# User-Centric Data Publishing

A toolkit to help you publish data openly, in ways that work for your users

## Contents

# About

- What is this toolkit for?
- Who is this toolkit for?
- Toolkit status: beta version?

*GRAPHIC DESIGN: Front page design please, perhaps with image showing different groups using data for different things (environmental concerns, education, gender fairness etc) - to convey the "social good" aspect*

# Introduction

Access to data is vital in tackling the big challenges we face – from the earlier detection and treatment of disease to reducing pollution in urban spaces.

Data also has an important role to play in driving economic growth, by supporting the creation of new technologies, products and services.

[In order to help help organisations realise the value in opening up and collaborating around data, Microsoft's Open Data Campaign xxx]

[something from the ODC, IDSIP, our partnership on why data openness/sharing is important].

[talk about why we developed this toolkit; why user centricity is so important]

[give on paragraph about each section/steps; why important; what it covers]

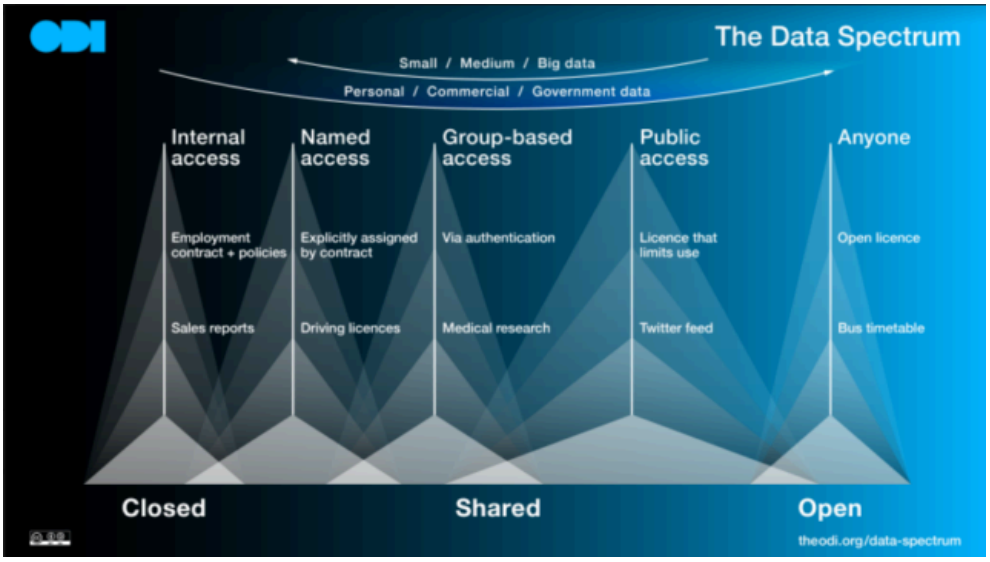
# Building the foundation for open data

It is helpful to think of data as existing on a spectrum of rights to how we can access it, from closed, to shared, to open.

Open data is data that is available for anyone to access, use and share. It is published under an open licence that allows it to be used for any purpose. Some data cannot be made open because it contains sensitive information about individuals or groups. It may still be possible to share that data with specific organisations, so long as there are appropriate safeguards in place.

Shared data is data that is only available to certain people or groups, such as researchers. Data that is shared will typically be made available for specific purposes that are defined by, for example, a data sharing agreement. There are many ways in which data can be made accessible, or 'shared', with others.

Closed data is data that is held privately within an organisation, like employment contracts and policies, or sales reports.




[Placeholder The Data Spectrum graphic]

We believe that if data is made as open as possible – while protecting privacy, confidentiality and security – it will contribute to a world where everyone can collaborate around data to make better decisions, improve efficiency and even help tackle some of the world's most pressing societal challenges.

For more information about the core definitions and principles of open data, check out the following resources:

- 'About open data' resources list
- eLearning - Open Data Essentials

A useful starting point for those who want to publish open data can be to undertake an open data maturity assessment, to assess how well prepared you are to publish open data and to identify actions for improvement where necessary.

The Open Data Maturity Model can help organisations to undertake this type of assessment. The model supports the assessment of operational and strategic activities around open data, provides guidance on potential areas for improvement, and helps organisations compare themselves against one another to highlight their respective strengths and weaknesses, adopt best practices and improve their processes. Organisations can use the model to set themselves appropriate goals based on their current maturity, resourcing and anticipated benefits.

The model consists of five themes:

- Data management processes
- Knowledge and skills
- Customer support and engagement
- Investment and financial performance
- Strategic oversight

Each of these thematic areas is scored against five levels of maturity:

1. Initial — the desirable processes are non-existent or ad hoc, with no organisational oversight
2. Repeatable — processes are becoming refined and repeatable, but only within the scope of individual teams or projects. There are no organisational standards
3. Defined — processes are standardised within the organisation based on best practices identified internally or from external sources. Knowledge and best practices start to be shared internally. However the processes may still not be widely adopted
4. Managed — the organisation has widely adopted the standard processes and begins monitors them using defined metrics
5. Optimising — the organisation is attempting to optimise and refine its process to increase efficiency within the organisation and, more widely, within its business sector

In the next section, we will cover the key considerations for each theme. We will also provide you with a printable template to undertake your own assessment, which may be more appropriate for an in-person workshop setting. For those working independently, you may prefer to use Open Data Pathway - a free online tool, based on the Open Data Maturity Model, which will help you to determine your current open data maturity level.

For a deeper understanding of each theme, please refer to our full research report.

## Data management processes

Within a mature open data organisation, a number of business processes will underpin the effective management of datasets. These processes will support both the release and reuse of open data.

Strong data governance helps to ensure that an organisation effectively maintains its data assets. Managing data quality is important regardless of how data is subsequently licensed and shared. However some data management practices may be less applicable for open data. For example, managing access and security is not a concern given that open data is accessible to anyone, by definition.

Conversely, there are practices that are particularly relevant for open data. These include issues such as:

- anonymisation and aggregation of data to remove sensitive information
- redaction of personally or commercially sensitive data
- the adoption of best practices that ensure that published data can be easily reused by third-parties

With this in mind, this theme highlights data management and governance practices that are particularly relevant to open data. But it is recommended that this assessment is made in the context of a wider evaluation of data governance within the organisation.

**Key considerations:**

- Building a process to support the release of data
- Developing and adopting standards for data
- Developing data governance
- Managing sensitive data

For the full range of considerations, refer to the 'data management processes' section of the maturity model.

| 1. Initial | 2. Repeatable | 3. Defined | 4. Managed | 5. Optimising |
|---|---|---|---|---|
| Little or no published open data.<br><br>Datasets that are published include [...] | Specific projects or products may have defined a repeatable process for [...] | There is a repeatable organisation-wide standard release process for [...] | All datasets are released according to the standard organisational process. | The organisation collects and monitors metrics on its release process |

Example assessment criteria - Building a process to support the release of data

## Knowledge and skills

A mature open data organisation will understand the benefits of openness and transparency and apply those principles appropriately.

Supporting the development of this culture, a mature organisation will ensure that staff have the necessary skills and expertise in a number of different areas. This knowledge will range from:

- a common understanding of the value of open data and its application to the organisation
- operational skills required to support data governance and publishing
- strategic understanding, at a senior level, of how to use open data to further the goals of the organisation

An organisation will support employees in developing the skills necessary to deliver on its open data strategy. This may include offering training on key topics such as licensing, technology or data governance.

Access to information is also an important component of developing internal expertise with open data. The organisation will ensure that all organisational standards, datasets and relevant data documentation are accessible to the staff who need it.

**Key considerations:**

- Developing open data expertise
- Knowledge management

For the full range of considerations, refer to the 'knowledge and skills' section of the maturity model.

| 1. Initial | 2. Repeatable | 3. Defined | 4. Managed | 5. Optimising |
|---|---|---|---|---|
| The organisation does not provide any direct training or support for [...] | No shared understanding around open data in the organisation, although some areas [...] | The organisation is aware of where further support and understanding is required [...] | The organisation is actively building a shared understanding around [...] | Knowledge and understanding of open data exists at all levels in the [...] |

Example assessment criteria - Developing open data expertise

## Customer support and engagement

Publishing high quality open data involves more than making data available online: reusers will need support in correctly interpreting and using that data. At a minimum this will require that data is published with enough documentation and supporting metadata that reusers can understand the structure, scope and provenance of a dataset.

A mature organisation will also consider additional ways that it can support its reusers such as via a help desk function or similar services. Growing a community around a dataset might involve helping interested reusers connect with each other to share experiences and benefits.

Open data is published for a variety of reasons. But regardless of whether data is published to meet legal obligations or as part of an open data business model, an organisation will need to plan how it will engage with its customers, monitor how its data is being used and identify where value is being created.

**Key considerations:**

- Developing an engagement process with data reusers
- Documenting your open data
- Building a reuser support process
- Creating open data community norms

For the full range of considerations, refer to the ‘customer support and engagement’ section of the maturity model.

| 1. Initial | 2. Repeatable | 3. Defined | 4. Managed | 5. Optimising |
|---|---|---|---|---|
| Little or no attempt made to identify potential re-users for released [...] | Some teams attempt to identify and engage with potential reusers [...] | The organisation identifies a repeatable approach for [...] | The organisation begins tracking the effectiveness of its [...] | The organisation is routinely tracking metrics relating [...] |

Example assessment criteria - Developing an engagement process with data reusers

## Investment and financial performance

Publishing open data, especially over the long-term and to a high standard, requires ongoing investment in both people and infrastructure. These costs may be offset by financial benefits from opening data, such as through exploiting new business models, or efficiency savings from deduplication of data curation efforts or easier integration with third-parties.

Similar benefits accrue from the use of third-party open datasets. An open dataset might replace a costly licensed alternative. Costs to maintain an internal dataset might be reduced by switching to an open collaborative model for managing the data.

A mature open data organisation will quantify both the costs and benefits that relate to its open data practice. By valuing both published and unpublished datasets, the organisation will be able to prioritise and justify its ongoing investment.

**Key considerations:**

- Ensuring financial oversight
- Developing dataset valuation processes
- Building open data into procurement practices

For the full range of considerations, refer to the 'investment and financial performance' section of the maturity model.

| 1. Initial | 2. Repeatable | 3. Defined | 4. Managed | 5. Optimising |
|---|---|---|---|---|
| Data releases are unfunded and done as exceptional expenditure. | Individual projects may include open data publication costs as [...] | Project funding and operational costs routinely include long terms [...] | The organisation actively monitors the financial costs and [...] | The organisation looks for efficiency savings around [...] |

Example assessment criteria - Financial oversight

If you are thinking about how to sustainably provide access to data that your organisation publishes, why not check out our resources on Sustainable Data Access?

Resources:

Designing sustainable data institutions (report)
Data institutions: Reducing costs and improving sustainability (report)
The Sustainable Data Access Workbook (tool)
Sustainable access to data (podcast)
Making data institutions financial sustainable (podcast)
Putting the sustainable data access workbook into action (podcast)
Developing a sustainable business model without compromising commitments to open data (use cases)

## Strategic oversight

Given the impacts of open data practise on internal processes and the potential financial benefits and investment required, a mature open data organisation will ensure that its adoption of open data is closely aligned with its wider organisational objectives.

Both internal and third-party open datasets will be recognised as assets that should be carefully managed. Close alignment with company strategy will ensure that objectives for delivering on the value associated with open data will be reflected at a senior management level.

**Key considerations:**

- Shaping open data strategy
- Asset catalogue management

For the full range of considerations, refer to the 'strategic oversight' section of the maturity model.

| 1. Initial | 2. Repeatable | 3. Defined | 4. Managed | 5. Optimising |
|---|---|---|---|---|
| The organisation has no strategy or policy with regards [...] | Individual business units identify benefits to open data for their [...] | The organisation defines an open data strategy. This will [...] | The organisation has aligned delivery on open data policy [...] | The organisation is using open data as a key element of its [...] |

Example assessment criteria - Open data strategy

# Ethics and transparency

In considering how mature your organisation's data practices are around collecting, maintaining, publishing data, it is also important to be aware of how the impacts of your work with data may affect others. Organisations should consider data ethics - a branch of ethics that evaluates data practices with the potential to impact on people and society - in data collection, sharing and use.

The Data Ethics Maturity Model is a tool for anyone who collects, uses and shares data. It helps assess and benchmark how widely embedded data ethics culture and practices are across your organisation, and helps you to consider what your future ambitions might be.

The themes of the Data Ethics Maturity Model align with the themes of the Open Data Maturity Model, but break down in a slightly different way:

- Organisational governance and internal oversight – highlights the need for an organisation to have a clear strategy around ethical data practices, and a leadership with responsibility and capacity to deliver that strategy
- Skills and knowledge – highlights the steps required to create a culture where ethical data practices are embedded by identifying the knowledge sharing, training and learning required within an organisation

- Data management risk processes – identifies key business processes that underpin ethical collection, use and sharing of data, focusing on identifying and assessing risks of harm to individuals and communities
- Funding and procurement – highlights the need for organisations to invest in embedding ethical data practices, and to reflect requirements in procurement processes
- Stakeholder and staff engagement – addresses the need for organisations to engage both with communities reflected in, or impacted by, data they are collecting, using or sharing, and organisations they are sharing data with or using data from
- Legal standing and compliance – reflects the need for organisations to abide by relevant laws, regulations and social norms to avoid harmful impacts from collection, use and sharing of data.

The model uses the same maturity levels as the Open Data Maturity Model:

1. Initial — the desirable processes are non-existent or ad hoc, with no organisational oversight
2. Repeatable — processes are becoming refined and repeatable, but only within the scope of individual teams or projects. There are no organisational standards
3. Defined — processes are standardised within the organisation based on best practices identified internally or from external sources. Knowledge and best practices start to be shared internally. However the processes may still not be widely adopted
4. Managed — the organisation has widely adopted the standard processes and begins monitors them using defined metrics
5. Optimising — the organisation is attempting to optimise and refine its process to increase efficiency within the organisation and, more widely, within its business sector

You can find the full Data Ethics Maturity Model available here.

# ~~Making data user-centric~~ A user-centric approach

This section offers insights, tools and guidance for data publishers to meet the needs of data users more effectively and to encourage productive engagement with data users..

USER CASE SECTION ADDED (James to see)

**The Use Case**

Articulating how a data user will interact with a data product or data system is essential for understanding requirements and engaging in communication at the high-level stakeholder level and at the data consumer level.

A use case is typically a project management tool that provides a description of the ways in which a user interacts with a system or product. For data publishers, the same can apply; developing a strong user case can help to assess and adapt the effectiveness of your data offer. A use case should establish and define a few key components:

- ☐ **System**: the product or service (For example, a spreadsheet that contains traffic data or a dashboard with several automated charts.)
- ☐ **Actors**: a data user or anything else that interacts with the system. (The actor could be another system, a piece of hardware, an internal colleague, an external customer or an entire organisation. An actor either initiates interaction with the system, or provides a service to the system.)
- ☐ **Scenario**: a specific sequence of actions and interactions between actors and the system (Thinking about
- ☐ **Success scenario**: A use case outlines the success and failure scenarios that can occur when the actor(s) interact with the system. In this section, you'd establish the main success scenario, i.e., the most desirable outcome between the actor and the system. You would also establish the alternative paths, which explain what happens in the event of failure or error.

A user case can help to:

- Manage the **scope** of the publication or project
- Establish the **requirements** of each actor

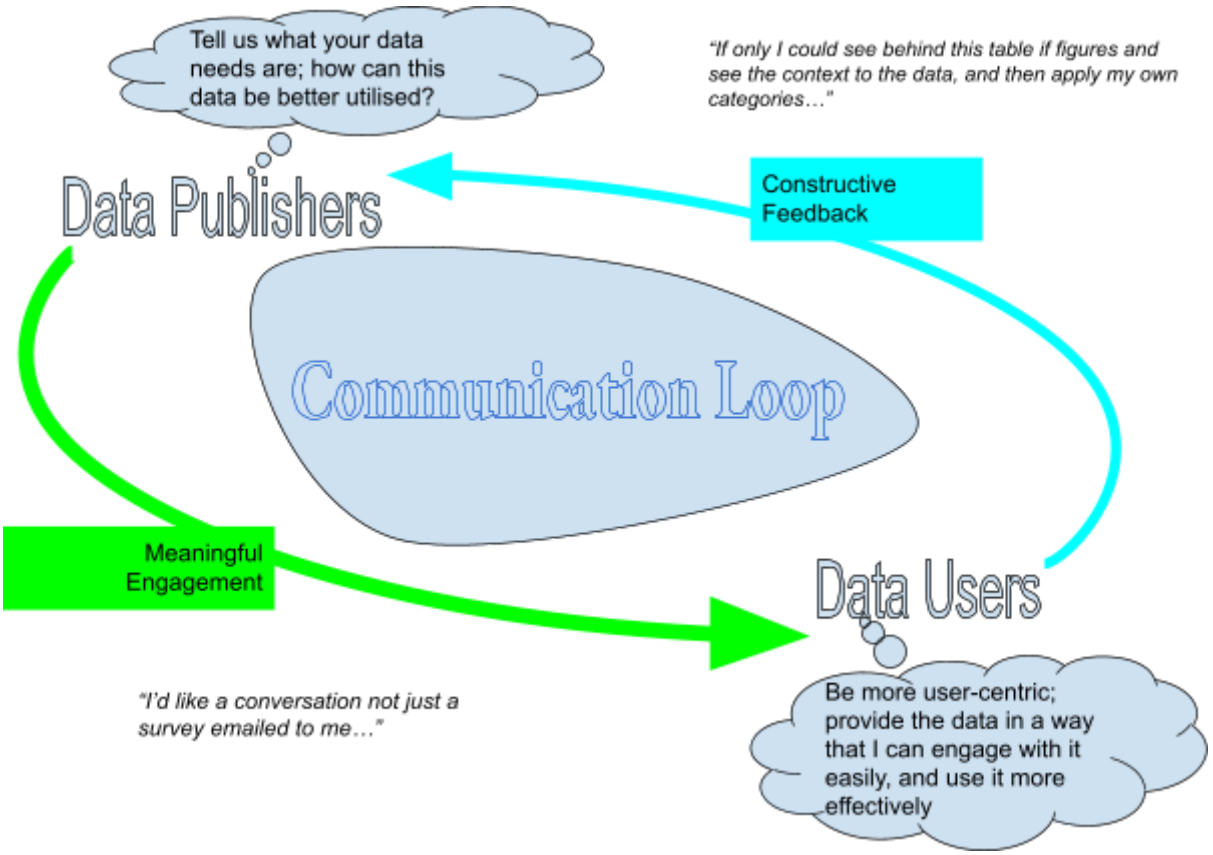
- Outline the ways a user will interact with the system, including which technological **functions** they may require - which in turn helps to communicate technical requirements to business stakeholders
- Visualise **system architecture**, which helps to identify how data should be presented and how data users view it, interact with it and download it
-

*GRAPHIC DESIGN: list above to be replaced with illustration please*

## The Communication Loop

The 'Communication Loop' describes the **two-way communication** necessary to a) fully understand what each party requires from successful open data publication, and b) to facilitate the best ways to meet identified needs. It ensures value is realised for everyone involved, and that there is transparency and openness within a productive open data ecosystem.

The potential to enhance the value of data and the potential to enhance the value of relationships can be realised through more effective engagement between data publishers and data users.  In many ways, this 'shared value' often manifests through working towards a common aim or having similar ambitions in addressing a common issue, such as improving child wellbeing or managing carbon emissions for example.




*GRAPHIC DESIGN: image above to be recreated in a more visually appealing way please*

**The value of data**

A 2021 report on the **value of data assets** describes that one of the ways to understand the value of data was through 'used-based' methods. This states that how data is used, in what context and for what purpose should be considered in understanding the value of data. They also review the "organisational characteristics" of data:

- Size of the firm: a key determinant of the capacity to extract value from data assets;
- Level of productivity in the firm: cause and effect of high-value data assets;
- Organisational culture: an enabler and a multiplier of the value of a data asset.

The report also provides a conceptual framework to assess the benefits and the costs of changes to data assets.

- **Value of data research reports 2022 - GOV.UK**

Another 2022 report on **measuring the value of data and data flows**, discusses different approaches to data valuation, and argues that the value of data depends to a large extent on the data governance framework determining how they can be created, shared and used. (link to data governance section)

- **Measuring the value of data and data flows | OECD Digital Economy Papers | OECD iLibrary**

**A common aim: data for social good (Collaboration with meaning)**

Using data to understand and address a social issue is likely to be a common aim for those interested in sharing data widely, and engaging with data users.

For consistency, we align our definition of 'social good' with Microsoft's Open Data for Social Impact Framework which describes it as: ***'a positive change towards addressing a societal problem, such as reducing carbon emissions, closing the broadband gap, building skills for jobs, and advancing accessibility and inclusion'.*** This definition enables us to focus on how effective data sharing can be used for a broad range of social impact aims.

Realising the value loop is an effective step towards identifying these common aims, and achieving them. This realisation begins with effective communication.

***Accelerate Aspirations: Moving Together to Achieve Systems Change*** is a comprehensive report by Data.org on key trends in the emerging field of data for social impact.

Some of the key takeaways from the report and relevant to the data publisher/user relationship include:

- Data for social good is an emerging 'sector' that has the potential to accelerate positive impact. However, efforts in this space tend to focus on "project-based interventions instead of **advanced data strategies** that focus on systemic solutions, risking the long-term impact and sustainability of their efforts."
- There are increasing opportunities to extract powerful insights through data sharing, particularly cross-sector data sharing - but this requires **stronger collaborative relationships** where there is agreement and protocol for how to share data, share infrastructure, and potentially share talent at scale.
- This new 'data for social good' sector requires "more diverse and interdisciplinary **purpose-driven data practitioners** who can drive change locally" and who can then capitalise on increased job opportunities, increased training and investment and inspired leadership.

### *What does good communication in the value loop look like?*

Two-way communication offers an outlet for all parties to express their concerns, ideas, opinions and needs. This is what helps to build and grow thriving partnerships. A main component of the value loop is instigating feedback.

Two-way feedback in the value loop should be aimed at improving the partnerships between data publishers and data users. It should produce insights that improve what data is shared, how it is utilised and how it can be further developed.

It is the responsibility of all parties to share constructive feedback, reflect on the feedback and respond accordingly. The best feedback helps others understand their strengths and suggests opportunities to build on those strengths, as well as highlighting areas for improvement.

| ***Suggestions for good two-way communication*** |
|---|
| 1. **Focus on the shared value/improvement for all.** Address the situation that can be improved, not the person or their identity or their organisation. No one wants to feel attacked but most people want to be part of affecting positive change/part of the solution. |
| 2. **Be open.** Sometimes feedback can challenge the way we are used to doing things. But feedback alone is useless if we are unwilling to let go of preconceived ideas about how things should be done. There is always room for improvement if we are willing to accept it. See the value of reducing organisational blind spots and commit to adapting attitudes and ways of working. You can play a role by offering to trial new ideas with colleagues or raise them with seniors/decision-makers in your organisation to see if it is something that can be taken forward. |
| 3. **Take responsibility for growth.** Applying a growth mindset is important because a growth mindset says that constructive criticism is valuable for learning. A growth mindset says we ought to reframe challenges as learning opportunities. We should aim to take ownership and responsibility for the value loop - ensuring everyone benefits from improvements. |
| 4. **Ask for feedback often.** Inviting feedback regularly creates reassurance that people are being listened to; a culture of open communication is important and strengthens the value loop beyond a tick box exercise or an annual survey. This is also key for staying ahead of the curve when it comes to upcoming challenges or changing data requirements. |

***What might this communication look like?***

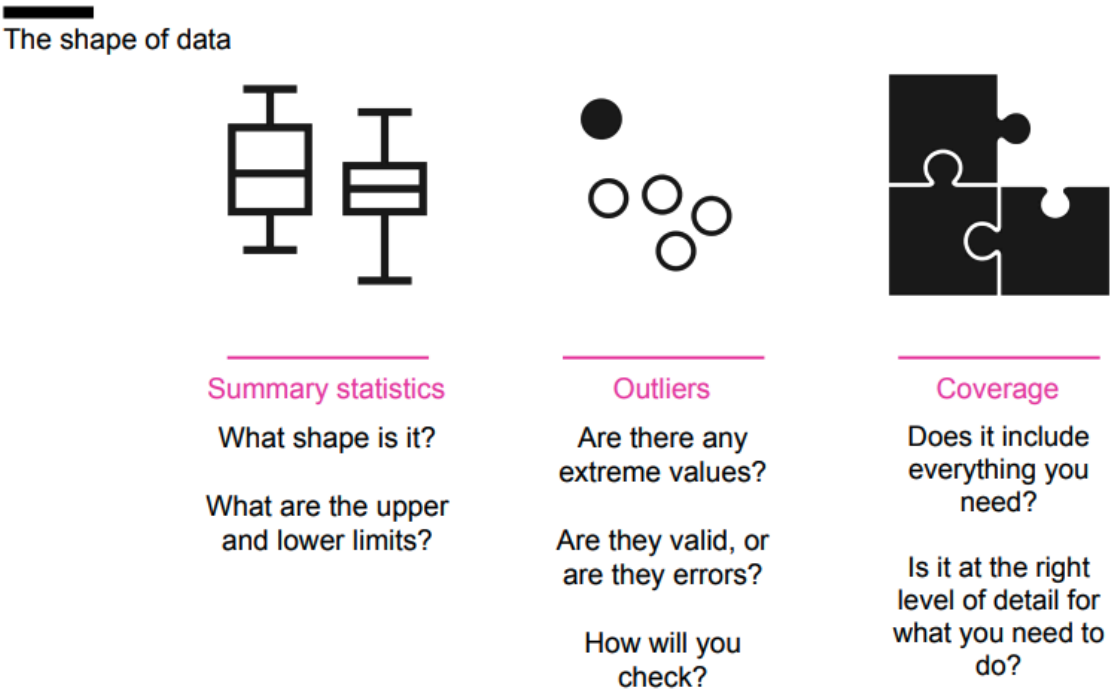
The following gives examples of what we might ask to stimulate a conversation for useful feedback between data publishers and data users.

| **Data Publishers** | **Data Users** |
|---|---|
| How do you use the data we publish? | How can I use the data you publish? |
| Is the data easily accessible for you? (format, permissions, visibility) | Can I access the data in these formats? (raw data, designed charts, editable charts, Excel tables, other) |
| How often do you need this data? | How often will this data be updated? Will |

| | |
|---|---|
| | historic data be amended? |
| Is the data disaggregated in ways that are useful to you? Are there specific time stamps or time periods required? | Can the data be disaggregated in different ways? (Financial year vs calendar year, or weekly vs monthly, per household vs per person) |
| What data challenges are you anticipating in future / What do you expect of those publishing data? | What data challenges are you anticipating in future / What do you expect of data users over time? |
| How do you prefer to engage? (Online forum, monthly emails, quarterly surveys) | How do you prefer to engage? (email feedback, invite to team meetings) |
| How could the data offer be improved? | How could my use of the data be improved? |

*GRAPHIC DESIGN: Simple table required but more visually stimulating…perhaps to show a two-way conversation…*

Data users may need help to consider the following to understand their data better.



*GRAPHIC DESIGN: already designs by ODI*

The following sections aim to help you create and/or improve the value loop between data publishers and data users.

# The User Journey

Our research with data users highlighted that many felt open data was still sometimes challenging to access and make use of. Although they may have been freely available online, some data users felt that it was difficult to actually extract the specific data they required, due to issues with the templates and formats, due to time periods and geographics regions that didn't quite fit their requirements and other misalignments.

## Problem examples

- Data is not explained
    - No description of metric
    - Unclear unit of measurement or time frame
    - Unclear distinction between whether figures given represent actual counts, or percentages or proportions
    - Minimal supporting text or context for data
    - Charts with unclear labels, colours or keys
- Data is not accessible
    - Raw data in large spreadsheets not easily downloaded or able to view on screen
    - Static PDF documents that can't be edited or text can't be selected
- There is no feedback loop
    - Users cannot review or comment at point of access/no identifiable contact information
    - No outlet for group feedback or peer support

Common errors in data also include:

- Dates. Mixed date formats is British and American dates
- Multiple representations. Abbreviations and expanded forms ie Vice-President and VP
- Duplicate record detection. Data may be duplicated to speed up searches across multiple domains
- Summation records. Data containing the sums (or other formula) of the data
- Redundant data. This isn't required so errors are often common with its input
- Numeric ranges. Often used to anonymise data, but can make searching difficult
- Spelling errors. Can lead to awkwardness when querying and visualising data

## Improving the user journey

You can't have a user-centric approach without knowing who your users are. You need to understand your audiences to establish how the data might help them. You may want to target a specific audience because you know they drive demand, or you may want to appeal to a wider range of people because the data has many different applications. Once you identify who your audiences are, you can start thinking about their data needs and goals.

- **Understanding who your current audience is**. By understanding who these people are, you can tailor content to their needs, provide specific types of service and support they're seeking, and ensure your product or service will resolve any challenges they're experiencing. This is how you foster strong, long-term relationships between your audience/customers/data users and your organisation as well as a sense of loyalty and advocacy over time.
- One effective method is to use analytics tools to track how people are interacting with your website or social media accounts. This can give you a sense of who your audience is (sector, profession, role, location), what they're interested in (standard metrics, benchmarking, bespoke analysis, insights, examples, guidance), and what kind of content they respond to (downloadable, interactive, technical, infographics).
    - ***TOOL: Audience Definition & Segmentation***

***Adapted segmentation categories for data users: psychographic, behavioural and geographic***

| Geographic | Behavioral | Psychographic |
|---|---|---|
| • Urban, rural, peri-urban<br>• Region, district, community | • Relevant behavior<br>• Stage of change/readiness to change<br>• Frequency of behavior<br>• Consistency of behavior<br>• Duration of behavior | • Benefits sought<br>• Values<br>• Activities<br>• Interests<br>• Attitudes, opinions<br>• Personality<br>• Preferences |

*GRAPHIC DESIGN: would prefer the above table in a pie chart or something more interesting than a table*

Audience segmentation typically enables people to match audiences, messages, media, products and services based on the specific needs and preferences of the audience. In the case of data users, we might want to segment audiences by skill set or seniority, rather than socio-demographic factors.

- Possible segments may include:

| Example audiences: | | HIGH | MED | LOW |
|---|---|---|---|---|
| | | **Data Skill** | | |
| HIGH | **Seniority/ Influence** | Senior Researcher, Data Manager | Head of Programme | Policy Lead, Journalist |
| MED | | Academic | Business Development | Project Managers |
| LOW | | Data Officer, Analyst | Assistant researchers | General member of the public |

*GRAPHIC DESIGN: simple table.*

- Different approaches to different segments:
    - **Data skill**: For audiences with substantial data skills in terms of analytical proficiency, you may want to publish data with more detailed descriptions using applicable technical language, as well as options to take the raw data and analyse it in bespoke ways. They may want to know about any outliers in the data or any limitations and quality concerns. Experienced analysts will have a good understanding of how to utilise the data and may want to manipulate and present it in novel ways. For those with limited data skill, it may be more user-friendly to publish data stories with clear concise descriptions of the data (what the data says and its potential implications). See section below.
    - **Seniority or Influence**: More senior audiences are likely to want a good understanding of the data and its potential implications in the most summarised and concise way possible, in order to use it to drive decision making.
    - **Audience Knowledge of topic**: Audiences with intimate knowledge of specific topics may want more deep dive information alongside the data or are likely to require more detailed descriptions and analysis
    - **Audience Attitude:** Attitude and motivation may also shape how your audiences perceive and use the data you publish. This includes things like being motivated by seeking funding, improving performance, responding to statutory requirements, and also being safety conscious (from cyber safety to protection of organisational reputation).

    - **Consider user needs at different points in the user journey**

      ***TOOL: Mapping User Journey.***

***Mapping User Journey***

***Mapping user motivations and touchpoints at different stages of engagement***

| | Initial contact | Delivery | Finalisation |
|---|---|---|---|
| User aim, intention or motivation? | *EXAMPLE: User needs data about the residents of a specific area* | | |
| User action / initial touchpoint | *User performs google search using the area name; looks for open data that includes geographic info* | | |
| What needs to change/ improve? Why? | *Google search doesn't pick up a specific database but does show organisation name.* | | |
| How will change be made? | *Allow dataset titles to be searchable via Google search using e-labels etc* | | |

- This tool could be used in a data user feedback workshop, for example.

*GRAPHIC DESIGN: Ideally, this would be an interactive sheet that people can populate themselves online, on the Gitbook platform. Or it can remain a simple table that can be printed and downloaded*

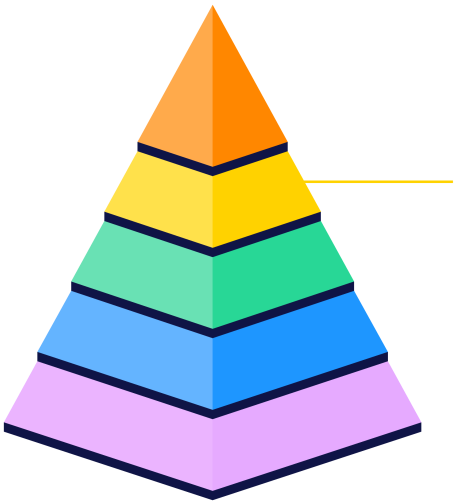
- **Engagement at specific points in the user journey**

1. Co-production of data needs/analytical innovation (at beginning of project or new data publication)
2. Building in Feedback throughout (publication points/user deadlines)
3. Good follow-up/learning points at the end of a project/data cycle/defined time period
4. Data showcasing, share value/realising value beyond the partnership (engaging wider audiences)

- From data to information

While some users may want raw data (actual figures that can be used to create bespoke graphs and charts for example), others  - particularly those non-analysts, those with limited analytical skills or those with limited time for analysis - may want information, insights or stories, alongside (or instead of) actual raw data.

- Data = a number, statistic or measure (quantitative); a fact, an opinion or experience (qualitative)
- Metric = data with specific parameters, type of measurement
- Information = metrics with context
- Insight = information with implications
- Story = connected insights



*GRAPHIC DESIGN: the five points on the list need to be illustrated as a process that becomes more refined at each level*

Turning raw data into something we can use to make a decision or form an opinion means deciding what to **show**, what to **emphasise**, and what to **filter out**.

For more info:
A four step approach to storytelling with data

## Building communities around open data use

## Targeting intended audiences

### Methodologies

- **Focus groups** with data users to discuss responses to your data publication or service. This gives you a starting point to work from when learning about your audience on a deeper level — you already have some understanding of the real pain points and challenges they experience and what they need from your product or service. Risks: anecdotal; small group may not be representative of typical audience.
- **Surveys** can generate a substantial amount of quantitative data about what your audience have experienced and how they might  rate or rate specific aspects of a data publication: ease of access, applicability, timeliness etc. Surveys are also great because you can elect to make them anonymous, which often incentivizes participants to be completely honest — meaning, you get the real data and results that matter. Risks: limitations of depth of questions, open text responses are time consuming to analyse.
- Email mail outs/website comments box. Risks: assumes healthy response rate.

### Stimulating demand for Open Data

- WPD input
- Innovative analysis
- Predicting trends and changing needs over time

# Planning for impactful open data initiatives

When planning to improve access to data, organisations must understand the data landscape in which they operate. The term ‘data landscape’ can refer to a specific market, sector, or community, as defined by the goals of your publishing initiative. For example, your initiative might be focused on a sector, such as ‘connected and autonomous vehicles’; a cross-sector challenge, such as ‘childhood obesity’; or a place, such as a specific city. A clear understanding of your landscape can help you effectively prioritise and plan activities for delivering impact against social, economic and environmental challenges.

To help organisations conduct a review of the landscape in which they operate in and to plan effectively for delivering impactful data initiatives, we have developed the Data Landscape Playbook. The playbook is designed for organisations which are considering designing and delivering projects or programmes of work which focus on collecting, using, sharing or publishing data with the aim of addressing a common social, environmental or economic challenge. We refer to these organisations or collectives as ‘data access initiatives’ throughout.

The playbook consists of five sections:

1. 'Play one: Explore the problem and how data can address it' provides guidance on how to identify how the problem your initiative is seeking to solve can be addressed by designing or strengthening data infrastructure in order to improve access to data.
2. 'Play two: Map the data ecosystem' takes you through steps to identify and list stakeholders and map them against the key data and value exchanges within the initiative’s data ecosystem. This brings to light the gaps, barriers and opportunities in the data ecosystem to which your activities can align.
3. 'Play three: Assess the policy, regulatory and ethical context' offers guidance and tools to help you understand the ethical context of your project, and assess any legal, regulatory and policy considerations that might impact the data access initiative.
4. 'Play four: Assess the existing data infrastructure' takes you through steps to assess the relevant data infrastructure, including understanding what data assets you will need, how to access them and how to look for available open standards. It will also encourage you to reflect on the data skills relevant to the initiative.
5. 'Play five: Plan for impact when designing your data initiative'

In this section, we explore how using the Data Landscape Playbook can help you to plan for more impactful open data publishing.

## Play one: Explore the problem and how data can address it

Successful data access initiatives are underpinned by a clearly defined vision that outlines the impact you want to achieve in a specific community, market or region. By improving access to data and involving stakeholders across the initiative’s activities, organisations leading these initiatives are better prepared to

address complex challenges and produce desired outcomes. Exploring the role of data within the problem can help you achieve this impact.

Having a clearly defined problem to be addressed by improving access to data and tackling a specific social, environmental or economic challenge will help focus your activities, communicate your vision to funders, and mobilise the resources and people needed to build or strengthen data infrastructure.

Activities in this play include:

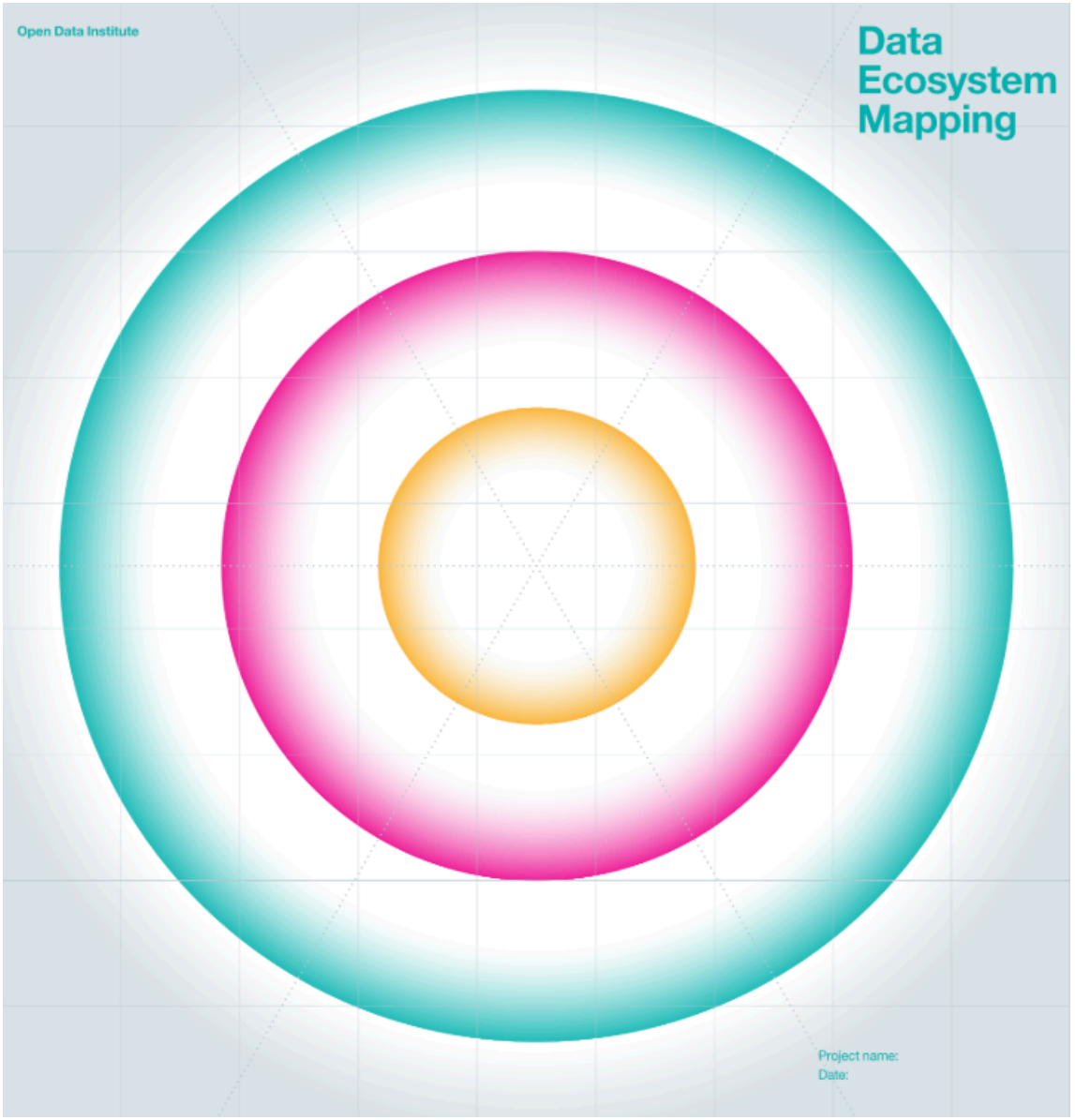
- Exploring how your identified problem can be addressed through data
- Deciding what type of data infrastructure the initiative aims to support
- Carrying out initial research and engagement

By the end of this section you should be able to:

- Define how improving access to data can help address your problem
- Identify what research needs to be undertaken in order to carry out other plays, such as stakeholder mapping and logic modelling
- Decide the 'use case' and how it might be addressed through strengthening data infrastructure

## Play two: Map the data ecosystem

Data access initiatives commonly require engagement with multiple stakeholders as they address complex challenges through collecting, sharing/publishing or using data. This need to collaborate, involve and sometimes influence stakeholders to improve access to data can be made easier by understanding your initiative's data ecosystem. A data ecosystem map can support you devising your initiative's roadmap to building data infrastructure by showing how the data and value flows, identifying key stakeholders, and highlighting gaps.



[Placeholder Data Ecosystem Mapping image]

Activities in this step include:

- Engaging with key stakeholders
- Mapping data flows and value exchanges in the ecosystem
- Identifying gaps, barriers and opportunities in the data ecosystem that your initiative can address

By the end of this section you should be able to:

- Understand the key stakeholders to engage
- Have a clear understanding of the data ecosystem
- Define an action plan to tackle your problem statement

# Play three: Assess the policy, regulatory and ethical context

This section will help you identify and understand important aspects of the wider context that impact the access, use and sharing of data by your initiative. The key question you want to answer here is whether the policies and rules around data are conducive to building the data infrastructure you need to tackle the problem you have identified.


[Placeholder Data Ethics Canvas image]

Activities in this section include:

- Understanding the legal, regulatory and policy context of the initiative
- Understanding the ethical issues that might impact the initiative

By the end of this section you should be able to:

- Inform your theory of change based on the policy and ethics landscape
- Plan any additional actions to comply with regulations
- Address or mitigate any negative ethical consequence that the initiative might not have considered

# Play four: Assess the existing data infrastructure

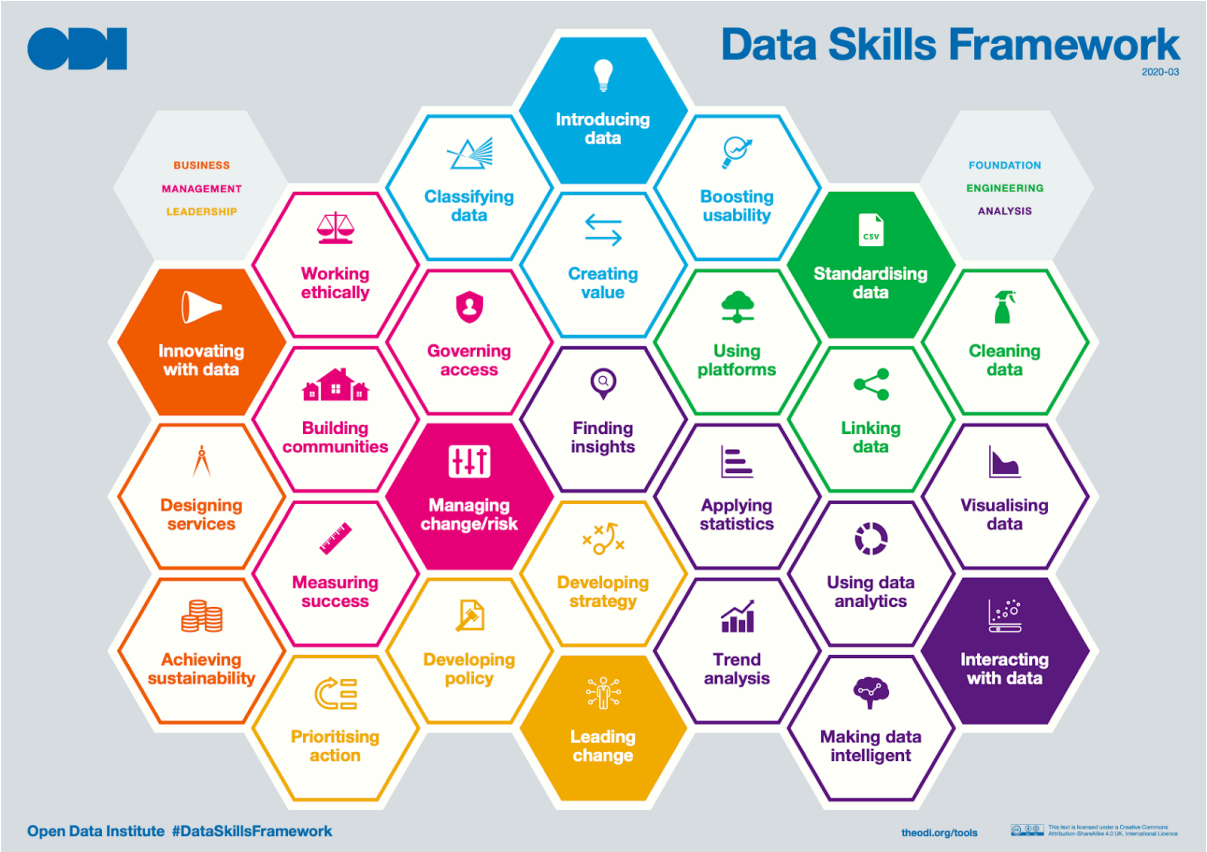
As part of their efforts in collecting, using and sharing data to tackle specific problems, data access initiatives need to build or maintain strong data infrastructure. For example, by creating new data assets, developing or reusing open standards, creating guidance and policies to make better decisions to address sector-specific challenges, or building new technologies to support the management and use of the data.

In 'Play one: Explore the problem and how data can address it', we went into depth about the type of data infrastructure your initiative might be building, improving and maintaining, including:

- Data assets (such as datasets, identifiers and registers)
- Standards and technologies used to curate and improve access to data assets
- Guidance and policies that inform the use and management of data assets and the data infrastructure
- Organisations that are responsible for stewarding data
- Communities involved in contributing or maintaining data infrastructure, and those who are impacted by decisions that are made using it to tackle a specific problem

In 'Play two: Map the data ecosystem', we provided guidance on how to identify organisations stewarding data and communities involved in, and impacted by, the building of data infrastructure to tackle the problem. In 'Play three: Assess the policy, regulatory and ethical context', we recommended steps to assess policies governing the use and management of data. In this play, we provide guidance on how to assess data assets, standards and skills to improve access to data. If you are interested in what technologies could help your initiative, you should check out our research looking into the common technical infrastructure of shared and open data.

Assessing existing data infrastructure and deciding whether your initiative needs to provide or strengthen tools that might tackle the problem includes thinking about how this data will be managed, collected and shared.




[Placeholder Data Skills Framework image]

Activities in this play include:

- Making a data inventory
- Adopting or building open standards
- Assessing skills needed to strengthen or build data infrastructure

By the end of this section you should be able to:

- Assess skills and literacies of the initiative
- Evaluate the available data assets and standards to tackle the problem

# Play five: Plan for impact when designing your data initiative

Defining the impact of the initiative – and the inputs, activities and outputs it needs to deliver to achieve that impact – will help you in designing a roadmap to address the problem you are trying to solve, and illustrate how and why your initiative is expected to bring about the desired change in its given context.

**Situation:** OpenActive (OA) wants to adopt a new standard for data about sport and exercise classes in the UK to increase public activity

| Inputs | Outputs | Outcomes | Impact |
|---|---|---|---|
| • Funding from Sport England<br>• Existing research and evidence<br>• Connections between partners and stakeholders in the sport industry and the expertise at the ODI | **Activities:**<br>• User and desk **research**<br>• Survey the data landscape (inc **ecosystem mapping & gap analysis**)<br>• Develop **use cases**<br>• Engage and align stakeholders (internal and external) including **training**<br>• Technical development of the **four OA standards, open source software and tooling** to support data publishing<br>• Develop a **roadmap** for the future including theory of change, and sustainability plans<br><br>**Participation:**<br>• Data stewards (gyms, sports teams etc)<br>• Data users (app developers etc)<br>• End users (public)<br>• Policymakers and regulators<br>• Decision makers (Sport England)<br>• People impacted by use of data<br>• Funders and sponsors (Sport England) | **Short term:**<br>• Data publishers see the value in adopting a standard<br>• The coalition of partners are aligned around, and committed to, a common vision, roadmap and KPIs for the future<br>• First set of standards are released with accompanying user guides (activity standards)<br><br>**Medium term:**<br>• More, and better quality, data is shared by activity providers<br>• More people use the service to find sports and activities<br>• Second set of standards is released with accompanying guides (booking standards)<br>• Organisations outside the ecosystem show interest in developing new products<br><br>**Long term:**<br>• People are making bookings thanks to the booking standard<br>• Strong alignment and commitment across the actors in the sports ecosystem<br>• A sustainable business model is implemented ensuring future stability of the standard<br>• New apps and services are developed utilising data under the OA standard<br>• OpenActive is standard practice for the sports and physical activity sector beyond March 2021 | • The UK public is more active, more people are playing sports and doing activities |

[Placeholder logic model example image]

Activities in this step include:

- Planning an impactful initiative
- Identifying the assumptions, external factors and consequences
- Sketching your evaluation framework

By the end of this section you might be able to:

- Define the main activities and inputs needed to build or strengthen data infrastructure
- Design the outputs that will contribute to an impactful initiative

# Publishing guidance for new data publishers

Having readied yourself to publish data at an organisational level, considered what your users need and thought about the impact you want to create with data, it is now time to consider the practical aspects of publishing data for others to access, use and share.

In this section, we'll explore how to make the data you aim to publish fit-for-purpose, what tools can help you to publish data in a cost-effective way and what the minimum considerations are for a new data publisher .

## Open data licensing

Our definition of 'open data' is data that anyone can access, use and share, but importantly, open data must be published under an appropriate open licence for it to be truly open.

An open licence is one that places very few restrictions on what anyone can do with the content or data that is being licensed. You can choose to make your content or data available under one of three levels of licence:

- A public domain licence which has no restrictions (technically, you waive your rights to the content or data).
- An attribution licence that says that reusers must give attribution to you.
- An attribution and share-alike licence which says that reusers must give attribution and share any derived content or data under the same licence.

A good licence must be clear on three aspects:

- What the user can do.
- What the user must do.
- What the user cannot do.

There are a number of templates available with standard terms included, for example:

- Creative Commons – templates and standard wording for open and non open licences for creative content and data.
- Open Data Commons – templates and standard wording for open database licences
- Also see further examples of standard licenses

| For further guidance on data licensing, why not check out our 'Data licences' resource list? |
|---|

## The FAIR principles of data access

An open data licence is not enough to guarantee that open data will be useful. In order for us to meet the expectations our users have of the data we publish, the data must be fit for purpose. The FAIR data principles are a set of guiding principles that help organisations to make sure that the data that they

publish openly meets the expected standards of quality that others might have. Under the FAIR principles, data must be findable, accessible, interoperable and reusable.

There are several sources that describe what each of the FAIR principles means, but one of the more accessible examples that we've found comes from the UK's Geospatial Commission, in their assessment of the UK's national geospatial data assets (figure 4):

- **Findable:** Findability is simply the capability of something being found and in recent years this means data discovery through a web search engine. Data should be easily and repeatedly discovered by experts and non-experts. Having found data, it is easy to determine whether the data is appropriate for their intended use. To support this any data should always have good discovery metadata and be published somewhere that it can be easily searched.
- **Accessible:** Accessibility refers to reducing the hurdles from data discovery to obtaining the data. To be accessible, the terms for using the data should be clear and simple with registration and authentication to data services minimised. Data should be designed with accessibility in mind, wherever possible turned into products that are tailored to common usage and delivery methods using open and widely used standards and formats
- **Interoperable:** Interoperability is the ability of different geospatial systems to accurately and unambiguously exchange data. Data needs to be designed to support interoperability, which means some consideration must be given to potential systems that may want to consume the data that is published. To enable interoperability, geospatial data should be available in multiple standardised open formats. These formats should contain information to support transformation of the data between common datums, projections and systems. Data should also contain, where appropriate, persistent, globally unique identifiers (for example Unique Property Reference Number (UPRN)) at the record or feature level.
- **Reusable:** Reusable is the ability for value to be derived from using the data. In practical terms this means supplying information that supports the Findable, Accessible and Interoperability aspects of the data. Usability information is typically documented as part of the data set metadata. Metadata should be written such that both a specialist and non-specialist user can understand whether the data is suitable for their purpose. This includes clear advice on the constraints there may be on its suitability for common applications outside of its primary intended use and what mitigations may be appropriate to apply.

## FAIR assessment tools

There are a number of freely available tools that you can use to help you consider whether the data you are looking to publish meets the expectations of the FAIR principles:

- F-UJI is a web service to programmatically assess FAIRness of research data objects (aka data sets) based on metrics developed by the FAIRsFAIR project.
- SATIFYD is a tool created by Data Archiving and Networked Services (DANS) which will show you how FAIR (Findable, Accessible, Interoperable, Reusable) your dataset is and will provide you with tips to score (even) higher on FAIRness.
- The FAIR Data Self Assessment Tool has been developed by the Australian Research Data Commons to help you assess how FAIR your research dataset and give you practical tips on how to improve the dataset's FAIRness
- The FAIR data assessment tool (FAIRdat) is a prototype tool designed by the UK Data Service to help you score the 'FAIRness' of a particular dataset.
- For guidance specifically on accessibility, check out our 'Accessibility' resources list

# Data quality

Even when published under an open licence and adhering to FAIR data principles, data must still be quality in order to be usable. The quality of data is determined by a set of practical, legal, social and technical requirements, defined by community driven standards.

In addition to the open licence, there are three legal requirements that need to be considered by open data publishers. You must:

- Protect sensitive information like personal data
- Preserve the rights of data owners
- Promote the correct use of data

In order to meet the practical requirements of quality open data, publishers must:

- Link to the data from the relevant website
- Update the data regularly if it changes
- Commit to continue to make the data available

There are three recommendations that define the technical aspects of open data:

- The format in which the data is published
- The structure of the data
- The channels through which the data is available

For data use to be sustainable, it is important to have an engaged community of users. The best datasets have:

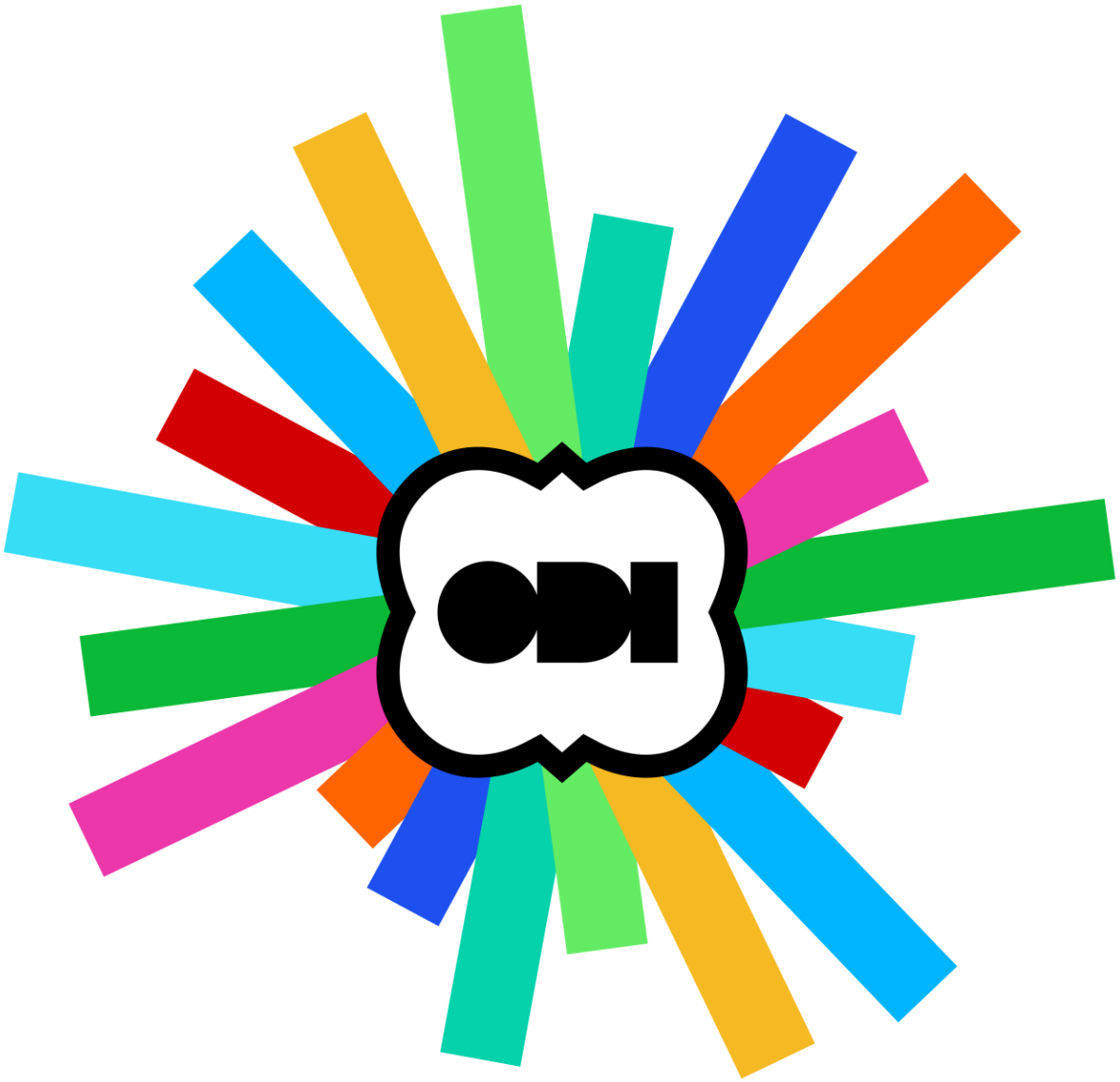
- Active support channels
- Discussion groups and forums
- Published how-to guides on working with the data

## Tools and frameworks to help you assess open data quality

Sir Tim Berners-Lee, the inventor of the World Wide Web and one of the ODI's founders, suggested that there is a 5 step deployment scheme for open data, referred to as the 5-stars of open linked data. This framework helps data publishers to consider the legal and technical marks of quality in open data.

The 5-stars scoring system is as follows:

⭐ - Be available on the Web, in any format, under an open licence
⭐⭐ - Be in the form of structured data (e.g. Excel instead of an image of a table)
⭐⭐⭐ - Be in a non-proprietary file format (e.g. CSV file instead of Excel)
⭐⭐⭐⭐ - Use URIs to identify internet resources
⭐⭐⭐⭐⭐ - Link your data to other data to provide context



[Placeholder Open Data Certificates image]

To support publishers with assessing all four markers of data quality, we provide a tool called Open Data Certificates. This free online tool helps you assess and recognise the sustainable publication of quality open data against established best practice guidance.

## Publishing data on the web

Publishing open data can be as simple as uploading a file to a commonly used publication platform, such as Microsoft OneDrive and then sharing the link with people. If you want to make open data available online in a minimum viable way and this meets the needs of your users, we recommend using Github – A popular code and data hosting site where you remain in control. If you are new to Github, register for an account at github.com and then head to octopub.io which is a tool created by the ODI that makes it easy to publish data on GitHub. Login to octopub using your Github account and follow the step-by-step guide to publishing a dataset.

You may instead want to publish the data that you have on an existing open data platform. Many organisations, including most governments, have their own data platforms. There are also a number of community platforms such as data.world, AWS public datasets and OpenStreetMap that allow for user contributions. These platforms often have specific restrictions on the type or format of data you need to provide, but offer additional functionality for reusers.

| For further guidance on where to publish data, check out our 'data publishing platforms' resource list |
|---|

If you are new to publishing data and don't have an existing publication method, check out our guide to 'Publishing open data in times of crisis'. Originally published with the aim to help others publish data that could be useful in addressing the Covid-19 pandemic in a quick and efficient manner, this checklist is also useful to data publishers that want to publish data openly, but don't have a lot of capacity to do so.

Hello Anthony & Gretchen

Thanks for agreeing to some time to review this resource we have been working on. I wanted to (virtually) introduce myself and give some background for Friday.

Publishing Data for Social Good: A toolkit
Guidance to help you publish data openly, in a way that works for your users.

The resource is still being finalised and has not yet had input from our graphic designers or copyeditors, so please excuse any unfinished elements. The resource is predominantly for smaller projects and organisations, who likely do not have much experience of publishing data, and limited resource for engaging data users. Language and scope of content have been intentionally pitched at this level, so please consider this as we review it.

https://open-data-institute.gitbook.io/wpa-data-for-social-good-v1-unpublished/

# Backlog - Not for copy edit

## Engagement with data communities

- Aims: attract new audience, raise awareness of your data service/platform/publications
- Methods: comments and engagements on your blog - and other people's, promoting on customer support channels, raising awareness on community web pages and forums, and social media profiles.
    - **Data Ecosystem Mapping**. A tool for documenting and mapping data ecosystems, to identify barriers to and opportunities for sharing data. https://theodi.org/article/mapping-data-ecosystems/
    - Data.Org **Resource Library**. Guides to help you understand why your organisation should leverage data for social impact and how to get started. https://data.org/library/
    - Carbon Brief's **Global South Climate Database** is a publicly available, searchable database of scientists and experts in the fields of climate science, policy and energy. https://www.carbonbrief.org/global-south-climate-database/

## Planning for impactful open data initiatives

- Impact / goals
    - **UKRI's Impact toolkit for economic and social sciences.** Aimed at social science researchers applying for funding, it supports in defining and communicating impact. Straightforward online advice and guidance. https://www.ukri.org/councils/esrc/impact-toolkit-for-economic-and-social-sciences/
        - ***Sector specific examples:***
        - Arts and Culture audience insight (possible diversity angle) https://www.artscouncil.org.uk/impact-and-insight-toolkit https://impactandinsight.co.uk/
        - Scotland Data Toolkit: Raising Attainment through Effective Use of Data https://education.gov.scot/improvement/Documents/excellenceandequitydatatoolkit.pdf
        - Data for Impact for physiotherapists (telehealth, recover time etc) https://www.ppssupport.com/hc/en-us/articles/360019506557-Using-the-Data-For-Impact-Toolkit-

        - For investors to manage impact assessment tools https://impacttoolkit.thegiin.org/
    - **Logic Model Templates for Data Initiatives.** Logic models are tools designed to help people plan impact projects and communicate those plans to others. They provide a structured way of thinking about how to build a programme of activities that will help to address a specific problem or challenge – essential for designing a good data access initiative.
- Skills
    - The **Open Government Data Toolkit** is designed to help governments, Bank staff and users understand the basic precepts of Open Data, then get "up to speed" in planning and implementing an open government data program, while avoiding common pitfalls. http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html
    - **Triage, open source machine learning toolkit** to help data scientists, machine learning developers, and analysts quickly prototype, build and evaluate end-to-end predictive risk modelling systems for public policy and social good. Requires coding on Github https://github.com/dssg/triage
    - The **Data Skills Training Program Implementation Toolkit** is designed to provide both small and large agencies with information to develop their own data skills training programs. US Government resource. https://resources.data.gov/resources/data_skills_training_program_implementation_toolkit/
    - **Data Skills Framework.** The Data Skills Framework shows how technical data skills can be balanced with other skills to support successful data innovation. https://www.theodi.org/article/data-skills-framework/

## Infrastructure (people, processes, tech, etc)

- **Local Government Benchmarking Support** by LG Inform. Enables area insights and performance benchmarking and comparison to improve services across UK local authorities. Includes full list of metrics themed by public service with access via automated analytical platform that produces automated geographical reports. Local government membership required for most areas. https://lginform.local.gov.uk/dataAndReports/explorer
- **Data Maturity Assessment**. Once assessment is complete, various tools are signposted to aim improvement, such as 'Cybersecurity for Non-Profits' https://datajourney.data.org/
    - Similar, Data Maturity Framework http://www.datasciencepublicpolicy.org/our-work/tools-guides/datamaturity/
- **GSMA Big Data for Social Good (BD4SG).** The mobile industry is harnessing big data to help public agencies and NGOs tackle health and environmental issues. Toolkit is extensive repository of documents and addresses

sustainability, realising scale and policy etc via links to wider resources. https://aiforimpacttoolkit.gsma.com/

- **The Microsoft Planetary Computer** puts global-scale environmental monitoring capabilities in the hands of scientists, developers, and policy makers, enabling data-driven decision making. There is a repository of applications being used, available here: https://planetarycomputer.microsoft.com/applications
-

## Ethics

- **ICO Data Sharing Info Hub and Code of Practice**. Data sharing information hub provides clear guidance and practical tools for organisations and businesses on how to share personal data lawfully, while protecting people's personal information. https://ico.org.uk/for-organisations/data-sharing-information-hub/
- **Ethics & Algorithms Toolkit**. Practical toolkit for cities to use to understand the implications of using an algorithm, clearly articulate the potential risks, and identify ways to mitigate them. (For those building or acquiring algorithms in the government sector and beyond) https://ethicstoolkit.ai/
- **IBM's Trustworthy AI**. A collection of tools for assessing bias in datasets, algorithms and machine learning, or assessing privacy risks of machine learning. Requires knowledge of coding and Github. https://research.ibm.com/topics/trustworthy-ai
- **Gender Diversity Dashboard**. The tool from *Women in Data Science and AI* research project at The Alan Turing Institute offers insights into gender inclusion within online tech workplaces. The Dashboard was designed for use by technology companies, running internally on top of GitHub and Slack, to measure and highlight in real-time potential differences between the treatment of men and women. https://www.turing.ac.uk/research/research-projects/women-data-science-and-ai
- Aequitas, an **open source bias audit toolkit** developed by the Center for Data Science and Public Policy at University of Chicago. Assessing unintended bias of predictive tools http://www.datasciencepublicpolicy.org/our-work/tools-guides/aequitas/

- Sustainability
  - The **ODI's Sustainable Data Access Workbook** is a practical set of activities for organisations to undertake to make more informed decisions about their current and future revenue models. The workbook is particularly useful for data institutions – organisations that steward data on behalf of others.
- Internal advocacy / leadership
  - **Principles of open data maturity**: ODM measurement and assessment covering: policy, impact, portal, and quality. Open Data Maturity, Data Europe 2022 Report

https://data.europa.eu/sites/default/files/landscaping_insight_report_n8_2022.pdf

- **Nesta Data Sharing Toolkit**. This toolkit provides useful guidance and resources for private and public organisations to prepare for and design data-sharing initiatives. https://www.nesta.org.uk/toolkit/data-sharing-toolkit/
  - Digital Dubai example https://www.digitaldubai.ae/docs/default-source/publications/data-sharing-toolkit.pdf?sfvrsn=91fb2fa1_12
- **Data Pitch Data Sharing Toolkit**. This toolkit has been developed to help organisations that want to generate value by sharing data or facilitating data sharing. Explains the concept, challenges, and processes to enable successful data sharing, and provide resources and recommendations. https://datapitch.eu/datasharingtoolkit/
- **CABI & ODI Data Sharing Toolkit.** Data Sharing Toolkit which could contribute to unlocking greater food security in Sub-Saharan Africa and South Asia through better access to information on soil health, agronomy and fertilisers. Includes seven eLearning modules with supporting case studies, checklists, cheat sheets and guides to help demystify how to use, collect, share FAIR and safeguarded data for the benefit of a country's agricultural economy. https://www.datasharingtoolkit.org/

## Publishing 101

What does the criteria for good open data publishing look like?

- 5-star open data
  - https://5stardata.info/en/
- Q-FAIR/O
  - https://www.go-fair.org/fair-principles/
    - Quality
    - Findable
    - Accessible
    - Interoperable
    - Reusable
    - Openly licenced
  - Examples of Q-FAIR applied to specific areas
    - Geospatial Commission UK advises the government on the most productive and economically valuable uses of geospatial data. https://geospatialcommission.blog.gov.uk/2021/06/25/byte-ing-back-better-introducing-a-q-fair-approach-to-geospatial-data-improvement/
    - Open Transport. Transport & Mobility Data https://opentransport.co.uk/2021/04/22/make-transport-mobility-data-fair/

        - https://fairtoolkit.pistoiaalliance.org/
        - https://fairsharing.org/
- **How to publish open data: a list of tools and advice.** This is a curated list of links which we think are most helpful in publishing open data. It includes useful links and advice about platforms, licences, metadata and standards. https://github.com/theodi/data-publish-list#about-open-data
- **Data Pitch Data Sharing Toolkit**. Developed to help organisations that want to generate value by sharing data or facilitating data sharing. Derived from experience with national and international initiatives, such as the Smart Cities Innovation Framework Implementation (SciFi), the European Data Incubator (EDI), as well as several recent pilots for data trusts in the UK.https://datapitch.eu/datasharingtoolkit/

- Security + privacy
    - **Anonymisation and open data: an introduction to managing the risk of re-identification.** 'Anonymisation and open data: an introduction to managing the risk of re-identification' discusses the risks and opportunities around open data, personal data and anonymisation. https://www.theodi.org/article/anonymisation-report/