NGED VALUE OF DATA

ELISA EVANS, GREG JOHNSTON

DIGITAL

JANUARY 2023

CONTENTS

1. INTRODUCTION

Data Best Practice guidance, and therefore presumed open data, has become a regulatory requirement through the RIIO2 price controls and will be implemented in ED2 from April 2023. However, the specific value to consumers, innovators, networks, and others has so far also been presumed. This paper is the culmination of a series of interviews Energy Systems Catapult (ESC) has conducted on behalf of National Grid Electricity Distribution (NGED) to begin to articulate the value of open data to a variety of stakeholders.

Through this investigation we hoped to demonstrate specific value to companies utilising NGED's open data, be it in time or resources saved, or indeed in value added to their products, services, research, or other work. To do this Energy Systems Catapult conducted interviews over many weeks with stakeholders comprising broadly of the following groups: innovators, academics and industry. This was done in parallel with a survey, responded to by ten organisations. The high-level insight derived from those interviews and survey, and resulting recommendations are set out below.

1.1. INSIGHT

There were a few themes that emerged from the interviews conducted. Principally, these can be categorised as:

- Visibility of the platform and data available is low
  - Many respondents mentioned that they were until recently unaware of the existence of the Connected Data Portal, or NGED data availability in general and many in their extended networks likely remain unaware.
- Historical, time series data is a high priority innovation enabler
  - A wide variety of use cases can be enabled by different types of time series data related to topics such as outages or headroom.
- Value can be difficult to articulate
  - Not many of the respondents were able to articulate with clarity a financial value the product, service, or project they were working on has. This seemed to also be reflected in a more general difficulty data users had in building more robust business cases due to a lack of available data.

1.2. PORTAL RECOMMENDATIONS

To maximise the value data users receive from the data portal, NGED should implement the following recommendations.

1) Prioritise the discoverability of the Long-term Development Statements (LTDS) data, along with any associated metadata and other contextual information for publication as soon as practicable.
2) Display the Connected Data Portal more clearly and higher up on the National Grid Digitalisation & Data website[1]
3) On the Connected Data Portal homepage, FAQs including Application Programming Interface (API) Guidance and how to subscribe to update notifications on datasets should be more prominent. Video tutorials of worked examples would be a richer medium to educate users, especially to the uninitiated.
4) Highlight a standard publishing format that users can use when citing Connected Data Portal data to aid in rediscovery of data referenced elsewhere and allow others to derive further value.

[1] https://www.nationalgrid.co.uk/digitalisation-and-data

2. REGULATORY AND INDUSTRY BACKGROUND

Presumed open data, and open data triage have become central to Ofgem's approach to data and digital regulation of the energy sector[2]. This approach to regulation, emergent from the recommendations of the Energy Data Taskforce[3] (EDTF), has created the regulatory mechanisms for presumed open data projects and initiatives to access funding through the price controls and draw a line from enabling initiatives to value for consumers. NGED's ED2 submission to Ofgem focussed on placing customers and stakeholders at the heart of digitalisation, setting out the case to data consumers by highlighting a value proposition centred around increasing availability and use of good quality and accessible data. This highlights the need to understand customer and stakeholder needs and expectations and therefore the importance of stakeholder engagement and collaboration to maximise the value of open data. This paper is a natural springboard to demonstrate some of the learning we have gained through this customer centred focus.

In addition to presumed open approach to data, the digitalisation of the energy system is growing in pace, driven by a combination of interventions by government, industry, and private enterprise (or a combination of) seeking to create digital infrastructure to enable a net zero future. Some of these reflect recommendations made by both EDTF & the Energy Digitalisation Taskforce[4] (EDiT), and others are meeting challenges identified by different organisations. In the aggregate, somewhere within this ecosystem of new digital services, a new critical national digital infrastructure for the energy sector will emerge. One where data; open or shared, will be available between a variety of actors in the energy sector and beyond to be utilised by some regulated or administrative process, or by private enterprises trying to create or acquire value from the energy sector.

Some of this digital infrastructure already exists, for example Open Energy[5], winner of governments' Modernising Energy Data Access competition. Other infrastructure is being built currently, such as governments' Automatic Asset Registration competition[6] or the different data sharing platforms being constructed as a result of funding acquired through the network price controls. Taken all together, the variety of digital infrastructure being built will construct a complex landscape of data availability where presumed open data becomes a governance mechanism to reduce that complexity at an industrial scale. Though this may only be realised if standard metadata, data licensing and interoperability of data assets is solved for at scale. At some level, this drive has been built on an assumed value of open data to the energy sector.

The question of "what value" open data offers to a variety of organisations is inherently difficult to answer at an industry wide scale, given the wide range of potential use cases for any given data asset, in any given state of quality, granularity or timeliness. It is, however, a question largely unanswered by existing literature within the energy space. This paper sets out some findings based on interviews with a subset of data users of NGED's Connected Data Portal (CDP).

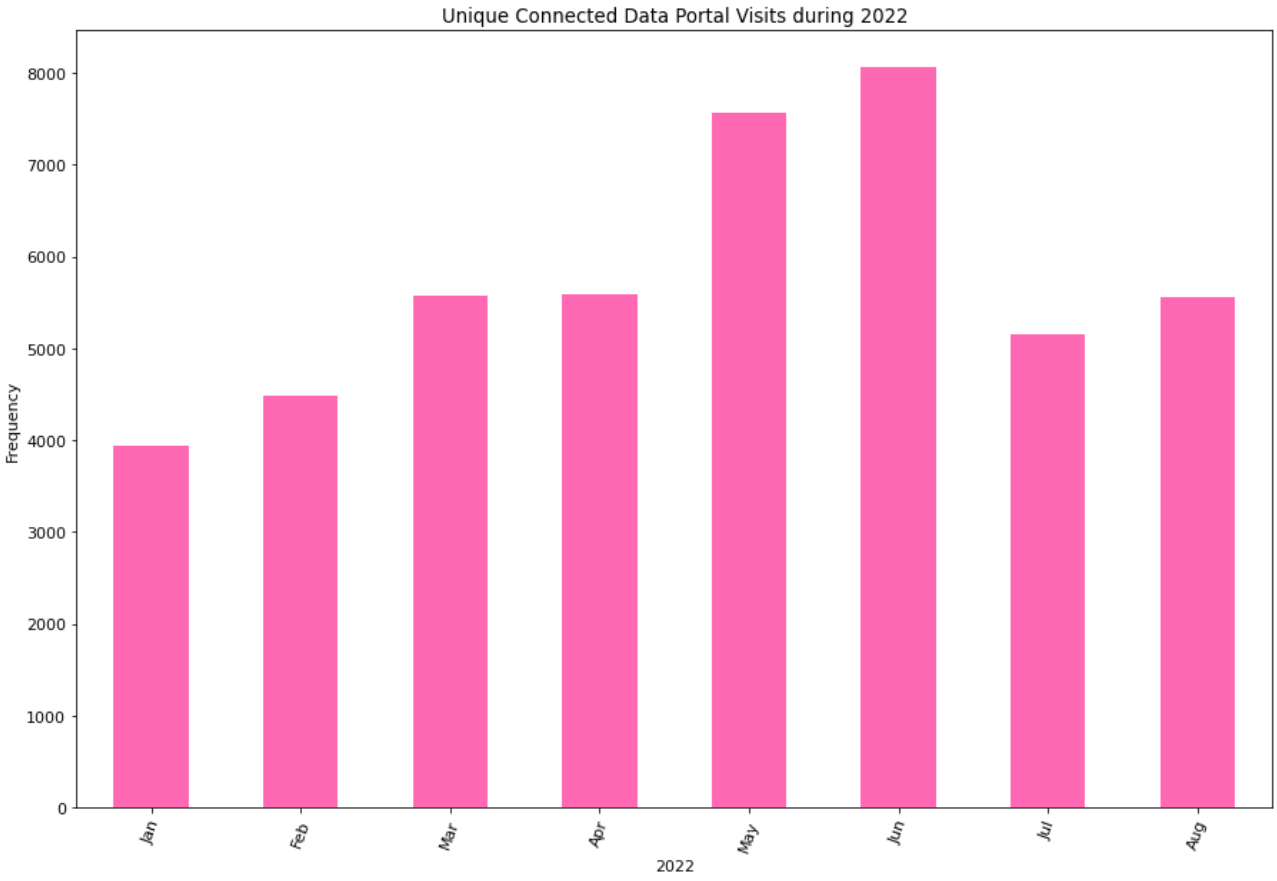
[2] https://www.ofgem.gov.uk/publications/decision-data-best-practice-guidance-and-digitalisation-strategy-and-action-plan-guidance
[3] https://es.catapult.org.uk/report/energy-data-taskforce-report/
[4] https://es.catapult.org.uk/report/delivering-a-digitalised-energy-system/
[5] https://openenergy.org.uk/
[6] https://www.gov.uk/government/publications/automatic-asset-registration-aar-programme

3. THE CONNECTED DATA PORTAL

NGED's publication of open data has been to date hosted on their Connected Data Portal[7] (CDP) and as of mid-September 2022 it has six different data groups comprising of 88 data assets in total. The data groups are set out as Connections, Demand, Flexibility, Future Energy Scenarios, Innovation and Systems and Network.

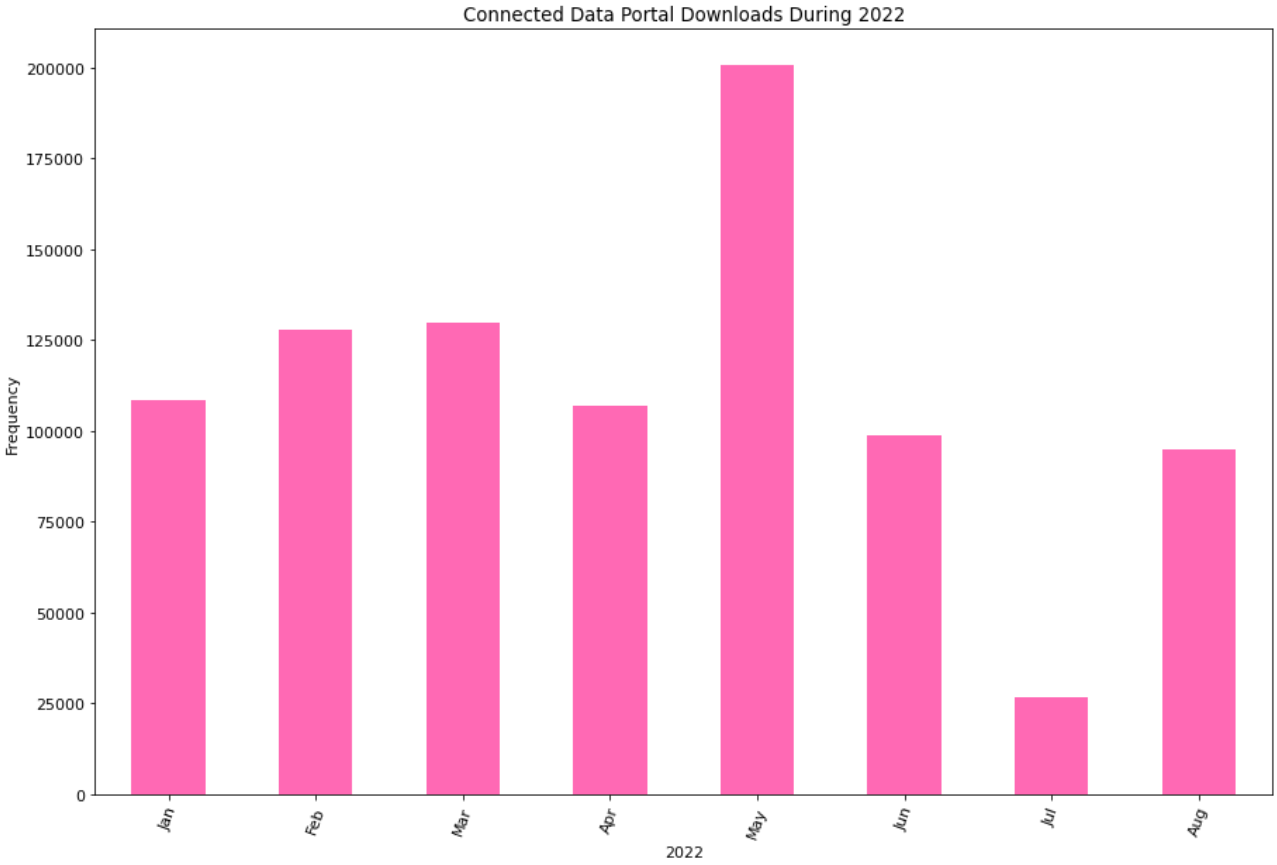
The CDP is NGED's response to regulations placed on the organisation, in conjunction with expectations set out by Ofgem and its wider stakeholder base to surface data about the status and efficacy of operation of the network that NGED manages. The portal attracted 45,902 visitors from January 2022 to the end of August 2022, as set out in Figure 1 below, along with 893,650 downloads of datasets, set out by date in Figure 2.

Figure 1



[7] https://connecteddata.nationalgrid.co.uk/

Figure 2



Most commonly, data users were downloading data about distribution demand, network assets and low carbon technology or generators connected to the distribution network. In conjunction with the currently available data, data users were also making requests to NGED about making more data openly available. Most commonly, this comprised of requests for network asset data & generation, these comprising 67% of the new data access requests.

Surfacing this data does not come without cost to NGED, or indeed any other network undertaking a similar approach and therefore understanding the value of the data being surfaced to parties outside of the network is a priority purpose of this report. At scale, platforms such as the CDP and their associated governance can reduce costs to data sharing through consistent approaches to data licensing (reducing legal overheads), interoperability of data (reducing transformation of data to meet request parameters) and the people capacity to manually cater for every actionable request for data (making APIs and options for direct download). The wider benefits of having a consistent approach to ***finding*** data about energy is also vitally important. Efforts to link CDP's data into the Open Energy[8] solution has improved the visibility of the data available to data users – visibility being a key recommendation of the EDTF report.

The benefits described above can be categorised as value to the wider sector, or ecosystem of organisations and individuals interested in energy systems data.

[8] https://data.openenergy.org.uk/dataset-list?s=western+power

4. INSIGHT & VALUE

With thousands of individuals and organisations potentially using NGED open data, it was only realistically possible to reach a small subset of data users. The insight generated by the interview process, supplemented by a survey of data users - details of which can be found in the appendix, provide the basis for this section.

4.1. HOW ORGANISATIONS ARE USING NGED OPEN DATA

Our interview process broadly focussed on innovators, industrial and academic NGED data users. Below sets out some of the uses these groups have made of data emergent from NGED.

4.1.1.INNOVATORS

There are a wide range of innovators working with open data in the energy sector. Our interviews with a subset of these who are using NGED's data provided insight into what is being used and why.

The Institute for Environmental Analytics (IEA) noted that they had a project looking at developing a proof of concept to monitor the environmental risk to physical assets on the grid based on extreme weather events. This was reliant on monitoring data for network assets as well as fault data, in combination with weather data associated to that part of the network. While they were successful in showing success in this approach with Columbian partners having received all relevant data assets, NGED's open data was only available for establishing network areas (polygons) and was not sufficient to do more than a demonstrator in *how* different polygons of the network were exposed to risk based on weather events. The lack of asset data hampered the project's ability to be specific to devices and the risk the network was exposed to with extreme weather events.

OrxaGrid participated in NGED Data Science Challenges 1, 2 & 3. Having taken the learnings for each of these challenges to heart, and using data input from Data Science Challenge 2, the team developed a product which takes different inputs and provides optimised charging/discharging profiles of battery to reduce grid import demand overall. This is focussed primarily on EVs and what happens to the grid when they connect. They expect this could be taken forward with installing charge points with the aim for providing fleet management services and optimising infrastructure in public places.

The open access of energy network data through an online portal has allowed NOCO Energy to browse through data and iterate different data processing approaches to extract value quickly and freely. Removing the human interaction element of data acquisition (i.e. back and forth emails) has drastically accelerated the rate of innovation and enabled creative experimentation.

Broadly speaking innovators responded that their needs were not sufficiently met by *any* of the networks they tried to surface data from. While NGED's data was helpful, often it did not have sufficient granularity, timeliness or the data simply wasn't available to enable them to do what they are attempting.

4.1.2. INDUSTRY

An industrial respondent, a large generator – noted that they had made expansive plans for the regeneration of a decommissioned energy site into a net zero hub. This work had to be underpinned by NGED data, to understand how the community they sought to build could be

served for a fully electric power offering. This was due to identifying the limitations the geography posed, making any hydrogen retrofit prohibitively expensive. Using a broad spectrum of NGED data, having partnered to gain access to some data which has subsequently been made available on the CDP, the group modelled the power grid to the best of their ability and created something of a digital shadow of the area. This enabled them to test different net zero scenarios for the location and build a business case.

One responded noting in interview that they have a handful of clients in the NGED region, and the time saved for each study they conduct on behalf of clients is massive. They stated that a solved power model of the part of the network they are concerned with gives them back a day in time, each time they agree to conduct a study for a client. When conducting 10-15 studies a month, each study requiring specialist knowledge and capability, it stacks up costs and time that can be used otherwise.

Another respondent, with a large energy analytics platform noted their organisation took part in NGED's Data Science Challenge 3 as a challenge and training opportunity for staff to test their modelling and data science skills. Their team managed to gain new (to them) usable modelling techniques, practiced modelling, and noted that applying data science to a real-world problem presented a good learning and development opportunity for staff. The organisation also identified a new demand forecasting service they could develop further based on their learnings from this challenge and may consider it in the future as a business development opportunity.

Of the industrial partners we spoke to, largely they remained unaware of the availability of data. A few noted opportunities had been missed due to data which is now available, was not available when required earlier in the process. There is some way to go to solve for the visibility of data for these organisations.

4.1.3. ACADEMIA

We spoke to several academics across the UK, who provided insight into the variety of ways in which they used NGED open data. A subset of the examples is provided here for context.

One academic noted that they were using secondary substation data to gain visibility of industry and wider power systems problems. By downloading secondary substation data and applying machine learning and other statistical methods to the data teaching resources were created. This enabled the academic to set tasks for students helping them to learn how to create live forecasts, testing methods for operational use by innovators/industry.

Another reported they used NGED's data science challenge data to infer high resolution data from low resolution data, using the Plymouth substation dataset. The respondent noted that high-quality low-resolution data has high utility as it is useful to train and test models. More broadly the data science challenges offered a real example of what was possible with network data. The same respondent reported that high voltage (HV) substation data had been used to build a predictive model for reactive power in conjunction with weather data to explain issues elsewhere on the network.

In addition to academics running research projects, NGED data also supports PhD students undertaking their thesis work. One respondent undertaking a PhD sponsored by an industrial partner reported that using historical time series data, particularly related to low voltage (LV) was vital to his project. The student seeks to forecast future demand response and extrapolate future trends and, given constraints with time and resource is only able to utilise openly available data to undertake this research.

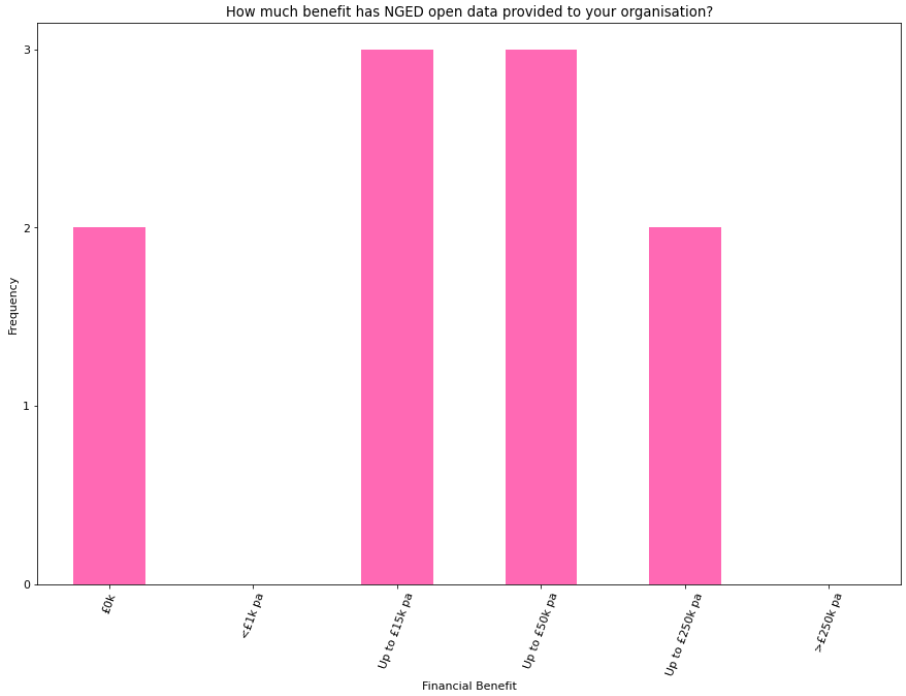
The range of academics we spoke to note the utility in both the usability of data from a teaching perspective and in identifying areas for further impactful research driven by real data and challenges of the physical system and constraints. ESC's paper on Academia to Industry noted[9] an inaccessibility to valuable data, domain expertise, and the major problems facing the industry can hinder the relevance and impact that research can make. Enabling open data through the CDP begins to solve some of those challenges.

4.2. WHAT VALUE IS BEING GENERATED BY NGED OPEN DATA

The spectrum of organisations that we spoke to found it difficult to articulate in pounds and pence what value the data has had to them, though section 4.1 sets out generally the kinds of work that have been supported and the practical value to them. NGED and ESC conducted a survey to try and gather a wider range of input from data users. While our survey was only responded to by ten organisations, and therefore provides very limited statistical value, it does start to articulate some experiences of the sector and provides a launch off point for further research.

On specific value derived, our respondents managed to articulate that the value of NGED's data to them was mostly answered as up to a value of **£15,000** or up to a value of **£50,000** equivalent per annum, expressed as effort/costs avoided or revenue gained as a direct result of using NGED data. Figure 3 below highlights these results. This range of figures representing 60% of the responses of the survey suggests that **there is likely a non-trivial benefit** to the organisations who require NGED data to develop their products and services. Two respondents noted the value to be between £50,000 and £250,000, with the remaining two respondents noting no value to their respective organisations.

Figure 3



[9] https://es.catapult.org.uk/report/data-science-from-academia-to-industry/

The benefits articulated by respondents of this data noted that the value can be categorised as the following:

- 60% noted that NGED data informs decision making for their service/product or business
- 50% said it helps realise commercial value
- 40% suggested it supports learning and development
- 40% are using the data for important processes; and,
- 30% are using the data for research purposes

One organization, Advanced Infrastructure[10] was able to articulate specific value during the interview processes, with CEO Christopher Jackson noting that "Accessing secondary substation connectivity data from the networks (*including NGED*) is vital to our business, and we estimate this data would create a business model worth approximately £2m per annum. If NGED's primary substation data could be added to our Carbon Intensity product, this would enable a potential further £500k of annual market value". Advanced Infrastructure's product is a web based local area energy planning tool and mapping visualisation suite.

4.3. IMPROVING THE VALUE OF EXISTING DATA ASSETS

The benefits of open data extend beyond the United Kingdom, the same types of data users (academics, industry, innovators etc) exist internationally looking to solve similar problems. OrxaEnergies are based in India but took part in the data science challenges and have been able to approach UK clients based on what they learned from the portal data.

Aiding discoverability would help users realise the value of existing data assets. If there was a way of providing feedback to NGED akin to a rating system, this may help other users identify datasets they might be interested in. Listing the 5 most downloaded datasets is informative but doesn't necessarily inform a user that previous users found this valuable.

Depending on constraints due to privacy/security, having further information alongside a dataset that helps explain the context of the data, could provide further enrichment. This information may be available on the portal already such as the information provided on a substation reading plate but in an obscure location such that the association is not made.

4.3.1. HISTORIC AND TIME SERIES DATA HAS UTILITY

Several data users across the organisational groups noted that historic data available alongside the current data is required for back testing and deeper analysis. They noted that current data has limited use without the historic retrospection. Particularly where respondents identified their desire to create forecasting models, where the ability to back-test those models is fundamental to creating a robust product or service. Other use cases for historical data included looking at the transmission power flow at a 5-minute level to be able to determine periods of curtailment across different parts of the network. Historical data would enable back-testing of models and potentially allow generators to identify areas of the network where they can more likely connect with minimised curtailment arrangements.

[10] https://www.advanced-infrastructure.co.uk/

In line with the data best practice expectation to store, archive and provide access to Data Assets in ways that ensure sustained benefits[11], NGED should enable use cases by providing historical data related to triaged requests on the platform. This therefore extends out to providing future data in a continuous series, and where distinct changes are required to that data over time, NGED should develop a mechanism to notify data users of a change of metadata or data structure.

Responses to the survey conducted noted that their applications of NGED data required both an extended historical data set for their purposes, as well as a high resolution. 60% of respondents noted they **required time series data for more than a year**, and **100% of respondents noted that they require half hourly data or more granular**.

4.3.2. THE LONG-TERM DEVELOPMENT STATEMENTS DATA IS CRITICAL

Respondents highlighted the critical nature of the LTDS data in their business models. Power modelling the network was represented as a challenge by several respondents, many making the same case that to work out *where* on the grid they should focus their energy and resources, they need a reliable grid model on which they can overlay other data assets to inform decision making.

One respondent noted that it took several weeks to build a model from the existing LTDS data available, due to missing bus names, lines, and other contextual information such as no electrical rating data on the reinforcement plans. Finding where on the grid is suitable to connect renewable generation, along with demand intensive activities such as rapid EV chargers are high priority for the UK's net zero ambitions and energy security. Anything that can be done to accelerate organisations finding suitable locations, connecting, and then utilising the network should be prioritised.

Beyond the pale of individual organisations solving for investment or business case decision making, the wider energy industry is testing and trialling Digital Twin technologies, with ESC[12] working with BEIS on one such, National Grid's Virtual Energy System[13] being another, and CReDo[14] the climate resilience demonstrator all being efforts to bring together disparate data assets and solve for whole system challenges. These will all require a level of engagement with LTDS data.

[11] https://www.ofgem.gov.uk/sites/default/files/2021-11/Data_Best_Practice_Guidance_v1.pdf
[12] https://es.catapult.org.uk/report/beis-energy-system-digital-twin-demonstrator/
[13] https://www.nationalgrideso.com/future-energy/virtual-energy-system
[14] https://digitaltwinhub.co.uk/credo/

5. OPPORTUNITIES

This section sets out some identified opportunities for NGED to improve aspects of the user experience for their data users.

5.1. SPECIFIC DATA TO TRIAGE

Across the interview process several use cases were discussed by a variety of data users. A short list of high potential data assets is set out below that should be triaged by NGED. It would be expected that data triaged should, where possible, reflect the need to integrate with LTDS data and then consider the utility of historical, time series data

- LIFO (last in first out) data being made openly available about generators on the network
- Providing data which shows, via a time series, what the headroom available for each substation is at as granular a time series as possible.
- Connections data, highlighting what is in the queue and what is coming down the line.

5.2. QUALITY OF LIFE IMPROVEMENTS

Several respondents identified some quick win data quality improvements that would make data users lives easier to navigate and use the data assets NGED provide. These include:

- Embedded capacity register improvements. Data substation IDs don't always match up with the LTDS information and a matching exercise to resolve this should be undertaken.
- A reference system for data would be beneficial. Often people will use available data, with that being used to create products, or research, where the data is subsequently taken offline or replaced. A unique ID for each dataset created would help users, and NGED find specific data later.
- Survey respondents noted that location and weather data were important considerations for their purposes, and NGED should consider how their locational information associated with assets can be married with other locational and weather data.

5.3. STEWARDSHIP OF STRANDED DATA ASSETS

Data assets, particularly those where NGED has been a partner but not running anything after project close, would benefit from "**Data Stewardship by common interest groups**". Community groups, much like in the open software space can maintain and enrich data assets for the public good. NGED enabling stewardship of specific data assets, beginning with focussing on those legacy data assets where innovation projects have closed, would be of benefit to the research and open data communities. This could be enabled by innovation projects nominating a data maintainer as part of project close, or for those already "stranded" assets, NGED checking compatibility of the licensing arrangements for those projects with open data publishing and ongoing management.

This could be facilitated in the form of academic partnerships or industry sponsored training programmes.

6. CONCLUSION

The pace of the transition to net zero in the UK could well be defined by the ability of private enterprise to surface data enabling them to plan, cost, and deliver energy systems projects as efficiently as possible. The CDP offers a useful interaction between enterprises of all scales and the networks, enabling them to self-serving their data needs and facilitates better decision-making across the value chain. This report sets out several recommendations, as well as insights into how the portal is already delivering real value to the energy sector and beyond.

6.1. ACKNOWEDGEMENTS

We would like to thank our interviewees for providing insightful, actionable input for this work. Particularly Advanced Infrastructure who helped exemplar just how important and valuable the journey towards presumed open data has been for the innovator community. Many thanks to all our respondents to the survey we conducted as part of this work. We hope others build on the insight of this paper and continue to make data openly available as we make the transition to a net zero economy.

7. ANNEX

7.1. DATA TABLES

Table 1: Connected Data Portal unique visitor sessions

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
3,932	4,483	5,565	5,593	7,566	8,058	5,145	5,560

Table 2: Connected Data Portal data asset downloads

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
108,388	127,773	129,774	106,913	200,630	98,669	26,580	94,923

LICENCE/DISCLAIMER

Energy Systems Catapult (ESC) Limited Licence for [NGED Value of Data]

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

Using information under this ESC licence

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

You are free to:

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

You must, where You do any of the above:

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

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

Exemptions

This licence only covers the Information and does not cover:

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

Non-endorsement

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

Non-warranty and liability

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

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

Governing law

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

Definitions

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

OUR MISSION

TO UNLEASH INNOVATION AND OPEN NEW MARKETS TO CAPTURE THE CLEAN GROWTH OPPORTUNITY.

ENERGY SYSTEMS CATAPULT
7TH FLOOR, CANNON HOUSE,
18 PRIORY QUEENSWAY,
BIRMINGHAM, B4 6BS.

ES.CATAPULT.ORG.UK
@ENERGYSYSCAT