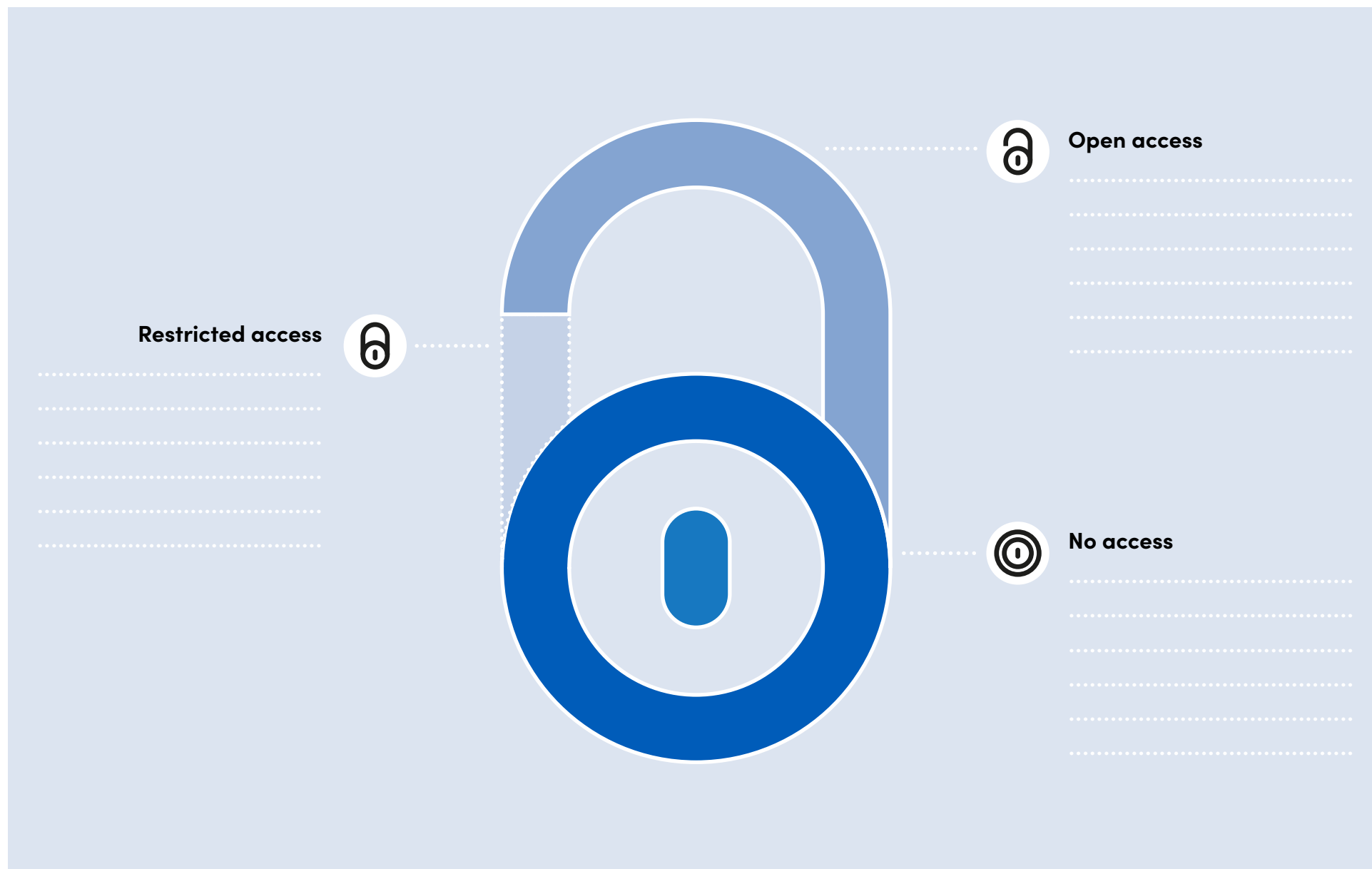


Restricted access

Access is regulated under specific terms, depending on the sensitivity of the data and governance arrangements. Some examples of data access types are:

- **USER REGISTRATION** e.g. The SeaDataNet portal and all metadata services are public domain. However, a user registration is required for submitting requests for datasets and for downloading datasets from the distributed data centres, which is arranged via the [Common Data Index \(CDI\) service](#). The user registration is required to ensure that users agree with the [SeaDataNet data policy and its associated User Licence](#), which rules all dataset deliveries via SeaDataNet. Moreover, it gives SeaDataNet partners insight in its users and their data requirements.
- **LICENCES DEPENDENT ON APPROVAL (I.E. TIERED ACCESS, MEMBERSHIP)** e.g. Access to the [Dementia Platform UK Data Portal](#) can be requested through an application process, limited to the variables needed for the scientific overview of the proposed project. The proposal is circulated to the data guardians of the cohort data requested. The analysis platform is housed within a separate remote desktop, which will appear as a window on the researcher's personal host computer. Data cannot be downloaded from the analysis platform. Summary tables for the purposes of reporting may be downloaded following additional approval.
- **VALUE EXCHANGE (MONEY/TOKENS)** e.g. any data marketplace, such as [Qlik DataMarket](#).

► How is data accessed?



B: PROJECT FOUNDATIONS

It is critical to recognise the important role that people play in supporting (or hindering) the success of data sharing initiatives.

At the end of the day, something as sophisticated as a data sharing initiative will only succeed if the people involved are working constructively in partnership with each other.

Investing in open and clear communication and effective working relationships is as critical as any other aspect of this work to ensuring success.

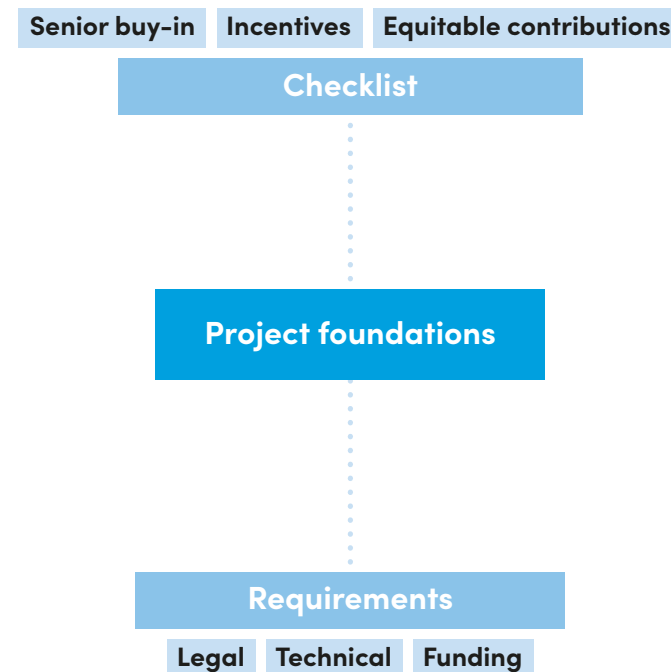


Project foundations

Once the key decisions are made, a data sharing model, or perhaps a number of possibilities will emerge. The important question then becomes how to turn that vision into a reality. There are two sets of considerations in this phase.

- A checklist, like the one on the following page, will help you go through the considerations that need to be addressed before initiating the partnership.
- The following section on requirements sums up additional requirements to consider, including legal, funding and technical requirements.

As previously mentioned, it is likely that in the design phase there will be a need to go back to the decision matrix while tackling a specific element of the project foundation and vice versa.



Checklist

What to do

To establish a good collaboration environment there are three elements that need to be addressed.

If you can't tick all three boxes, you might want to consider how to identify blockers and solve them before starting the process. This will avoid more issues down the line.

Tools

[Nesta's Partnership Toolkit](#) is a very useful resource to support this element of the process, which identifies the practical steps that help create a successful partnership, write an effective partnership agreement and get stakeholders' collaboration off to a good start.

The [EAST framework](#), developed by the Behavioural Insights Team from its experience of applying behavioural insights over the past few years, sets out four simple principles for influencing behaviour – make it easy, simple, attractive and timely (EAST).

□ Senior buy-in to work together

Are you engaging with people senior enough to make decisions and unlock issues when they arise? What signals have you been given that there is senior buy-in, on both sides? Do you think this can last the test of time?

□ A clear incentive for all parties to be involved

For a data sharing partnership to work, all parties must benefit. You might have to help your potential partner to understand how they will benefit from the partnership. Remember that they will need to sell the idea internally, no matter how senior they are. Think outside the box and seek input from others who bring a fresh perspective. To help them understand the business case for partnership, consider whether they might benefit from these incentives:

LEGAL INCENTIVES

Regulatory measures can be taken by the government to compel data sharing.

An example is [Open Banking](#), whereby the UK government

has compelled banks to make their data open through APIs that authorised third-party organisations can use to develop personalised financial services.

ECONOMIC INCENTIVES

Value that directly or indirectly affects the bottom line by increasing revenue or reducing costs (such as efficiency gains, enlarging customer base or creating a competitive advantage) or from the direct commercialisation of data (e.g. any data exchange platform or market).

Sometimes an economic incentive to solve a particular problem can be created by announcing a challenge prize, such as in the case of the [Taiwan Presidential Hackathon](#).

NON-ECONOMIC INCENTIVES

If the benefits that arise from sharing data are not strictly economic nor result from regulations, they fall into this category. These include considerations on the common good (i.e. if the value generated by sharing data benefits society at large). An interesting case is [MIDATA.coop](#), a Swiss co-operative that gives people control over their

medical data. One of the things that makes MIDATA different from other data storage platforms is that it does not use monetary rewards, but wider societal benefit to encourage data sharing, as they consider financial incentives the wrong incentive for people to share their health data.

Another important set of incentives in this group respond to the principle of reciprocity (i.e. people or organisations participate with the aim of helping and receiving advantages at the same time, of which reputation is a good example).

□ Equitable contributions

Each party involved in the data-sharing arrangement must be able to offer something to the partnership; however, making an equitable contribution does not mean making an *equal* contribution. Examples of contributions include money, time, resources, expertise, connections or data. What is required for a partnership to succeed is for all stakeholders to be clear and happy that the contributions brought into the partnership are fair and valuable to all parties.

Requirements

What to do

Each of these will vary considerably from initiative to initiative and will require ad hoc advice from legal and technical experts.

Make sure that everything decided in this sphere is aligned with the design decisions.



Regulation

Navigating this will require an ad hoc analysis of the project at multiple levels and will require:

- Investigating what the relevant legal/regulatory requirements are.
- Ensuring that the project complies with these requirements and justifying this to relevant authorities.
- Being aware of any evolving legal requirements and communicating these appropriately to relevant stakeholders and, when appropriate, to the public.

For example, the data protection regime in the local context where the project is held will influence the structure of the partnership. For global technology companies, though, the problem isn't the imposition of a single regulatory regime, but rather the need to consider many, potentially conflicting ones to maintain their global businesses.



Technical

Together with individual evaluations of data maturity and technological infrastructure audits, other considerations should include:

- **DATA QUALITY, STANDARDS AND SHARING FREQUENCY**
Poor data quality and systems that are not interoperable pose challenges where ongoing data sharing is required. It is important to understand whether the project will require a one-off sharing, or whether it will require ongoing, routine data exchange. This will have a direct impact on the technical architecture that is needed for supporting data sharing.
- **IS THE DATA SHARING ARCHITECTURE GOING TO BE OUTSOURCED OR DEVELOPED AD HOC**
Commercial pre-built solutions are available on the market, but might carry constraints in the way data is handled. Tailor-made data sharing technical architectures can also be developed in-house or contracted from external consultants and suppliers.



Funding

Different types of data sharing partnerships will need funding structures and mechanisms designed to support adequately the type of partnership.

- **INVESTMENT OF ONE LEAD ORGANISATION**
Such as a government/private company or third-sector organisation.
- **EQUAL OR TIERED FUNDING**
One, some or all partners form the oversight of the initiative, each contributing a predetermined amount (as per their tier).
- Tiers may be organised based on the level of input and/or incorporate a 'cost-free' tier, where partners can document their interest and support through data provision or consultancy where necessary.
- **EXTERNAL FUNDING**
e.g. a grant, sum of money from a government or other organisation for running a particular data sharing project or pilot.
- **COMMERCIALISATION AS PART OF BUSINESS MODEL**
with income generated through selling data created (e.g. [23andMe](#)).

Copyright © Smart Dubai. 2020.
All rights reserved.

No part of this work may be reproduced
or transmitted in any form or by any means,
electronic, manual, photocopying, recording
or by any information storage and retrieval
system, without prior written permission
of Smart Dubai.

Designed by soapbox.co.uk

