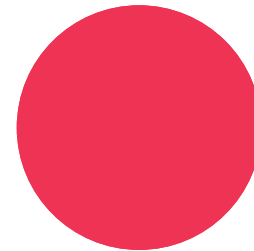# Energy Data Visibility Personas and user journeys

Discovery

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

Oversight

Measures whether the desired outcomes are being achieved by the service associated regulatory and policy levers

"If one [Organisation] does [share a set of data] and one doesn't you would look at why not. Is one giving sensitive data? or is one not contributing open data?"

Context of use

- Independent party responsible for overseeing the system to ensure that the outcomes required by the policy and regulation are being achieved
- Reports on progress towards goals
- Understands why differences exist between the data shared by the various organisations in the system
- Takes corrective action if rate of delivery against outcomes looks to be compromised.

Needs

Transparent set of KPIs to show movement towards desired sharing outcomes:

- visibility of data, by type and organisation sharing
- method of access for data by type 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

Other Duties

- Role may also include helping to promote the output from the service (discoverable, searchable and understandable data) across the industry
- The individual from the role should be from a neutral and independent organisation whilst understanding data in order to meet the service user needs.

Regulator ●━━━━━━━━●━━━━━━━━● Policy

Energy Statistician

This person 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.

"I'm a statistician, and not an energy expert, and it's not easy to understand what's out there… we've not tapped into these rich datasets enough because the lack of knowledge that's held us back."

Context of use
- Data is used to understand renewable energy sources and prevalence of low carbon heat generation
- Uses public sources such as BEIS and Ofgem. Carries out a Google search or talks to people for additional sources of data
- Tracks how progress is being made towards targets that will work towards meeting the energy and climate change ambitions and goal of net zero 2045
- Creates public reports and dashboards used from internal policymakers to government and the wider public

Needs
- Needs the regional level data in order to get a more accurate picture for how the country is meeting their net zero ambitions
- Needs to know what data is available and how to access it so they are not missing valuable information
- Needs the data to have context and definitions in order to understand what the data means
- Needs to understand where data has come from and what has been done to the data set in order to enable them to trust and have confidence in using that data

Pain points
- Data is currently aggregated at GB level, which makes it difficult to disaggregate data to a specific regional or national context
- Data is difficult to find, even when a source of data is known:
  "I've been doing this for two years. And I think it was just last week, I came across something that BEIS publishes, and has been there for ages. And I'd never I never stumbled across it. And I've been trying to get in touch, when trying to come to us with often try and work out where some of this data where we could go to this data published all along by BEIS."
- If data lacks definitions, it is difficult for someone who is not an energy expert to understand what the data means
- Definitions are not consistent across sources leading to increased errors

Provider ●————————————————● ○ Consumer

High understanding of energy system data ●————————————————● ○ Low understanding of energy system data

Data specialist ●————————————————————————————● Non data specialist

Energy Statistician

This person 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.

"I'm a statistician, and not an energy expert, and it's not easy to understand what's out there… we've not tapped into these rich datasets enough because the lack of knowledge that's held us back."

Steps

| Finds a data source through conversation or department knowledge of data sets | Spends time to understand the data and what it is telling them | Disaggregates data to try view it at a country, then regional level | Work to match data sets and complete gaps in order to combine and use data from different sources | Publish data in app show this can be used by policy people and the public. |

Insight

| "Finding stuff on the BEIS website is really, really tough. If you stumble across the right landing page, you're fine, but finding stuff is really, really difficult." | "Not having sufficient detail surrounding what the data means… most people in the industry will know that means - but we're not very good in the energy industry is making things simpler and clearer." | "[The reason data can't be published at a granular enough level] is that most data is collected at an aggregate level, e.g. most meter readings are aggregated into a top-down view at GB level. They don't really have that granularity embedded in the way they're deriving that information, so it's quite hard for them to make that information more granular since it's not built for that purpose. | "Inconsistencies in units leads to increased errors, and it's also quite time-consuming to bring it all together." | |

Opportunities

| Improve the discoverability of data sources | | Find out whether source data could be made open / whether it exists. Explore approaches to mitigate risks and make data open. | | This could be done in one place, with data broken down by country/region to save duplication of effort. |

Economic statistician

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

"So, with my job, I've got 3 days to look at the whole economy. I've really not got that much time to look at electricity - 30 minutes -, and I have to do that month after month on a very tight schedule."
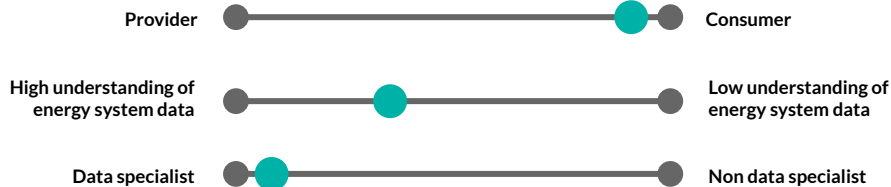
Context of use
- Calculates GDP to an agreed process and supplies to a number of sources including the bank of england
- Raw data is supplied by a number of sources including BEIS for the energy data. Energy Data supplied by BEIS is collected by a number of methods including survey. Some processing is carried out by BEIS.
- Further processing is done by the statistician to produce the output each month
- Energy data is used alongside data from other industries to look at the wider economy

Needs
- Confidence that the data is available when I need it every month to replace current Service Level Agreement (data is supplied by BEIS via email to an agreed timeline every month)
- Operational Point of contact to resolve any immediate issues with quality of the data supplied
- Provenance of the data (collection method, processing steps done prior to receiving it, assumptions made) all published in a quality assurance document.
- 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

Pain points
- Relies on the same data they have always received as have limited time to look at energy data alongside all of the data from the other industries making up the economy.
- Changes to the industry over time mean that renewable data is a large part of the generation mix. The raw data is not split out by generation asset type (renewables, nuclear etc)
- Autogeneration (people that generate for themselves then sell the excess to the grid) are not included in the raw data they are forecasted
- Level of forecast (calculated) versus actual data. Data is not granular enough for example to account for whether a nuclear power station is out of action for maintenance (may be possible to meet this need using existing BMRS data)
- Would love to receive cost to make a kW hour of electricity

Provider ———————————●— Consumer

High understanding of energy system data ——●———————— Low understanding of energy system data

Data specialist —●———————————— Non data specialist

Economic statistician

User journey for finding and validating a new data source

Steps

New data source required to support existing analysis	Identify potential suppliers of data by raising a market engagement notice (internal process)	Or, Investigates new sources of data through network of contacts	Obtains a sample of the new data source	Evaluates new data source to see if it is fit for purpose (checking for consistency with existing data source)	Make decision whether to use new source of data

Insight

"Of course, you know, if that was a much wider catalogue available where you are actually able to see what else is available out there just beyond what you have currently using. I mean, that would be great."

"Today procurement who run the marketing engagement notices spot duplicate requests."

"Key thing will be to have contacts of people that you can contact if you need help."

"What it is, where does it come from probably some quality information, sample sizes. Source - Is it a survey preparation, if it is free or not"

"I guess probably also some information about how to get access to it."

Opportunities

Improve the discoverability of data sources/ removing waste and cost

Access to a person is also helpful when attempting to find the right data source

Important attributes of the data: What it is, where it comes from including whether its a survey, quality info, price, how to get access, frequency of update

Economic statistician

User journey for completing monthly analysis of GDP

Steps

Receive energy data set from BEIS (monthly via email) in line with service agreement	Check quality of data and raise any issues to supplier	Complete statistical analysis (combining energy data with all other data)	Publish key indicators	Discuss & explain output with key stakeholders (Bank of England etc.)	Repeat every month

Insight

"The prices they use to calculate renewable volumes is sub-optimal. So the data from 2010 was looking at obligations purchases (renewable permits bought) as a proxy for price..that is a very misleading proxy now"

"You should get to an idea of the capital cost and maintenance cost for building a wind farm. We don't have that data at the moment, but we know BEIS does."

"So the other opportunity is when prices are updated, which is also every 5 years. So we currently use data from 2010, and we're trying to use data from 2015."

"I'll be asked a very direct question, but the Bank of England is, do you trust your data? And that's as simple as that. And in that the higher the forecast content, the less you trust it."

Opportunities

Discovery alternative sources of data in order to make improvements to the analysis

Having a route to raise data quality issues is a key part of using the data

Terms needs to be understood across the energy industry and by other users of the data.

Understanding the source of the data & processing means that questions can be answered

Data granularity is an issue in places (gap in the generation data for "autogenerators" could alternative sources be found?

Academic / Researcher

Use 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

"The vast majority, 80% of my time is finding the data and putting it in the right shape. And then typically, once it's in the right shape it's easy to query."
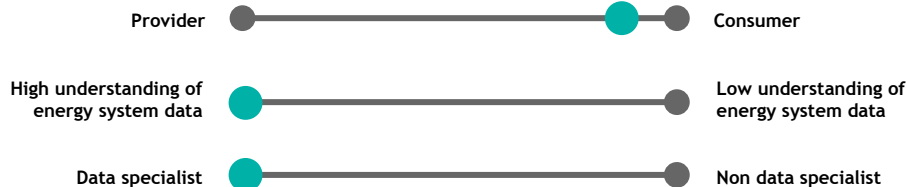
Context of use

- Use aggregations of operational metering data, open-source or commercially-obtained PV generation data, and weather data to better model and forecast PV out-turn for the United Kingdom
- Find datasets through Google, or via contextual knowledge built through conversations in the industry

Needs

- Need better definitions of datasets to aid their discoverability, searchability, and understandability -
  and to better understand the current data landscape
- Need clear owners assigned to datasets to enable collaboration, accountability, and feedback loops
- Feel that the industry needs to "start somewhere" and facilitate exploration and innovation, which will lead to natural improvements in data quality
- Need more granular data on renewable assets on the grid / distribution networks. Where data is aggregated, definitions on methodology are needed

Pain points

- Existing descriptions of the data are very context-specific, and not effective at unlocking the utility/value of the underlying data beyond that context; this also translates into difficulty in finding valuable data that may already exist, e.g. on the Elexon portal / BMRS
- The process of getting access to granular asset owner (i.e. owner of low-voltage assets) data involves lengthy conversations with numerous organisations
- Given that similar data is often aggregated across different organisations, users often have to unpick data overlaps and duplicated identifiers, and fill in gaps in the datasets. This increases the amount of time spent in *cleaning data* vs. *utilising* energy sector data

Provider ———————————— Consumer

High understanding of energy system data ———————————— Low understanding of energy system data

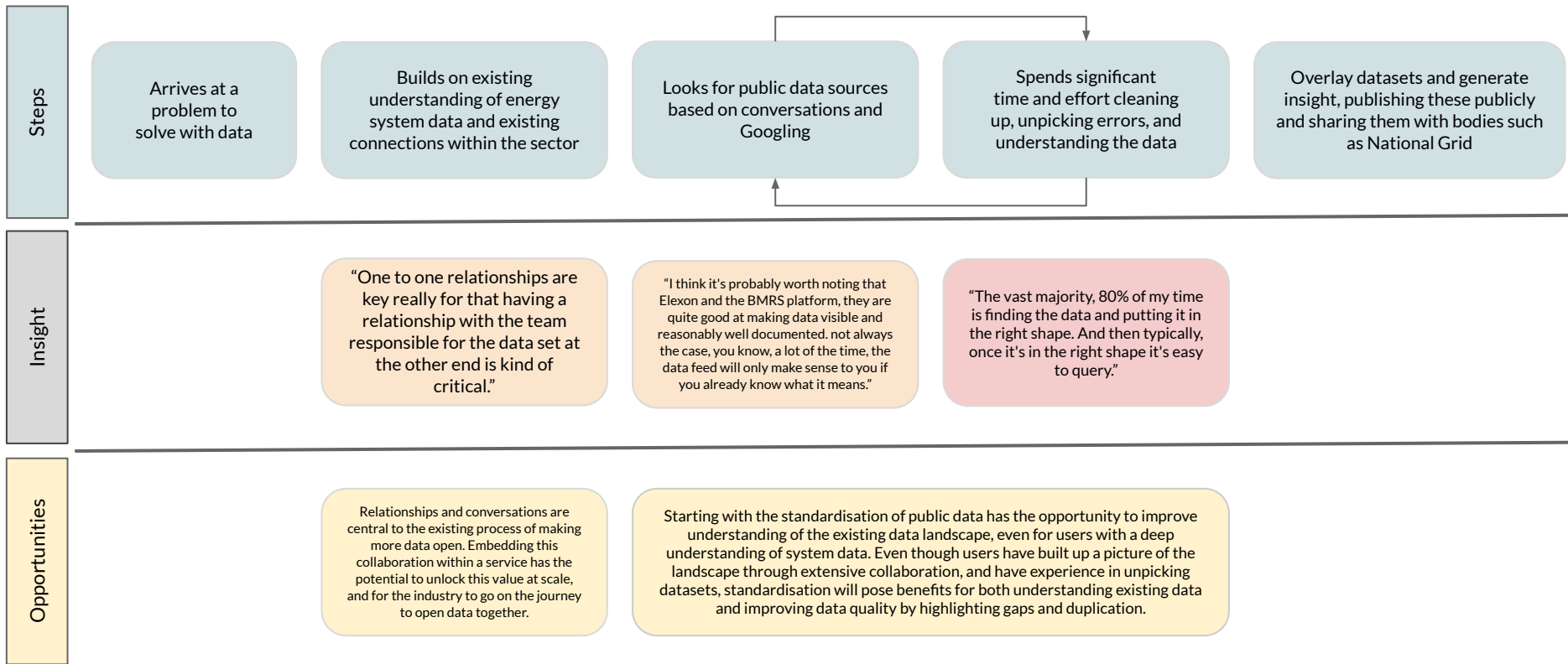Data specialist ———————————— Non data specialist

Academic / Researcher

Use 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

"The vast majority, 80% of my time is finding the data and putting it in the right shape. And then typically, once it's in the right shape it's easy to query."

Steps

| Arrives at a problem to solve with data | Builds on existing understanding of energy system data and existing connections within the sector | Looks for public data sources based on conversations and Googling | Spends significant time and effort cleaning up, unpicking errors, and understanding the data | Overlay datasets and generate insight, publishing these publicly and sharing them with bodies such as National Grid |

Insight

"One to one relationships are key really for that having a relationship with the team responsible for the data set at the other end is kind of critical."

"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 data feed will only make sense to you if you already know what it means."

"The vast majority, 80% of my time is finding the data and putting it in the right shape. And then typically, once it's in the right shape it's easy to query."

Opportunities

Relationships and conversations are central to the existing process of making more data open. Embedding this collaboration within a service has the potential to unlock this value at scale, and for the industry to go on the journey to open data together.

Starting with the standardisation of public data has the opportunity to improve understanding of the existing data landscape, even for users with a deep understanding of system data. Even though users have built up a picture of the landscape through extensive collaboration, and have experience in unpicking datasets, standardisation will pose benefits for both understanding existing data and improving data quality by highlighting gaps and duplication.

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.

"So, what I'm looking for is a template, so for BEIS and ESC to come together and say 'this is a template; this is how we want a GDN [Gas Distribution Network] to share their data.'"

Context of use

- Manage sensitive data that is considered central to the operation of Critical National Infrastructure
- Operate in a complex context with numerous cultural, technological, regulatory, and commercial barriers and risks that must be overcome on the journey to open data

Needs

- 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
- Believe in an incremental and collaborative approach to change, with some users noting that visibility could be a valuable driver for open data
- A balance between incentives and regulation, with one user noting that regulations should set a baseline, and incentives should be rewards
- To understand how their 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 enable consistency in sharing

Pain points

- There is little room in regulated budgets to invest in open data. There are associated concerns in regulated monopolies around cost to the end consumer.
- Some users note that the definition of "open" is unclear, leading to inconsistencies in what data should and should not be made open - this is somewhat evident in the inconsistent data within the current landscape
- Struggles with the quality of the data they hold in the first place, with them either not having the physical capability / resource to collect higher quality data or not having relevant data - e.g. smart meter data - shared with them
- There's a fear that data visibility will expose low data quality within organisations; however, visibility of being 'good' was also seen as a possible driver for improvements
- Lastly, there are concerns around security in making data open, particularly given that a lot of these assets are considered CNI and due to the potential risk introduced by the misinterpretation of complex data

Provider	Consumer
High understanding of energy system data	Low understanding of energy system data
Data specialist	Non data specialist

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.

"So, what I'm looking for is a template, so for BEIS and ESC to come together and say 'this is a template; this is how we want a GDN [Gas Distribution Network] to share their data.'"

Steps

Begin by understanding and cataloguing the data they hold internally	Understand how existing data can align to prescribed standards (DCMES)	Conduct a triage process to understand what data should be made visible	Understand and address various risks associated with open data	Build a business case for making data visible	Publish data and iterate based on feedback.

Insight

"In the immediate term 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."	"Issues that has come up - definition of what is open data - grey and wooly. The best definition I have found is that shared assets should be presumed open, but we're also Critical National Infrastructure, so how do we classify?"	"Internally, cybersecurity concerns, and concerns around asset and network security since we're critical national infrastructure... people have taken the data and made some very ill informed assumptions"	"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."	"It's an evolutionary approach... I'm keen to get data out, get some feedback and criticism so we can learn from that but the data is open... it's relatively small numbers that are engaging, they're embedded in the industry... the next set of data will be for more users... those who might not have the energy background, clear explanations.. guides and information in a consistent way."

Opportunities

Standards can be developed collaboratively as opposed to being prescriptive, striking a balance between organisational realities and the needs of end customers.	Opportunity to collaboratively define what data should be open and a consistent triage process, guiding & supporting providers on the journey to open data	Opportunity to define mechanisms for access management, and consider how datasets can be presented in a way that reduces risk of misinterpretation	Opportunity to define investment structures that drive innovation and are proportionate to the costs organisations undertake on the journey to open data.	Collaboration with consumers is clearly important to improving data quality and making more data more open - as previous research has also confirmed. Collaboration with end consumers must be at the centre of the end service."

Trade

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

"We use a massive amount of data and the inconsistency in data is one of the great bugbears but also, I guess, sources for possible competitive advantage as well."
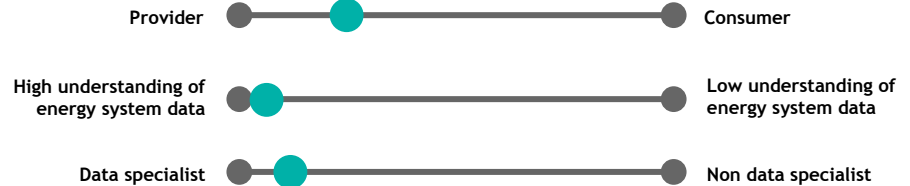
Context of use

- Team that looks at the buying and selling of gas and power in short term timescales.
- Uses data from a wide range of internal sources and external sources, sometimes using services like Enappsys which reformat data that is publicly available via platforms like BMRS
- Uses various models to forecast energy prices and where they think demand and supply are going to be for both gas and power.

Needs

- To understand how their data is being used in order to make sure it is being understood and used correctly; this is usually codified in contractual terms and via NDAs
- Easy access to data from a number of sources in order to make informed half hourly trade decisions
- Data to be consistent and up to date in order to make the data easy to understand and consume
- Appropriate information surrounding data so decisions can be made on what it contains and whether it would be useful to them; noted this should be like the basic principle of "label your axes." At the same time, where information surrounding data is available, it is difficult to navigate, understand and consume.

Pain points

- Mismatches in quality and consistency mean that a lot of time is spent unpicking data before analysing it. Sometimes, unexpected changes or updates to datasets might also lead to significant disruption to existing processes.
- Some of the settlement reports are so big that you can't open them in Excel or in a CSV - they're not easy to consume
- Noted that the learning curve and density of existing information surrounding energy data was incredibly daunting and complex, and not what one would describe as very easily understandable
- At the same time, If things are made more simple, they might lose their competitive advantage gained from the deep understanding of the data they have accumulated

Provider ———●——————— Consumer

High understanding of energy system data ●——————————— Low understanding of energy system data

Data specialist —●————————— Non data specialist

Commercial manager

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

"I think the industry is crying out for leadership and direction... we've had BREXIT, now COVID-19 and we need leadership to say this is what we're going to do and this is when we'll do this."

Context of use
- Support investment decisions to connect new assets to the local electricity network (Electric Vehicle charging, photovoltaic cells, battery storage) on behalf of their customers
- Support decarbonisation objectives and cost reduction goals for their customers

Needs
- 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 their customers; they see a strong need for standards to this end
- Believe in an incremental approach to standards and data sharing, realising that improvements in data quality and richness won't be achieved overnight. View starting with a "base set" of data and a collaborative and use-case driven approach as central to this change.
- Strongly feel the sector needs regulatory leadership in order to achieve open data and long-term change.

Pain Points
- Time and effort taken to find out whether a project can go ahead (often six weeks) due to lack of trusted data being available.
  - Often require rounds of questions with DNOs or meetings needed to get answers. This delays customer's progress towards their climate change objectives.
- Data and processes across various distribution networks vary in quality and consistency, resulting in experiences of varying quality when trying to connect consumers to the grid.
- Queue for connecting on substations (first in first out) does not take account of whether a project will create additional capacity on the network or drive decarbonisation. Believe that DNOs can make better and more innovative use of data to drive intelligent queue management.

Provider ———————●——————— Consumer

High understanding of energy system data ——●———————————— Low understanding of energy system data

Data specialist ————————●———————— Non data specialist