hippodigital

ONS-Data Visibility: Survey insights

27th May 2020

Users and Methodology

User research

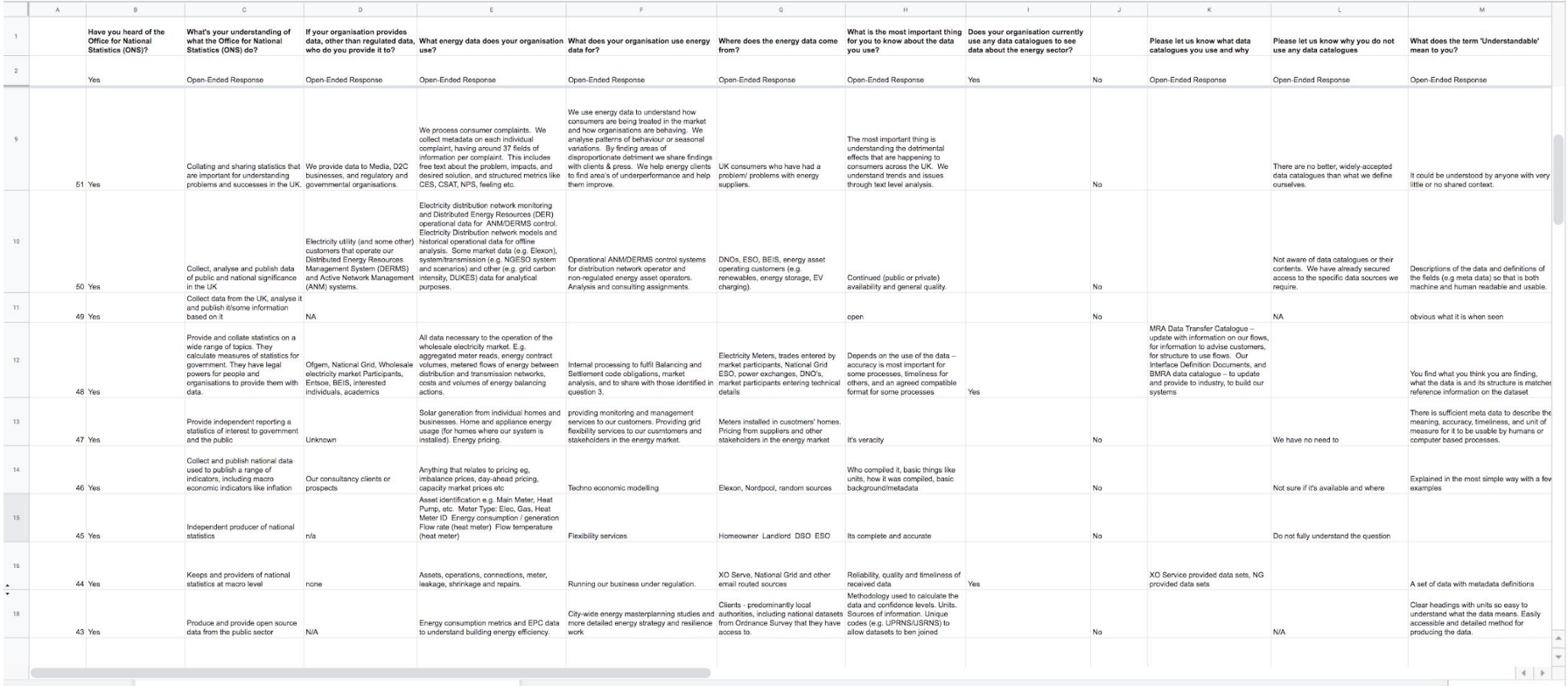
Users

- Survey via survey monkey
- 56 participants

Objectives

- 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’

Responses: https://docs.google.com/spreadsheets/d/1VhMdYQqQHDdmVW4rErGu49vEMN_BmyRX87F4o1cK_P0/edit#gid=232170016



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Have you heard of the Office for National Statistics (ONS)?	What's your understanding of what the Office for National Statistics (ONS) do?	If your organisation provides data, other than regulated data, who do you provide it to?	What energy data does your organisation use?	What does your organisation use energy data for?	Where does the energy data come from?	What is the most important thing for you to know about the data you use?	Does your organisation currently use any data catalogues to see data about the energy sector?		Please let us know what data catalogues you use and why	Please let us know why you do not use any data catalogues	What does the term 'Understandable' mean to you?
2		Yes	Open-Ended Response	Open-Ended Response	Open-Ended Response	Open-Ended Response	Open-Ended Response	Open-Ended Response	Yes	No	Open-Ended Response	Open-Ended Response	Open-Ended Response
9	51	Yes	Collating and sharing statistics that are important for understanding problems and successes in the UK.	We provide data to Media, D2C businesses, and regulatory and governmental organisations.	We process consumer complaints. We collect metadata on each individual complaint, having around 37 fields of information per complaint. This includes free text about the problem, impacts, and desired solution, and structured metrics like CES, CSAT, NPS, feeling etc.	We use energy data to understand how consumers are being treated in the market and how organisations are behaving. We analyse patterns of behaviour or seasonal variations. By finding areas of disproportionate detriment we share findings with clients & press. We help energy clients to find area's of underperformance and help them improve.	UK consumers who have had a problem/ problems with energy suppliers.	The most important thing is understanding the detrimental effects that are happening to consumers across the UK. We understand trends and issues through text level analysis.		No		There are no better, widely-accepted data catalogues than what we define ourselves.	It could be understood by anyone with very little or no shared context.
10	50	Yes	Collect, analyse and publish data of public and national significance in the UK	Electricity utility (and some other) customers that operate our Distributed Energy Resources Management System (DERMS) and Active Network Management (ANM) systems.	Electricity distribution network monitoring and Distributed Energy Resources (DER) operational data for ANM/DERMS control. Electricity Distribution network models and historical operational data for offline analysis. Some market data (e.g. Elexon), system/transmission (e.g. NGESO system and scenarios) and other (e.g. grid carbon intensity, DUKES) data for analytical purposes.	Operational ANM/DERMS control systems for distribution network operator and non-regulated energy asset operators. Analysis and consulting assignments.	DNOs, ESO, BEIS, energy asset operating customers (e.g. renewables, energy storage, EV charging).	Continued (public or private) availability and general quality.		No		Not aware of data catalogues or their contents. We have already secured access to the specific data sources we require.	Descriptions of the data and definitions of the fields (e.g meta data) so that is both machine and human readable and usable.
11	49	Yes	Collect data from the UK, analyse it and publish it/some information based on it	NA				open		No		NA	obvious what it is when seen
12	48	Yes	Provide and collate statistics on a wide range of topics. They calculate measures of statistics for government. They have legal powers for people and organisations to provide them with data.	Ofgem, National Grid, Wholesale electricity market Participants, Entsoe, BEIS, interested individuals, academics	All data necessary to the operation of the wholesale electricity market. E.g. aggregated meter reads, energy contract volumes, metered flows of energy between distribution and transmission networks, costs and volumes of energy balancing actions.	Internal processing to fulfil Balancing and Settlement code obligations, market analysis, and to share with those identified in question 3.	Electricity Meters, trades entered by market participants, National Grid ESO, power exchanges, DNO's, market participants entering technical details	Depends on the use of the data – accuracy is most important for some processes, timeliness for others, and an agreed compatible format for some processes	Yes		MRA Data Transfer Catalogue – update with information on our flows, for information to advise customers, for structure to use flows. Our Interface Definition Documents, and BMRA data catalogue – to update and provide to industry, to build our systems		You find what you think you are finding, what the data is and its structure is matched reference information on the dataset
13	47	Yes	Provide independent reporting a statistics of interest to government and the public	Unknown	Solar generation from individual homes and businesses. Home and appliance energy usage (for homes where our system is installed). Energy pricing.	providing monitoring and management services to our customers. Providing grid flexibility services to our cusmtomers and stakeholders in the energy market.	Meters installed in cusotmers' homes. Pricing from suppliers and other stakeholders in the energy market	It's veracity		No		We have no need to	There is sufficient meta data to describe the meaning, accuracy, timeliness, and unit of measure for it to be usable by humans or computer based processes.
14	46	Yes	Collect and publish national data used to publish a range of indicators, including macro economic indicators like inflation	Our consultancy clients or prospects	Anything that relates to pricing eg, imbalance prices, day-ahead pricing, capacity market prices etc	Techno economic modelling	Elexon, Nordpool, random sources	Who compiled it, basic things like units, how it was compiled, basic background/metadata		No		Not sure if it's available and where	Explained in the most simple way with a few examples
15	45	Yes	Independent producer of national statistics	n/a	Asset identification e.g. Main Meter, Heat Pump, etc. Meter Type: Elec, Gas, Heat Meter ID Energy consumption / generation Flow rate (heat meter) Flow temperature (heat meter)	Flexibility services	Homeowner Landlord DSO ESO	Its complete and accurate		No		Do not fully understand the question	
16	44	Yes	Keeps and providers of national statistics at macro level	none	Assets, operations, connections, meter, leakage, shrinkage and repairs.	Running our business under regulation.	XO Serve, National Grid and other email routed sources	Reliability, quality and timeliness of received data	Yes		XO Service provided data sets, NG provided data sets		A set of data with metadata definitions
18	43	Yes	Produce and provide open source data from the public sector	N/A	Energy consumption metrics and EPC data to understand building energy efficiency.	City-wide energy masterplanning studies and more detailed energy strategy and resilience work	Clients - predominantly local authorities, including national datasets from Ordnance Survey that they have access to.	Methodology used to calculate the data and confidence levels. Units. Sources of information. Unique codes (e.g. UPRNS/USRNS) to allow datasets to ben joined		No		N/A	Clear headings with units so easy to understand what the data means. Easily accessible and detailed method for producing the data.

Raw data

https://docs.google.com/spreadsheets/d/1VhMdYQqQHDdmVW4rErGu49vEMN_BmyRX87F4o1cK_P0/edit#gid=232170016

Have you heard of the Office for National Statistics (ONS)?

What's your understanding of what the Office for National Statistics (ONS) do?

Office for National Statistics (ONS)?

All participants answered ‘Yes’ to having heard of the Office for National Statistics, and was able to describe what the department does.

‘Collects and publishes official statistics about the UK.’ - P21

'Independent statistics authority for UK.' - P55

‘Collating and sharing statistics that are important for understanding problems and successes in the UK.’ - P51

'The UK's largest independent producer of official statistics and the recognised national statistical institute of the UK. The main responsibility of the ONS is to collect, analyse and disseminate statistics about the UK's economy, society and population.' - P25

What is the most important thing for you to know about the data you use?

Provenance

Provenance was the most important thing for people to know about data.

'Whether it is fit for my purpose - so why was it created, by whom, how, when, to what standard and accuracy (spatial and attribute), how complete is it, when it was last updated, license terms etc.' - P03

'Need good quality metadata and descriptions/explanations to accompany the data.' - P02

'Providence (source, accuracy, currency, etc.)' - P55

'The provenance - who collected it, how and if it has been processed.' - P01

Accuracy and reliability

Users needed to be confident in the accuracy and reliability of the data in order to use it

'Its complete and accurate.' - P45

'Confidence that the information we have is accurate, protected, and suitable for analysis, i.e. correctly labelled and obvious.' - P25

'The confidence factor i.e. how accurate is it. is it real metered data, calculated etc. Then ideally location (coordinates, NOT just high level postcode area) to be mapped.' - P36

'That it is timely and accurate and accessible. Regarding distribution companies - getting more granular data would be good.' - P14

Does your organisation currently use any data catalogues to see data about the energy sector?

Please let us know what data catalogues you use and why

Please let us know why you do not use any data catalogues

Current use of data catalogues

36 users do not use a data catalogue for the energy industry, for most this is because are unsure what a catalogue is, or what catalogues exist

‘Don't know where it is, don't know what is available’ - P56

‘I don’t know what they are.’ - P54

‘Not sure if it's available and where.’ - P46

‘There are no better, widely-accepted data catalogues than what we define ourselves.’ - P51

‘Not aware of data catalogues or their contents. We have already secured access to the specific data sources we require.’ - P50

‘I did not know they existed.’ - P22

‘I do not even know what you mean by data catalogues.’ - P42

Current use of data catalogues

18 users use a number of data catalogues for the energy sector

‘Elexon's BMRS has some catalogue functionality.’ - P02

‘DTC, MDD and UK Link data formats’ - P25

‘Industry MDD (elexon and SPAA) and use MRA Data Transfer Catalogue website’ - P31

‘BMRS data catalogue is one example.’ - P27

‘There are many. UKERC, OFGEM, ONS, BEIS, ESO, DSO's and many more.’ - P29

‘https://github.com/awesomedata/awesome-public-datasets#energy https://github.com/pangeo-data/awesome-open-climate-science’ - P33

‘XO Service provided data sets, NG provided data sets’ - P44

Recommendation:
At the beginning of Alpha, carry out some usability research to explore the functionality of these data catalogues with users.

What does the term 'Understandable' mean to you?
What does the term 'Searchable' mean to you?
What does the term 'Discoverable' mean to you?
Are there any other adjectives that describe what data needs to be in order to make it useful to you?

‘Discoverable’

Users defined ‘Discoverable’ as 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

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

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

‘Searchable’

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

‘Understandable’

In this context, users defined ‘Understandable’ as data that is clear and well defined, making it easily understood by 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

Other adjectives

As well as discoverable, understandable and searchable, users defined other characteristics, for example as interoperable, functional timely and publically available. There are a number of characteristics that will help make the data usable.

'Accurate, up to date and relevant' - P54

'Documented Reliable Accurate' - P01

'Interoperable - we need to be able to link to the dataset and load it into our internal systems for analysis' - P53

'Accurate useful timely flexible relevant' - P04

'Timely - made available in time frames appropriate to its potential uses (e.g. access request responses and close to capture date)
Historically copious - providing long-range
Granular - providing enough detail to support a number of analytical uses **Foresightful** - encompassing genuinely valuable uses to others beyond its originators.
Publicly available - assumed open for legitimate uses.' - P50

'Functional - That data has the most appropriate set of labels. For example, one field does not have too many different options that are similar. These things can render data lakes less useful.
Concise - that we try to be concise with the number of fields (few overlapping fields); and that we try to be concise with the number of labels in a field (ensure that we chose the minimal number of labels to express everything).' - P51

Do you think a centralised data catalogue would be of value to the UK energy industry? Please explain why

(10 users did not answer these questions)

Centralised data catalogue

The majority of users felt a centralised data catalogue would be beneficial to the energy sector to reduce time taken to find and use valuable data

‘Would help with both ease of search time taken to search and would also highlight useful data that one was not already aware of’ - P52

‘Data about the energy sector is difficult to acquire and often understand.’ - P47

It would be helpful to have a consistent way to describe this across the industry. - P53

To help people find, use and cross reference data from different sources, and for data managers to send out centralised updates on definitions of data - P48

It is very hard to keep track of all the changing sources and locations. It takes a lot of time and a lot of our public sector stakeholders trying to roll out local energy projects really struggle to find the data and time constraints so making this easier would be a significant benefit. - P36

Centralised data catalogue

One users felt a centralised catalogue was the wrong approach as there are other approaches for organising data from distributed data sources

> ‘Centralised is a very silly way to hold non static data that you aren't producing. Will always be backwards looking. Hard to incentivise people to improve it etc. ALSO VERY BACKWARD LOOKING when there are so many other ways to organise data in a distributed manner now.’ P42

Centralised data catalogue

5 users felt it depended on how the catalogue would be managed and maintained

‘It depends what data is being held. Is anyone interested (apart from a Supplier) the make and model of my meter? However, it might be of value that I have Solar panels and an EV - P31

‘Depends what you mean by centralised - I would favour the ability to search federated catalogues through a single search interface rather than physically building and maintaining a single central catalogue - avoids duplication of effort, more likely to be up to date and accurate if the data owners are responsible for maintaining their own catalogues to a common standard’ - P03

‘It is depends...will I have to pay for it? is there layers of governance? who maintains it and what's the SLA?’ - P19

‘I am not sure. Because I don’t know what they are.’ - P54

‘Only if you can identify how the data is going to be used. If data is not going to be used, there is no value but cost in making it available.’ - P21

Who would be the best party to manage the catalogues?
Please let us know why you chose that party

Who should manage the catalogue?

Five users felt BEIS would be best placed to manage the catalogue as they have an understanding of the sector and applications for the data, whilst being a government department therefore being required to be open and transparent

‘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

‘We have a close relationship with BEIS, so access wouldn't be an issue. Being a govt dept they should be open and transparent with their data.’ - P20

‘BEIS is best placed to provide the available outward facing filtered datasets. ONS is best placed to provide the repository of all raw data sets.’ - P39

Who should manage the catalogue?

Four users felt Ofgem should manage the catalogue as they are responsible for regulation the industry and therefore are experts in the sector

‘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

Who should manage the catalogue?

Overall, 19 users felt the Office for National Statistics are experts in data and statistics and therefore would be best to develop this service

‘Because they should be in charge for 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

‘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 capacity 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

‘Energy data is just one set of data utilised at macro level, the true benefits arrive when this data is correlated with other data sets that only ONS can provide.’ - P44

‘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 has the experience in data collection & analysis.’ - P34

Who should manage the catalogue?

Some users thought the Office for National Statistics would be best placed as they are independent from the energy industry

‘Energy Systems Catapult! they did the groundwork and should be running the next phase of work. **The problem is** that there will be **so much work to maintain** them that maybe ONS is best, but they aren't experts in the field. OFGEM have a conflict of interest. BEIS's remit might change in the future. so **on balance ONS**.’ - P29

‘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

‘Independence from industry.’ - P47

Who should manage the catalogue?

18 users felt it could be managed by a number of different parties both commercial and public were given as options. Some users felt it was less important who manages the catalogue as the governance in place to manage is more important

‘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

‘Energy Networks Association’ - P53

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

‘Should be commercially open to any party to provide as a service competitively.’ - P09

‘Any public org like BEIS/Ofgem but with a dedicated team of people who understands energy AND DATA.’ - P46

‘I don't mind which org manages the catalogue.’ - P33

Who should manage the catalogue?

Some users wanted a catalogue to be decentralised with individual organisation managing their own data sets, however they wanted their to be shared standard and governance

‘Decentralised management where each data owner is responsible for the management of their own data items in the catalogue.’ - P48

‘The parties who generate it, as long as they share standards and identifiers then there are many ways to bring this together.’ - P42

‘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