CATAPULT Energy Systems

Energy Data Management Canvas

A tool and process to help you to define and document the use and governance of data in your project



INDUSTRIAL STRATEGY

UK Research and Innovation

Contents

Introduction 1

What is a data management plan? 2

Process Overview 4

1. Understanding (Data) Stakeholders 5
2. Data Sources & Collection 7
3. Data Storage 9
4. Data Governance 12
5. Data Sustainability 17
6. Conclusions & Next Steps 20
7. Appendix 21

DISCLAIMER
This document has been prepared by Energy Systems Catapult Limited. For full copyright, legal information and defined terms, please refer to the "Licence / Disclaimer" section at the back of this document.

All information is given in good faith based upon the latest information available to Energy Systems Catapult Limited. No warranty or representation is given concerning such information, which must not be taken as establishing any contractual or other commitment binding upon the Energy Systems Catapult Limited or any of its subsidiary or associated companies.

Introduction

Smart Local Energy Systems (SLES) use digitalisation and other innovative techniques to make, move, store, trade, use and conserve energy intelligently at a local scale. By innovatively integrating technologies and using novel business models, they can deliver for the needs of a specific area to create greener, more flexible localised energy systems.

Successful integration of local energy projects requires collaboration and partnerships with a wide range of stakeholders including Local Authorities, network owners and operators, technology developers, investors, community groups and the end users. Effective digital, data and information management are key factors in achieving and demonstrating the benefits that these kinds of projects can offer to these stakeholders.

The Prospering from the Energy Revolution (PFER) Programme is supporting the development of SLES projects with the Energy Revolution Integration Service (ERIS), bringing learnings from across the programme together to provide recommendations for what is needed to accelerate the development of more local energy systems.

This is the energy data management canvas, produced by KnowNow Information in collaboration with Energy Systems Catapult for the ERIS. It was developed to support Smart Local Energy System projects; however, it could be applied to any project where data plays an important role.

The goal of this document is to guide you through a process of putting together an appropriate **data management plan** for your project to help you to understand the overall picture and ask the right questions. We have taken a broad view of what constitutes data, and the questions in this document are (for the most part) applicable to qualitative data, as well as quantitative.

This canvas was developed by Richard Gomer, Chris Cooper and Jake Verma. To get in touch, please contact energydata@es.catapult.org.uk .

What is a data management plan?

A data management plan is a document that clearly sets out out how you will collect, store, use, govern and exploit data for a project, process, or activity.

Who is this canvas for?

This data management canvas has been developed based on experience working with projects in the PFER programme. As such, it is tailored to energy system use cases and uses those use cases to form the examples, but the principles are industry agnostic and could be applied across multiple sectors.

Why do I need this canvas?

A data management plan is used to manage your data like you would an asset. If managed correctly, data can be a valuable asset, delivering fascinating and useful insights. Manage it incorrectly, and data can be confusing, contradictory, or more trouble than it's worth. Data management plans are used to make sure data collection, use and publication is as efficient and robust as possible to get the maximum value for your project.

How should I use this canvas?

Read it through first and familiarise yourself with its content. Find a group of people in your project that will need to be involved in either developing or reviewing the data management plan.

Sometimes it will be helpful to deviate from the process that we have set out. The key thing is to think carefully and critically about your data, using the process and questions here as a starting point. It is quite normal, and expected, that you will approach the task iteratively, adding more detail and advancing your thinking each time.

Why did you develop this canvas?

Some of the common challenges ESC has identified are:

- Multi-party consortia will be less able to understand and benefit from their research if they don't have a unified approach to data management
- Questions such as what data, where to store it, how it's kept up to date and how it's governed are all simple questions that don't start out with common answers
- Not having common answers to data questions leads to data silos, which inhibits the use and dissemination of the data – in a research context this can impair learning and potentially the delivery of the product or service that is the subject of the research
- GDPR challenges become unmanageable as the "right to be forgotten" is not implementable in a systematic way.

Most projects involve multiple partners. Your data management plan will need buy-in from all of them, especially where partners must co-operate to derive value from the data that is available. It can help to involve all those stakeholders in your discussions - either initially, or as you iterate.

What can I use this canvas for?

Using this data management canvas, you can tackle all the above issues in a way that your stakeholders can see and support. You may also be able to use it to prove further value of your project. Following this process will help you to identify important aspects of your plan, and to work with data in a way that meets your overall project aims and objectives. By recording your thinking

as you go and the answers to key questions, you will find that documenting your data management plan is easier too.

Why should I use this canvas in particular?

You don't have to stick to just this canvas! There are other guidelines that will help you to develop a data management plan to support a project proposal or request for funding. However, this data management canvas covers the whole lifecycle of the project and goes into detail about often overlooked factors such as the value of data, stakeholder needs and detailed governance principles.

What should I do after I've completed my data management plan?

Once you have produced a data management plan, it is recommended that you consult the Energy Data Best Practice guide to understand how you can leverage the value of your data across the wider energy industry, and beyond, by applying the Energy Data Taskforce principles and following its recommendations.

Feedback: Please send us feedback on this document and the process it describes. This is our first version and we anticipate revising this guidance to help future projects, based on your experiences of using it.

Process Overview

To support and overcome these challenges, ESC and KnowNow Information have built on the Energy Data Best Practice principles to create a data management canvas and workshop framework that will help the projects to implement and benefit from best practice guidelines.

Good data management processes make standard activities easier, such as reporting and business insight, but this benefit often conflicts with the upfront time required to set up the necessary processes. Projects are required to demonstrate their impact across a wide range of topics. By implementing a common set of principles, this process can be more efficient and effective.

Stage 1: Ultimately, all the data that is collected and will be to satisfy a set of known needs and objectives that belong to the project stakeholders. Understanding stakeholders will help to identify those needs, and to identify the start of the "**data threads**" that will run through the data management plan. The threads that are identified here will be used in the later stages of creating the plan.
Stage 2: In this stage you will identify the other end of the data thread, possible data sources and data collection mechanisms that might allow you to meet those needs and objectives. You will also need to think about the costs and risks involved in data collection, and how minimising your data will help to control those.
Stage 3: In this stage you will identify how the data will be stored and processed
Stage 4: Identify the organisational processes that will surround your data processing. This data governance is about the rules and processes that you set up to control and assure the data that you are using. The aim of Data Governance is to mitigate risks, comply with legal requirements, and make sure that you can adapt
Stage 5: This stage will help you to identify what data has value now, what data may have value in the future, and to balance cost against value.

You can find a canvas document in the appendix to guide your **data threads.**

In the canvas you will see the following colour coded sections to help you.

Tips – These are small pieces of information that provide additional help

Examples – These are situations from real world projects that the authors have worked on

Exercise – These are activities that it is highly recommended that you conduct to create your data management plan

1. Understanding Data Stakeholders

Goal: The goal of stage 1 is to understand your stakeholders and what their data needs are.

Ultimately, all the data that you collect and process during your project will be to satisfy a set of known needs and objectives that belong to the project stakeholders.

Understanding stakeholders will help to identify those needs, and to identify the start of the "data threads" that will run through your data management plan. The threads that you identify here will be used in the later stages of creating your plan.

1.1. Create your Stakeholder Personas

We start by creating a set of "personas" that describe the project stakeholders and what their ultimate needs and objectives are. Stakeholders are the people or organisations who are involved in, or impacted by, your project. You must understand what their role is, and what they care about.

Exercise: Use the **example personas**, and the **blank persona template**, to describe your data stakeholders and identify what they want to get from the project. You don't need to understand exactly what data they will need, that comes later. At this stage focus on their high-level goals and motivations.

Any project stakeholder is potentially a data stakeholder! Of all the stakeholders you have identified, some will have needs that require you to collect, process or provide data. Those stakeholders are *data stakeholders*. Some data stakeholders will be interested in the data itself, but others will only be interested in the outcomes that the data helps you to create.

Example:

As an electric vehicle charge point operator, I need to know when EVs are charging so that I can reduce the costs of procuring energy from my supplier.

Tip: **Bootstrapping vs business-as-usual**

Your data stakeholders can change as your project progresses. You should consider the stakeholders in your immediate project, as well as those who are involved longer-term. For instance, during an innovation project it might be important to a funder to prove the social value of a new service. Later, stakeholders might be more concerned with assessing the scalability and market demand for a product.

1.2. List your Data Threads

Based on your stakeholders, you should be able to identify a set of needs that can, or might, be met using data of some kind. Some needs will be shared between stakeholders.

Each need now becomes the start of a "data thread". Each thread links a stakeholder need to all of the other aspects of your data management plan. The rest of this data management canvas will help you to plan the collection, storage, governance and sustainability of the data threads that you have identified.

You can add, remove, split or consolidate threads later. In fact, if you are new to this data management canvas, you could start by considering just one or two threads to begin with and add more once you are comfortable with the process.

What's next?

By now, you should have an understanding of the different stakeholders involved in your project and some idea of the data threads that will run through your data management plan. You probably also have some ideas about the data you need to collect, how you need to process it, and some of the implications of doing so - but the rest of this guide is designed to help you consider those ideas in more detail.

2. Data Sources & Collection

Goal: The goal of stage 2 is to identify and describe sources of data that will meet your data needs.

In stage 1, you identified one end of your data threads: stakeholder needs and objectives In stage 2 you will identify the other end, possible data sources and data collection mechanisms that might allow you to meet those needs and objectives. You will also need to think about the costs and risks involved in data collection and how to size your data collection to help to control those.

2.1. What data could you collect or obtain?

There are potentially many sources of data that you could use during a project, sometimes even multiple different sources that would help to answer the same stakeholder need. In other cases, finding data that meets a particular objective (like demonstrating social value) might require you to think outside the box.

For each data thread, consider:

- What are some data sources we might use to satisfy the stakeholder's needs?
- Do we need to collect that data ourselves, or work with a third party to get access?

Exercise: One exercise that might help you to identify relevant data sources is "Data Tool Cards" Use a mix of the 'data' and 'inspiration' cards to help ideate possible data sources that might satisfy your stakeholder needs.

Tip: Don't forget that the research and proving stage of your project might be different to the ultimate "business as usual" that you envisage. For instance, if you are trying to show an impact on a specific group of people, you might need to speak to them to understand what benefits they want to see and how you need to design around their needs.

2.2. Quality, Costs & Risks

The quality and format of different data sources might help you to choose which ones are most appropriate for your project. For the different potential data sources, think about the following questions.

- What are your requirements for data quality (for instance accuracy, coverage and sensitivity)? And do any of your stakeholders need data that conforms to a particular standard or in a specific format?
- How would you check whether the data meet specific quality or standards criteria? How will you deal with data gaps?
- Have you made any assumptions about the quality, format or availability of that data? How might you test those assumptions?
- Do you need to pay for access to the data? Or will it cost money to collect that data either directly or indirectly?

Costs can arise because you need to pay a third-party for access to a dataset, or because there are costs incurred in collecting data. Collecting data could involve employing researchers, deploying sensors, or modifying software to collect additional telemetry. You may not know exact figures at this point but having an idea of the general costs involved in collecting data can help you to choose which data sources are most practical for you.

2.3. Obligations

Obligations can arise from contracts, data licences, or legal regulations. Common obligations include the conditions that are imposed when you obtain data from a third party (either under contract or via licensing conditions), or obligations to data subjects that arise from the GDPR when you process personal data. Personal data is any data about a living person that could lead to their reidentification.

Tip: The ways that you collect and use data in your project need to be ethical, which - on top of your legal obligations around personal data - includes making sure that you avoid discriminating. Use the Open Data Institute's Data Ethics Guide to help you consider ethical aspects of your data management.

2.4. Minimise your data (and costs, and risks)

Finally, you should think about how data collection can be minimised. Minimising data collection minimises the associated costs and risks. It is particularly important to minimise personal data because the law requires you to do so.

- Where can your threads "double up" on data collection to minimise collection costs and risks? For instance, can two or more threads all make use of a single data source? Having more than one reason to store and process data is better than one!
- What is the minimum amount of detail that you need to collect in each thread in order to serve the stakeholder's needs?
- If the data is personal data, could you accomplish the same task without using any personal data?
- If you have many customers or deployments, do you need to collect data from all of them or could you just collect a sample?

What's next?

Each data thread should now have a comprehensive list of data sources that you will need to fulfil the stakeholder's objectives you identified in stage 1. You should have a clear idea as to how you will collect the data and what your obligations with those data. In stage 3, you will consider the data storage requirements for each thread.

3. Data Storage

Goal: In stage 2 you identified where data could come from. In stage 3 you will identify how that data will be stored and processed.

3.1. What do you need to store?

You don't just need to store your 'raw' data, it is common to have some processing in between data collection and long-term data storage. In practice, most projects need to strike a balance between the richness of the data that they store and the costs and risks of doing so. The data that you collect as part of a project will always be part of an identified data thread.

It is important that there is a purpose behind your data storage as this is key to understanding the value of your data you are collecting and what you will need to store. Data that has multiple purposes is likely to have more value than data that has a single purpose. The number of purposes a datapoint has could help prioritise what is stored versus what is not.

Example: Time series data storage

There are some rules of thumb to help you decide what data to store when collecting time-series data for operational monitoring:

- Sample every 10-15 minutes
- Capture the average, minimum, maximum
- Only record delta changes
- State the parameters of collection so you can see where measurements may be incorrect

Make sure you align this to your stakeholders needs and data threads!

Tip: Use the threads you identified earlier to help you decide which data need to be stored; all your stored data should relate to at least one thread that you have identified. Could any of the data that you are collecting be processed prior to storage to reduce costs and risks?

Tip: What metadata will you need to store? Storing metadata such as the source, collection date and version of your datasets will help you to interpret it and manage it responsibly. Section 3 of the **Energy Data Best Practice Guide** includes more information about metadata. In the case where you are being provided data that you will not store yourself you may be able to simply store the metadata.

3.2. What are your storage requirements?

Some data is stored for many years for financial, regulatory or legal reasons. However, many data elements in projects aren't required for such long periods. Likewise, some datasets need to be accessed regularly, while others are accessed rarely (or never). The lifetime and access-pattern of your data assets will influence the type of storage that you choose.

Exercise

This exercise will help projects identify the requirements for their data storage. Answer the following four main questions when determining your data storage requirements:

- How much data is being stored?
- How long is it required for?
- How often will I need to access the data once it is stored?
- How fast does my storage need to respond?

3.3. Types of Storage

Use the following table to match the requirements from the previous exercise to a type of storage.

How much?		How Long for?		How Often?	
Big >1TB	Disk/Tape	**Forever**	Tape/Low-cost disk	**Virtually never**	Disk/Tape
Medium <1TB	Memory	**Year**	Disk	**Now and then**	Disk
Small <100GB	Memory	**Week**	Memory	**Frequently**	Memory

The majority of solutions are likely to use a cloud hosted solution. Depending on the cloud platform chosen it will be possible to choose a cost-effective data storage medium. Don't fall into the trap of thinking that cloud storage is cheap, and that it doesn't matter what storage medium you use. Different cloud technologies tend to use one of the "conventional" media under-the-hood. For instance:

Media Type	Cloud Equivalent
Memory	Amazon Elasticache, Redis, memcache
Disk	Amazon S3, Amazon EBS, OpenStack Swift, Azure Blob Storage
Tape	Amazon S3 Glacier, Azure Cool Blob Storage

3.4. How much risk of loss or corruption can we bear?

Different storage media have different cost implications and risks that need to be factored when deciding what data to store and where. The world of data has two considerations.

1. Is the data needed to deliver a service, complete a transaction and/or fulfil a contract?
2. Is there a compelling reason to keep the data for posterity? Examples of the second consideration could be record-keeping for regulatory compliance or creating long-term datasets to study external trends or business performance.

One risk inherent in storing data is that it can be corrupted or lost. Keeping backups can reduce the risks from data loss and corruption. How many times and how many versions of a backup are maintained should be dependent on the risk and importance of the data.

3.5. Data Backups

Follow the rule of three for your data backups:

- Have a minimum of 3 backups
- Locate your backups in 3 different places

- Have a minimum 3 generations of backups

There are two questions to ask when considering your backup scheme:

1. **What is the longest I can survive/operate without having the data available post an event that disables my system?**
   This is called the **Recovery Time Objective (RTO).** The quicker demand for recovery typically equates to a more expensive solution.
2. **What is the point data can be lost so it does not impact the service, contracts or safety?**
   Does the data have to be recovered from the point of failure? Or can any loss in data be recreated from other sources or does it not matter? Once that point in time is known, a suitable solution can be provided. This is called the **Recovery Point Objective (RPO)**.

Exercise

This exercise will help projects identify the type of storage media that is the best fit in terms of risk. Answer the following questions regarding each of your data sets:

- How long before I need that data in the event of a failure? (RTO)
- Can I do without that data? Or can it be recreated? (RPO)

3.6. Minimise your data storage

Is it practical to have a second by second or even a minute by minute view of events that could occur? Yes, it could be done, but consider whether this will increase the accuracy of a prediction or the quality of an outcome obtained.

Just good enough is best as the cost to be 'perfect' likely outweighs the marginal benefit obtained. Over time as the system beds in and the prediction algorithm improves as the amount of data generated becomes ever greater, this in itself is more important.

Example: Smart meters generally collect measurements on a 30-minute time threshold parameter. For a project collecting this it is 48 daily measurements and 48 opportunities to influence behaviour change. Over a 30-minute period change can be significant and incentives to shift behaviour or outcomes are most likely to be successful.

What's next?

At the completion of stage 3, you will have a set of technical requirements for what is stored, for what length of time and whether this data needs to have multiple copies. In stage 4, you will think about the organisational controls that go around your technical infrastructure.

4. Data Governance

Goal: The goal of stage 4 is to identify the organisational processes that will surround your data processing so that it is safe, secure, and legal.

In stage 4 you will put together a framework for governance. This will build on the stakeholders identified in stage 1, the purposes identified in stage 2, and the storage requirements from stage 3. Governance will ensure the project has ongoing sustainability and compliance.

Data Governance is about the rules and processes that you set up to control and assure the data that you are using. The aim of Data Governance is to mitigate risks, comply with legal requirements, and make sure that you can adapt. Recall in stage 2 that some data sources may have come with obligations to other parties - part of Data Governance is ensuring that those obligations are met.

Tip: Documentation is an important aspect of data governance. Many of your governance decisions will be translated into documents like policies and contracts. You will not write all of those documents now, but at the end of this stage you should identify what documentation you will need to produce and have started to create lists of rules and processes that will form that documentation.

4.1. Which rules apply to your data?

Governance needs to identify the rules that apply to your data and put in place a process to make sure that those rules are obeyed. Rules can come from various places:

4.1.1. Personal Data / GDPR

If you process any personal data, it is the law that you comply with data protection rules. Personal data is any data that relates to an identifiable living individual. You should:

- Identify if and where you are processing personal data
- Identify a "lawful basis" for each purpose that you are using personal data - this will depend on which stakeholder needs it meets. If you cannot find a reasonable legal basis, you should not collect or process the data.
- Identify the data controllers for each dataset - a data controller is a party that makes decisions about how data is processed.
- Identify any data processors for each dataset - a data processor is a party that processes data on behalf of a data controller. This will include any third-party storage or compute providers.
- Identify the documentation that you need to produce. At a minimum, you will need to produce a public "privacy notice" that explains why and how you are processing personal data. If you are using the "consent" basis then you will also need to create specific notices for getting consent to each purpose from the data subjects.

Tip: The GDPR requires you to find a Lawful Basis when you process personal data. Find out more about the Lawful bases on the ICO website

4.1.2. Licensing Conditions

Data that you obtain from third parties often comes with licensing restrictions - rules about how it must be used. For instance, some datasets cannot be passed on to third-parties or can only be used for specific purposes.

- These restrictions are enforced via contract and intellectual property laws, and so they could be almost anything.

- You should identify any datasets that are coming from third parties and check the licences or contractual terms that have been applied.

4.1.3. Funding Conditions

Your project funders might have imposed restrictions on how data must be used. Like licensing conditions, these could be almost anything, so check the terms of your funding agreements. Common conditions include:

- Seeking ethical review of your data handling. It is common to require external oversight of data handling, especially if your project is funded by a scientific research council. You might need to appoint some external ethical advisors or seek assistance from an established body like a university ethical review board.
- Making certain data available. Some funding aims to create datasets that can be reused, so you might be required to publish your datasets during, or at the end of, the project. Such a requirement should take into account your other obligations - for instance, you should never be required to publish personal data!

4.1.4. Audit and Accounting

Your project or business might need to account for the value of your data assets, or to demonstrate that your processes comply with the law. Internal audit is especially common in the public sector, and if you are processing personal data then the GDPR requires that you document how your processes meet the legal requirements. The different organisations involved in your project might also have requirements that one another are audited, especially where they are sharing personal or sensitive data between one another.

- Do you have any explicit internal audit requirements?
- Do any of the project partners require other partners to be audited for things like GDPR compliance?
- Do you need to estimate the value of your data assets for accounting purposes?
- Would external audit of your data management add value by reducing risk or demonstrating to customers or investors that you are handling data properly?

Exercise

In this exercise you will identify the rules you will need to abide by with your data collection. For each data thread, make a (high-level) list of the rules you need to obey. You can use the **Governance Checklist Template.** Later, those rules should form the basis of your more detailed governance arrangements; and you can use them as a checklist to validate your governance arrangements.

4.2. Responsibility & Due Process

Once you understand the rules that your data management needs to obey, you will need to establish the organisational controls that comply with those rules, and which ensure that compliance continues into the future.

The process of governance is based on a set of principles that are owned by a community, which collectively undertakes to be a steward of the principles. We have identified some key questions for you to answer. You can use the **Responsibility & Due Process Template** to record your answers to these questions.

4.2.1. Steward & Ownership

Organisationally, who will take ownership and responsibility of governance?

Could you convene a "data board" or steering group?

4.2.2. Principles

How could you describe the core values that describe your project's approach to data? For instance, is it important to you to be a fair or ethical data handler?

What are the rules of thumb for handling data in your project? Is the value of the data understood? Refer back to your stakeholder needs.

4.2.3. Collaboration

Which partners need to work together for your data management to be effective?

How will multiple data owners and data cultures work together to deliver your plan?

4.2.4. Transparency

Which elements of your governance processes and data management plan need to be communicated to other stakeholders, like your customers, funders, or the general public?

How will you make that information available to them?

4.2.5. Process

What steps will you document and follow when handling data?

4.2.6. Tools

What tools will support your data governance process?

4.3. Controlling Access and Use

Access to data, and use of data, are likely to form part of your governance. As well as the technical measures to control access to data sets, you need to identify a process for deciding who gets access to each dataset. The use of data is distinct from access - a party that has access to a dataset for one purpose probably cannot be prevented by technical means from using that data for a different purpose, but your governance rules should be clear about why each party is using data and provide some assurance that it is not being used for anything else.

Access control:

- How will access be controlled technically?
- How will the integrity of those technical controls be assured? For instance, what requirements will you have around passwords, penetration testing, and software updates?
- How will the project ensure that those controls are in place and maintained? For instance, will there be a contractual agreement of some kind? Could there be an audit process to check that different parties are implementing controls correctly and as agreed?

Usage control:

- How will parties be held accountable for their data use? For instance, will there be contractual restrictions in place, and will parties' data use be audited?

Exercise

In this exercise you will identify how you will need to give access to different users. For each data thread identify the parties that will have access to it. Write down details about:

- What access they'll need to the project's data,
- What usage they should be making of the data, and
- The ways that each organisation will be made organisationally and legally accountable for their data handling.

4.4. Publishing & Re-use

The Energy Data Best Practice Guidance (EDBPG) encourages you to publish data in a way that is Open and Discoverable, so that others can re-use it.

Your governance process needs to incorporate a process for identifying, triaging and publishing data. How will you follow the EDBPG? Who will take responsibility for those processes?

4.5. Response Process

Part of governance is responding to external events. You should know the process for dealing with foreseen and unforeseen problems such as security breaches, data protection breaches, data loss, or any specific contractual rights that could be exercised by a partner or external party.

Consider:

- What external events can you foresee that might need you to respond?
- How will you respond to those events?
- How will you respond to unforeseen events that impact on your data management? Who will take responsibility for co-ordinating a response across partners, and how will that work?

Typical events that you will need to consider:

- Unauthorised access to data through organisational or technical failure (security breach)
- Accidental or malicious disclosure of personal data (GDPR breach)
- Hardware or software failure that results in data loss or unavailability
- Complaints about personal data handling (GDPR)
- Subject access requests or objections to processing of personal data (GDPR)
- Suspected misuse or mishandling of data by a project partner

Exercise:

Draw out the flow for responding to a subject access request from a person that your project holds personal data about. Be sure to include how you will identify data about that person, who will coordinate the response among your project partners, and who will need to be involved.

Extend your flow to include an objection to processing from the same data subject. What project partners will need to be involved, what actions do they need to take, and how is your response coordinated?

Example:

A participant in a heat-as-a-service trial would like to leave the project and asks to be forgotten. There are three project partners, all who hold personally identifiable data about the participant. The project partner that is the point of contact consults their data catalogue and finds the data controller for the other two partners. They have established that an email is sufficient to raise a right to be forgotten request. Each of the data controllers reviews where that participant's data is being held and deletes it, confirming to the point of contact that they have done so before they confirm to the trial participant.

4.6. Review Process

Stakeholders, objectives, processes and the law can all change, so an important aspect of any data management plan is to make arrangements for ongoing review, especially if the financial returns and wider societal impacts are not meeting expectations.

The review process is an aspect of governance. You could create a 'decision support' portal to help you review your data governance arrangements as your project progresses. This could even be modelled, and what-if scenario planning is provided as a function. This portal will require ongoing curation and guardianship to maintain relevance, trust and usefulness.

Exercise

In this exercise you will look at your review process to keep your data management up to date with the other partners on a project. For your data management plan, answer the following:

- Which partners will take responsibility for reviewing the data management plan?
- How will those partners co-ordinate?
- How often will that review take place?
- Where concerns exist or new risks are identified, how will they be raised?
- Will partners be audited with respect to their data use and adherence to the data management plan?

What's next?

By now, you should have identified the rules that your project needs to follow for individual data threads and how they meet stakeholder needs. In stage 5, you will think about how to make your data processing sustainable into the future.

5. Data Sustainability

Goal: In stage 5 you will identify where the data has value through its lifecycle and how your data management fits within a sustainable business model.

Stage 5 builds on the foundations of the previous stages and could reinforce or question why certain data threads are required.

The data that allows a system to operate and deliver value in an ongoing, safe and secure manner has some intrinsic sustainability to the business or project when it is fit for purpose. Over time requirements of data collected can change resulting in the data no longer fit for purpose.

It is necessary for an ongoing data management approach that puts data sustainability at the forefront. Not all data elements are equal and understanding the purpose of the data will be key. This stage will help you to identify what data has value now, and in the future and to balance cost against value.

5.1. Costs & Cost Recovery

A key driver of being sustainable is cost management. Ever expanding disk/storage is not necessarily a desirable requirement. Technology requires management and the first consideration is to contain and reduce cost as much as possible. The rules here need to be strict from day one: the designers need to justify the data that they want to keep and the length of time that they want to keep it.

A technique to help contain data management overheads is to assume that all data will eventually be archived, which means that the cost will be finite. Capturing useful and compelling information about the project could mean that some data is actually useful for a long time.

The value of conducting trend analysis could well see small yearly increments grow over time, causing some datasets to become large. The challenge here is ensuring that this data is stored appropriately and that completed reports and analysis are made available in an open and transparent way.

5.2. Value Creation

Understanding where value is accrued with data depends on the assessor of value and what they determine to do with this. To help measure value, this stage can be used to identify where this data has an impact by providing a worked example. Once the determined measurement of value is known a number can then be attributed to it.

Something to bear in mind is that data will change value over time and that is impacted by its use case. Understanding the role of each data element over time will help projects put a suitable value on the data they generate. You can refer back to your stakeholder needs from stage 1.

The stakeholders will value data that is just in time and relatively accurate. The challenge is that most actors involved may all be working to different time thresholds.

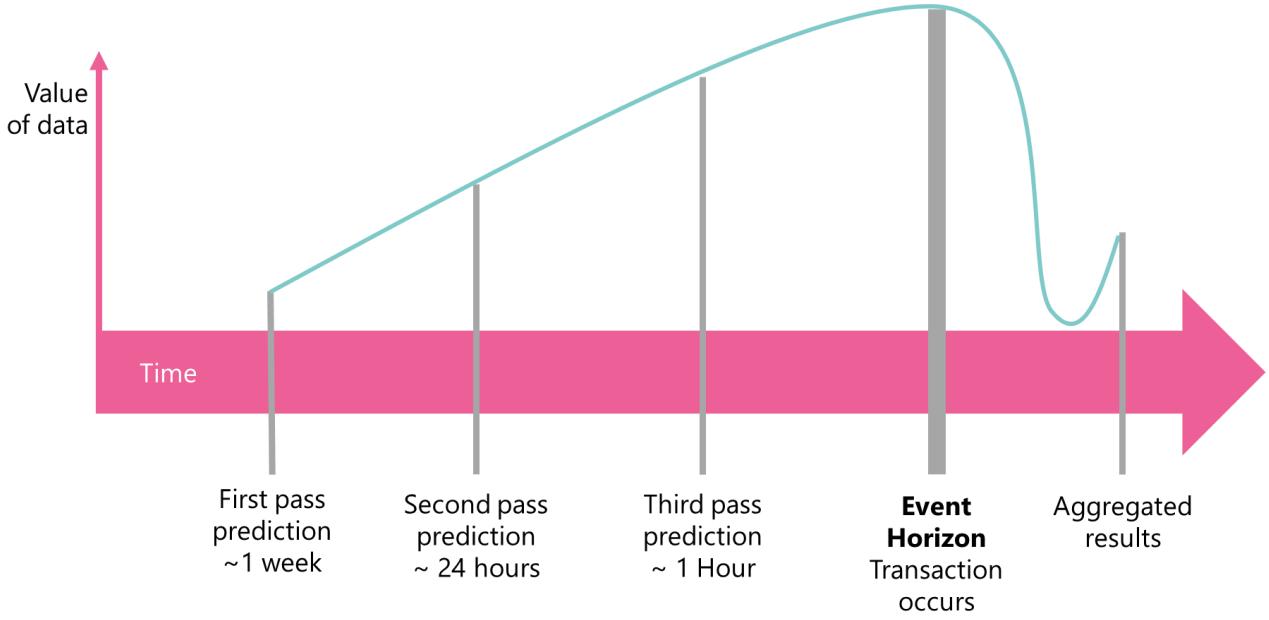
The time threshold is defined as the amount of time it takes to change behaviour. Perhaps that behaviour is to take advantage of the energy being created, or the price being offered to persuade a user to consume or not to consume, to generate or even store energy.

For different actors, the motivations and the thresholds are all different. Depending on the season, the weather (both here and now as well as forecasted) and the prevailing conditions of the users all impact the time at which the behaviour will be fixed with no change in outcome.

Example

This is a worked example which involves a using predictions of energy usage to help make decisions on whether to buy, sell or store energy.

The graphic below describes the relative value of data over time before the 'event horizon'. This is the point at which the data from a number of sources goes from high value to minimal or no value.



Thinking about this can help projects identify what data is useful and when as data can be normalised across time. There are three key moments in time for this worked example.

1. The time before the event horizon. This can be sub-divided into a number of predictions.
2. The actual event horizon.
3. After the event horizon, where amalgamated data provides evidence for trends and which can help make predictions more accurate.

In this example, a generator makes a prediction 1 week out, which gives some confidence over the pending weather forecast (important for solar and wind-based energy generation). The weather also drives patterns of user behaviour and therefore demand. This first prediction could be used to influence user behaviour by offering incentives, for example a financial incentive to maximise the gain for the generator, storage party, or the consumer. The prediction has some value to those parties.

The second data pass is 24 hours out and the weather forecast more accurate. Consumer's plans are a little less fluid with commitments made. The ability to influence is reduced, but the overall value of the data is greater. This is because the 24-hour prediction is signalling the intent of the plans to be exercised by the actors involved. There is greater value because with more knowledge and less variability in the expected outcome there is less risk. Resources can, therefore, start to be planned to be available, for example for storage, generation or consumption.

The final prediction is very short term. This could be anywhere from an hour to minutes in advance. The shorter the time to go to the event horizon the less likely an actor can react as it will take time to take alternative action. Some of the lag between the decision to change to actually making a change can be facilitated by automation. However, when people are the key actors they struggle to respond outside of a ten-minute window.

At the event horizon the energy generation, storage, release, consume moment occurs. No matter how many predictions are made the result will have zero impact. The time threshold for an event horizon will be the time it takes for a transaction event to occur.

After the event horizon, the value of data accrues as the previous trade can be added to the prediction algorithm to help assist a more accurate prediction for the next trade. The larger the dataset to drive a prediction the more valuable the whole dataset is. This is because the prediction will grow in accuracy and be able to influence the outcome expected (partly because the mass of evidence says that outcome will occur).

Exercise

In this exercise you will think about the factors that affect your project and what data will provide value *before* the event horizon. When thinking about the value of data as the event horizon approaches, consider the data elements that have more value as they get closer to this point in time.

You should consider the following questions:

- What is the forecasted behaviour of the elements of the system?
- Will this impact how you consume a service?
- Will this impact how you supply a service?
- What will the likely operations be during the time period?
- What are the opportunities for influencing decision making?
- What is the risk that this forecast will be incorrect? How accurate is it? What is the likelihood of the forecast changing?
- What is the optimum outcome to be achieved here?

Exercise

In this exercise you will think about the factors that affect your project and what data will provide value *after* the event horizon. Ask each stakeholder involved in delivering the project to undertake a data journey. This journey can be captured as a role play, animation, storyboards as the data moves from creation to processing to a post-processing.

- Who else might be able use this data now that the primary purpose has come to an end?
- Now that it is no longer required can this data be archived with a data trust or repository?
- Who takes ownership of the data?

The end result will be a recognition that other actors and stakeholders may have a use for the data created and the value of data can start to be validated.

What's next?

After stage 5 is complete the project will have justified why certain data is required based on the value that different data provides.

6. Conclusions & Next Steps

Bringing together all the stages will mean your project will have:

- Stage 1 - Known stakeholders that each have a use case involved with the system
- Stage 2 - Identified the data sources and purpose identified
- Stage 3 - Fit for purpose data storage requirements
- Stage 4 – Produced a governance process and principles
- Stage 5 - Maximise the value of the data available

Following the process in this guide should have helped you to consider the value and lifecycle of data within your project, as well as the organisational controls and processes you need to put in place to manage it safely, legally, and effectively.

By now, you probably have a better idea of the work you need to do - defining your policies, perhaps convening a data management board, and going through the process again to define more data threads, or to consider those that you have started in greater detail.

Treat your data management plan as a living process and keep revisiting it. Make data management something that you discuss with all your project partners and use this guide to help structure those conversations if you need to. Often good data management can save you huge amounts of time and energy, particularly during a project closedown when questions are asked about impact and success.

7. Appendix

7.1. Process Overview - Data Threads

A high-level process outline that can be used to understand the process and elements of each data thread and keep notes of key points, realisations and “to-dos”.



data-thread.pdf

7.2. Stage 1 – Personas



persona-template-blank.pdf



persona-owdc.pdf



persona-ukri.pdf

7.3. Stage 4 - Responsibility and Due Process Template



org-controls.pdf

7.4. Stage 4 – Governance list



governance-list.pdf

LICENCE/DISCLAIMER

Energy Systems Catapult (ESC) Limited Licence for Energy Data Management Canvas

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 Energy Data Management Canvas, 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

© 2021 Energy Systems Catapult

Energy Data Management Canvas

Stakeholder Needs & Outcomes

Understand your stakeholders and what their data needs are.

Data Sources & Collection

Identify and describe sources of data that will meet your data needs.

Storage

Identify how data will be stored.

Governance

Identify the organisational processes that will surround your data processing so that it's safe, secure, and legal.

Sustainability

Identify where the data has value through its lifecycle and how your data management fits within a sustainable business.

Project

Date

Stage 1



What's the stakeholder's name?

How would this stakehold explain their role in the project?

What are this stakeholder's objectives, what do they hope to achieve through the project?

How might this stakeholder describe their mission?

What are some questions this stakeholder might ask?

Stage 2

What data could help this stakeholder to meet their objectives or answer their questions?

Project

Date

Stage 1



What's the stakeholder's name?

Outer Wessex District Council

How might this stakeholder describe their mission?

Delivering public services that promote a greener, healthier and happier outer Wessex.

How would this stakehold explain their role in the project?

OWDC owns a significant stock of social rented housing. OWDC will recruit and communicate with households, and contract a provider to install equipment in participating homes.

What are this stakeholder's objectives, what do they hope to achieve through the project?

OWDC hopes that the project will identify ways that energy data can help tenants to save money on their energy bills, and measure the impact of council-funded energy efficiency measures such as insulation.

What are some questions this stakeholder might ask?

How much have our tenants spent on energy?
Which energy- and money-saving interventions are most cost-effective?
Can this project scale-up to our whole housing portfolio?

Stage 2

What data could help this stakeholder to meet their objectives or answer their questions?

Impact statements from project households

Weekly energy spend data from intervention + control households

Installation time/cost records from contractors

Project

Outer Wessex Says Yes to Energy Saving (OWYES)

Date

June 2020

Stage 1



What's the stakeholder's name?

Innovate UK

How would this stakehold explain their role in the project?

To provide funding, and ensure that the funding is spent effectively and as agreed by the project consortium .

What are this stakeholder's objectives, what do they hope to achieve through the project?

InnovateUK hopes that the consortium will develop and prepare to exploit commerically-viable technology.

How might this stakeholder describe their mission?

Work with our partners to ensure that world-leading research and innovation continues to grow and flourish in the UK

What are some questions this stakeholder might ask?

Has this project spent money as agreed?
Are the ideas and technology investable?
How has the project met the evaluation criteria of carbon reduction, job creation and air quality improvements?

Stage 2

What data could help this stakeholder to meet their objectives or answer their questions?

Aggregate household spend data

Aggregated carbon emission data (hourly energy use + grid carbon intensity?)

Impact statements from households

Project partner spending reports, proof of equipment installation

Project

Outer Wessex Says Yes to Energy Saving (OWYES)

Date

June 2020

Responsibility & Due Process

Stewardship & Ownership

Principles

Collaboration

Transparency

Process

Tools

Project

Date

Governance Checklist

What rules does your project need to comply with? Complete the list.

Thread:

GDPR Privacy Notice

GDPR Personal Data Inventory

Licensing Conditions?

Funding Conditions?

Project

Date