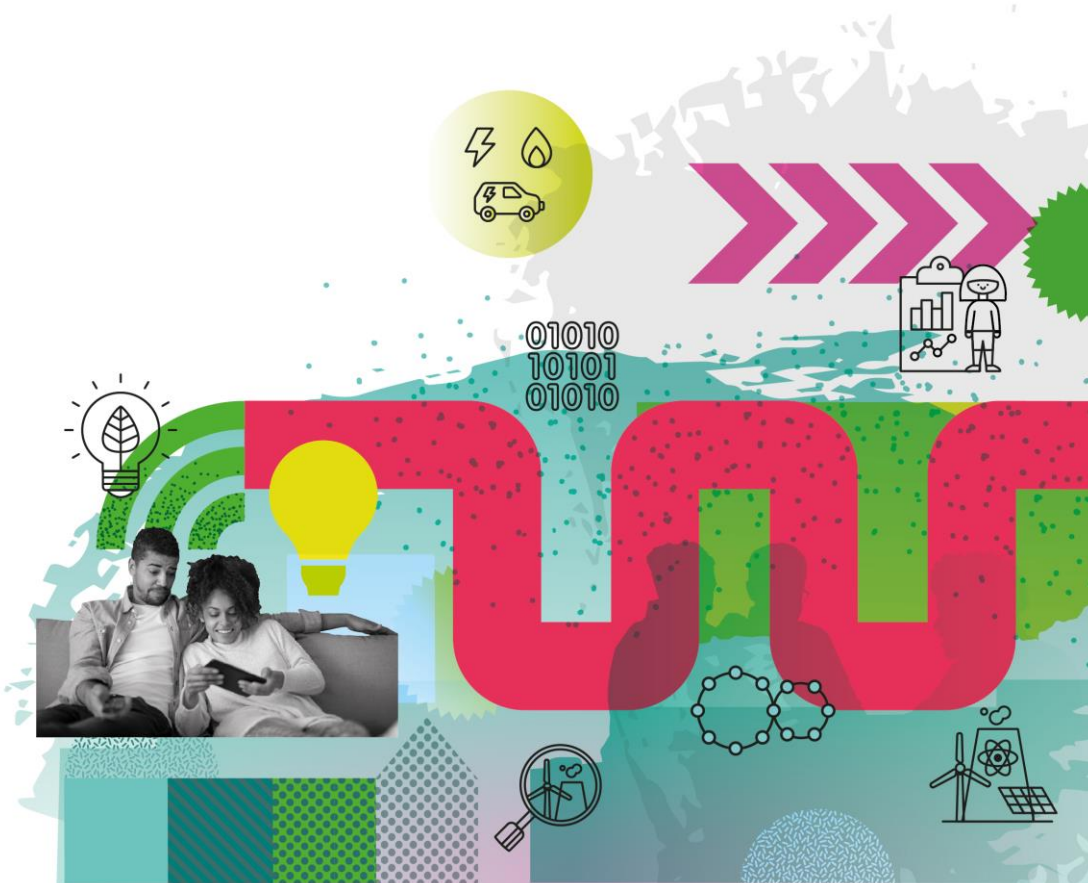
Data Catalogue

Energy Data Taskforce Appendix 2

Energy Data Taskforce

13/06/2019



Contents

1. Introduction

The Taskforce recommends that a **Data Catalogue** be established to provide visibility of the data that exists across the sector through a common metadata standard. This needs to encompass all Energy System Datasets across Government, the regulator and industry. Government and Ofgem should mandate participation though regulatory and policy mechanisms.

Recommendation 3: Visibility of Data

A Data Catalogue should be established to provide visibility through standardised metadata of Energy System Datasets across Government, the regulator and industry. Government and Ofgem should mandate industry participation though regulatory and policy frameworks.

1.1. Why a Data Catalogue

Visibility of Energy System Data is currently very poor. New innovators and experienced incumbents alike struggle to identify if a dataset exists, which organisation owns it and if they can gain access. This creates several negative impacts including stifling innovation, obscuring data quality issues and creating inefficiency when the same data is collected many times over.

The Data Catalogue is designed to address the problem of data visibility by requiring organisations holding Energy System Data to contribute metadata about their datasets. The Data Catalogue collates this into a single, searchable location and enables parties to quickly and easily find out if data exists and who holds it. Where data is open or publicly available it can be directly linked but in other cases a request for access can be created via the Data Catalogue.

The Data Catalogue offers the following benefits:

- Greater transparency of data which is being collected and by which organisations and the outcome of its openness triage
- Identification of data quality/granularity issues and sector-wide data gaps
- Increase adoption of metadata standards and information management best practice
- Enable objective evaluation of increasing data openness by organisations and comparison of the capabilities and performance of similar organisations
- Support market demand-based prioritisation for open data

1.2. The Case for Intervention

Visibility of metadata is an enabler for innovation and efficiency and drives a number of benefits as described above. However, it is unlikely that a solution such as this would be spontaneously created by the market without external leadership for the following reasons:

1. **Participation**: The value of the Data Catalogue comes from having a complete view of all datasets. This requires all energy system organisations to participate which is difficult to achieve without a clear obligation and sector leadership

2. **Risk**: Without complete sector participation there would be a disadvantage to being first mover. Giving away information without the promise of gaining access to other's information is risky
3. **Costs**: There are costs to develop and maintain the Data Catalogue which should rightly be shared across the entire sector. Creating a private contractual arrangement to facilitate pooling of costs would be very difficult.
4. **Ownership**: There is no obvious industry organisation which has enough independence, mandate and the right set of skills to own a platform such as this.

External intervention is required to unlock the benefits a data catalogue can provide.

1.3. Data Catalogue Requirements

The core functionality of the data catalogue is to collate standardised metadata for datasets relating to the Energy System and make it available for all so data can be discovered easily. However, there are a number of additional features which can be implemented via the data catalogue which will increase the value of the catalogue for the sector.

Data Request: The principle of Presumed Open advocates that all data should be as open as possible, but it will take some time to assess the sensitivity and determine how open each dataset can be. By integrating a simple 'click to request' function users will be redirected to the data source if it is open or publicly available or given the opportunity to make a data request if it is not.

Openness Reporting: In order to track the progress towards openness it is important to have a low impact way to audit the relative openness of organisations. The Data Catalogue makes it possible to track the number of datasets an organisation is reporting and compare this to similar actors across the sector and then measure the openness of those datasets. This data can then be used to identify organisations which are not complying with the principles and offer support or pursue enforcement.

Market Pull: Through analysis of the data requests being made it will be possible monitor the number of parties who would like to gain access to each dataset and use this to evidence market pull for the data. This will enable the data manager (and the regulator) to evidence demand for datasets and prioritise the data openness triage accordingly. Clearly there will be a need for additional layers of prioritisation as some datasets hold the key to unlocking significant benefit for the Energy System but create little demand now.

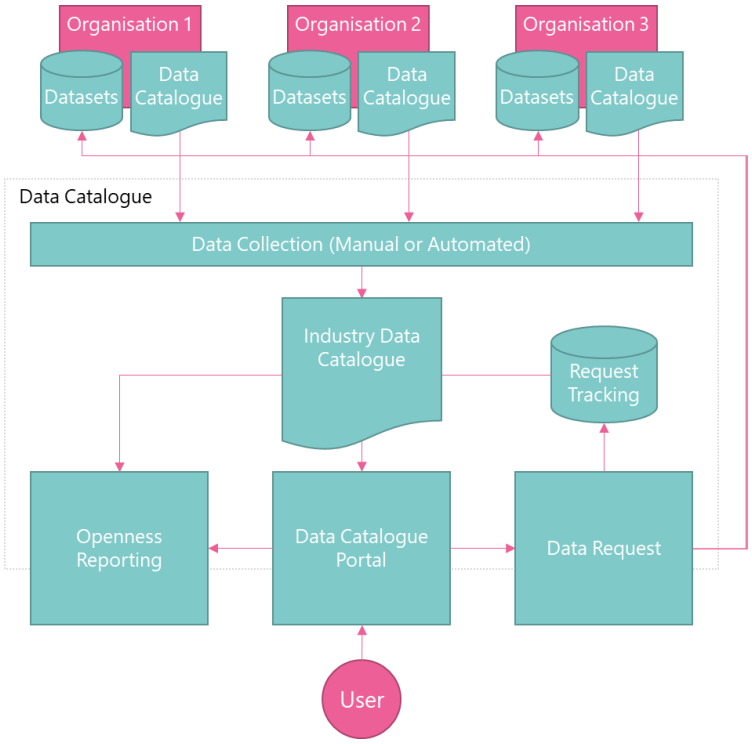
Peer Review: Implementation of Presumed Open requires that organisations triage their own data to determine how open it can be. This will inevitably lead to disparity across the sector and it is important that peers are able to review the decisions of others and challenge where they believe the data is unnecessarily closed or where an issue has not been identified and the data is too open.

Metadata Standardisation: The Data Catalogue will need to collect a large amount of metadata from a range of organisations across the sector. In order to make this information manageable and useable it is important this the metadata is in a standard format. We propose that Dublin Core 'Core Elements' metadata standard is utilised as this is a common format used across a number of sectors including industry initiatives such as the Energy Data Centre which can be used to archive Energy System Data.

Common Glossary: To complement the metadata standard it is necessary to develop a common glossary such that data can be categorised and it is possible to find common datasets across organisations.

2. Data Catalogue Design

The diagram below illustrates the functional blocks required to deliver the desired functionality of the Data Catalogue. The individual elements have been described in more detail below.



2.1. Data Collection

The first element is the data collection. This involves gathering metadata from individual organisations about the datasets which they hold. Metadata should be provided in a Standard Format and contain information about the data itself and, if the dataset is publicly available, a link to the specific location where the data can be accessed.

Initially, this could be implemented manually but for the Data Catalogue to be sustainable this should be migrated to an automated mechanism such as Rest API. As organisations update their internal data catalogues these changes can be automatically pushed / pulled into the Industry Data Catalogue.

2.2. Industry Data Catalogue

Once data has been collected it can be collated to create a central Data Catalogue with data from many organisations across the sector. This is the master dataset and there should be efforts to ensure it is accurate and up to date.

2.3. Data Catalogue Portal

The Data Catalogue Portal is the interface which enables users to view, search and access metadata from across the sector. There should be no requirement for users to register but this may be offered as an optional function which provides additional functionality or automation e.g. automatically populating information on data requests or providing a search history. The Data Catalogue Portal should be the main interface for the service which links to the Data Request form and Openness Reporting function.

2.4. Data Request

When a user identifies a dataset which they would like to access it should be possible to either click through to the data source or create a data request directly from the Data Catalogue. The data request can initially be as simple as standard email template to the data manager within an organisation but this could be enhanced to enable a user to manage their requests, monitor the outcome of requests and make a complaint directly to the regulator where the outcome is deemed unsatisfactory.

2.5. Request Tracking

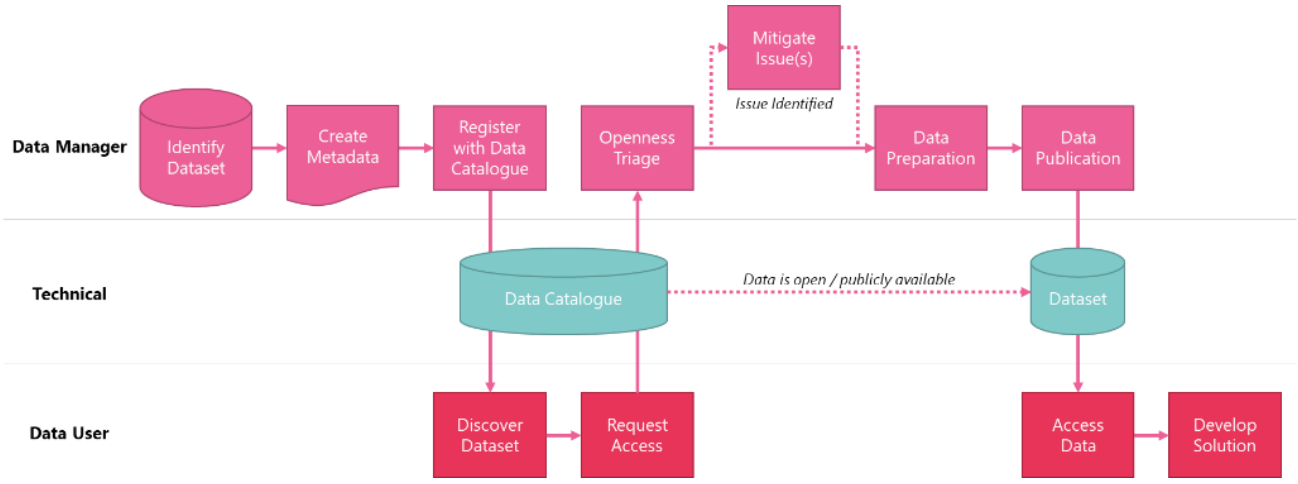
The Data Catalogue should track the total number of requests for data in order to evidence the impact of open and public data but also track the outcome of specific data requests. This will provide an enriched evidence base for the regulator when considering if organisations are complying with the Presumed Open principle.

2.6. Openness Reporting

By aggregating data from the Industry Data Catalogue and Request Tracking it will be possible to build up a very good picture of both organisational and industry wide data openness. The openness reporting function should be an automatically generated (and updated) report that is accessible to any interested party.

2.7. Data Journey

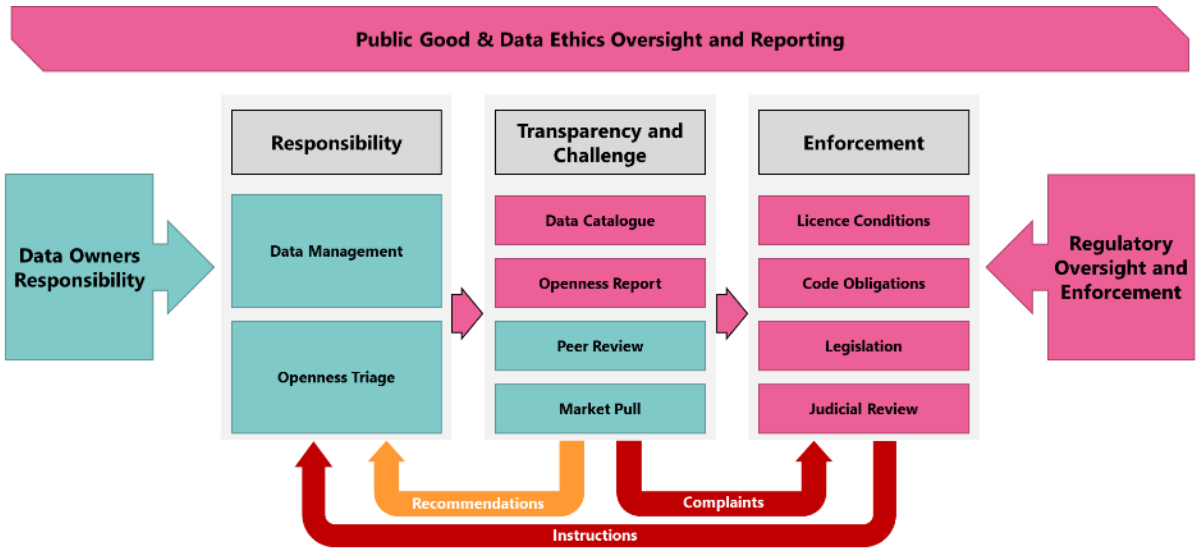
The diagram below illustrates a typical 'data journey' from identification within an organisation through to third party accessing the information and developing a solution with it.



3. Governance

The Data Catalogue is an important enabler of the overall governance system. We outline below the main features which enable Transparency and Challenge to form a powerful technique for driving openness whilst ensuing risks are adequately managed.

The Governance process is based on 3 pillars: Responsibility (of data managers), Transparency and Challenge and Enforcement. Over the top of this there is space for a cross sector oversight function which ensures risks within energy and beyond are managed at a national level.



Data Catalogue: Creating a high level of dataset visibility through the registration of metadata for Energy System Datasets. This is an enabler for the following points.

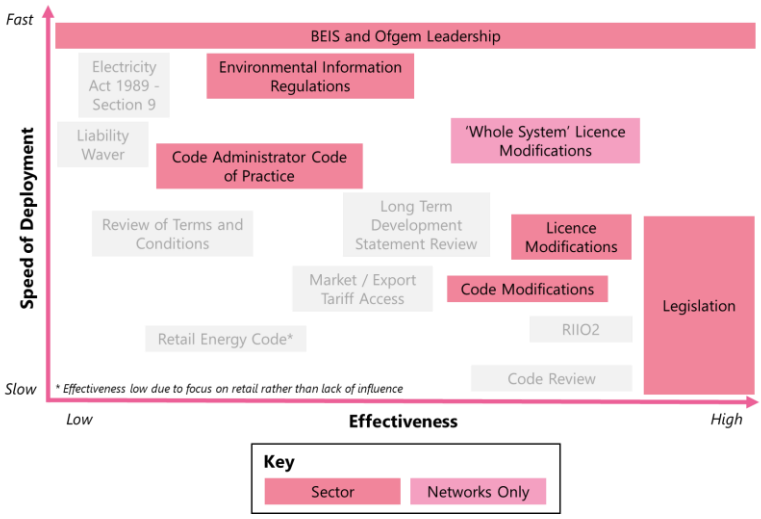
Openness Report: The Data Catalogue will provide access to key metrics about organisation and sector datasets and openness. This information can be periodically collated and published as a report for Ofgem and BEIS on the data openness of the sector.

Peer Review: The Data Catalogue enables peers to review the openness of datasets and challenge where appropriate. This enables organisations to benchmark their openness and provide feedback where new issues are identified.

Market Pull: Requests for access to non-public datasets can be tracked, enabling market pull to be quantified and used to form a priority for opening.

4. Delivering the Data Catalogue

In order to successfully deliver the Data Catalogue, it will be important to ensure that organisations participate via the following mechanisms:



1. **Leadership from BEIS and Ofgem:** The Taskforce recommends that BEIS and Ofgem endorse the recommendation to implement a Data Catalogue and are clear on policy and regulatory intent in this regard. This will encourage many Energy System actors to participate with the Data Catalogue before any required regulations are in place.
2. **Environmental Information Regulation:** The Taskforce recommends that Energy System organisations utilise regulation 4 of EIR 'Proactive Dissemination' to make data more open. Additionally, individuals and organisations can make data requests to obtain specific datasets which have not proactively been made available.
3. **Whole System Licence Modifications:** Ofgem are currently considering licence changes for electricity networks and have just concluded a consultation to which the Taskforce responded. **The Taskforce recommends** that Ofgem consider taking this opportunity to obligate organisations to deliver dataset visibility.
4. **Code Administrator Code of Practice: The Taskforce recommends** that industry consider adopting the principles and/or integrating the Building Block projects into the CACoP in order to encourage adoption of the recommendations.
5. **Licence Conditions: The Taskforce recommends** that Government and Ofgem consider using licence conditions to ensure licensed actors are compelled to register their metadata with the Data Catalogue.
6. **Codes**: The Taskforce recommends that code modifications be considered and if appropriate, implemented to increase the number and range of actors required to participate with the Data Catalogue. Discussions to date have indicated that Code Administration Code of Practice, licence modifications and Significant Code Review and the Energy Code Review could be options for embedding such requirements.
7. **Policy Change**: **The Taskforce recommends** that Government consider acting to address a number of legislative barriers to data release which could be reduced and overarching obligations which could be created. Key areas for review are:
    a. Changing Ofgem/Secretary of State duties to embed requirements to explicitly promote greater transparency, develop high quality data sources and publish more data; to avoid data being deprioritised in the future
    b. Extending the general duties of network companies to reflect that data is a key component of infrastructure and should be treated as such

5. Implementation Recommendations

The Taskforce recommends that the Data Catalogue is based on the **Dublin Core 'Core Elements' metadata** standard. This enables a minimum level of standardisation without being overly burdensome with the option to extend the standard over time. To compliment this there will be a **common glossary** to enable actors to effectively categorise their data.

The Taskforce recommends that the Data Catalogue should be developed and managed by an independent, trusted party with a strong track record in data management. The Taskforce recommends that the **Office for National Statistics (ONS)** be considered for this role. This would enable the Catalogue to benefit from the ONS's existing security, governance and quality standards and their data analysis expertise. In addition, the ONS's statutory role would enable energy data to be informed by wider economic data sets.

Work to date suggests that a Minimum Viable Product (MVP) Data Catalogue can be ready within 6 months. Government, Ofgem, Code Operators and Settlement Agencies have indicated their agreement to log data with the Catalogue.

LICENCE/DISCLAIMER

Energy Systems Catapult (ESC) Limited Licence for **Data Catalogue: Energy Data Taskforce Appendix 2**

ESC is making this report available under the following conditions. This is intended to make the Information contained in this report available on a similar basis as under the Open Government Licence, but it is not Crown Copyright: it is owned by ESC. Under such licence, ESC is able to make the Information available under the terms of this licence. You are encouraged to Use and re-Use the Information that is available under this ESC licence freely and flexibly, with only a few conditions.

Using information under this ESC licence

Use by You of the Information indicates your acceptance of the terms and conditions below. ESC grants You a licence to Use the Information subject to the conditions below.

You are free to:

- copy, publish, distribute and transmit the Information;
- adapt the Information;
- exploit the Information commercially and non-commercially, for example, by combining it with other information, or by including it in your own product or application.

You must, where You do any of the above:

- acknowledge the source of the Information by including the following acknowledgement:
- "Information taken from Data Catalogue: Energy Data Taskforce Appendix 2, by Energy Systems Catapult";
- provide a copy of or a link to this licence;
- state that the Information contains copyright information licensed under this ESC Licence.
- acquire and maintain all necessary licences from any third party needed to Use the Information.

These are important conditions of this licence and if You fail to comply with them the rights granted to You under this licence, or any similar licence granted by ESC, will end automatically.

Exemptions

This licence only covers the Information and does not cover:

- personal data in the Information;
- trademarks of ESC; and
- any other intellectual property rights, including patents, trademarks, and design rights.

Non-endorsement

This licence does not grant You any right to Use the Information in a way that suggests any official status or that ESC endorses You or your Use of the Information.

Non-warranty and liability

The Information is made available for Use without charge. In downloading the Information, You accept the basis on which ESC makes it available. The Information is licensed 'as is' and ESC excludes all representations, warranties, obligations and liabilities in relation to the Information to the maximum extent permitted by law.

ESC is not liable for any errors or omissions in the Information and shall not be liable for any loss, injury or damage of any kind caused by its Use. This exclusion of liability includes, but is not limited to, any direct, indirect, special, incidental, consequential, punitive, or exemplary damages in each case such as loss of revenue, data, anticipated profits, and lost business. ESC does not guarantee the continued supply of the Information.

Governing law

This licence and any dispute or claim arising out of or in connection with it (including any noncontractual claims or disputes) shall be governed by and construed in accordance with the laws of England and Wales and the parties irrevocably submit to the non-exclusive jurisdiction of the English courts.

Definitions

In this licence, the terms below have the following meanings: 'Information' means information protected by copyright or by database right (for example, literary and artistic works, content, data and source code) offered for Use under the terms of this licence. 'ESC' means Energy Systems Catapult Limited, a company incorporated and registered in England and Wales with company number 8705784 whose registered office is at Cannon House, 7th Floor, The Priory Queensway, Birmingham, B4 6BS. 'Use' means doing any act which is restricted by copyright or database right, whether in the original medium or in any other medium, and includes without limitation distributing, copying, adapting, modifying as may be technically necessary to use it in a different mode or format. 'You' means the natural or legal person, or body of persons corporate or incorporate, acquiring rights under this licence.

Energy Systems Catapult supports innovators in unleashing opportunities from the transition to a clean, intelligent energy system.

Energy Systems Catapult
7th Floor, Cannon House
18 Priory Queensway
Birmingham
B4 6BS

es.catapult.org.uk
info@es.catapult.org.uk
+44 (0)121 203 3700