

A guide to the

Open Data Maturity Model

Assessing your open data
publishing and use

Table of Contents

How to view and download the Open Data Maturity Model	3
Executive summary	3
Foreword	3
Introduction	5
<ul style="list-style-type: none">• Background and history• Developing the model• Acknowledgements	
Overview of the Open Data Maturity Model	8
<ul style="list-style-type: none">• How is the model structured?• How does the model relate to existing work?• What is the scope of the model?• How can the model be applied?	
The five maturity levels	10
The five organisation themes and related activities	11
The two aspects of open data practice - publication and reuse	12
How to perform an assessment	12
<ul style="list-style-type: none">• Setting an assessment process• Prioritising and aligning your activities	
Theme 1. Data management processes	15
<ul style="list-style-type: none">• Building a process to support data release• Developing standards and adoption• Developing data governance• Managing sensitive data	
Theme 2. Knowledge and skills	20
<ul style="list-style-type: none">• Developing open data expertise• Knowledge management	
Theme 3. Customer support and engagement	23
<ul style="list-style-type: none">• Developing an engagement process with data reusers• Documenting your open data• Building a reuser support process• Creating open data community norms	
Theme 4. Investment and financial performance	27
<ul style="list-style-type: none">• Ensuring financial oversight• Developing dataset valuation processes• Building open data into procurement practices	
Theme 5. Strategic oversight	30
<ul style="list-style-type: none">• Shaping open data strategy• Asset catalogue management	

How to view and download the Open Data Maturity Model

The Open Data Maturity Model is available to explore online and download as a Google Doc, Excel file and PDF at: <http://theodi.org/guides/open-data-maturity-model>

Executive summary

The Open Data Maturity Model is being developed by the Open Data Institute and the Department for Environment, Food & Rural Affairs to **help organisations assess how effectively they publish and consume open data**.

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.

The model is based around **five themes**, each representing a broad area of activity: **data management processes, knowledge and skills, customer support and engagement, investment and financial performance and strategic oversight**. An assessment grid helps organisations identify their levels of maturity for each of the activities.

Organisations can use the model to set themselves appropriate goals based on their current maturity, resourcing and anticipated benefits. To achieve the full, long-term benefits of open data, organisations must take steps beyond basic data publication, the assessment of open data publishing and consumption is a strong starting point.

While the model has been initially developed for a UK public sector audience, it can also be applied to any type of organisation with little or no modification, whether they are already publishing or consuming open data, or are planning to do so.

Foreword

The first challenges that many organisations face when becoming open data publishers are technical, largely centred around how to publish data in easy-to-use formats with clear licensing.

If reusers are to have reliable, sustainable access to published data then publishers will need

to consider these challenges as well as the strategic, financial and operational impacts of making their data open.

Understanding this organisational change is an important aspect of becoming an effective open data publisher. Similar changes take place as organisations begin to reap the benefits of reusing open data to reduce costs, increase efficiency and to drive innovation.

Organisations are at different stages in this open data journey. Some are only just beginning to publish data, while others have already undergone significant changes towards a more “open by default” model. Many are seeking guidance about how to take the next steps in their journey.

The goal of this Open Data Maturity Model is to highlight issues commonly encountered by publishers, and provide a means for organisations to assess and improve their effectiveness as publishers and consumers of open data. The maturity model provides a framework for understanding the different areas of organisational change towards open data, while also identifying the benefits of that change.

Introduction

The Open Data Maturity Model is being developed to support organisations in assessing their effectiveness in publishing and consuming open data.

While the model has been initially developed for a UK public sector audience, it can also be applied to any type of organisation with little or no modification, whether they are already publishing or consuming open data, or are planning to do so.

The model:

- supports assessment of the effectiveness of an organisation in its operational and strategic activities around open data
- provides guidance to organisations on potential areas for improvement
- compares organisations to highlight their respective strengths and weaknesses, support wider adoption of best practices and help improve processes

With these purposes in mind, the model can be used by both data managers working within individual business units and senior management with oversight on data management and governance practices.

The model should not add unnecessary burden to organisations that are publishing or reusing open data, rather, they can use it to set themselves appropriate goals based on their current maturity, resourcing and anticipated benefits.

To achieve the full, long-term benefits of open data, organisations must take steps beyond basic data publication, the assessment of open data publishing and consumption is a strong starting point.

Within the UK public sector the Open Data Maturity Model should be particularly relevant to organisations responsible for managing aspects of the UK National Information Infrastructure¹. These organisations should have a clear view of their open data maturity and have appropriate targets for development.

¹ National Information Infrastructure, <https://www.gov.uk/government/publications/national-information-infrastructure> accessed on 2015-03-19

Background and history

The Open Data Maturity Model was developed as a joint project between the Open Data Institute² (ODI) and the Department for Environment, Food & Rural Affairs (Defra)³.

The Defra Network transparency panel identified a need to measure the effectiveness and transparency of the Defra network organisations as open data publishers. The panel felt that better understanding of relative maturity would help to drive the organisational change required to promote open data publishing.

Defra and the ODI were awarded funding to develop a maturity model and a prototype assessment tool under the Release of Data Fund⁴ administered by the Cabinet Office and the Open Data User Group.

Developing the model

We began to develop the Open Data Maturity Model by drawing on the results of a series of requirements workshops attended by data managers and open data experts from the UK government and wider open data community.

The attendees discussed a number of key themes thought to be components of data governance and management practices. Each theme was reviewed to identify key challenges, outputs (such as policy documents), and evidence for progression.

The requirements workshops were supplemented with additional research on data governance and management practices. Some comparative research on the design and development of maturity models was also drawn on during the analysis.

We published draft versions of this document and the assessment grid for public comment. We reviewed feedback and incorporated it into the final documents.

We intend to revise and extend the model in future based on experience with applying it in real-world use. An online tool will also be provided to support organisations in assessing their maturity.

2 Open Data Institute, <http://theodi.org>, accessed on 2015-02-04

3 Department for Environment, Food & Rural Affairs, <https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs>, accessed on 2015-02-04

4 Breakthrough Fund and Release of Data, <https://www.gov.uk/government/publications/breakthrough-fund-and-release-of-data-fund>, accessed on 2015-02-4

Acknowledgements

Defra and the ODI would like to jointly thank all of the attendees of the requirements workshops for their contributions to the development of the maturity model. We would also like to thank those who provided feedback on the draft documents.

The funding and support for the project from the Cabinet Office and the Open Data User Group (ODUG)⁵ were essential in making the model a reality.

⁵ Open Data User Group, <https://www.gov.uk/government/groups/open-data-user-group>, accessed on 2015-03-20

Overview of the Open Data Maturity Model

A ‘maturity model’ generally provides a framework that allows an organisation to assess how well its processes conform to industry best practices. The model acts as an independent benchmark that allows organisations to score their maturity, usually in a number of related areas.

The Open Data Maturity Model has been designed to specifically focus on how open data practice impacts on an organisation.

A completed assessment against this model will give an organisation a **maturity score** for a number of important activities, namely **how data is released**, **how it is governed** and **how datasets are valued**. The score will reflect the maturity of the organisation’s processes in a specific area and can be used to identify areas of improvement and set measurable targets.

How is the model structured?

The Open Data Maturity Model consists of 15 organisational activities, e.g. data release processes or adoption of community norms, which are each explained in this paper.

The activities are grouped into five themes that categorise the activities. These have been derived from the categories used in a “balanced scorecard”⁶ to assess and monitor organisational performance. This reflects the goal to assess the variety of ways in which open data practice may impact on an organisation.

Activities can also be grouped according to whether they relate to the publication of data, reuse of data or both of these areas. These are referred to as aspects.

The overall structure of the model is reflected in the assessment grid that accompanies this paper (see Appendix). We recommend that you review the grid alongside this document.

An organisation will assess its maturity against each of the activities in the model, producing a maturity **score** from 1-5 for each activity. These five **maturity levels** use names drawn from similar five-point schemes in other maturity models: initial, repeatable, defined, managed, and optimising.

Each of the maturity levels has a definition which describes the key characteristics of an activity occurring at that level. For lower scores, activities are likely to be ad hoc, while at higher levels processes will be more refined.

6 Balanced Scorecard, http://en.wikipedia.org/wiki/Balanced_scorecard, accessed on 2015-01-30

To help motivate organisations to progress through the levels of maturity, each of the activities is associated with a description of the **beneficial effects** that follow from increased maturity in the activity.

The various elements of the model are covered in more detail later in the paper.

How does the model relate to existing work?

While there are other efforts^{7,8}, to assess maturity of data management and governance practices, these existing models do not adequately address the issue of open data.

The Open Data Maturity Model has been developed with reference to existing maturity models that measure other areas of organisational maturity⁹. Organisations that are already applying other maturity models should find that the Open Data Maturity Model aligns well with other approaches.

The model also complements the Open Data Certificates, which provide publishers with feedback on how effectively they are publishing individual datasets¹⁰. The Open Data Maturity Model assesses effectiveness at the organisational level. Mature organisations will be able to routinely publish datasets that are likely to gain a higher level of certification.

What is the scope of the model?

The Open Data Maturity Model is not intended to offer a prescriptive description of exactly how organisations should publish or reuse open data. Rather, the model focuses on the general behaviours that an organisation should exhibit and the processes it should adopt.

For example, the model highlights data governance as an important organisational activity. It suggests that a robust, well-defined and widely deployed data governance process is a characteristic of a mature open data organisation. The model also notes that a well-designed data governance process will involve clear ownership over data and cover elements such as data quality management. However, it does not recommend a specific data governance process that organisations should adopt.

7 IBM Data Governance Council Maturity Model, https://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf, accessed on 2015-03-25

8 Stanford Data Governance Maturity Model, <http://web.stanford.edu/dept/pres-provost/irds/dg/files/StanfordDataGovernanceMaturityModel.pdf>, accessed on 2015-03-25

9 See, for example: Capability Maturity Model, http://en.wikipedia.org/wiki/Capability_Maturity_Model, accessed on 2015-01-30

10 Open Data Certificates, <https://certificates.theodi.org>, accessed on 2015-03-19

Individual organisations should implement these processes based on their needs. It would be overly prescriptive for this model to recommend processes that could be usefully applied in every organisation.

While we expect that best practices will eventually emerge for many of these areas of activity, that detail is outside of the scope of the maturity model.

How can the model be applied?

The structure of the model allows it to be applied in several ways:

- It can be used to produce a single maturity score providing a summary of an organisation's overall maturity
- It can produce scores for individual themes, allowing an organisation to focus on broad areas that may need specific improvement
- It can produce scores for individual activities, giving a more detailed assessment of an organisation
- It can be used to assess maturity as an open data publisher or as an open data consumer

Importantly, we intend the model to be actionable: it should be possible to use the model to both set and monitor goals for organisational development, and as a source of guidance on implementing improvements.

The five maturity levels

The Open Data Maturity Model is based on five levels that represent the different states through which an organisation will pass as it matures. Advancing to the next stage involves creating, developing and refining specific business activities and processes.

Setting aside the details of individual activities, the levels themselves can be characterised as follows:

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.

Broadly, the levels represent an increasing level of sophistication in the organisation:

- moving from ad hoc uncontrolled processes to those that are repeatable, standardised and well-managed
- moving from a reactive to a proactive approach within a particular area of activity
- moving from isolated expertise, e.g. individuals championing open data, through to wider organisational support

The five organisation themes and related activities

The model is based around five themes. Each of the themes represents a broad area of activity within the organisation:

- **Data management processes** — identifies the key business processes that underpin data management and publication including quality control, publication workflows, and adoption of technical standards.
- **Knowledge & skills** — highlights the steps required to create a culture of open data within an organisation by identifying the knowledge sharing, training and learning required to embed an understanding of the benefits of open data.
- **Customer support & engagement** — addresses the need for an organisation to engage with both their data sources and their data reusers to provide sufficient support and feedback to make open data successful.
- **Investment & financial performance** — covers the need for organisations to have insight into the value of their datasets and the appropriate budgetary and financial oversight required to support their publication. In terms of data consumption, organisations will need to understand the costs and value associated with their reuse of third-party datasets.
- **Strategic oversight** — highlights the need for an organisation to have a clear strategy around data sharing and reuse, and an identified leadership with responsibility and capacity to deliver that strategy.

Each of these themes have been broken down into several activities that describe the behaviours and processes that the organisation should carry out.

The assessment grid identifies how organisations at different levels of maturity will carry out each of the activities. The majority of this document provides more detail on the themes and activities.

The two aspects of open data practice - publication and reuse

We expect that a mature open data organisation will be both a consumer and publisher of open data, although the overall balance is likely to vary across organisations. Some organisations may primarily publish rather than reuse data, while others may primarily consume data and have little or no published data.

Some activities within the maturity model apply to both the publication and reuse of data, while others are clearly focused on one of those aspects of open data practice.

To help organisations focus on the elements of the model that are most applicable to them, the assessment grid identifies whether the individual activities are associated with one or both of the following aspects:

- **Data publication** — addresses the organisational activities and processes that support the creation and management of datasets that are made accessible under an open licence.
- **Data re-use** — is concerned with the processes that support the effective reuse of third-party datasets.

How to perform an assessment

The Open Data Maturity Model provides a framework for assessing organisational maturity across a range of activities. As part of carrying out a maturity assessment, an organisation should attempt to score itself against all of these activities. This will produce a complete assessment of open data maturity and support benchmarking with other organisations. However we do not expect that all organisations will obtain the maximum score in all activities.

Some elements of the model will be more applicable to certain organisations. For example, an organisation handling sensitive personal data will likely focus on gaining greater maturity

in desensitising data. In contrast, an organisation that manages only non-personal reference data will not require a high maturity in this area.

Some organisations may be creating and managing data as part of their primary task or goal, while others may only be generating data as a side-effect of other actions. These differences will affect how important it is to attain a high level of maturity and, accordingly, the level of investment appropriate to support improvement.

Organisations should determine the target maturity that is appropriate for their specific goals and purpose. As an organisation gains value from its open data practice this may justify additional investment, and a subsequent raising of targets to help unlock further benefits. Assessment and improvement should be an iterative process.

If an organisation is transparent about its current open data maturity and targets it may help the wider community understand what it can expect from that organisation in terms of its open data practice. Transparency can also help support benchmarking across organisations.

Importantly, the model is not meant to hinder the release or use of open data. For example, a low maturity score for a data release process should not preclude publishing open data. The activities described in the model should provide a roadmap for improvement not a list of tasks to complete before data is released. Indeed, it is desirable for organisations to experience low levels of maturity so that they develop processes that are appropriate for the organisation. Attempting to leapfrog to high maturity levels may result in imposing processes that do not take into account the culture, environment or particular needs of the organisation.

Setting an assessment process

The following process outlines a recommended approach for conducting a maturity assessment:

- 1. Identify an organisational lead** — a thorough assessment will likely require input from across the organisation but there should be a clear lead who coordinates the assessment.
- 2. Identify the scope** — the maturity model can be used to assess individual departments or a whole organisation. We recommend beginning with an assessment of the whole organisation.
- 3. Identify key participants** — which people in the organisation may need to be involved to help answer specific questions or support the evaluation?
- 4. Assess and score each activity** — using the assessment grid, review each of the activities and identify the level of maturity achieved by the organisation. To qualify at a

specific maturity level, the organisation should exhibit all of the described behaviours.

5. **Set appropriate targets** — having conducted a baseline assessment, identify appropriate targets for improvement. This will involve either maintaining or improving the score for specific activities.
6. **Develop action plan** —based on the results and the targets, identify a plan for implementing improvements.
7. **Circulate results** — share the results, targets and action plan within the organisation, including to those involved in supporting the assessment. Senior management support and review will be essential in helping to implement improvements. An organisation may also wish to share its results more widely.
8. **Set date for next assessment** — the action plan should set a date for a further assessment. This will allow the organisation to monitor its progress. We recommend conducting regular annual assessments.

Prioritising and aligning your activities

The assessment grid and this guidance document describe a number of activities. The order in which these are presented reflects a rough progression from operational concerns (e.g. technology, standards) through to strategic issues (e.g. finance and policy). However there is no underlying assumption that any of the activities have a higher priority or value than others.

Also, in practice, the activities will relate to one another. For example, developing an internal asset catalogue to help create strategic oversight may also progress good data governance and data release processes. Similarly, developing a dataset valuation process may help inform financial planning and prioritising releases.

An organisation may choose to assign its own priorities to the activities in the model:

- When **conducting an assessment** it may be useful to prioritise the review of certain activities, e.g. to review well-defined and understood areas first or to align with other organisational priorities.
- When **setting targets and developing an action plan** the organisation may wish to prioritise certain activities for improvement. Some organisations may prefer to implement changes in a top-down style (perhaps focusing first on strategy and oversight) while others may prefer a bottom-up approach.

Theme 1. Data management processes

Activity	Aspect	Level 1 - Initial	Level 2 - Repeatable	Level 3 - Defined	Level 4 - Managed	Level 5 - Optimising	Beneficial effects
Data release process	Publication	Little or no published open data. Datasets that are published [...]	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, [...]	Reduce overheads associated with data releases.

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 re-used 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.

Building a process to support data release

A mature open data organisation will **have a well-defined process to support the publication of open data**. The process will address the technical aspects of publishing both new datasets and updates to existing datasets in a timely manner.

The **release process** will be well documented and address key issues such as:

- the **technical infrastructure** used to support a release, e.g. a specific data platform or portal
- the **creation and maintenance of dataset-specific metadata**
- the **internal processes and workflows** that support review and packaging of data for release
- the **syndication of datasets** and/or metadata to third-party data catalogs and platforms

The organisation's standard process will describe the key steps involved in a release, identifying responsibilities for ensuring releases happen in a timely manner and to a high standard. Individual datasets may be released using a methodology that adapts the organisation-wide standard based on the needs of the specific dataset or product.

Key metrics that may be collected about this process include:

- the **number of datasets released**
- the **number of datasets released to schedule**
- the **number of datasets being regularly updated**
- the **mean time between internal updates being made to a dataset and those updates being shared with others**

Moving from ad-hoc approaches to data releases to a repeatable process will bring a number of benefits, including:

- **simplifying release of new data** through reuse of existing workflows and tools
- making it **easier for reusers to find and use a variety of datasets** from the organisation
- **clarity around whether datasets are being** released in line with corporate objectives

Developing standards and adoption

Mature organisations will **benefit from both the use of open standards for formatting data and the adoption of industry standard identifiers** in their datasets.

These benefits for publishers include:

- **available open source tooling** to support managing data
- **easy recruitment of specific technical expertise**
- **reduced burden of maintaining and documenting bespoke standards**
- **increased adoption and use of published data**

Adoption of standards also has **benefits for reusers** of open data:

- **datasets published in industry or de facto standard formats are easier to consume**
- **datasets that use common identifiers**, such as codes for geographical areas, are **more easily compared and combined**
- reusers **can easily find relevant data** and documentation by looking up identifiers
- reusers **can more easily process datasets** when they are published using standard metadata and packaging formats

When standards are widely adopted it benefits everyone as costs can be lowered through the use of common, standard tools. Datasets that are linked together can be more easily analysed and combined more flexible ways.

A mature open data organisation will **define the technical standards that it will use when publishing its data**. These standards will address:

- the data formats (e.g. CSV, JSON, KML) used to structure the data
- the access methods (e.g. an API) used to make data available
- the means by which the data is structured, e.g. through use of standard schemas

The standards may refer to a single set of technologies or recognise that certain types of data (e.g. geographical or statistical data) might be best published using type-specific standards.

The mature open data organisation will also **recognise that as standards evolve it may need to revise its data publication practices in order to best support both new and existing consumers**. The organisation will therefore track industry trends and monitor how other organisations are publishing similar and interrelated datasets.

A mature organisation will also **research the ways in which reusers would like to process data**. This might lead the organisation to publishing data using several different access methods. For data analysis the ability to download data in its entirety is often essential, whereas real-time access to data may be better served by development of an API.

Developing data governance

Releasing high-quality open data requires organisations to apply appropriate levels of data governance to their datasets, addressing issues such as data quality, integrity, monitoring and protection across the entire data lifecycle.

A mature open data organisation will **implement a data governance process across all of its datasets**. This includes datasets produced internally and those it is consuming from third-party sources.

Organisations may find that publication of open data will highlight existing data governance issues. There are many examples of inconsistencies and inaccuracies in data being uncovered only after the data is exposed to wider review. This often leads to improvements in not just the quality of the data but also review and improvements in the underlying data management practices. The resulting improvements to data quality can bring particular benefits when the organisation uses the data within its own processes and decision making.

A mature organisation will **treat a third-party open dataset as if it were its own asset**. This will involve providing appropriate feedback to the original source on data quality issues. In the case of crowd-sourced or collaboratively maintained datasets, the organisation will likely invest time in resolving these issues for the benefit of itself and other reusers.

Managing sensitive data

Datasets may include a variety of sensitive information such as commercially sensitive data that an organisation may not be happy about sharing with competitors, or personal information about its customers, employees, or other individuals.

A mature organisation will **conduct risk and impact assessments prior to the publication of sensitive datasets**. For example it may conduct a Privacy Impact Assessment¹¹. It will also have procedures to mitigate risk prior to release, including:

¹¹ Privacy by design, <https://ico.org.uk/for-organisations/guide-to-data-protection/privacy-by-design/>, accessed on 2015-03-23

-
- have an anonymisation policy and **know about the process of anonymising data**
 - **redacting commercially sensitive figures** or company names
 - subsetting of data to **remove sensitive fields**
 - aggregation in order to **summarise data**
 - **obtaining appropriate consent** from individuals whose personal data may be released
 - **considering the data environment** such as other related datasets, people who have access to the data or people who have a motivation to exploit the data beyond its intended uses.

A mature organisation may also **seek independent, external validation of its approach to anonymising highly sensitive datasets**. For example, it may engage a third party to perform a motivated intruder test.

In addition to identifying standard approaches to desensitising, the organisation will properly educate its staff about their use ensuring that techniques are applied correctly and effectively.

Theme 2. Knowledge and skills

Activity	Aspect	Level 1 - Initial	Level 2 - Repeatable	Level 3 - Defined	Level 4 - Managed	Level 5 - Optimising	Beneficial effects
Open data expertise	Publication & re-use	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 [...]	Reduce reliance and cost of using external expertise.

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.

The skills development support provided by organisations for health & safety is a good analogy for how an organisation may develop both a broad understanding of good data management practices as well as identified, accessible expertise.

All organisations offer health & safety training as a standard element of staff induction and key staff members are identified as health & safety experts. “Data health & safety” – understanding the risks, responsibilities, and actions required to ensure that data is well managed and shared effectively – should become an element of staff development in mature open data organisations.

Expertise in topics such as licensing and data standards will be identified and made available across the organisation.

Developing open data expertise

A mature organisation will **ensure that staff have the necessary training and support required to deliver on their individual responsibilities relating to open data**. This is likely to be one aspect of wider awareness of the benefits of good data governance and its application to company strategy.

It will be important that employees have training to ensure that they have

- an **understanding of both the risks and benefits** of using and publishing open data and how that affects their own role and responsibilities
- an **understanding of the organisation's open data policies and strategy**
- the ability to **apply the appropriate level of data governance** in their own projects
- **specific training on individual topics**, e.g. technical standards, best practices, licensing issues

Ideally this education programme will begin at staff induction (“data health & safety”) and be offered as part of continual staff development, e.g. via internal training and refresher courses.

A mature organisation will also work to develop skills at all levels of the organisation, supporting both operational needs and strategic oversight.

Where staff identify additional support is required then access to the necessary expertise, whether from other parts of the organisation or externally, will also be made available. In larger organisations this may translate into dedicated staff responsible for managing open data programmes and technical platforms.

Knowledge management

Knowledge management is an important process in any organisation. From an open data perspective, knowledge management is important for two reasons.

First, staff need to be able to find and use documentation on organisational standards and policies. Many of these policies will also be of interest to third-parties such as potential reusers of published data. A public open data policy and strategy will help identify the level of organisational commitment to open data and identify how third-parties can engage with the organisation.

Second, both internal and external users need to have access to the necessary documentation required to support the use of published data. This is particularly important for organisations that undertake research projects or similar activities that may produce a number of datasets. For data to retain its value in the long term, people who use it need to know about how it's collected, how it's processed and who owns it.

A mature organisation will maximise the value of its data assets by ensuring that appropriate information is captured and shared throughout the lifecycle of a dataset. This information should be easily available to both internal and external users.

Theme 3. Customer support and engagement

Activity	Aspect	Level 1 - Initial	Level 2 - Repeatable	Level 3 - Defined	Level 4 - Managed	Level 5 - Optimising	Beneficial effects
Engagement process	Publication	Little or no attempt made to identify potential re-users for released [...]	Some teams attempt to identify and engage with potential re-users [...]	The organisation identifies a repeatable approach for [...]	The organisation begins tracking the effectiveness of its [...]	The organisation is routinely tracking metrics relating [...]	Greater insight into re-user needs & activities.

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 a 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.

Developing an engagement process with data reusers

A mature open data organisation will **define a process for engaging with reusers of its data**. For organisations that are already familiar with publishing data or supporting external customers this type of engagement will be an evolution of existing practices. In other organisations this might be a new process and may require the organisation to identify new roles and communication channels.

Organisations with existing customer engagement and marketing programmes will need to consider how publishing open data may result in new audiences for their data. Cost-reduction and network effects may surface additional demand for and opportunities for using data.

An engagement process will typically include a range of activities¹²:

- identifying both **potential and actual reusers of a dataset**
- identifying the **key stakeholders in the re-user community** in order to prioritise and focus engagement activities
- engaging with that community to **help them understand how a dataset might be used** and to highlight relevant new data releases
- a communications plan to **help promote the activities of both the publisher** (news, events, etc) **and the community itself** (e.g. share success stories, insights into the data, and case studies)
- reflecting experiences back into the organisation to help demonstrate value, build business cases, etc.
- running events such as hack days or challenges to **encourage and incentivise use of data** and to generate case studies for wider sharing

Ideally the process will be metrics-driven, allowing an organisation to monitor and improve its outreach activities. Metrics gathered by the organisation might include basic indicators such as:

- the number of **users accessing or using data**
- the number of **users contributing to discussions or data updates**
- the number of **applications developed on published data**

Not all datasets will necessarily need or require a detailed engagement activity. One step in defining the engagement plan for a dataset will be identifying the appropriate level of engagement necessary and appropriate to the ongoing investment in the data.

Documenting your open data

Data cannot be used effectively if it is not properly documented. A mature open data organisation will **ensure that all datasets are published with a standard set of supporting documentation that follows a consistent template structure**. This documentation will be created and maintained throughout the lifecycle of the dataset, rather than just prior to publication.

The level of documentation required will usually vary depending on the complexity of the individual dataset and its method of collection and analysis. Typically documentation will include:

12 see also, the ODI guide on engaging with reusers: <http://theodi.org/guides/engaging-reusers>

-
- **descriptive data** — providing a high-level summary and supporting metadata to describe the dataset including title, description, keywords, licence, rights statements and attribution requirements
 - **access information** — describes how to access the data, location of archives and mirrors, etc
 - **indicators** — summary statistics that provide insight into the size, rate of growth, quality and update frequency of the dataset, and data quality
 - **relationships** — pointers to other data sources that were used to construct the dataset, e.g. references to standard code-lists and other complementary datasets
 - **scope & coverage** — a description of the contents of the dataset, the types of entity it describes, its geographic focus, and the time period to which it applies
 - **provenance** — how data has been collected and processed prior to publication
 - **technical documentation** — a description of the data formats with reference to formal schemas that define the fields in the datasets; sample records may be used for illustrative purposes

Data delivered via an API rather than a dataset download will also be accompanied by a specification of the API implementation.

Creating and maintaining this documentation will form an important element in both the process of releasing a dataset and supporting its subsequent use. Ideally documentation will be managed using a content management system that enables its rapid production, review and publication.

In a mature open data organisation, reference **documentation will often be supplemented with additional material**, such as tutorials on using or accessing a dataset, sample applications, or pointers to useful tools and resources that support effective reuse.

The effort put into providing both the core and any additional documentation will be appropriate to the value and investment in the dataset.

Building a reuser support process

A mature open data organisation will have **well-defined contact points that will allow third-parties to engage with the organisation about its open data practice**. For example, they will allow reusers to request that the organisation open up data, or provide input into the prioritisation of planned releases.

For data that is already released, a mature organisation will also provide support for reusers

in other ways:

- an **open forum for raising questions about a dataset**, e.g. with a support team or others in the reuser community
- a **means to report errors, data quality, or privacy issues with a dataset**
- **communication on issues that may impact reusers**, e.g. significant changes to a dataset; reported data quality issues; retiring of a dataset; or a decision to not release an identified dataset

The organisation will ensure that the support available for a dataset is clearly defined so that reusers can have appropriate expectations about the level of support available.

Creating open data community norms

A mature open data organisation will **support not only its reusers but will also act as a good citizen in the wider open data community**. For example an organisation will be transparent about the open data it uses, ensuring that appropriate attribution is given to those sources.

The organisation will also:

- seek to **share its experiences with individual datasets**, providing feedback to their publishers on ease of use, data issues, etc
- **be transparent about the open datasets** it uses to help highlight the value of those datasets in the community; this will include ensuring that sources are attributed
- **create and share case studies** that highlight the impact and benefits of working with open data

Theme 4. Investment and financial performance

Activity	Aspect	Level 1 - Initial	Level 2 - Repeatable	Level 3 - Defined	Level 4 - Managed	Level 5 - Optimising	Beneficial effects
Financial oversight	Publication & re-use	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 [...]	Costs of publishing and using data clearly [...]

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.

Ensuring financial oversight

A mature organisation will **actively monitor the financial costs and benefits of both publishing and using open and closed (restrictively licensed) data**. This will include ensuring that project funding and operational costs for new and existing projects include the appropriate levels of investment to support data publication where applicable.

The organisation will recognise the difference between the ongoing costs associated with general data governance and those associated specifically with open data publication. This will avoid incorrectly attributing to open data publishing efforts the entire cost of improvements to data governance and similar best practices.

The organisation will also use this financial insight to look for efficiency savings in its publication and use of data such as through reduction in data management overheads or cost savings from adoption of open data.

Developing dataset valuation processes

A mature organisation will **recognise that investment in open data publication should be driven by a cost/benefit analysis**. High value datasets, namely those directly aligned with delivery of corporate objectives, may require higher levels of investment than other secondary datasets.

The organisation will have a methodology for assessing the value of a dataset to both the business and end-users. The methodology may involve engaging with end-users in order to identify the level of interest or demand in datasets, and how that data might be used.

The lack of a data valuation process should not hinder the release of open data. But by defining a process an organisation will be better able to prioritise datasets for release and to ensure appropriate levels of investment are made in their governance and ongoing publication.

A mature organisation will also **be transparent about its approach to valuing datasets, sharing the metrics it uses with those who have interest in its data**.

Similarly, where an organisation is re-using datasets it should be able to quantify the financial benefits associated with their reuse, for example by comparison with the costs associated with licensed alternatives. The organisation will work to maximise the value it derives from data assets and look for opportunities where switching from a licensed dataset to an open dataset might provide additional benefits, such as affording greater transparency or the ability to switch to less costly service providers.

Building open data into procurement practices

For an organisation to understand its rights to publish and reuse data it is essential that there is clarity about how data is created and managed throughout the organisation¹³. In some cases data assets may be created or managed by third-parties such as by contractors, partner organisations, or in Software-as-a-Service infrastructure.

A mature open data organisation will ensure that:

- **contracts used for sub-contractors clearly describe the intellectual property rights** associated with any data delivered or created during a project or service, where possible granting those rights to the organisation

¹³ See the ODI guide on embedding public data into procurement of public services <http://training.theodi.org/Procurement/Guide/>

-
- procurement processes have been updated to **ensure that, where appropriate, potential contractors provide information about how they will deliver open data** as part of a project
 - contractors **provide clarity around the provenance of any data they use** or supply so that the organisation is clear on any rights relating to derived datasets
 - procurement practices **take into account the whole-life costs** of a service to favour suppliers that provide easy access to data required to move suppliers

This will ensure that the organisation is proactively clarifying intellectual property rights, where necessary, rather than reactively undertaking costly rights clearance after contracts are awarded.

Theme 5. Strategic oversight

Activity	Aspect	Level 1 - Initial	Level 2 - Repeatable	Level 3 - Defined	Level 4 - Managed	Level 5 - Optimising	Beneficial effects
Open data strategy	Publication & re-use	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 [...]	Clarity within the organisation and externally about open data strategy.

Given the impacts of open data practice 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.

Shaping open data strategy

A mature open data organisation will **have an open data strategy that clearly describes its ongoing commitments and policies relating to its open data practice.**

As an internal resource, an open data strategy will set out the organisation's strategy for building open data into its products and reporting. It will also identify the roles and responsibilities of those involved in publishing open data on behalf of the organisation. The strategy will reference the key processes highlighted elsewhere in this model, that is for procurement, data governance, release processes, etc.

A mature organisation might **publicly publish its open data strategy, making a stronger commitment to open data.** A public strategy will give reusers insight into how the organisation prioritises its datasets for release and describe how to engage with that process. A good open data strategy will also identify how the organisation's policy may evolve over time.

The mature organisation will also **attempt to measure its progress against its strategy.** For example it may include commitments that set out targets for gaining and improving the levels of dataset certification according to the Open Data Institute's certification progress.

The open data strategy will have a clear owner who will have both the budget and responsibility for its delivery. The individual processes referenced in the strategy, such as data governance, will also be owned by specific roles in the organisation and their responsibilities will be clearly defined.

Asset catalogue management

A mature open data organisation will **treat data as an asset**. The organisation will maintain a catalogue that lists both its own internal datasets and those it uses from third-parties. The catalogue will include all significant internal data, not just that published as open data.

The catalogue will support the reporting necessary to deliver the appropriate oversight over data governance and release processes and, at a higher-level, progress against the organisational strategy. The catalogue may also support discovery of cost savings by identifying duplicate or overlapping data sets and opportunities to reuse existing datasets in new products and services.

For datasets that the organisation owns, the catalogue may include information such as:

- the **risks** associated with releasing a dataset, e.g. whether it contains personal or sensitive information
- the **value** of the dataset, as determined through the organisation's data valuation process
- the **details** of ongoing financial investment associated with its release (or re-use) as open data
- whether the dataset is **planned for release**
- a **summary of the key metrics** collected by the engagement process, etc.

For datasets that the organisation reuses, the catalog may include information such as:

- the rights that the organisation has to use the data
- for commercial data, the costs associated with reuse
- for commercial data, the risks associated with a price increase
- the risks associated with a dataset becoming unavailable or unmaintained
- the ways that the data is re-used by the organisation



Authors:

Leigh Dodds

Andrew Newman

Editing:

Anna Scott

Kathryn Corrick

Production:

Phil Lang

Design

Adrian Philpott



This text is licensed under a Creative Commons Attribution - ShareAlike 4.0: International License