Energy Data Visibility

Discovery report

Office for National Statistics

June 2020

Hippo Digital Ltd

Point of contact: Seb Greenwood / Katie Lambeth-Mansell

Email: seb@hippodigital.co.uk / katie@hippodigital.co.uk

Contents

Executive summary

Why we carried out this Discovery

This Discovery is part of a stream of work under the Modernising Energy Data programme. The Energy Data Task Force report's recommendation 3 outlined how the energy sector's data needs to be more open. In response to that recommendation we agreed to investigate the following in our discovery:

This Discovery focused on recommendation 3 with the objective of evidencing and identifying recommendations to take forward into an Alpha to test how best to make data "*discoverable, searchable, and understandable*".

Our approach to Discovery

The Hippo Digital and ONS team followed an agile based approach to the delivery and worked collaboratively as one team to support delivery of the project.

We conducted a 6 week Discovery consisting of:

- Gaining insights from 64 individuals through a survey and interviews
  - 21 in-depth interview sessions, which included:
    - individuals who were either consumers, providers or both
    - Individuals who were at various seniority levels within their organisations
    - individuals who were at varied levels of energy system data understanding
    - individuals who considered themselves as data specialists and non data specialists
- Conducting a Survey with 56 respondents
  - Objectives were to:
    - understand how ONS are viewed by the UK energy industry
    - understand what data people are using, how they access the data and what is most important about it
    - explore awareness of existing catalogs and if these are meeting the users needs
    - explore if a new data catalog is perceived to be the right thing and who would manage this
    - understand how people define the terms 'Understandable' 'Discoverable' and 'Searchable'
- Conducting high level reviews of digitalisation strategies published by the regulated networks

● Conducting high level reviews of open data sources from organisations

Key research findings

User needs

Persona	Need
As a Data Consumer I need	To find data that is right for the task so that I can complete my task at hand and deliver value to my organisation.
	To access data sets that are relevant to me
	To have the ability to request access to data putting forward a business justification.
As a Data Provider I need	To understand how my data is being used and to ensure that complex data is not misinterpreted and misused.
	Leadership, guidance and collaboration on the journey to open data
	Value for money for the end Consumers of energy for any investment made.
	To control and manage access to sensitive data as a means of addressing security risks
	A consistent triage process across similar organisations to understand what data is and isn't "open" and consistency in sharing
	An approach to standards, architecture, and governance that doesn't present significant operational or financial overhead.
	Better innovation investment structures to drive investment in open data
As an Overseer of the Service I need	A transparent set of Key Performance Indicators to show movement towards desired sharing outcomes.

What's preventing user needs from being met?

Our research found that users were looking to find and access the right data for the task to bring value to their organisations. We heard from users that today data is difficult to discover, search and understand, verifying the findings from the EDTF report. This is caused by the data being aggregated, defined and presented in a way that reflects the organisational silos and complexities of the energy sector. The same or similar data is also available in many locations. In addition, there is data that is either not collected or not shared.

The complexity of the landscape means that consumers of data have a very steep learning curve, and investment of time, to find the data they need, often relying on experience and conversation. Importantly users don't see discovering data as separate from accessing data. They wish to have the least friction across the entire journey.

Our research indicated that organisations were in different states of maturity and readiness for publishing data following the presumed open principles. A number of barriers in place that prevented data sharing such as: the culture and skills within data provider's organisations, fears around the misuse of data, the level of infrastructure investment and the funding mechanisms needed. Organisations were also concerned about delivering value to end consumers of electricity and felt that an iterative approach to delivery would be most appropriate.

We heard of projects within organisations which have been initiated to overcome these barriers. These included culture change initiatives and working with open Data experts like the ODI to work towards open data. We heard that both telling good news stories and the process of making the currently published data more visible enabling comparison across organisations were levers to drive more data sharing.

In the future, the increasing decentralisation of the energy sector will lead to an increasing number of data sources that will call for mature data management across the sector. These data sources will need to be better understood to enable effective regulation of the system and unlock innovation. This makes the problems outlined above very important problems to solve now.

Summary of recommendations

From the Discovery's research and analysis we have the recommendation that the project should proceed to Alpha, prototyping a service that will:

- Make existing public energy sector data discoverable, searchable and understandable
- Facilitate collaboration amongst data providers and consumers
- Seek to validate the hypothesis that the service can act as a lever to expand the volume of energy sector data that is discoverable, searchable and understandable

Recommendation	High level detail	Discovery theme
Start with public data	We recommend that the service's initial objective is to have all data that is currently publicly available searchable by a "base" set of standard dataset attributes, in order to prove that this will result in the public data being more discoverable, searchable, and understandable. Test the broader assumption that the improved understanding of the data landscape and comparability across providers driven by this being situated within a usable and collaboration-focused service will establish the basis of the market pull that is envisioned as a broader function of the service.	Discoverable Searchable
Start with a simple set of attributes and evolve	We recommend that datasets have a set of characteristics that support consumers to make the decision as to whether they are right for the task. Datasets should be described and categorised in a meaningful way to support that activity	Discoverable Searchable Understandable
Business glossary	We recommend a common set of business descriptions need to be developed to enable the datasets to be described consistently	Searchable Understandable
Visibility and access	Data consumers do not view visibility and access as separate concerns. Therefore, work in Alpha will need to understand how the end-to-end service can enable this access, and develop the right security, architectural, and governance foundations to enable this.	Searchable
Evolutionary and collaborative approach	We recommend structures need to be in place for organisations to collaborate across the industry to address complex concerns surrounding areas such as standards, security, architecture, and governance. Existing efforts towards collaboration being driven by bodies such as the ENA can be leveraged to this end.	Service Condition Lever

Problem landscape

A great deal of work has previously been done in order to help the energy industry understand what is needed in order to modernise energy data, and this Discovery is part of a stream of work under the Modernising Energy Data programme.

The Energy Data Taskforce (EDTF) report carried out extensive work, and found that there are a number of problems holding the modernising of energy data back:
- Discoverability of data that is already available is poor
- A lack of common data standards for interoperability hinders efforts to make data more open
- Existing data is not enough to meet the needs of data consumers as it's not collected in a granular manner by modern standards
- Data is often not simple to understand and use
- Electricity and gas data is treated as separate while achieving largely similar outcomes

These problems with energy data contribute to preventing a full picture of the existing energy sector from being formed and understood.

However, these problems will be exasperated in the future as it is expected that data outside the remit of existing energy sector parties will account for the far greater share of growth, and managing the challenges of increasing reliance on ungoverned data will be a key theme of modern energy data.

The Energy Data Task Force report's recommendation 3 outlined how the energy sector's data needs to be more open, and this Discovery focused on the objective of evidencing and identifying recommendations to take forward into an Alpha to test how best user needs may be met to make data "*discoverable, searchable and understandable.*"

Research around improving data visibility in the energy sector, which was completed under a previous initiative by Public, found the following themes on open data, and this Discovery's research and analysis has provided further insights and evidence around these:
- The need for common standards and taxonomy
- The role of regulation and compliance
- The need for organisations to retain ownership of their data
- The need for the service to support people in bringing data sets together to form new insights
- The need for data to be free and built in a modular way

hippodigital

To help this Discovery begin to understand the problem landscape we reviewed the previous research and analysis, and collated the following high level views of the problems the discovery may look to validate.

Energy data today	- Is aggregated and defined in very context-specific ways - Reflects organisational silos instead of the underlying value of the data - Is difficult to discover, search for, and understand - Is collected but not shared, or not collected at all - Has a lot of perceived and tangible barriers to sharing data
Energy data in the future will	- Have a large number of new data sources - Be increasingly decentralised - Require better data management and visibility - Need to enable effective regulation and innovation

Our approach to Discovery

The Hippo Digital and ONS team followed an agile based approach to the delivery and worked collaboratively as one team to support delivery of the project.

The plan below set out our approach to delivery -

Sprint 1	Sprint 2	Sprint 3	Sprint 4	Sprint 5	Sprint 6
Inception Set conditions for success	**Research, analysis & options** Understand the problem and how it can be solved			**Evaluate & agree** Assess how possible solutions deliver value and meet need	**Conclude** Discovery findings

Our approach to the delivery of Discovery was also aligned to the GDS service manual as described below.

Defining the problem: Problem statement

In order to define the problem we framed the challenge within a problem statement. It is not uncommon within a Discovery phase to be presented with an assumed solution(s). Our role was to interrogate the solution and define the problem we're going to be working on. We developed and agreed on the problem statement (outlined above) which supported the team to understand where to focus our research and analysis. Allowing us the scope to provide strong recommendations at the end of the Discovery phase.

A problem statement is a clear concise description of the issue(s) that need(s) to be addressed by a problem solving team. It is used to center and focus the team at the beginning, keep the team on track during the effort, and is used to validate that the effort delivered an outcome that solves the problem statement.

Problem statement format should be made up of:

- the **specific problem** the team is addressing
- the **result** of the problem
- the change of situation which means the problem becomes a higher risk (CATALYST)
- the risk that's a result of the problem and change of situation (RISK)

SPECIFIC PROBLEM:
The energy sector's data needs to be more open, defined as the data being "Discoverable, Searchable, Understandable".

RESULT OF PROBLEM:
However how the energy sector can achieve this is unclear.

CATALYST:
Open data across the energy sector is seen as a key enabler to meeting net zero ambitions, delivering a more competitive energy market, regulating a decentralised energy model, along with the intermediate ambitions of building the collaboration, knowledge and skills to establish an open data culture in the UK.

RISK:
During this Discovery, we (the Energy Data Visibility Discovery) will evidence what energy sector organisations understand by data being discoverable, searchable and understandable, and what is currently impeding the energy sector in successfully sharing data openly.

Using this evidence we will identify potential systems, conditions and levers that could be introduced to influence the sharing of data openly, and what needs to be in place to make data discoverable, searchable, and understandable to data consumers."

Understanding users and their context: User mapping

We mapped out our users and defined objectives for our user research. We wanted to understand consumers' data needs, providers' motivations and blockers with regards to data sharing, and the complexities in the current landscape of discovering, searching for, and understanding energy data.

Understanding constraints

Within the Discovery phase we identified assumed constraints and conducted our research to understand and evidence these (refer to research plan). Below is a list of some which we provide further detail on later in the report -

- Technical
- Financial
- Cultural
- Regulatory

Understanding constraints supports identifying possible risks within the Alpha phase of the project and opportunities of mitigation.

Next steps: Using insights to shape recommendations

The research within the Discovery phase allowed us to gain valuable insights. These have shaped a variety of the recommendations and improvements which can be tested within the Alpha phase.

Sharing knowledge and insights: Show and tells

The team began the project by embedding GDS design principles, working in the open and transparently.

The team presented show & tells throughout the Discovery enabling stakeholders to understand and collaborate on the project. Links to the team's collaboration tool (Miro boards) have been available to all stakeholders who were interested in the team's work.

Creating a method for delivery: Flexible multi-disciplined team

As outlined in the GDS service manual, the team was made up of a variety of disciplines to support and deliver the project.

Business Analyst
Helen Walker

User Researcher
Fern Williams

Service Designer
Ameya Krishnamoorthy

Delivery Manager
Temoor Malik

Senior Digital Specialist
Seb Greenwood

Senior Solutions Architect
Colin Bull

Assistant Project Manager
Elinor Phillips

Project Lead
Lee Hood

Project Manager
Richard Sayer

Link for reference to GDS Discovery phase -
https://www.gov.uk/service-manual/agile-delivery/how-the-discovery-phase-works

Research, objectives and methodology

In total, the Discovery gained insights from 64 individuals through a survey and interviews. Initial contacts for research were provided by the Department for Business, Energy and Industrial Strategy. Following this, participants were chosen from our participant backlog built from the research sign-up in the survey . We gained permission to share participants' contact details for research in the wider Modernising Energy Data programme.

Gathering qualitative data: In-depth interviews

We carried out in-depth interviews with 21 people from a number of organisations with a range of roles (see table below for more detailed information about our users). Initially when segmenting users, we were thinking about where they were in the energy process e.g. generation, transmission and whether they were predominantly a provider of data or consumer of data. All interviews were conducted remote moderated via Google Hangouts or Zoom.

The objectives of the in-depth interviews were to:
- Understand the current landscape for data, what data do organisations currently use and how do they access that data
- Are there any barriers to access
- Understand motivations and levers for sharing and making data visible
- Understand the benefits and challenges to making data visible

Analysis

For each interview we took notes during the session and used the otter.ai platform to transcribe our sessions in order to support our analysis. We used the Miro platform to take notes during research and to conduct the affinity sorting exercise. After theming our insights, we created our personas and user journeys, before applying our assumptions to the wider themes of discoverable, understandable and searchable.
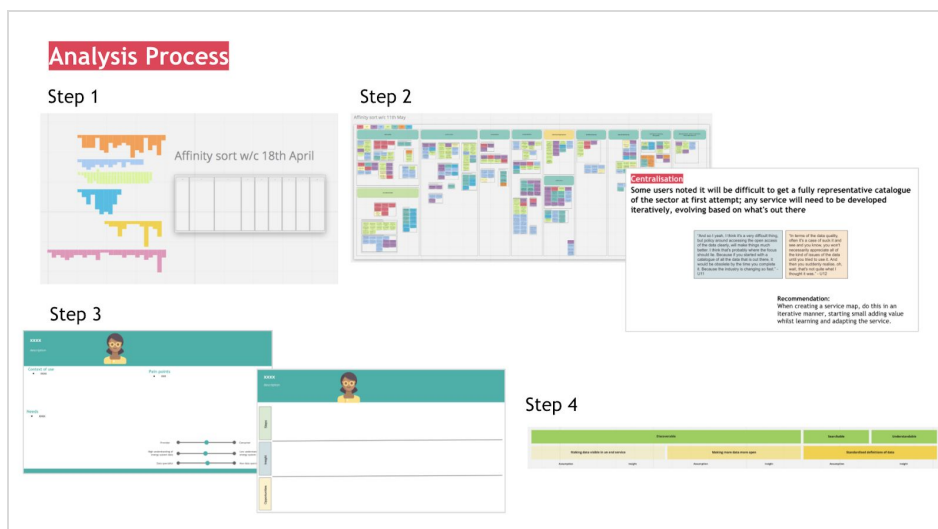
Gathering qualitative and quantitative data: Survey

We carried out a survey via Survey Monkey which was completed by 56 respondents.

The objectives of the survey were to:

- Understand how ONS are viewed by the UK energy industry
- Begin to understand what data people are using, how they access the data and what is most important about it
- Explore awareness of existing catalogs and understand if these are meeting the users needs
- Explore if a new data catalog is perceived to be the right thing and who would manage this
- Understand how people define the terms 'Understandable' 'Discoverable' and 'Searchable'

Please see the attached research plan (separate file) for more detailed information.

Adding context: Background research

We reviewed the digitalisation strategies published by the regulated networks to understand what work was underway towards presumed open data particularly that could help overcome any barriers that had been identified.

We also reviewed the data sources that the organisations had already published openly.

It should be noted that the list of initiatives will not be an exhaustive list as it is based on a limited number of participants and is limited by the level of detail presented within the strategy documentation.

hippodigital

Findings

Summary of user needs

This table outlines the high level types of users and their needs.

Persona	Need
As a Data Consumer I need	To find data that is right for the task so that I can complete my task at hand and deliver value to my organisation.
	To access datasets that are relevant to me
	To have the ability to request access to data putting forward a business justification.
As a Data Provider I need	To understand how my data is being used and to ensure that complex data is not misinterpreted and misused.
	Leadership, guidance and collaboration on the journey to open data
	Value for money for the end Consumers of energy for any investment made.
	To control and manage access to sensitive data as a means of addressing security risks
	A consistent triage process across similar organisations to understand what data is and isn't "open" and consistency in sharing
	An approach to standards, architecture, and governance that doesn't present significant operational or financial overhead.
	Better innovation investment structures to drive investment in open data
As an Overseer of the Service I need	A transparent set of Key Performance Indicators to show movement towards desired sharing outcomes.

Here is a breakdown of all of the personas, for more information see the personas and user journeys attached document (separate file).

Oversight
Measures whether the desired outcomes are being achieved by the regulation and policy levers

Energy Statistician
Works in a government department and is responsible for using data from a number of sources to help policy makers understand progress and make new decisions and policy.

Economic Statistician
Energy data sets are one of many used to enable short term economic output indicators (ie gross domestic product (GDPO) to be calculated on a monthly basis

Academic / Researcher
Uses operational data from across public, commercial, and open-source data sources to better model renewable energy out-turn and move the UK to decarbonisation faster

Data Provider
Are on the journey of making data from their organisation open, attempting to strike a balance between the risks, benefits, and costs of open data.

Trade
Manages the short term trading team looking at the buying and selling of gas and power in short term timescales.

Commercial manager
Works for a large company and is involved in selling low carbon systems to corporate and public sector clients

High-level insights

Our discovery undercovered a number of insights which we've grouped into the following themes .

- Right data for the task
    - Discoverable and understandable
    - Searchable
    - Complexity
    - Access to data
- Conditions and systems
    - Collaboration across providers and consumers
    - Industry collaboration
    - Incremental improvements
    - Overseeing the service
- Benefits of data sharing
    - Better decision making
    - Driving Innovation
- Levers for data sharing
    - Visibility of data
    - Sharing stories of success and best practice
    - Incentivisation versus Regulation
    - Political Pressure
    - Restructuring/Breaking up organisations
- Barriers and costs of data sharing
    - Organisational culture
    - Technical capability, governance, and data quality
    - Concerns around security and use of data
    - Concerns surrounding privacy are a barrier to opening some data up
    - Lack of understanding and alignment around what "open data" means
    - Costs of standardisation
    - There is little room in regulated budgets to innovate

Below are the details of the themes and their insights:

Right data for the task

The key needs for Consumers of data were to be able to:

> *"find the data that was right for the task and gain access to that data so that value could be delivered to their organisation."*

This section of insights includes what we learnt in relation to this covering: what discoverable and understandable mean to users along with more functional needs covered by searchable.

We have also included insights in relation to access to data and what we have learnt about the landscape that makes that hard today.

Discoverable and Understandable

Our research showed that for data to be valuable to users it needed to be both discoverable and understandable. Users outlined that discoverable means the ability to find if data exists or not, with relative ease.

> *'I can find it! This may not be the technically correct answer but even organisations that have been doing this for years like National Grid ESO put data on strange parts of its website that are difficult to locate.' - P14*

> *'I am able to find the dataset I need with relative ease.' - P01*

> *'That, I can find items of interest that I didn't know existed before running the search.' - P33*

> *'I can find if something is there or not.' - P21*

Users expressed understandable as meaning data that is clear and well defined, making it easily understood by a wide range of individuals with different levels of knowledge.

> *'A dataset needs to be well defined by meta data, covering things like its granularity (including spatial, periodicity, timeliness, asset types, customer types, etc), quality, reference to relevant standards (eg. Common Information Model), etc. This meta data will accurately define the dataset and enable a common understanding.' - P53*

> *'Data can be understood easily by a wide range of individuals and organisations with different levels of knowledge and made available in a variety of ways. Also that it relates to easily the subject matter and can be analysed to suit a purpose.' - P52*

'Clear headings with units so easy to understand what the data means. Easily accessible and detailed method for producing the data.' - P43

'It is clear what the data refers to and means, and it can be used.' - P14

Insights

Data sources are not easily discoverable and often users are relying on their knowledge and awareness of data sources within the industry. Most users have learnt where to find the data, or they are learning from others within their organisation. This was especially problematic for some consumers who have more difficulty than others finding and understanding the data.

"I'm building up my knowledge of what exists and where to find it." - U10

"[In terms of finding data] we just speak to people - and we usually end up speaking a lot to BEIS for example. We also speak to our policy colleagues, and that also prompts further discussion." - U08

"For example, I worked on a Northern Irish project recently, which I hadn't worked on before. So I just contacted somebody who I knew had worked on a large project before and they were aware of what data was available. But obviously, it's not the most thorough way of doing it because you're probably going to miss stuff." - U09

Searchable

From the survey results, users defined searchable as the ability to search keywords/terms and find the data or find relevant information within a data set.

'Could be simply able to be searched (as in a catalogue) or it could be interpreted as able to be searched for (ie found) as in a data set.' - P03

'That data can be searched in order to find relevant data for the task at hand.' - P15

'It's possible to search a catalogue to find data that is required - keywords and standardised taxonomy are key. It's possible to access data in a format that means that the attribute data can be searched and interrogated.' - P06

'The ability to return specific information from a dataset based on a given set of criteria.' - P27

Insights

In order to be searchable, users needed to know a number of things about the data in order to identify that the data is right for the task at hand.

"And so in terms of the data quality, often it's a case of suck it and see and you know, you won't necessarily appreciate all of the kind of issues of the data until you tried to use it. And then you suddenly realise, oh, wait, that's not quite what I thought it was." - U12

"And some sample data, so you can get a feel for what it is… To understand if the data can be reused for what I need it for."

This theme also strongly came through in our conversations with other projects with similar problems in different industries/departments of government.

"Then there's also the issue about using that data. So once I've, I've selected the date range, identified it I select it is the right data. It's about that ability to interpret that information. So is the data I've just come across? Is it right? Is it the right thing for the task I need to do now." - ONS

"How do you make information more findable - they [people] use google to hunt. The way people were describing data was meaning google was not prioritising the data that was available - they did an SEO activity to look at this." - GC

"Sometimes it's hard to see how that would help because when you're used to data and tech you sometimes over complicate things. Non technical or data people just want to be able to find the information they need... they don't want to learn how to do sophisticated searches to do this." - GC

Our research identified that users wanted to know the following characteristics about data sets in order to ensure the data set is right for the task they want to complete.
- Description
- Organisation lineage and ownership
- Aggregation method
- Update frequency
- Date last updated
- Confidence/Quality level
- Format
- Units
- Risks
- Contact
- Keywords
- Licence conditions

A breakdown of the evidence to support this can be found in the Appendix to this document.

Complexity

The complexity of the landscape has led to the use of services that simplify data that is already open.

> *"I was gonna say there are companies out there whose business model is creating access to what is essentially already public data, but in a more structured and useful way." - U11*

Insights

In user research, we also identified users mentioning services like Enappsys that have built business models around simplifying data that already exists, which speaks to the fact that data in the current landscape is not discoverable, searchable, and understandable.

> *"And then we also have a lot of data that either comes directly from, I guess the BMRS, Balancing Mechanism Reporting Service, either direct or through third party providers such as Enappsys, who precis and reformat a lot of BMRS data." - U01*

Access to data

The current experience of accessing data is disjointed, with users having to go to many different sources to obtain similar data from across the sector, and that users wanted as few barriers as possible to accessing data.

> *"Especially distribution companies... they've all agreed to do it (publish registers) to the same basis but annoyingly it's on their website, not central." - U03*

> *"When I'm aware of data from another job, I just wish i could go and get it." - U03*

Insights

When conducting research with both data consumers and providers, we found that users currently went through numerous journeys and platforms to obtain similar data from across the sector, resulting in a cumbersome user experience that adds to the difficulty in finding relevant data from across the sector. As a result, we found that consumers' expectations from a solution tended to revolve around centralising disparate data in a meaningful way.

> *"Currently you might have to go to multiple websites, but if there is one place and everyone knows about that, that would be better" - U02*

> *"Developing a single hub, which has already done that standardisation and alignment would obviously enable those different sources to put together with less additional effort. Yeah. So be able to more quickly do things with them and produce answers." - U02*

"I think this whole idea of just having it in one place makes it a lot easier to manage, trace. And if the catalogue in the front had information on a data dictionary, or information on what's in the dataset, then that'd be really useful. And it'd be somewhere where you're not having to jump through a million different links on a website to find the data you need. It'll be quite easily found again and again when you need to refer back to it." - U09

"A catalogue suggests to me that there's a lot of different pages that bring together a lot of different data sets but they'd all be interconnected; showing different things in a standardised form. Whether that's in an old-school Excel sheet or a more modern database." - U09

Our user research identified that users wanted to access data with as few barriers as possible. While these barriers may exist for valid reasons, services should seek to only establish these barriers where necessary in line with the principle that data should be "presumed open."

"Any data the market puts out needs to be publically available ... I just want to get the data, or see the data!" - U03

"Unless someone can come up with a valid reason for a password, there should not be one" - U03

Conditions and systems

The problem statement for this Discovery stated we would look to "*identify potential systems, and conditions that could be introduced to influence the sharing of data openly*" and this section provides more detail and insights around system and conditions.

Collaboration across providers and consumers

We found that a dialogue between data providers and consumers was central to the journey towards opening up data of appropriate quality

> *"I think a lot of time you're leveraging relationships to get your hands on data" - U11*

Insights

There were several benefits that were mentioned associated with that dialogue:
- Feedback to providers provides a mechanism for improving the quality of published data
- Dialogue enables the right data to be opened up for consumers
- Curation of the data sets could evolve to include the consumers of the data

An interesting insight on this is that collaboration and engagement with providers were often the primary means through which consumers built up their knowledge of energy sector data, with knowledge from conversations being accumulated over time to form bespoke pictures of the data that is out there.

> *"[in terms of finding data] we just speak to people - and we usually end up speaking a lot to BEIS for example. We also speak to our policy colleagues, and that also prompts further discussion." - U08*

> *"But yeah, it's more like, I mean, for example, I worked on a Northern Irish project recently, which I hadn't worked on before. So I just contacted somebody who I knew had worked on a large project before and they were aware of what data was available. But obviously, it's not the most thorough way of doing it because you're probably going to miss stuff." - U09*

> *"A lot of awareness of data sources comes from ad-hoc conversations, and some of it is just Googling. So, for example, the P114 data from Elexon, unless you're a network licensee or in the system or very involved with the data, you're unlikely to know that it exists. We found out about it when we spoke to a colleague at National Grid who said it's a very good proxy for demand data. And then we had contact with Elexon, who said we could only have access because we were affiliated with National Grid, who are a network licensee. So, a lot of the time it's a very convoluted process." - U11*

"Having a relationship with the team responsible for the data set at the other end is kind of critical." - U11

"Not many of these data sets or companies have any systematic mechanisms for kind of feeding back about the data… I think it would, it would be nice to see some, I guess better better data management and handling and version control. And you're making use of a more modern approach to curating these data sets so that there can be that formal feedback mechanism." - U13

Relevant initiatives

Collaboration does seem to be at the heart of most organisations' journey towards open data. Here are some examples:

- UK Power Networks' have established an Open Data Feedback Forum alongside their open data portal. This enables all users to engage in discussions around the quality of existing data and to also make requests for more data to become available.
- Western Power has a feedback form for their Energy Data Hub. Although feedback forms have a disadvantage that the community of users do not benefit from conversations that have gone before.
- Organisations such as Cadent and Northern Powergrid have also used mechanisms such as the Customer Engagement Group to ensure that the needs of data consumers are considered in their journey to open data. Organisations like Wales and West Utilities note a similar commitment to engaging with consumers, noting that they currently share data on request and that they have ambitions of publishing a "cross-sector data platform" to share data with consumers in 2020.
- Western Power Distribution appears to be putting continual engagement with stakeholders at the heart of their Presumed Open Data project. This includes working with a wide range of data consumers on identifying clear use cases, rapidly publishing datasets on their Energy Data Hub. They also intend to galvanise activity around a "data science challenge" in the future. This is a highly innovative example of collaboration throughout the journey to open data.

Industry collaboration

Insights

We found that collaboration with other industry participants to build the right solution was essential as a condition for success. We also found that collaboration was needed across a range of roles (including architecture and analysis) and involving all impacted parties.

"And I think collaboration is also about getting the right people involved. So getting analysts involved as much as policy makers in these workshops that keep happening, for example the EDTF workshops." - U08

"So, shockingly, no distribution company was involved in the EDTF recommendations." - U19

"What you've seen in the financial space is …. even though it's a commercial area, there's been this initiative for open data, open data banking has really taken off really well. And this is where this is where commercial companies have got together collaboratively driven by market forces and got together to form an open group and it's been very successful" - U18

'It has to be done together in a collaborative way, it can't be done otherwise. I mean, there is no programme of work that will be able to do that on its own- CG2

"It's all about trust and collaboration, and understanding common purposes for the data. It's also about building local and regional stakeholder groups that meet early and often to deliver the change needed" - CG4

Relevant initiatives

The National Underground Asset Register which is a programme that involves 35 organisations told us that effective collaboration and building trust around a very targeted and narrow scope has helped them on their journey.

The Energy Networks Association / Data Working Group has been mentioned as a forum that industry participants were involved in in order to create the triage process for open data, and for driving relevant efforts such as aligning various networks to the Common Information Model standard for their data.

Incremental improvements

Insight

We found both data providers and data consumers talk of the need to "start somewhere," and view the visibility of a "base set" of data as a good place to start. We also found many people talk about the need for an incremental, but consistent, approach to delivery. Participants were keen to ensure that the solution delivered value for money for the end customer.

Similarly organisations expressed a concern on the amount of time that may be spent up front cataloguing their data to a certain standard, only to find the data is outdated, or of insufficient quality, or doesn't meet user needs by the time of its publication.

"For me, I think this is about incremental improvement. The gap between where we are today and where we need to be is too big, so we need to do it incrementally." - U19

"Because if you started with a catalogue of all the data that is out there, it would be obsolete by the time you complete it. Because the industry is changing so fast." - U11

"But a lot of that [digitalisation] will come at a cost: a cost that consumers must be willing to pay. So, the impact we have equates to 15-16% typically, on the customer's bill, and we need to ensure that there's that value to the customer." - U19

"Brilliant. And I guess yeah, that's really just sort of a, an ongoing journey of engagement of what stakeholders and third parties want to see. And I think that's why I'm really keen that it's an evolutionary piece, there's a real danger that you disappear and say, we're canvassing for more opinion, we're canvassing for more opinion. And then three, three years later, you think you're outputting a perfect data set that's unusable. I'm keen that we get data out. And maybe there is some, you know, improvements and tweaks and criticism on it that we can sort of, hopefully by doing that you engage with third parties and take them on the journey as well. And so that's what I'm, that's why I'm keen to do and I think, yeah, we just, we just need to be ready for that and are doing that now. And I think that's providing some good value." - U20

Relevant initiatives

The strategy review showed evidence of initiatives making data open, which was also noted in user research as providing a fruitful basis for the collaboration between providers and consumers mentioned previously. Examples being the publication of capacity maps by a large number of network organisations, and the publication of standardised system wide resource registers across the sector.

Western Power Distribution's Energy Data Hub, the UKPN open data platform, and National Grid ESO's Data Portal are also using an iterative approach where a base set of data is released and feedback is used to iterate and improve data quality.

The disadvantage of the work that is going on is that even more portals may exist for users to go to in order to obtain energy data, which was a noted pain point in user research.

Overseeing the service

From our survey, users felt the Office for National Statistics or a neutral party would be the best placed to oversee the development of the service.

'Their longstanding experience as data curators. They have full sight of the level of effort, and are dedicated to providing that kind of service. Their remit is clear and won't change. They are well known and trusted. BEIS would need a whole new capability to run it, and may find themselves subject to a change in focus/ funding. OfGEM would also need to develop a new capability. They would need to drive the policy / regulation to require data to be made available, and that's where their focus should be, in collaboration with established data curators.' - P06

Insights

Users felt the ONS have the experience with data and therefore the capabilities of the service, whilst being a neutral party that could work alongside BEIS and Ofgem to ensure policy and regulation were working towards the same goals.

> *'Neutral.' - P49*

> *'ONS are seen as more independent than Ofgem or BEIS, as those parties have goals they wish to achieve that raw data may or may not support.' - P32*

> *'I don't know - the identity of the party is less important than the rules governing the management.'- P10*

Benefits of data sharing

This section highlights a few themes relating to the benefits of data sharing that were identified during our research.

Better decision making

Better data leads to more reliable analysis and better decisions

> *"They will make better investment decisions the more data they can see." - U03*

Insights

Users explained that one benefit of increasing data visibility would be for the industry to better optimise their processes and make better decisions as they would have a fuller picture to do so. If everyone is using the same source of data then there will be more consistency across the analysis, reducing processing cost and making work more reliable.
It is noted that opening up data as well as sharing data would be needed to obtain some of these benefits. This is consistent with the findings of the MED report.

> *"Huge benefits to data sharing for the industry. a lot of the forecasting we do, a young, agile company would be able to do it better." - U17*

> *"Well, I guess like I said, about the consistency between when different consultancies are carving out some of the work and comparability between them. I think when you're working as a whole looking at overall trends, it makes sense that **everyone is using [the] same data**, I think you'd end up with **fewer errors,** because they'd be flagged consistently. And so they could be corrected rather than having a million different sources, and then some correction and so on. And it would just make everyone's work more reliable." - U09*

Driving innovation

Data is a driver of innovation.

> *"If you're a big company, then you'd have better data than outsiders. Open data is about levelling that information asymmetry, so that we can have more innovators provide new services to the Grid. It [data sharing] would also have more of an impact on monitoring, and help us drive changes with evidence." - U07*

Insights

We found that a key benefit of data sharing is allowing innovation. Other industry sectors have demonstrated this well e.g. Transport for London. This is consistent with the findings of the MED report.

> *"Huge benefits to sharing. TFL shared openly in APIs and the response from the community... so much cool stuff was made from streams of that information. Cycle movement, trends around the*

city. It was a lot of work for TFL.... but they got so much valuable stuff for free... like you would from a hackathon!" - U10

Levers for data sharing

One of our Discovery objectives was to learn what levers and conditions could help on the journey to open data. The levers have been identified either during our research interviews or as part of the stakeholder mapping work which is being carried out alongside this deliverable which aims to identify the influential players in the energy sector.

The definition of a lever that we have used is:

> A lever acts on the system as a whole to create a change. If you apply a lever in one place something moves elsewhere. Use of the word lever also infers that a change should be happening in a timely manner.

Visibility of data

We found that some organisations are concerned or embarrassed about either the gaps or the quality of some of their data.

> *"Benefits of open data- hanging your washing out in public don't want to see the tshirt you've had that's faded and rubbish. Force us to get our act together on data quality" - U18*

Insights

Users recognise that effort will need to go into improving the quality of the data. There was recognition that making it easy to compare the level of sharing of data across organisations would provide a lever for change. Our recommended approach to the service, which drives comparability through making existing data discoverable, searchable, and understandable, capitalises on this lever.

> *"Benefits of open data- hanging your washing out in public don't want to see the tshirt you've had that's faded and rubbish. Force us to get our act together on data quality" - U18*

> *"Challenges - being frank. A huge amount of asset data is very old. Large percentage of the data is poor quality for various reasons. Challenge is [we] would be embarrassed to put this data out. It might be incomplete...or missing compulsory fields (last time inspected) ....data quality process work to clean it up" - U18*

> *"It shouldn't be the lowest common denominator driving this." - U19*

> *"Would force us to get our act together" - U18*

Sharing stories of success and best practice

We found that publishing data sharing success stories was thought to be a positive means of driving data sharing.

> *"If other people are doing it, you will do it because you see some benefit in it. Seeing a benefit is very important." - Innovate UK*

Insights

Organisations would benefit from the good news stories and it may have an impact for the organisations should they be seeking investment. Our approach to the service seeks to capitalise on this lever, looking to test the assumption that a clear demonstration of such "success stories" in a visible way will be a driver for data sharing and visibility.

> *"InnovateUK have talked about plaudits and positive Use Cases that BEIS and InnovateUk will publish so that people are held up as 'leading organisations' and the desire to be seen as a leading organisation will be stimulated. That can have a really big impact when attracting investment. " - U18*

> *"Demonstrating benefit/value e.g. Octopus - "if we are doing this and getting comms opportunities from it, then others should." - Innovate UK*

Incentivisation versus regulation

Our research did not find a consensus as to whether incentivisation or regulation would be most effective to drive data sharing.

Insights

Ultimately, work across the Modernising Energy Data programme in Alpha and beyond will need to consider what the right "balance" between incentivisation and regulation is, and how both can be used as levers to drive the necessary change needed to achieve data visibility and move the industry to a state where data is "presumed open." The insights above have already started forming a picture of where some interventions to this end may be helpful, e.g. changes to the delivery of regulated price controls.

> *"It's both, incentivisation is great, it's a reward system, regulation is more penalty... incentivisation will make some organisations go above and beyond... what's regulated should be the fundamental stuff and what's incentivised is the stuff that should go above and beyond." - U20*

> *"You've got to have an incentive to share the data - regulation doesn't give you an incentive to innovate or work collaboratively." - U18*

"If you look at incentives, DNOs have a customer and stakeholder engagement incentive to ensure communication.... it's a competition and they get ranked for money. but they all win, they all get money." - U15

"My gut feeling is it has to be regulated. the cost would be smeared on customers but we'd be able to add standards and it becomes part of the bill. we would know the data is maintained to the right standard and quality." - U15

"If there is enough of a show of them doing it of their own accord, regulation will not be put into operation. [Name] does a lot of work in this area "let us appease regulators enough so that we are not told what to do" The leading ones will do it but not the rest of the sector" - Innovate UK

"The driving force behind this is regulation. E.g. in order to provide a good distribution service you have to disclose certain information to resource services- such as where cables are, what is buried under the tarmac. All sorts of ways – market structure, safety, consumer needs – relates to 'open data' - Innovate UK

Political pressure

Insights

We found that political pressure was one approach that was being used already to drive the sharing of data.

"Political pressure. This is a large part of what we have been working on. There has been no change to regulation but there has been a change in Governmental pressure." - Innovate UK

Restructuring/Breaking up organisations

Insights

We heard that an effect of breaking up organisations - for example requiring there to be separate legal entities for generation and retail - was that data could no longer be shared internally and exclusively.

"National Grid has been split up and split up. A bit like BT. Under the same group umbrella but all quite separate and not permitted to share information as they did before. This has all been to do with reducing unfair advantage and reducing monopoly. This causes a forced stop in sharing data but what it also means is that of course, if they are happy to share the data with everyone, openly then their other division can see it too" - Richard Dobson

"This forced "air gap" has also happened in the big suppliers where they have both retail and wholesale. Divisions have been forced between them to ensure fair pricing, stopping them sharing data that would give them an unfair advantage. The same applies, however, that is they are happy to share openly with all then the other division can also see it" - Subject Matter Expert

Barriers and costs of data sharing

Across the sector, there are numerous barriers that get in the way of sharing data and moving the industry as a whole to a landscape where data is "presumed open." The barriers we identified within these 6 weeks cover quite a wide range, and have been considered as part of our overall recommendations for the design of the end service.

Organisational culture

Organisational culture was mentioned in several of our interviews as a barrier to sharing data.

> *"Crucially, it's a massive cultural change that we need to undertake. That's one of the biggest challenges that we, and indeed most distribution companies, will face." - U19*

Insights

Data sharing advocates within organisations talked about having to persuade their colleagues of the benefits of making data open. Data was seen as potentially commercially sensitive. Although some participants saw this being less of an issue with regulated monopolies.

> *"I had this conversation with somebody in it yesterday. And I said that we're trying to make all our data open. And I was really struggling to persuade him what the benefits of it were." - U18*

> *"This is probably the biggest one [culture], because at the moment, everybody has the impression that data should be confidential. You know, so we're working, we're working from one end of the scale, and we're gonna work our way all the way back to the open side of the scale." - U18*

Relevant Initiatives

The National Underground Asset Register (NUAR) work also invested effort in understanding and resolving concerns around issues such as security to help smooth implementation. This included attending key stakeholder meetings. Support at all levels of the organisation from the CTO, head of IT through to advocates in the data team was mentioned as helping to address this issue in one interview.

Similarly, references across the digitalisation strategies that speak of building agile and data-driven cultures are a positive step to this end. Similarly, organisations committed to assigning leadership and advocacy for data sharing and best practice, such as SP Energy Networks, who have assigned regional 'data leads' and 'data champions' and have made data a board-level priority according to their digitalisation strategy, can also be seen as driving the right cultural and organisational changes.

The hope is that such initiatives will work towards building cultures that better understand the

value of data visibility, and appropriate mitigations to this end, and are more likely to drive the data sharing and visibility needed to achieve open data across the sector.

Technical capability, governance, and data quality

Insights

Our research has shown that some organisations feel that they do not have all of the organisational, technical, and governance capability to be successful in a data-driven world. However, there are actions being taken within organisations to establish the capability and governance required to enable data sharing and improve data quality and governance to this end.

> *"We're in the process of identifying challenges as we go along. legacy systems to store data make it complicated. there's often disparate data that now must be combined." - U20*

> *"We need physical network infrastructure for data we're not collecting... internally, we don't have a huge team of data scientists." - U20*

> *"Lots of organisations will be the same. shadow systems... we need to improve the data governance... The data from last year is not good enough to plan. We need to look at the data now." - U20*

> *"Legacy data around assets in the ground for 100's of years. Don't know much about them (no leaking!). Issues with granularity and the quality of data." - U18*

> *"A single source of truth from the legacy systems has been difficult... how we improve the source level of information.. accuracy and immediacy." - U20*

> *"A big part of it is the data governance, what is where and what do we need to facilitate that sharing." - U20*

> *"Other things holding back sharing - capacity - people that will be needed. Will be a big project. Need to recruit people with necessary skills" - U18*

Relevant Initiatives

A review of the strategies from across the sector outlined actions organisations are taking towards enabling data sharing. Presented here are some examples of how organisations are building the right organisational, governance, and technical capability to enable data sharing:

- National Grid ESO - Strategy outlines the behavioural shifts that will be needed to support new ways of working
- SGN - Strategy recognises the level of change that needs to happen within their organisation to support being both data driven and agile ways of working. Working with

Deloitte Digital, Microsoft and others with mature digital processes to bring in best practices

- Western Power is working with the ODI to open up their data iteratively
- Northern Powergrid are working with the ODI Leeds to deepen their understanding of open data and are publishing their Distributed Future Energy scenarios using an open data approach co-developing regional forecasts with their stakeholders
- Northern Gas Networks have worked with DAMA - an international data management body - to establish a well-defined data governance process called DICE to ensure the appropriate management and quality of their data
- Various organisations across the sector are also working towards building the skills and technical capability necessary to be able to collect, manage, and analyse high-quality data. Scottish Hydro-Electric Transmission, for example, has a dedicated cross-organisation IT Group that provides specific expertise in data and infrastructure, and have noted they have specialist teams across Data and Analytics, Technology and Architecture, and Cyber Security. Similarly, organisations such as Cadent are establishing "central data teams" as part of becoming a "data-driven organisation."
- Alongside the above, numerous innovation projects are underway to improve the quality of data organisations across the sector provide, such as Western Power's innovative LCT detection project. Work towards modernising legacy systems and getting a clear view of internal data, such as Scottish Power Distribution's work on developing an internal Network Asset Management System, are also valuable to this end.

Concerns around security and use of the data

We found that there are concerns around the risks that opening up more data would cause if the data is used incorrectly.

"More data is dangerous if you don't know what to do with it." - U20

Insights

Numerous security concerns surround the fact that a lot of data pertains to critical national infrastructure, and there was a need to understand and agree how data was being used before it is shared

"Internally, there's the security concerns, cybersecurity concerns, and concerns around asset and network security since we're critical national infrastructure." - U19

"So, mainly it's about our obligation to public safety. So, there have been cases in the past that people have taken the data we've provided and made some very ill-informed assumptions. Ultimately, what dictates the energy system is the laws of physics. So, if they arrive at the wrong answers from the data we provide, that's got a serious risk." - U19

"Historically, we've been quite protective of our data. There's always been a sense of caution around what data we provide and how we provide it". - U19

"And there's a whole sea of regulation that we've got for this, but we've also got a great track record of providing data to businesses and customers, usually via bilateral agreements." - U18

"Knowing who is using the data would be good for cost-benefit" - U03

Relevant Initiatives

As mentioned previously the NUAR project has experience of resolving specific security concerns by building effective collaboration. Similarly, the work being done by the ENA Data Working Group to create a standardised triage process across organisations can also be leveraged here as part of addressing broad concerns around governance and security. Lastly, existing patterns such as the need to register on to platforms where data is made available such as the Elexon Portal, Western Power Distribution, and UKPN's Open Data portal also outline potential security protocols that might need to be considered & developed to open standards as part of sharing data.

Concerns surrounding privacy are a barrier to opening some data up

We found that concerns around privacy served as a barrier to opening up certain data that would be useful.

Insights

While this is a valid concern under the "presumed open" model, concerns such as these will need to be raised and addressed as part of embedding a consistent triage process and getting shared agreement on the "presumed open" model within Alpha.

"We consume a lot of data through smart meters and all of that... but a lot of the data we hold is personal so from a practical point of view there is only so much we can do." - U15

"So, one really good example is Smart Meter data, but we as a distribution company can get aggregated consumption data, but we can't get individual consumption data because there's concerns around personal data." - U19

"Consumer privacy/personal privacy. People have seriously said that we may have to decide between personal privacy and helping the climate." - Innovate UK

"With more data, there is a risk that if you disclose ..where your consumers are located and information about that, there is a risk that bigger companies essentially eat up smaller companies." - U07

I think as well between the different providers is often a big disparity between and how they perceive data ownership. And some of them will say, you know, we don't own the data, we're just collecting it and curating it on behalf of our clients. And it's only the kind of asset owner that

actually owns the data. So we don't have the rights to share it with you, et cetera. And, and I think, yeah, when it goes down to ... the domestic level, different companies that have metering in place, have different views on whether they actually have the right to share that data, or indeed sell it." - U11

Lack of understanding and alignment around what "open data" means

In user research, we identified that there was significant uncertainty surrounding what "open" data - as per the presumed open model - means for data providers.

> *"An issue that has come up is that the definition of what is open data is grey and woolly. So we're working on classifying all of our data, but guidance around classification levels would be useful" - U18*

Insights

This has resulted in organisations spending a significant amount of effort, uncertainty and time spent on trying to establish a triage process and understanding how their data should be categorised, licenced, and shared. Organisations have noted a need for alignment, consistency, and precise guidance to this end to enable effective data sharing.

> *"But then the challenge is also that we are critical infrastructure - so how do we classify our data sets into categories?" - U18*

> *"And then obviously a lot of our data is shared via bilateral agreements, so we're also looking at the licence conditions surrounding data" - U18*

> *""a robust triage process... we don't want to create seperate data sets for each user... we'd create a uniform set of data with everyone's needs in mind." - U20*

> *"So, as a data working group, we're looking at how we can make the triage process consistent across various distribution networks." - U19*

Relevant Initiatives

In user research, we found that the Energy Networks Association's Data Working Group was working towards bringing consistency into how the triage process worked across its constituent network organisations; existing initiatives such as these can be capitalised on within Alpha as part of embedding a shared understanding of the "presumed open" model and a consistent triage process as part of developing the service.

Costs of standardisation

Alignment of data to a standard poses a range of challenges to providers, chief amongst which is the cost introduced by having to map internal standards for data to a different external standard.

Similarly, the level at which standardisation occurs is also an important consideration, with benefit being an important consideration for this.

> *"People have to change their systems - there will be costs associated with the standardisation." - U03*

Insights

Standardising data at attribute level might present benefits in the near-term in a very specific domain, but might lock providers into an inflexible and rigid model that isn't able to keep pace with industry change and might not unlock value for the broader community of data consumers.

As such, these costs must be considered as part of developing an evolutionary approach to standards within Alpha. To this end, there was also a need again for clear use cases to be used as the basis for defining any standard, and for collaboration to be at the centre of defining any standards.

> *"In the immediate if it was a different standard to us, there would be some work for us to do… people need the data to look the same, it shouldn't matter where the data is from, you should be able to use it." - U20*

> *"But, at the same time, we need to understand the use cases the standards will need to align to so that we know what the standard needs to look like… And, if we don't do that, then we'll be doing it at massive risk to ourselves and others" - U19*

> *"All stakeholders together to get agreement will take time. Whose system is the right one?" - U02*

> *"Standardisation would be a good thing... However we've built systems around data hierarchies. changes could mean those systems need changing" - U05*

> *"You either need mapping removed from the internal representation to an external representation, or to update all the internal systems with the agreed standard." - U02*

> *"I guess the costs to providers is that data manipulation, gap-filling, and standardisation of data is quite labour-intensive. And I think there's also considerations around how costs are distributed…" - U09*

Relevant Initiatives

From a review of the digitalisation strategies published across the sector, numerous network organisations have noted that they are aligning their data to the Common Information Model as part of the broader activities being driven by the Energy Networks Association. As noted above, a consideration here is one of value, effort, and flexibility - there are open questions around whether the CIM presents value to the broader community of data consumers and is flexible

enough to keep up with the pace of sectoral change without requiring significant operational overhead to maintain.

Similarly, organisations such as National Grid ESO and Northern Gas Networks have noted that they are aligning their data to the Dublin Core Metadata Elements, and organisations who have already made a significant portion of data public such as Western Power will also be aligning their data to a metadata standard of some sort - albeit they haven't stated specific alignment to Dublin Core. While work in Alpha will be able to build on these efforts, the main challenge will be aligning providers at various levels of maturity to an agreed set of attributes, and ensuring that costs to this end are proportionately distributed.

There is little room in regulated budgets to innovate

Insights

A number of concerns have been raised about the costs and timelines associated with opening up data and the lack of funding for activities due to some participants being regulated monopolies.

> *"So, the real challenge is the budgetary issue of price controls and how we get funding for data access and making data open." - U19*

> *"We've we've had the last price controls a thing called GD-1, which is where they say right, basically they just say what we're going to do is we're going to pay you X number of pounds for maintaining a mile of pipe and we're going to pay you X number of pounds every time you have to repair a pipe and we're going to pay this much money for replacing pipes. In a nutshell, they were coming up to the next price control point, which is GD-2. And that doesn't come into effect until 2021." - U18*

> *"But we know that for GD-2, we are going to have to make our data a lot more open. But we won't get any, and we weren't getting any income from Ofgem, for making our data more open until we get to the far end of 2021. Yeah, but we already know that we again have to spend quite a lot of money to do this. But we're not going to be able to recoup that money until we get into the GD-2 price control period." - U18*

> *"There's no profit against the business case - how do we make cost savings.. hard to do" - U18*

Recommendations

Alpha overview

From the Discovery's research and analysis we have the recommendation that the Alpha deliverables should be to prototype a service that will:

- Make existing public energy sector data discoverable, searchable and understandable
- Facilitate collaboration amongst data providers and consumers
- Act as a lever to expand the volume of energy sector data that is discoverable, searchable and understandable

Summary of recommendations

Recommendation	High level detail	Discovery Theme
Start with public data and evolve	We recommend that the service's initial objective is to have all data that is currently publicly available searchable by a "base" set of standard dataset attributes, in order to prove that this will result in the public data being more discoverable, searchable, and understandable. Test the broader assumption that the improved understanding of the data landscape and comparability across providers driven by this being situated within a usable and collaboration-focused service will establish the basis of the market pull that is envisioned as a broader function of the service	Discoverable Searchable
Start with a simple set of attributes and evolve	We recommend that datasets have a set of characteristics that support consumers to make the decision as to whether they are right for the task. Datasets should be described and categorised in a meaningful way to support that activity.	Discoverable Searchable Understandable
Business glossary	We recommend a common set of business descriptions need to be developed to enable the datasets to be described consistently	Searchable Understandable
Visibility and access	Data consumers do not view visibility and access as separate concerns. Therefore, work in Alpha will need to understand how the end-to-end service can enable this access, and develop the right security, architectural, and governance foundations to enable this.	Discoverable Searchable Understandable

Evolutionary and collaborative approach	We recommend structures need to be in place for organisations to collaborate across the industry to address complex concerns surrounding areas such as standards, security, architecture, and governance. Existing efforts towards collaboration being driven by bodies such as the ENA can be leveraged to this end.	Service Condition Lever

Service vision

As such, we view the service as being built on interrelated foundations:
- Data is described, categorised and attributed in a meaningful and consistent way for users to understand
- This will enable users to discover what data is available
- As more data becomes shared, it is made visible in a way that is meaningful to users

The Alpha will look to test the hypothesis that these outcomes will form a feedback loop to drive energy sector data to be discoverable, searchable and understandable and increase the visibility of energy sector data that is not currently shared through network effects and market pull.
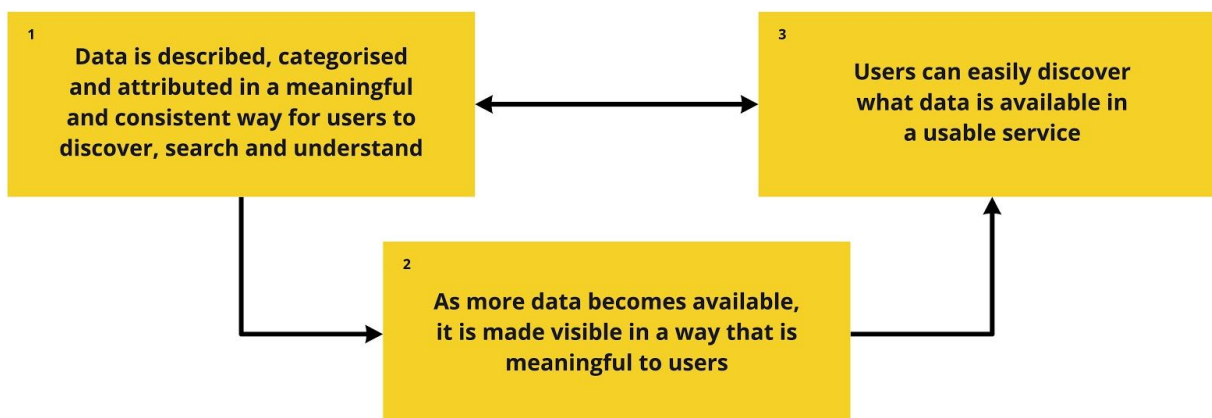
It is important to note here that users don't necessarily distinguish between the visibility of data and access to that data. They wish to have a friction-free journey to complete the task at hand. The wider programme will need to consider how to support the end to end goal of the user by delivering an architecture and operating model which addresses the wider needs of the user while addressing security concerns and access control. At a programme level, this outlines the need for collaboration across various programme workstreams to deliver the overall "end service."

To help achieve the service vision outlined above, we recommend an agile approach to the design and development of the service, developing from the base capabilities in a Minimum Viable Product to a fully capable service that supports the wider EDTF outcome of having open data across the energy sector.

Riskiest service assumptions to test in Alpha

From the evidence and insights we've gathered through Discovery's user research and analysis, we have been able to identify the following high-level assumptions that would need to be prototyped, tested and proven during the Alpha. These have been identified as the riskiest assumptions we'd like to test in Alpha, with risk being a function of how **important** it is for a given assumption to be true for the service to be valuable, and how **certain** we are that said assumption is true. **At the moment, these assumptions represent what we think the service will look like and consist of**. Our delivery approach in Alpha will look to test all of these riskiest assumptions as part of defining what the MVP should look like. We have also mapped these key assumptions to a range of low-fidelity prototypes in a separate slide deck attached to the report.

Assumption	Risk
The team will be able to find a set of attributes that aids energy sectors datasets' discoverability, searchability, and understandability for data consumers while still being acceptable and cost-effective for providers. An initial set of assumed attributes has been outlined below in Recommendation 2.	High Importance / Low Certainty
Providing users with the information needed to help the discoverability, searchability, and understandability of the datasets - as outlined in assumption 1- and informing users of how datasets relate to each other, will help flatten the learning curve for data consumers who are unfamiliar with energy sector data.	High Importance / Low Certainty
Building on Assumption 2, a better understanding of the existing data landscape will provide consumers with the information needed to make clear, well-defined requests for data that is missing - driving further visibility.	High Importance / Low Certainty
We believe that the service will be able to achieve collaboration and buy-in from across the industry around standards, security, architecture, and governance concerns in delivering its intended outcomes.	High Importance / Low Certainty
Using the service to highlight examples of innovation from across the sector, and illustrate how data sharing by data providers drives positive collaboration and innovation will: ● help connect innovators to data providers and other collaborators, creating a positive feedback loop of innovation ● encourage more data sharing, drive additional investment in improving data quality, and lead to the creation of rich new data sources	High Importance / Low Certainty
Given that users don't view visibility and access as separate concerns, the service will need to have appropriate considerations around security and governance to enable providers to share data. Our current assumptions to this end are as follows:	High Importance / Low Certainty

• Highlighting security considerations surrounding a dataset within its attributes will help mitigate some of these concerns • Providing a means within the service for providers to see how their data is being used, and by whom, will be a relevant prerequisite for providers to share data	
Enabling collaboration between providers and consumers within the end service will be central to enabling the market pull to make more data visible; in line with the assumption above. Our current assumptions around what these enabling features may look like are as follows: • Forums to discuss issues, gaps, and duplication across datasets, and - importantly - to discuss use cases for data that might feed into requests and visibility • Appropriate ownership and contact details to be provided as part of a dataset's standardised attributes to ensure that consumers can raise concerns and queries with data providers, and for comparability across providers to be achieved to drive the necessary market pull envisioned as being a function of the service • A mechanism of some sort wherein consumers can request data from providers, which ties in to the need for datasets to have clear ownership assigned to them and the broader insight that users don't view visibility as separate to access	High Importance / Medium Certainty
We believe that the visibility of requests within the service will be able to feed into developing a standardised "presumed open" data triage process across organisations, helping move the industry to a state where data is comparable across a wide range of providers. We believe that the work being done by the Energy Networks Association towards developing a standard triage process across network organisations can be leveraged to this end, and that this approach could be expanded to include other trade organisations such as the ADE. This is expanded upon in Recommendation 5.	High Importance / Medium Certainty
At a user experience level, there is a need for available data sources from across the sector to be centrally visible "in one place," so that users don't have the disjointed experience of moving across various platforms to find out what data sources are available. Therefore, we believe that there is value in bringing datasets together in a central location. Since users don't view visibility and access as separate concerns, the service should enable both visibility of and access to data. It is important here to not conflate the centralisation of visibility, which is an experiential concern, with a centralised approach to architecture, which is an infrastructural concern that will be considered within Alpha.	High Importance / High Certainty

Detailed recommendations

Recommendation 1: Start with public data

As user research has identified, a lot of users currently gain significant amounts of value from data that is public today, but because of the problems in the current data landscape, this data is difficult to discover, search for, and understand. So much so, in fact, that services like Enappsys have built their business model on better representing data that is already public via the BMRS platform, for example. Therefore, we recommend that the service's initial objective is to have all data that is currently public, listed and searchable by the service's standard dataset attributes, in order to prove that this will result in the public data being more discoverable, searchable, and understandable. In line with the vision and assumptions above, we'd seek to test the assumption that the improved understanding of the data landscape and comparability across providers - driven by this data being situated within a usable and collaboration-focused service - will establish the basis of the market pull that is envisioned as a broader function of the service.

Recommendation 2: Start with a simple set of attributes and evolve

From user research with providers and consumers, this is the "base" set of dataset attributes we've developed to test and evolve in Alpha. The evidence underlying the need for these attributes is outlined below in our findings surrounding the information users need about datasets.

Attribute	Description
Title	A descriptive title for the dataset
Description	Accessible, clear information about the data that a specific dataset contains
Organisational lineage	The organisations across which the given dataset has been processed
Owner	The owner / provider of a specific dataset. Work in alpha will need to unpick the complexities around data controllers and data owners,
Aggregation method	The methods through which data within a given dataset has been aggregated
Update frequency	The frequency at which a dataset is updated
Last updated	The last time a specific dataset was updated
Confidence level	The level of confidence associated with a specific dataset. Numerous measures exist towards calculating this, and a standardised process will need to be devised and agreed within Alpha
Format	The formats that a dataset is available in
Units	The units in which energy is described within a dataset

Risks	Considerations, risks, and constraints around making assumptions based on specific datasets
Contact	Contact information for the provider of the dataset to raise quality issues and gaps within datasets
Keywords	Keywords associated with the dataset, which will inevitably evolve over time and refer to the categories
Access conditions	Any controls, requirements, or pre-requisites surrounding access to data
Licence conditions	The licensing conditions controlling access to the data
Identifier	An identifier that "points" to the dataset, in line with best practice for open data

We assume that starting with a standard set of attributes at dataset level presents a number of benefits; these must be tested within Alpha:

- Standardisation of attributes at a dataset level minimises the requirement for standardisation at the data attribute level, since standardisation at dataset level will naturally lead to an understanding of common attributes at the level of individual data item attributes which can serve as "keys" on the basis of which datasets can be joined. Over time, this will evolve into a principles-based approach to standardisation that is flexible enough to evolve with the energy sector.

- That standardised attributes at dataset level are much less effort than standardisation at the level of individual data item attributes, easing costs of adoption for data providers in the first instance.

We also believe there is the added benefit here that this approach ensures that relationships and commonalities across datasets are truly *discovered* as an emergent property, therefore better reflecting user needs instead of organisational priorities. The common attributes at data level emerge as a result of the data being well defined instead of being based on assumptions in the first instance. This is a more efficient way of working. We believe the service as a whole will be a catalyst for this evolution to occur in a collaborative way between data providers and data consumers.

Recommendation 3: Business glossary

Users wish to search for datasets using language that they understand. In the current landscape, seemingly simple terms such as "power" can mean different things across different contexts within the industry. A common set of business descriptions need to be developed to enable the datasets to be described consistently. This is particularly important for completing the attributes of "Title" and "Description" in the table above. In Alpha, this business glossary should start being developed as part of the broader approach to developing the "base" set of dataset attributes described above.

Recommendation 4: Visibility and access

As noted above, users do not distinguish between visibility of and access to data; therefore, it is important that the security, governance, and architectural concerns associated with enabling this access is also considered here. As user research has identified, there are numerous security-related barriers to accessing data even when data is public. Similarly, the associated architectural and governance concerns associated with enabling this access will also naturally need to start being considered. Exploring how these concerns feed into the design of the service will be an important activity for Alpha. The need for a collaborative and evolutionary approach to addressing these concerns is outlined below.

Recommendation 5: An evolutionary and collaborative approach to standards, security, architecture and governance

As our roadmap and broader approach to the design of the service outlines, the capabilities underlying the service will need to evolve in a truly agile way. In particular, the complex concerns surrounding standards, security, architecture, and governance will need to be addressed through cross-industry collaboration, potentially leveraging existing working groups such as the Energy Networks Association's Data Working Group as drivers to this end. Going beyond existing working groups, however, collaborative working groups in the future will also need to more closely involve the end consumers of this data alongside data providers within the industry. The stakeholder engagement efforts noted within the published digitalisation strategies by network operators are positive indicators to this end.

The ultimate objective of this collaboration will be to strike a balance between the needs of data consumers and the organisational realities, barriers and concerns of the various data providers across the sector.

Standards

In making data discoverable, searchable, and understandable through standardised attributes, the long-term goal for the service will be to support the creation of an ontology for energy sector data, wherein rich and semantic relationships between existing and future data helps unlock the underlying value across the datasets.

To help achieve this, we recommend the approach to defining the attributes for the datasets must be collaborative and evolutionary. Establishing this collaboration from the start will ensure that the approach to standards developed as part of this service remains flexible, collaborative and valuable to both providers and consumers.

Achieving a balance between the needs of data consumers and organisational realities of providers will be essential, so the service's dataset attributes must be developed in consultation with both providers and consumers and be contextually appropriate for energy data. The

challenge here is one of ensuring any potential future costs for providers to adapt their metadata to the service's dataset attributes is not prohibitively high and evenly distributed whilst still driving the transformation needed to achieve the EDTF recommendations.

Security, architecture and governance

As noted above within our findings and personas, there are numerous concerns surrounding security and governance that serve as barriers to data sharing and moving to an energy data landscape where data is "presumed open."

Collaboration will also be at the centre of addressing these concerns within the end service. As the related work done in developing the National Underground Asset Register highlights, bringing organisations together is an effective way to define how security concerns can be addressed within a service that strives to make data more open.

It is important to note here as well that the "presumed open" model and the idea of an "openness triage" does note that data should be freely accessible only insofar as there aren't any valid commercial, security, privacy, or consumer interest concerns that would prevent this data from being accessible. Just because there are barriers to accessing data, doesn't necessarily mean that this data isn't "open" as per the "presumed open" model.

Alongside proper attribution on datasets highlighting where access controls and sensitivities surrounding a dataset exist, work in Alpha must consider the mechanisms that need to be in place to manage security concerns while enabling visibility and proportionate access to energy sector data, therein working towards achieving the vision that energy sector data should be "presumed open." With public data, some of these mechanisms already exist, e.g. having to register on a platform to access certain sensitive network data. Considering how security protocols such as these need to be incorporated into the end-to-end service experience - in alignment with open standards and best practice guidelines such as the NCSC cybersecurity guidelines - is an important task for Alpha that will lay the right foundations for security as more data is shared or created and then made visible.

From a governance perspective, we found in user research that there was a lack of understanding surrounding the "openness triage" process, a lack of consistency in triage across organisations, and a lack of a solid, context-specific definition of what data should be open / public / shared / closed to this end. Work in Alpha will need to begin addressing these concerns in a collaborative way, and help the industry and consumers arrive at a shared understanding of what "presumed open" truly means with respect to their data, which can then be complemented by the expected market pull effects driven by visibility of data across the industry in an end service. Here, associated concerns surrounding ownership of data and organisational lineage will also need to be agreed upon and considered.

We believe a central part of this broader triage that the service might enable is a value-driven approach to data sharing, wherein the service and the collaboration it enables becoming a driver for providers to understand the value of their data. This can then be compared against the cost of opening up this data, and the concerns embedded within the EDTF-recommended "presumed open" triage model, as part of prioritising how data should be shared.
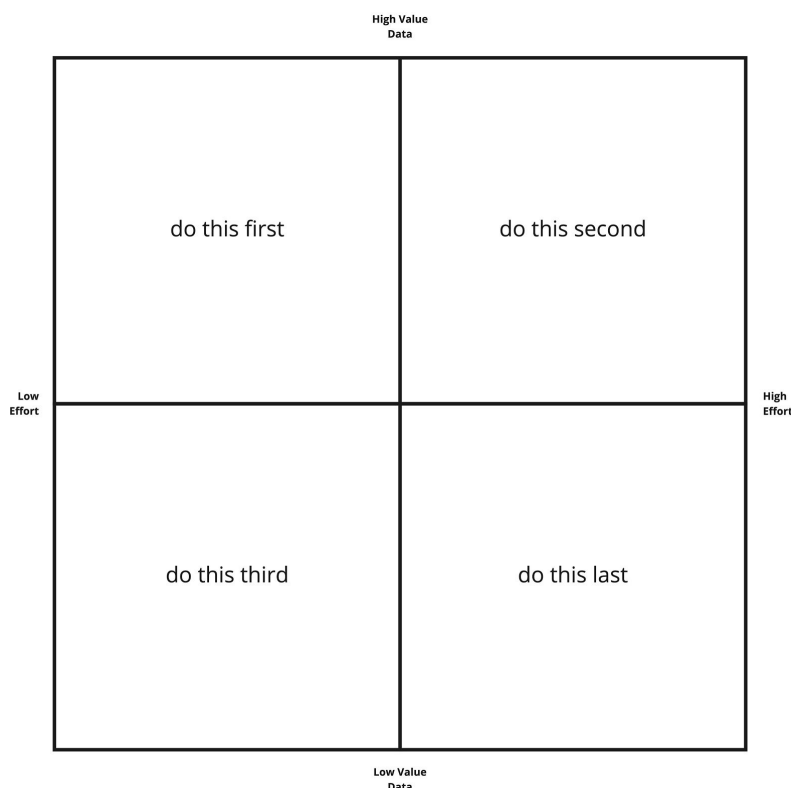
Lastly, given the numerous technical constraints, legacy technology, and the varying digital and data maturity across the sector, the approach to architecting access within the service will need to be tailored to these realities and the broader security and governance concerns outlined above.

Getting shared understanding around these themes of security, architecture, and governance will also be important for the party governing the service in an "oversight" capacity to effectively measure the success of the service, and make targeted interventions where actions are taken that detract from the jointly agreed "presumed open" model for industry data. Needless to say, the approach to governance and security will also naturally need to evolve over time in response to the generation and sharing of new data sources.

Service evolution

At a high level, this is how we currently envision the service evolving over time.

Action	Deliver an MVP that brings together public energy datasets using standard attributes that meet user needs, and has appropriate features to enable collaboration	Evolve standards and service in line with the changing energy system and as more data opens up	Maintain and operate a scalable service for energy system data, acting as an exemplar for other industries to build on
Outcomes	A better understanding of the data landscape Improved understanding of gaps / duplication in data across the datasets Visibility of existing data, improved understanding of gaps/duplication, visibility of innovation, comparability, and collaboration will create a "market pull" for more data to be made visible A solid security, governance, and architectural foundation to build on as more data is identified and introduced to the service	Ensures service keeps pace with rate of sectoral change and drives innovation, leveraging the infrastructure, visibility, and collaboration established by the MVP	Service drives agile regulation and sectoral transformation Service becomes a blueprint for other sectors to emulate Service can evolve to enable cross-sector data sharing

Over time, this approach has the potential to drive sectoral transformation, with rich, semantic, and interlinked data driving innovation and helping achieve the vision of a smart, flexible and decarbonised energy system, and potentially an approach which can then serve as a model for related sectors such as water and other utilities.

The following roadmap presents a more detailed overview of how we expect the various components of the service to evolve through Alpha and beyond.

				MVP			
		Alpha			**Private / Public Beta**		**Live**
Search / Utility Maturity	Basic attribute keyword search		Categorical / Faceted Search capabilities		Schema Projections	Knowledge base like search	
Data Access Capability	No Proxy Access (External Redirect)		HTTP / File (Proxy)			Streaming / Subscriptions (Proxy / Hosted)	
Catalogue Maturity	URL List	Basic Metadata	Basic Metadata Grouping		Canonical Metadata model	Full Ontology	
Underlying Data		Public Data			More data opening up	Full data visibility	

Catalogue Maturity	Access Capability	Search Capability
URL list	External redirect	Basic search based on provided dataset titles / strings; the utility this provides is that data is centrally visible, but not discoverable, searchable, or understandable in a meaningful way.
Fully attributed datasets and canonical metadata model	HTTP / file proxy	Categorical / faceted search capabilities, which means data is organised into logical categories based on relevant attributes, and users can search based on these relevant attributes.
Full ontology	Streaming (Proxy / Hosted)	Knowledge-base like search across dataset attributes, with well-defined relationships that users can search based on.

The roadmap offers room for a valuable discussion on where the MVP should lie in this evolutionary approach; in line with the table above, we believe that the MVP should have fully attributed datasets and an HTTP / file proxy access capability to be valuable for users. There has often been an assumption that a list of URLs / datasets on their own will offer sufficient utility for data consumers, and that this should form the basis of the MVP.

As a principle, MVPs are supposed to be usable, valuable, and feasible. In other words, an MVP should seek to deliver the highest amount of value at the lowest possible cost. While an MVP is not supposed to be a fully-featured version of a service, it does need to meet the three value criteria above.

- In being usable, the MVP will need to have the appropriate functionality and user experience to ensure that the service allows users to meet their objectives without experiential barriers.

- In being feasible, the MVP will need to lay a strong set of architectural, governance, and security foundations for the service to evolve from - as outlined in detail within Recommendation 5.

- Most importantly, in being valuable, the service should be preferable to the current experience of finding and accessing data. It is unlikely that a URL list will be sufficient to meet these three criteria of value based on the information uncovered in this Discovery; however, one can certainly test this assumption in Alpha.

As our user research highlights, users don't view discoverability and understandability of data as separate concerns. Users outlined that discoverable means the ability to find if data exists or not, which - in their view - fundamentally relies on the data being understandable, i.e. that it is clear and well defined, making it easily understood by a wide range of individuals with different levels of knowledge. Building off of this, searchability is a functional concern, where users want to search for data based on terms that are logical to them.

To this end, we believe that a mere URL list without a well-developed set of attributes is unlikely to deliver value for users since it wouldn't have the appropriate context that users need to truly obtain value from data, and would not enable search capabilities of sufficient quality. Given that the service exists alongside numerous levers in creating the right conditions for data visibility and market pull, it is very important to get the MVP "right" to drive the necessary evolution on the journey to open data.

hippodigital

Service value

To understand how the service will deliver value across user groups, a useful way to think about it is in terms of learning curves and incremental organisational change.

Our research has shown that currently there is a very steep learning curve to overcome to understand and use energy data and this is impeding people from getting value from the data without significant investment and cost. We recommend the work in Alpha seeks to prove that "flattening" the learning curve will result in reducing costs and improving the value attained through the data being discoverable, searchable and understandable.

This is how we envision our proposed approach to the service delivering value across our core user groups.

User group	Short-term benefits	Long-term benefits
Consumers with high understanding of energy system data	Provide the means for collaboration to increase visibility of missing data	Help them identify datasets they might not have been aware of, and spend less time unpicking data by simplifying its complexity
Consumers with low understanding of energy system data	Help them identify and understand the extent of datasets available, helping them increase their knowledge of the system and its data	Provide the means for collaboration to increase visibility of missing data
Providers	Understand how and why their data is being used, which was/is an important need across providers, and be able to collaborate on the journey to open data Drive organisational transformation, by introducing a standardised way to understand and manage data requests, replacing ad-hoc processes for managing requests and moving everyone towards a presumed open model. Smoothen the transition from an operational perspective, ensuring change in architecture, security, and governance is delivered in an evolutionary way across the industry	Understand what data they need to provide publicly, via a better understanding of demand

Workstream alignment

It is important that the various workstreams within the Modernising Energy Data programme are aligned as work on the service recommended here moves forward to prevent duplication of effort and effective service design and delivery.
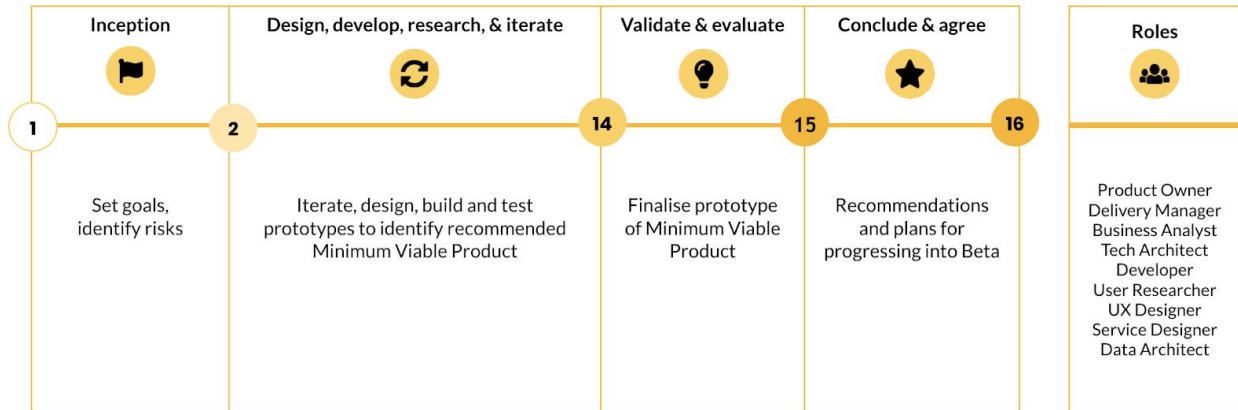
Most importantly, given that users don't necessarily view visibility of and access to data as separate activities, we strongly recommend that the various discoveries running as part of the Modernising Energy Data Access competition are aligned with the Energy Data Visibility workstream moving forward. This will minimise the inevitable duplication of effort across these programmes, and ensure that they work towards delivering a holistic, end-to-end service that is scoped based on data consumers needs. Here, other workstreams and "building blocks," such as the Digital System Map, might also be seen as part of the broader concern of visibility, whereas activities that generate new data such as the Asset Registration Strategy will have to account for the collaborative approach to standards being driven by the service.

To this end, we have started forming an initial view of how various workstreams across the Modernising Energy Data programme might relate to the broader foundations underlining our service vision. Given that this is a view formed over 6 weeks with a very high-level understanding of the programme, it must be subject to further scrutiny and change in Alpha.

Users can easily discover what data is available in a usable service	Digital Systems Map		Modernising Energy Data Access and Energy Data Visibility Service
As more data becomes available, it is made visible in a way that is meaningful to users	Asset Registration Strategy	Cross-Sector Digitalisation Strategies	Presumed Open Triage and Levers for Data Sharing
Data is described, categorised and attributed in a meaningful and consistent way for users to understand	Evolving data catalogue providing a "single source of truth" for sector data		

High level Alpha plan

Based on this Discovery's learnings and recommended prototyping and outcomes for Alpha we recommend that an Alpha of 16 weeks is needed supported by a team with the roles detailed below.

Inception	Design, develop, research, & iterate	Validate & evaluate	Conclude & agree	Roles
1 → 2		14	15 → 16	
Set goals, identify risks	Iterate, design, build and test prototypes to identify recommended Minimum Viable Product	Finalise prototype of Minimum Viable Product	Recommendations and plans for progressing into Beta	Product Owner Delivery Manager Business Analyst Tech Architect Developer User Researcher UX Designer Service Designer Data Architect

hippodigital

Risks and dependencies for Alpha

Risk	Mitigation
User research recruitment	Research participants were not available to the team within the first 2 weeks of the project. The team was able to adjust the delivery plan to support research once availability was confirmed. In the Alpha phase, visibility of research participants who can be contacted should be available immediately, allowing the team to make contact and engage during inception.
Programme workstreams	There are several projects which have been commissioned due to the Energy Data Task Force report. These projects are beginning on various dates and due for completion in their Discovery phases by July 2020. It is recommended within the Alpha phase, the workstreams are more aligned and working collaboratively to enable shared learning and mitigate duplication of work.
COVID19	The team have worked collaboratively and been engaged in the project, whilst working remotely. Hippo Digital was able to provide confidence if any team member is sick, Hippo Digital would provide relevant resources to cover the team member. Working remotely and risks associated with the pandemic will need to be considered within the Alpha.
Agile maturity	There is a mixed level of agile maturity across stakeholders involved within the Discovery. Alpha team will need to be aware of this and provide support and understanding to stakeholders where required.
GDS maturity	There is a mixed level of GDS service standard understanding across stakeholders involved within the Discovery. Alpha team will need to be aware of this and provide support and understanding to stakeholders where required.

Appendix

Glossary of terms

Common Information Model
Refers to the Common Information Model standard for network assets that various distribution and transmission network parties are aligning to collaboratively via the Energy Networks Association

Metadata
any data that helps describe other data

Business Glossary
a glossary explaining what terms and definitions in a body of data mean in a specific context

Lever
A lever acts on the system as a whole to create a change. If you apply a lever in one place something moves elsewhere. Use of the word lever also infers that a change should be happening in a timely manner.

Research and analysis artefacts

The participants

The following are individuals we conducted qualitative interviews with -

Role	Use case
Head of Short Term Optimisation & Dispatch	Uses data from a range of sources to help make trade decisions.
Modelling Development Manager	Data is used to look at every aspect that makes up the electricity demand currently, heat demand and potentially transport demand in the future.
Consultant	Provides support to organisations who need help understanding how to comply with regulation. Uses data to do modelling to inform client asset purchasing decisions
Data Management	Operates the balancing mechanism. Is a major data provider of settlement data.
Business Lead	Organisation operates the whole transmission system. Part of which involves looking at how to work closer with the distribution operators.
Head of Policy	Understand through qualitative research what trade members really want - members want more information available from the control room so you can properly understand: ● why things are dispatched differently, where the grid is going with the services it needs so we can plan portfolios and how much each service is going to be needed ● how much data is there to monitor National Grid Electricity System Operators's commitments - e.g. skipping and meeting obligations for using distributed assets on the grid.
Policy	Looking into how the industry can move from Distribution Network Operators to Distribution System Operators - a much smarter system that makes better use of flexibility.

Statistician	Analyses country level datasets in order to understand and present them back for policy and public use. Tracking where the country is relative to its energy and climate change ambitions in relation to various targets for where we want to be. e.g. refreshing our targets in line with net zero. Evaluate where the country is and needs to be in terms of low carbon and renewable electricity.
Analyst	Uses data in order to carry out heat demand mapping for heat network feasibility studies, or for forming energy master plans for developments. It's important to know about low carbon heat sources in an area so they can review where there is a supply and a demand to evaluate how the two can link together`.
Digital+Energy project manager	Energy master planning. City decarbonisation. Evaluating what interventions cities need in order to reduce their carbon footprint. Includes mapping wind farms or hydro planning to where people are. The maps produced show super and lower output areas in a city or nationally.
Researcher	Contracted by National Grid ESO to develop a monitoring service for GB photovoltaic generation on the basis of half hourly metered data. The data is then also taken up by the electricity industry as part of informing a benchmark for solar out-turn and that helps traders and operators down at a distribution level to inform their management of the system.
Researcher	
Researcher	
Macrocosmic statistician	Data about the energy industry is required to inform the Producer Price Index which is created by the Office for National Statistics.
DNO Commercial Manager	Provides low carbon energy sources to industry, uses data to create business cases for new assets, as they need to understand how an asset will impact the local network.
Technical Director	Collects and analyses granular data from meters connected to individual assets to help optimise energy usage
Data scientist / Co-founder	Uses data for forecasting what energy will be generated from renewable sources both solar and wind.
Enterprise Data Architect	Uses data to inform decisions about the running of the gas network.
Head of DSO	Uses data for making informed decisions about a number of aspects relevant to generation such as weather, staffing for a storm, planning in the local authority region.
DSO Digitalisation & Data	Uses data for making informed decisions about a number of aspects

Manager	relevant to generation such as weather, staffing for a storm, planning in the local authority region.
Macrocosmic statistician	Data is used to inform the monthly GDP. The energy industry is just one industry that data is collected from.

Summary of quotes regarding who should manage the data catalogue and why.

Department	Number of users	Quotes
Ofgem	4 users	'They have the most market knowledge, being the regulator, and this will be essential to providing a data catalogue that is useful to the businesses in the market. They should be using user driven feedback and agile practices to achieve this. User stories, short iteration and deliverable increments should be mandatory.' - P41 'Because they deal with all other energy matters.' - P22 'They can compel companies and they could also use it to help set policy.' - P08 'Regulator as the own makes most sense.' - P16
BEIS	5 users	'The organisation needs to be very experienced in the full life cycle of energy from infrastructure and production to consumer experience and consumption. However, this is a task that would be useful for a variety of sectors so it should be owned by a body that operates across sectors too.' - P51 'Industry knowledge and understanding of likely applications of the data.' - P43 'Data should be managed by domain experts. For me, BEIS is the main source of energy data, for national and sub-national energy statistics.' - P13
ONS	19 users	'Their longstanding experience as data curators. They have full sight of the level of effort, and are dedicated to providing that kind of service. Their remit is clear and won't change. They are well known and trusted. BEIS would need a whole new capability to run it, and may find themselves subject to a change in focus/ funding. OfGEM would also need to develop a new capability. They would need to drive the policy / regulation to require data to be made available, and that's where their focus should be, in collaboration with established data curators.' - P06 'Because they should be in charge of the data being collected, calibrated etc.' - P54

		'The have experience in, for example collating and presenting statistics and probably is the organisation which would have the last steep learning curve.' - P52 'Energy data could become part of the national record and curated (openly) for the public good in ways consistent with other public good data.' - P50 'ONS are seen as more independent than Ofgem or BEIS, as those parties have goals they wish to achieve that raw data may or may not support.' - P32 'Neutral.' - P49
Other (neutral party)	15 users	'A neutral party who can manage a third party to do the data collection and processing while ensuring the data remains public. I think the connection of a third party to actually do the data processing would be essential to ensure it stays up to date, given BEIS can be restricted by resource.' - P36 'Any public org like BEIS/Ofgem but with a dedicated team of people who understands energy AND DATA.' - P46 'I don't know - the identity of the party is less important than the rules governing the management.'- U10
Decentralised	3 users	'Decentralised management where each data owner is responsible for the management of their own data items in the catalogue.' - P48 'The individual catalogues would be best managed by the data owners. Some central body would need to manage the central search interface and have a governance role over the data owners - OFGEM or BEIS might have the right levers to enforce compliance to standards and maintenance of individual catalogues.' - P03 'The parties who generate it, as long as they share standards and identifiers then there are many ways to bring this together.' - P42

hippodigital

Evidence for data set characteristics

Characteristics

The table below has been created to set out characteristics and evidence for what users wanted to know about data sets in order to ensure the data set is right for the task they want to complete.

Characteristics	Evidence
Description	"It's an issue of not having sufficient detail surrounding what the data means, and most people in the industry will know that means because they're in the energy sector and are quite techy - but we're not very good in the energy industry is making things simpler and clearer." - U09 "Yeah, some form of catalogue would be. I mean, I think what's missing from the energy industry is any, they're not very good at documenting things. So unless you've worked in the energy industry, it's almost impenetrable in terms of discovering data and understanding terminology. Acronyms, they're a big fan of acronyms. Yeah, so until you actually kind of get involved in the work and start speaking to people, that there's not much opportunity to discover how things actually work and where the data sets are. And yeah, I think I think it's probably worth noting that Elexon and the BMRS platform, they are quite good at making data visible and reasonably well documented. not always the case, you know, a lot of the time, the date feed will only make sense to you if you already know what it means. And yeah, and I guess the MET Office is pretty good at that kind of stuff as well with the Midas data. That's that's very well documented, but it's also an incredibly mature data set that's been around for decades." - U12 "Where it's not standardised, you'd have a data dictionary explaining what's different and what's similar across datasets, so you can fill in the gaps. So it's having that consistent kind of rule about how it should be just repeated over time.". - U09
Organisation lineage and ownership	"Understanding where the data has come from" - U09 "Which organisation made it or what sources is it from" - U10 where that data has come from and what's behind that data - U08 'Why was it created, by whom, how, when.' - P03
Aggregation method	"How it's been processed to form the end data set" - U10 "That the data is high quality and can reliably be disaggregated to a Scottish level" - P20 "The data quality, the assumptions behind the data." - P13 'How and if it has been processed.' - P01

Update frequency	"That means that, you know, you know how often it's being updated. And if they keep a record of that... but there's no sort of note next to the data saying if any of it has been changed or updated, etc. - U09 "When will it be updated" - U10 'Source, frequency, trustworthiness.' - P08
Last updated	"Sometimes you'll get data through and it might tell you what year it was taken." - U09 'When it was last updated.' P03
Confidence/Quality level	"The quality of the data collected, and the methods used - e.g. actual metered data vs. surveyor data." - U09 "The higher the forecast content, the less you trust the data. A lot of industries, you can get away with it. But with electricity, everyone's an expert, so weather, energy - there's a lot of variables. And there's a lot of scenarios that are predicated on knowing the data really well, so forecasts mean very little without the quality of the underlying data being assured." - U21 'To what standard and accuracy (spatial and attribute), how complete is it.' - P03 'That is trustworthy - i.e. can be relied upon and has integrity.' - P52 'Accuracy and completeness, including relevant metadata.' - P40 'Confidence in data reflection of reality.' - P39
Format	"Generally, my team uses SASS - and it's quite easy to convert from CSV to SASS." - U08 "Quite a lot of datasets that they had are just not in the same format. So whether that's literally it's an Excel but they're all in different formats or they're not in a state that makes it easy for us to use it in our software." - U09 "We have a variety of tools that allow us to work with almost any file format." - U10
Units	"They won't all publish geographic location in the same way e.g. longitude, latitude." - U17 "Weather sets are really nice, it's always UTC but energy sets you don't know what time, so you have to just see. if it's off for an hour, you know the energy data is on BST." - U17 "Different formats of customer addresses - scottish addresses tenement blocks. different ways of expressing the same address." - U18 'Accurate columns labels (including units)' - P33

| Risks	"But it's not fully transparent. And because you know that there's inconsistencies, and it's not recorded, which makes it hard for us to use it. It'd be fine if they made the mistakes clear, because we'd know where we're making guesses." - U09
	"Any caveats about its accuracy." - U10
	"More data is dangerous if you don't know what to do with it" - U20
	"So, mainly it's about our obligation to public safety. So, there have been cases in the past that people have taken the data we've provided and made some very ill-informed assumptions. Ultimately, what dictates the energy system is the laws of physics. So, if they arrive at the wrong answers from the data we provide, that's got a serious risk." - U19
	'Its limitations in terms of applicability to deployment of local energy projects.' - P28
	'How reliable and up to date it is and any gaps.' - P06
Contact	"Contacts for people and what's included." - U14
	"I guess another thing is having that collaboration and that information exchange to enable collaboration across the parties using the data." - U08
	'Mechanism for feedback, which means that you could kind of work around the same data with other parties.' - U09
Keywords	'Keywords and standardised taxonomy are key.' - P06
How to access	"Free or not and how to access it." - U14
	'Continued (public or private) availability and general quality.' - P50
Licence conditions	"Can I use it for my own purposes? Can I use it instead of collecting my data and save money?" - U17
	'License terms etc.' - P03

Discussion guides

Our discussion guides were used as a guide to carry out our semi-structured in-depth interviews. These are in separate files and attached to the report.
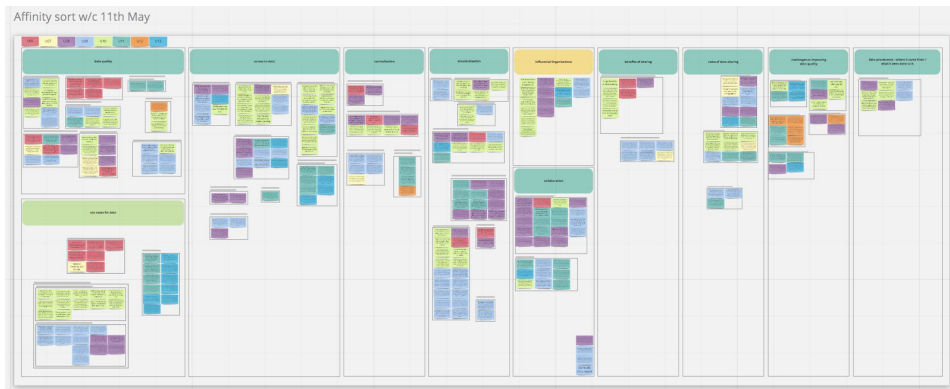
Analysis

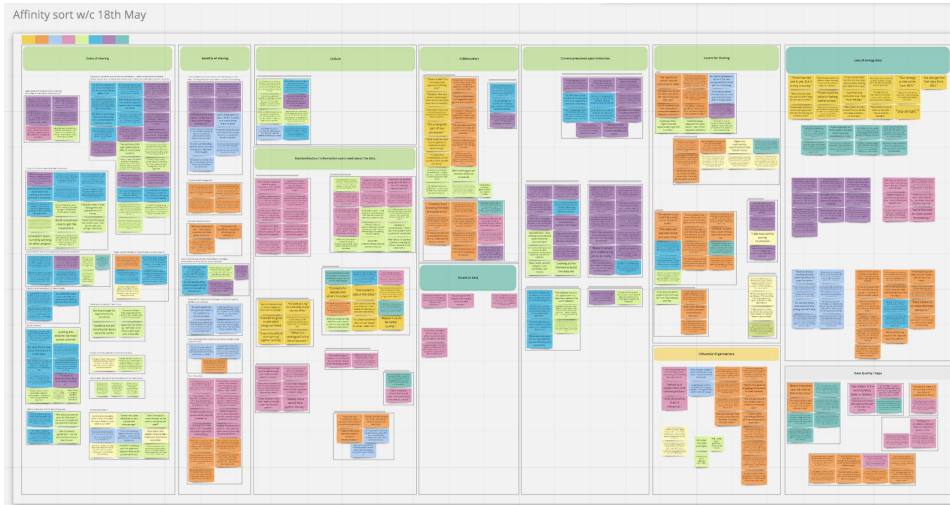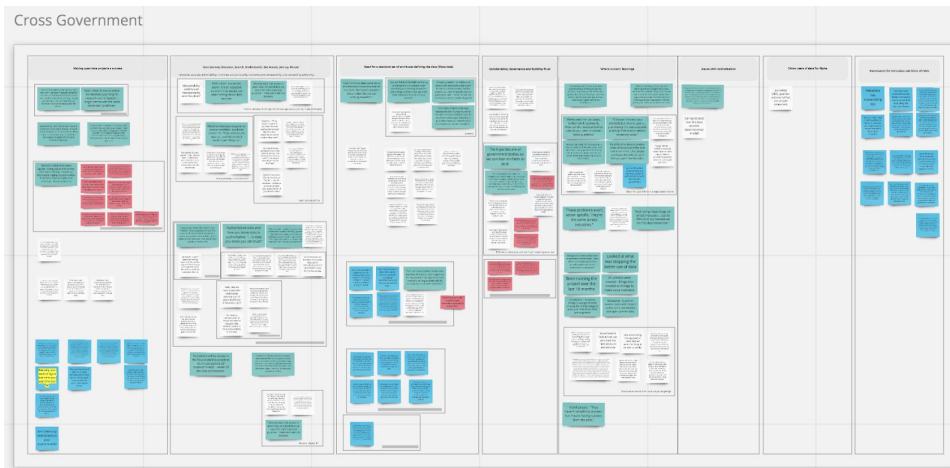ONS have access to the source Miro boards -

Round 1
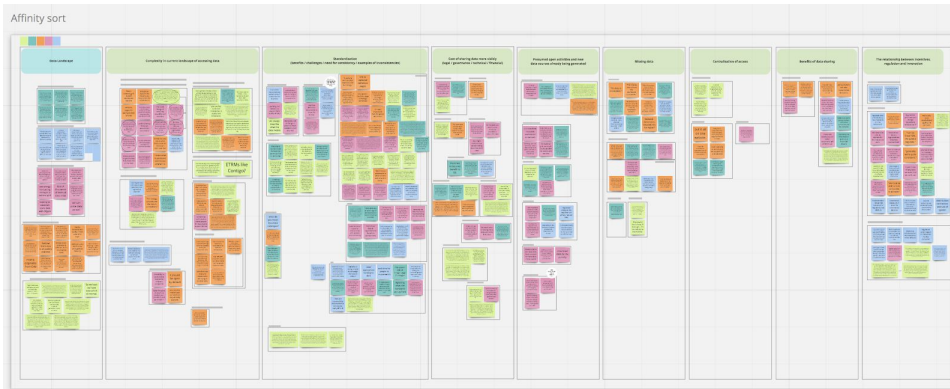
Round 2

Round 3

Cross Government

Research playbacks

These are in separate files and attached to the report.

Survey

The insights are within a separate file and attached to the report.

Personas

This table outlines the persona type, their needs and how it relates to our insights below. For more information see the attached personas and user journeys document.

Persona	Needs
Oversight: As a person overseeing the service, I need	A transparent set of Key Performance Indicators to show movement towards desired sharing outcomes. Visibility of data, by type of data and organisation sharing Method of access for data by type of data and organisation sharing Requests for new data sources and timeline to get access (via various stages in process) - Assumes there will be an audit trail of requests for data sources and an escalation process to resolve any issues
	Report on gaps in the datasets by organisation (and ability to compare across organisations) to enable a discussion around why the differences exist
	Track quality issues with the data & expected path and timeline to resolution
	Evidence to be able to challenge policy/ regulation as necessary
Energy Statistician: As an energy statistician, I need	Regional level data in order to get a more accurate picture of how the country is meeting their net zero ambitions
	To know what data is available and how to access it so that I'm not missing valuable information.
	The data has context and definitions in order to understand what the data means.
	To understand where the data set has come from and what processing the data has been through, so that I can trust and have confidence in using that data.
Economic Statistician: As an economic statistician, I need	Confidence that the data is available when I need it every month
	An operational point of contact to resolve any immediate issues with quality of the data supplied

	To know the provenance of the data (collection method, processing steps done prior to receiving it, assumptions made) (currently all published in a quality assurance document)
	An Operational Point of contact to help refresh or improve the quality of the data over time. There are regular review points in the cycle (2025 for next large review)
	To know what data is available in the event that I wish to refresh my data sources
	To know if there is a cost for the data so that I can make an informed decision about if the data set is worth the cost
Academic / researcher: As a researcher, I need	Better definitions of datasets to aid their discoverability, searchability, and understandability and to better understand the current data landscape
	Clear owners assigned to datasets to enable collaboration, accountability, and feedback loops
	The industry to "start somewhere" and facilitate exploration and innovation, which will lead to natural improvements in data quality
	More granular data on renewable assets on the grid / distribution networks
	Where data is aggregated, I need definitions on methodology to understand how the data has been processed
Provider: As a provider, I need	An approach to standards that doesn't present significant operational or financial overhead
	Leadership, guidance and collaboration on the journey to open data
	Better innovation investment structures to drive investment in open data
	To understand how my data is being used and to ensure that complex data is not misinterpreted and misused
	To control and manage access to sensitive data as a means of addressing security risks
	A consistent triage process across similar organisations to understand what data is and isn't "open" and consistency in sharing
Trader: As a person working in trading, I need	To understand how my data is being used in order to make sure it is being understood and used correctly

	Easy access to data from a number of sources in order to make informed half hourly trade decisions	
	Data to be kept consistent and up to date in order to make the data easy to understand and consume	
Commercial manager: As a commercials manager, I need	Timely access (self-service) to rich, consistent and accurate distribution network data to support investment decisions	
	Consistent data and processes for connection across a wide range of DNOs to ensure a consistent experience for customers.	
	The sector to have regulatory leadership so that open data and long-term changes can be achieved.	

Possible roles for future user research

The roles in this list were identified as future research candidates.

Role	Type	Level of understanding of Energy system data (assumed)	Data Expertise (assumed)	Rationale
Charity (climate or other user or energy data)	Consumer of energy data	Low	Low	Understand if we intend to target these roles with our solution (ONS recommendation)
Journalist	Consumer of energy data	Low	Medium	Understand if we intend to target these roles with our solution (ONS recommendation)
Innovator - outsider of energy	Consumer of energy data	Low	High	Identify the specific needs of data specialists wanting to innovate within the energy sector to improve sector knowledge and combine with other data sources
Generator	Data provider	N/A	N/A	To provide a breadth of data producer insight.
Paid for data provider	Data provider	N/A	N/A	To explore the appetite of cataloging commercial data within the service

In addition a broader range of stakeholders would need to be engaged to cover more solution requirements (Security, architecture, legal, operational stakeholders).

Low-fidelity prototypes

The attached slide deck (separate file) contextualises some of the key service assumptions through the use of low-fidelity prototypes. This will provide a useful basis for work in Alpha to build on.