"Data, data, every where; nor any drop to drink"

Introduction

There's been fantastic progress over the last few years on opening up energy data.

But I fear we're sleep-walking into a situation where there'll be tonnes of data but, for all practical purposes, it'll be so hard to use the data that it'll be useless.

Disclaimer: I'm no expert on these issues! I haven't studied knowledge engineering. I'm just a humble data user who sees the power of open energy data to reduce CO_2 emissions & passionately wants to see that full power realised.

(A tiny bit of background on me: I'm a computer scientist who's terrified of climate change! I did a PhD on disaggregating domestic smart meter data at Imperial College; then worked on forecasting wind power at Google DeepMind. In 2018 I left DeepMind to co-found Open Climate Fix (OCF). For the past 2 years I've split my time 50:50 between OCF and consulting for National Grid ESO. OCF partnered with Icebreaker One for the MEDA Open Energy project phases 1 & 2. And we're partnering with ESC & CAIC on a bid for EDVP.)

(This doc is based on [a Twitter rant I wrote on 2021-03-09]!)

On to the issues and some possible solutions…

Issue 1: Technical data standards

If data doesn't conform to common standards then it's like being at a party where everyone speaks a different language: lots of noise but no useful communication.

The [British Standards Institute] (BSI) are doing [great work on 'smart appliances'] but there's *loads* more to do! (e.g. standards for data recording power flows from DERs; distribution network congestion; energy forecasts; network topology; etc. etc.)

Governance of technical standards

To be blunt, a lot of decision-making in today's energy industry is far too slow and opaque. In contrast, the open-source software community has a long history of agile, inclusive decision-making which could be adapted to the energy community.

Technical decision-making in the energy industry, like data, should be 'presumed open'.

One of the great things about the energy industry is that people stay in the energy industry for many decades. But that also means that some senior folk simply don't have hands-on experience of advances in neighbouring industries. As such, technical decision-making would benefit hugely from a more diverse range of views: Include those who are on the frontline and have day-to-day experience of the technical issues under consideration, and include folks in neighbouring industries. For example, wouldn't it be great if computer science PhD students could help develop technical standards for the energy industry as part of their thesis.

Technical decision-making also needs to become much faster and more responsive. Don't wait until a complete draft of a new spec is ready before running a lengthy consultation that stays open for half a year and takes a day's work to respond to. Instead, release design proposals early, and let the community make comments on as much or as little of the proposed design as they wish. Evolve the design in full view. (Although only allow the core team to *accept* proposed changes.) The design process should be more like a real-time *conversation*. Everyone from core technical contributors to Ofgem to BEIS to interested lay people should interact with the design process in the same way. This is how hundreds of open source software projects are managed, including ones which run the vast majority of the Internet.

Several places to learn from:
- OpenStreetMap decides its schema in an entirely open and collaborative way.
- W3C standards (like DCAT) are openly discussed on Github.

Steps forwards: Governance of technical standards should be 'presumed open'

- Perhaps the most important aspect to get right from the start is the *governance* of the standards. If a single organisation completely controls the standard and can be a blocker to innovation (for example, this is an issue with CIM).
- Survey industry to collect their requirements for different data standards.
- Then review existing standards and assess whether any can be used as-is.
- In terms of engaging the community: one-on-one meetings & webinars are useful but by no means sufficient. Designing technical standards requires quiet, focused contemplation over long periods of time and can't be bashed out in an afternoon.
- Data standards should be chosen or designed in the open, using the same tools the open source software community use every day to collaboratively design software.
  - To be specific: Use GitHub!
    - Discuss on GitHub issue queue.
    - Suggest design changes using GitHub pull requests.
    - The core team can then select which changes to accept or reject; or can ask for tweaks to the proposal before accepting.
- We need diversity of opinion & to invite ideas from knowledge-engineering experts.
- The design process needs to be far more inclusive and agile than existing design processes. The process should be more like a *conversation*.
- Data *must* be machine-readable.
  - For example:
    - Simple Excel files aren't ideal but they're fine.
    - Excel files with complex formatting are bad.
    - PDFs are terrible.

Issue 2: Rich metadata

Data needs machine-readable descriptions to make it findable & interpretable.

The ENA DWG have settled on Dublin Core. Which is a great start but - crucially - doesn't answer the main questions I have when searching for data, nor will Google Dataset Search read it.

In my humble opinion, [DCAT] is the bare minimum we should be aiming at for energy data metadata. And even DCAT isn't sufficient. I think we should consider extending DCAT for the Energy system, and do so entirely in the open as a new W3C metadata standard.

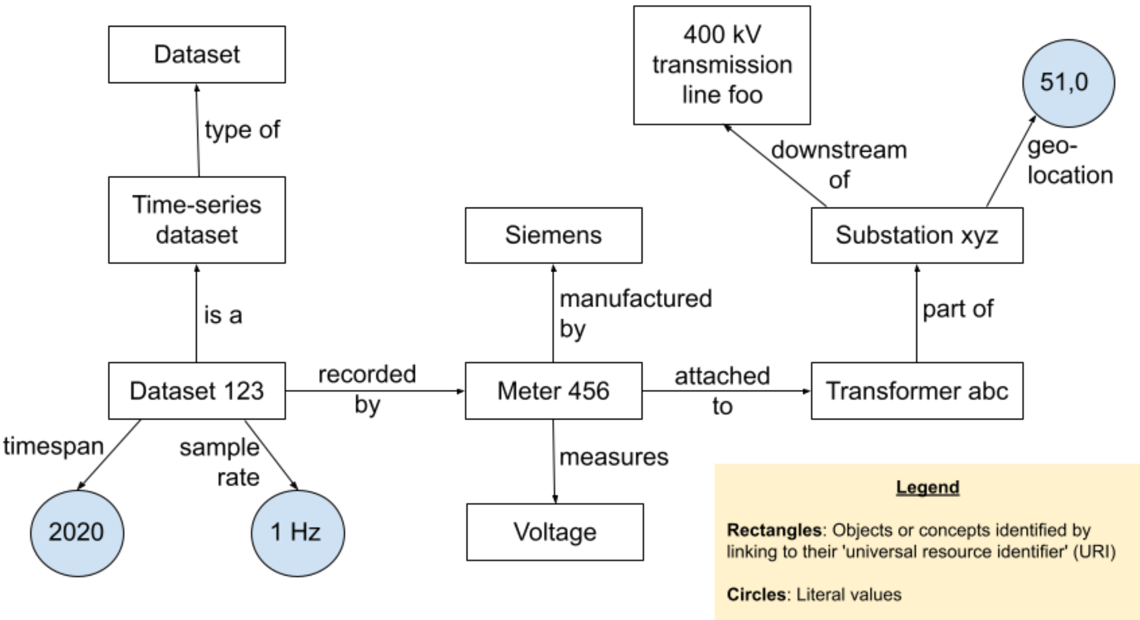
But I'm no expert on knowledge engineering! We desperately need to engage expert academics & practitioners who know how to design metadata standards! It's really hard to get right. And we're laying the foundations that will last for decades, so we've *got* to get this right!

The "steps forward" are basically the same as for "technical data standards", except focusing on standards for *metadata* instead of the *data* itself. (Although the boundary between "data" and "metadata" is fuzzy!) And the next issue (common naming conventions) is highly related to the issue of rich metadata.

Issue 3: Building a distributed digital representation of the energy system

Users want to find energy datasets using advanced queries such as "show me all datasets recording power flowing through substation *xyz* in 2020". These sorts of queries are only possible if energy metadata captures the relationships between energy datasets and energy assets.

The energy system can be described as a set of *assets* (datasets, transformers, transmission lines, companies, etc.) and the *relationships* between those assets (a dataset is recorded by a meter, that meter is attached to a transformer, etc.) The figure below gives an example of how an energy dataset can best be understood as part of a 'web' of assets:



In mathematics and computing, we call this collection of objects and their relationships a "graph" (as distinct from the more common use of the word "graph" meaning a visual representation of data). Graphs can be described and stored in many ways including "linked data" (whereby metadata files distributed across the Internet *link* to other metadata files, just as web pages link to other web pages) and a "graph database". Linked data and graph databases are well-established technologies. For example, Facebook uses a graph database to represent people and their relationships; and COVID researchers use graph databases to track contacts between individuals. Graph databases and linked data are used extensively in the biomedical sciences

Common names for energy assets

Today, different datasets use different names for the same things, and often don't agree on what constitutes 'a thing'. This makes it impossible to automatically combine datasets.

This is a HUGE problem!

Of all the issues on this list, this might be the hardest to solve completely.

Steps forwards: Linked data for the Energy system

There are two broad questions to answer:

1. *How* do we want to *structure* our digital representation of our knowledge about the world of energy? (This is similar to the 'foundation data model' recommended by the National Digital Twin programme)
2. Which categories of 'thing' do we want to represent? And, for each category, which instances do we want to represent? (This is similar to the "reference data library" recommended by the NDT)

I think along similar lines to the National Digital Twin programme: we need to select or define a 'proper' formal ontology for the UK energy system and use 'semantic web' / 'linked data' technologies to allow this information to be distributed across multiple organisations. But, unlike the NDT, I think the formal ontology for Energy needs to be as simple as possible because energy companies don't tend to employ people with PhDs in knowledge engineering!

I wrote much more about how to adapt "linked data" to the energy system in this blog post.

Issue 4: Explicit licenses

If data doesn't come with an explicit license, then users don't know what they can and can't do with that data, so they won't use the data! No one will spend time building a new service based on data that might be pulled at any moment.

Completely open data should be released under a license such as Creative Commons Attribution 4.0 (CC-BY-4.0) (as is now done by BNetzA, a network regulator in Germany)

The wonderful folks at @IcebreakerOne are looking at licenses as part of Open Energy, and hopefully they'll solve this problem! But it's not clear to me if they'll look at data licenses for *open data*; or just licenses for *shared data* (data which can't be made completely public).

Open Energy is focused on enabling secure access to data that's so sensitive that only regulated entities can view the data.

Issue 5: Data from small distributed energy resources

Small DERs have no obligation to report their state in real-time. So the energy system is completely blind to them. This is utterly bonkers. How are system operators supposed to optimise a system they can't observe?

Data from small DERs isn't personal if it's aggregated across geographical regions.

DER manufacturers want money for the data they collect. Which is fair. But we need mechanisms to do this at scale.

Options for steps forwards

There are several ways we could try to temp manufacturers to share their data:

1. Pay manufacturers for the data!
    a. This could be done as a 'data co-op' where manufacturers group together, and the data co-op sells data to grid operators and energy traders.

   b. Real-time data is probably far more valuable than historical data. So charge for the real-time data. But release anonymised historical data for free to encourage research and to advertise the usefulness of the data (hence, ultimately, increasing revenue for the live data).
2. Legislate that manufacturers must release data. This is happening in Australia: All new solar PV systems must be connected to the internet and send live data to the grid operator, and allow the grid operator to turn off the PV generation if there's over-supply.
3. Get users excited about data and then the users will put pressure on the manufacturers (unlikely!)
4. Provide such amazing data services (e.g. forecasting & analytics) that manufacturers are willing to give up their data in return for those free services.

Issue 6: Controlling access to data

A lot of energy data can be openly shared for free. This is especially true for data from deep within the energy system which describes physical infrastructure that's been paid for by everyone, and where the data doesn't reveal anything personal or commercially sensitive. It's been said before that this energy infrastructure is a little like our national parks: paid for by everyone, for the benefit of everyone, and accessible by everyone.

But there is also plenty of data which cannot be openly shared. There are two broad reasons for not openly sharing data for free:

1. **Sensitivity**: I don't want potential burglars to see when I'm on holiday by getting access to my home's real-time smart meter data! That data is private to me. But I'm happy for regulated entities to access that data if it helps the wider grid. The Icebreaker One Open Energy Directory is purpose-built for controlling access in this way. The Directory provides a digital certificate to prove identity. This digital certificate is a little like a passport. That certificate could be given to regulated organisations to prove that they are regulated by Ofgem. The dataset metadata could indicate whether an Open Energy digital certificate is required to access the underlying data. If it is, and if the user already has a certificate, then they should be able to seamlessly access the data, in the same way that you don't have to re-enter your password every time you visit Amazon.com.
2. **Money**: Some data will only be shared in exchange for payment. For example, if you buy a solar PV inverter today, chances are that it will 'phone home' to send live data to the hardware manufacturer. The manufacturers are keen to make money from all the data they collect (not because they're greedy but because profit margins are razor-thin in the solar industry, so manufacturers are desperate to claw back any profit they can). But, for solar PV data to be really useful, it needs to cover a high proportion of the solar PV systems in a country. So maybe manufacturers could club together to form a 'data co-operative' which aggregates data from multiple manufacturers, standardises the data, and anonymises the data by summing it up across multiple PV systems. Any money flowing into the data co-op will be given to the manufacturers. But, crucially, once the data is anonymised then the data can be given to anyone who has paid. The data isn't personal or commercially sensitive.

Data buyers could directly purchase data from the data co-op, just as you do when you buy from any other seller on the web.

Different access controls for different views of the same data

Smart meter data from individual homes is very private. But, if you aggregate that same data across geographical areas, then you can no longer infer any individual's private behaviour from the data, so the aggregated data can be shared openly.

Live, spatially aggregated PV power data might be worth money (because grid operators and energy traders want live data). But historical data should be given away for free to encourage research. This is standard practise in the weather forecasting community, where 'live' forecasts cost money, but historical forecasts are freely available. Giving away some data also helps to advertise those datasets and so, ultimately, giving away some data should increase total revenues for the data providers.

Describing access control in the dataset metadata

The dataset metadata should be able to express these concepts in a machine-readable way. For example, the metadata should be able to make statements like:

- "This dataset of real time, spatially aggregated PV data is available from PvDataCoOp for a yearly fee. But historical data is available for free from <link to other dataset metadata, which describes the historical version of this data>"
- "This dataset of raw smart meter data is only available to people in possession of an Open Energy Directory certificate. But spatially aggregated, realtime data is available for free from <link to freely available, aggregated dataset>"

MEDA isn't addressing most of these issues

Open Climate Fix had the huge privilege of collaborating with Icebreaker One on Modernising Energy Data Access Phase 1 and 2, and OCF implemented the alpha version of the Open Energy Search in phase 2. In general, we're huge fans of the general idea of a distributed web of energy data (as you can see from our enthusiastic blog post last year proposing the use of linked data for the energy system) and we passionately want to see this vision come to life.

Icebreaker One is focused on solving one specific problem in energy data: How to control access to data that cannot be openly shared. Icebreaker One's proposed solution is to adapt the Open Banking Directory, which provides access-control certificates to prove identity (like a passport). Icebreaker One have said that they aren't focused on defining metadata standards themselves, instead they are looking to the wider community to define the metadata standards.

Pulling it all together: Making Energy data visible

(*The text in this section is copy-and-pasted from some draft text I've been playing with for a blog post I'm thinking of publishing about the Energy Data Visibility Project… Hopefully this text is useful because it pulls together a lot of the ideas described above into a coherent "vision of the future" (if that's not too pompous a phrase!)*)

My favourite film is the anti-war 1983 classic, [WarGames](https://en.wikipedia.org/wiki/WarGames).   One of the main locations in WarGames is NORAD: the NORth American aerospace Defense command. It's a huge room filled with screens giving a real-time, super-detailed view of America's defense systems. NORAD is all about making America's defense data visible:

Let's do the same energy: Let's share enough energy data to enable innovators to build gorgeous, interactive, real-time "digital twins" of the entire energy system which display multiple layers of energy data.

This "digital twin" wouldn't be part of the EDVP itself, but the energy metadata defined by the EDVP should enable other innovators to build a map UI. Crucially, it's important to consider the 'digital twin' use-case now so we ensure the EDVP's metadata enables feature-rich use-cases like digital twins. At the very least, the metadata should capture spatial information, and should describe datasets in sufficient detail to enable the automatic ingestion of diverse datasets into a 'digital twin', and should describe relationships between assets and datasets.

These digital twins would go far beyond a static map of assets: These digital twins would also show the current state of each asset and their relationships to other assets. Click on a generator and the UI will highlight the wiring between the generator and the local substation, and show real-time power flowing from the power station. Select specific 'layers' of data (e.g. only show the solar systems or the transmission lines). Show network congestion just like how Google Maps shows road traffic congestion. Animate energy flows across time and space like this animation of solar PV generation across the UK. Use physical modelling to infer the state of assets for which we have no direct data. Enable users to run forecasts / simulations for many different scenarios of the future. Go back in time to study the state of the energy system in the past and to run "what if" scenarios.

The EDVP Beta (or the EDVP Alpha 'developed' MVP)

There could be 3 main interfaces to the EDVP:

- Textual search
  - Run rich semantic searches like "Show me all the time-series datasets which record solar PV power output at hourly resolution or better; and cover the south-east of England from 2015 to 2020. The data must be from individual PV systems instead of being aggregated; and must be released under a permissive open license like CC-BY-4.0."
- Map search
  - Click on an asset to see all the datasets associated with that asset
  - Click-and-drag to see all datasets for a geographical region
  - See the geographical distribution of datasets that satisfy a textual search (e.g. show me the spatial coverage of each solar PV dataset with data for 2020)
- API

Under the hood, the energy data visibility system will construct a knowledge graph (described in this blog post) which represents the entire energy system: the assets, the datasets, and their relationships. Once this knowledge graph is constructed, it's relatively easy to expose that data however the user wants.

The source data for the knowledge graph will be decentralised, and other people will be free to build their own knowledge graphs. The source data will consist of small, machine-readable files published on the web servers of the country's energy organisations. These files will link to each other (just as web pages link to each other) to represent relationships. These files will describe energy assets, energy datasets, and the relationships between all these entities. These files will constitute a "web of energy data".

These metadata files will be written using a standard format and a standard vocabulary. The vast majority of users won't create metadata files from scratch in a text editor. Instead, the metadata will be created a little like how HTML is created today: Most HTML on the web was created by people who have no idea how to write HTML from scratch. Instead they use tools (mostly on the web!) to create HTML (e.g. to write a blog, use Medium.com).

The technical standards for the metadata and datasets will be evolved rapidly by the community out in the open, just as open source software is designed by distributed teams in the open. Guardrails will be set by Ofgem but, within those guardrails, the community is free to update the technical standards as they see fit using simple online discussion tools.

Why do we need better data?

I spend half my time at National Grid Electricity System Operator, so I'll focus on operating the grid, but there are a huge number of other use-cases for sharing energy data.

Operating the electricity grid today is a little like trying to remotely manage a large team over a really glitchy, laggy, low-resolution video conferencing system, where half the team haven't even bothered to turn up.

As the manager, you also have an enormous responsibility: the cost of failure is truly catastrophic (the grid going down). So, to make up for the rubbish communications links, the grid operator has no choice but to run the system with tonnes of contingency (mostly 'spinning reserve'). In general, when your ability to observe and control a fast-moving, complex system is impared, you can't run that system anywhere near its optimal performance. You have to run the system with loads of slack.

On the electricity system, that 'slack' is very costly: It cost over a billion pounds to balance the UK grid in 2020 (that cost is paid by bill payers). And there's a huge carbon penalty too, because a lot of the spinning reserve comes in the form of fossil-fuel powered generators.

We could substantially reduce the costs of balancing the grid and substantially reduce carbon emissions 'just' by improving information flows. Better situational awareness means we can run the system closer to its optimal performance (with less slack). In fact, I'd argue that if we're to achieve net-zero, we *have no choice but to* get a lot better at sharing data. A net-zero grid will be far more complex than today's grid, and so will *absolutely require* far better information flows. Net-zero isn't possible unless we get a lot better at sharing data!

I can hear some people shouting "but the electricity system operator already has all this data!" So let me give some concrete examples of the ways in which the data feeds to the Electricity System Operator are broken today. And, to be clear, this isn't the ESO's fault: The ESO have to make do with whatever data they're given by the wider system. And these problems are present on grids around the world. In fact, the British grid is doing better than many!:

- The control room doesn't receive data recording the nation's true electricity demand (the total power demand of the nation's kettles, lights, computers, offices, factories, etc.). That's really weird, given that the control room's *main job* is to balance electricity supply & demand. Yet they don't receive good demand data. Yes, the control room gets real-time telemetry from grid supply points (the boundary between the transmission and distribution systems), but there's increasingly large amounts of 'embedded' generation: generation which, from the ESO's perspective, is 'behind' the grid supply points. Meter readings from each grid supply point represents total demand minus an unknown amount of embedded generation.
- The control room doesn't get live data from a large proportion of the nation's power generators. Again, remember that the control room's *main job* is balancing supply and demand! It's like we're asking them to do their job blindfolded! The control room gets absolutely no realtime data for most embedded generation (with the exception of PV, where the control room gets a real-time estimate for the nation's 1 million PV

systems. This estimate is inferred by the clever folks at Sheffield Solar from a real-time feed from about 1,000 PV systems. Those 1,000 PV systems represent 0.1 % of the nation's PV fleet). One of my jobs whilst consulting for ESO is to pull in Electralink's half-hourly MPAN data from embedded generation. But the Electralink data isn't realtime: it's at least 24-hours old. And the Electralink data doesn't capture data from meters which export less than 30 kW to the grid, so it doesn't capture domestic microgeneration. Oh, and the MPAN admin database doesn't record fuel type, so we struggle to figure out what type of microgeneration is behind each MPAN meter. The Embedded Capacity Register definitely helps, but the ECR doesn't capture DERs under 1 MW. Embedded generation breaks down (very roughly) as:

    - 13 GW of solar PV (with another 10 GW in the planning pipeline)
    - 7 GW of wind generation
    - 5 GW of thermal generation (CHP, gas turbines, diesel, etc.). This thermal generation is particularly worrying for the control room because a lot of it is price-sensitive (and the UK has a single price electricity). So, even though each individual unit might be small, large numbers of these generators turn on and off in unison in response to swings in the electricity price. So the control room might see multiple GW of embedded thermal generation magically disappear with absolutely no warning. This is bad.
    - 1.1 GW of batteries (with another 16 GW in the planning pipeline)

A system operator's main job is to ensure, at every moment, that electricity supply = electricity demand. Yet they don't receive reliable data for either side of this equation! So they have to run the system with loads of slack, and use the system frequency as the authoritative guide to the balance of supply and demand.

It gets worse: a system operator's job isn't *just* to balance national supply and demand: they have many other responsibilities such as ensuring there's enough inertia on the system, and ensuring that local constraints aren't violated (e.g. for voltage). This is all driven by the exact generation mix at each location on the system. But, today, no one knows the exact generation mix at each location on the system!

Sharing data across the grid will become increasingly important as the grid becomes more complex. In a few years time, we'll look back at the grid of 2021 and think "wow; the grid was so *simple* back then!". In Britain, there are an additional 16 GW of batteries in the planning pipeline that will do their own thing; and another 10 GW of PV; and DSOs will increasingly do more local balancing; and demand will become spikier and harder to predict as transport, heating and industrial processes are electrified; and as electricity gets converted to and from hydrogen and synthetic fuels. The conventional energy companies will increasingly need help from innovative startups to help manage and predict all this! And those startups will need access to data.

To summarise: We need much better real-time information about what each part of the grid is doing. Better data will substantially reduce costs and carbon emissions.

The rest of this document proposes how the Energy Data Visibility Project fits in with the wider landscape, and helps bring us a step closer to having a "NORAD" for energy data.

Why the EDVP alpha is so important

My interpretation of the EDVP alpha is that the project is about:

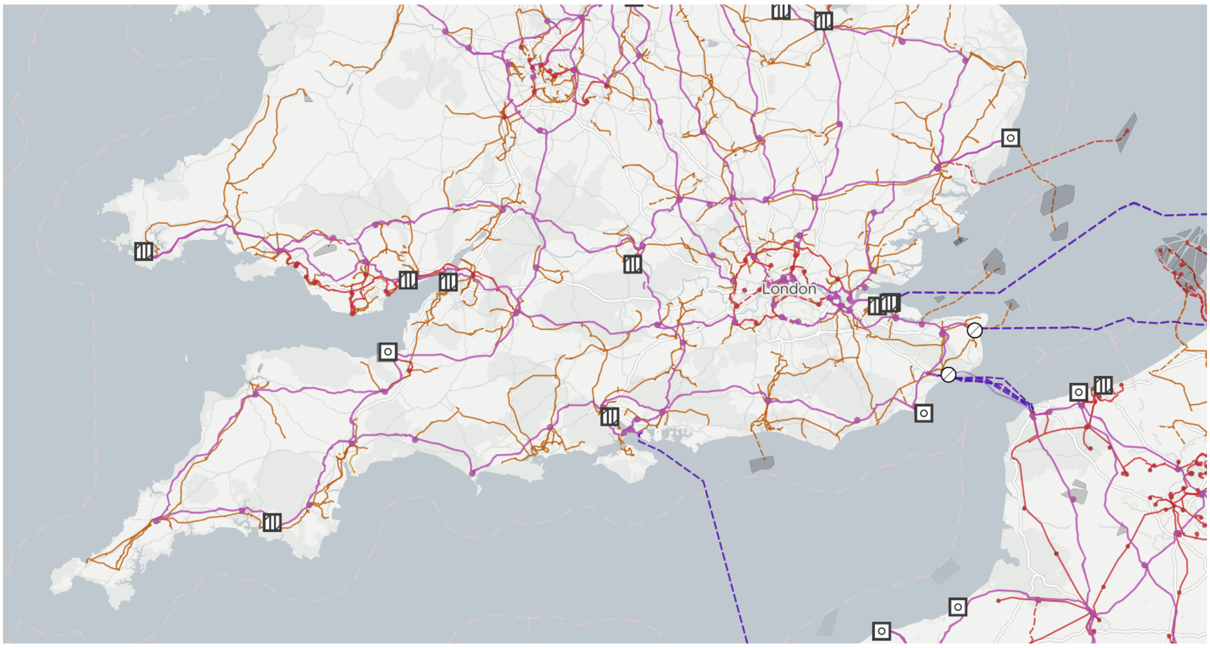
- Defining (or selecting) metadata standards for energy data.
- Converting some existing metadata into this new standard.
- Building a central representation of this distributed metadata.
- Building a front-end to easily search the metadata.
- Testing if this solution satisfies users' needs.

I think this is super-important because:

1. The opportunity isn't *just* to define a metadata standard that captures the dataset creator, date of creation etc. That's been done, and frankly isn't super-interesting. The opportunity here is to lay the foundation of a fully distributed digital representation of the entire energy system, and the datasets recorded by the energy system (as sketched out in the 'Linked Data for the Energy System' blog). A representation that's living and breathing. To take the best bits of "linked data" and apply them to the entire energy industry. To fundamentally transform the way we think about energy data, and to massively improve the chances of the wider ecosystem building innovative solutions to reduce CO2 emissions. To realise some of the things that, in Open Climate Fix, we've dreamt of since OCF's inception: open energy data that's super-easy to find and consume enabling cutting-edge research, and enabling a step-change improvement in situational awareness & energy forecasting (amongst many other use-cases).
2. Of course, the EDVP can only take the very first steps towards this goal during the EDVP alpha. But it's *essential* to get those foundations right.
3. Metadata standards stick for decades. It's unbelievably important to get this right before it's locked in; otherwise we'll be living with the consequences for the remainder of our professional lives. I don't see there being another opportunity like this in our lifetimes, I really don't't! (It's also essential that the metadata can evolve over time: Perhaps the most important thing to get right from the beginning is the governance of the technical standards for the data and the metadata.)
4. It's an alpha, so I fully expect the *code* to be thrown away. But the *ideas* will be influential (assuming they work).
5. And the timing is great:
    a. Ofgem, BEIS, ESO and the DNOs are all super-focused on data right now.
    b. If we don't demonstrate a much more ambitious vision in 2021 then far less powerful approaches to metadata may be locked in for decades to come, much to the detriment of the entire energy system, bill payers, and our ambitions for reducing CO2 emissions to net-zero.

Design ideas for the EDVP

Energy data is intrinsically spatial



(image from OpenInfraMap.org)

All energy assets exist *somewhere*, and that location is super-important. The questions that people want to answer with energy data are often related to the spatial positions of energy assets. For example, renewable energy developers ask questions such as: *where* to install the next renewable energy project; *where* is a cheap place to connect to the grid; *where* is land available. Local authorities want a list of all energy datasets for their region. Energy forecasters want to know the location of renewable energy generators.

As such, searching for energy data feels more akin to planning a road trip (remember those?!) or searching for a local restaurant on Google Maps than searching for a new laptop on Google Web Search. (That said, the EDVP should allow users to search using *both* an interactive map and a text search like Google web search. Whilst *most* searches will benefit from a map visualisation, some searches will be a better fit for a textual web search.)

Represent both energy assets and energy datasets

Energy assets and energy datasets are two sides to the same coin, a little like musicians and songs. Assets produce data (just as musicians produce songs). Data describes assets. When you search for music on, say, Spotify, you can search for the name of a new song you just heard on the radio. Then, to find similar songs, you can click on the musician's name to find similar musicians, and then find popular songs by those musicians. Now imagine an alternative world where songs and musicians are represented in two totally separate databases. The 'songs database' has no concept of 'musicians'; and the 'musicians' database has no concept of 'songs'. Not only is it more work to build two separate databases, but separating 'songs' and 'musicians' also makes the database far less useful. So, to return to energy: it will be less work *and* will result in a far more useful system if we build a distributed representation of *both* assets and datasets, and the relationships between them. This will also help to solve two problems at once: asset registration and data visibility.

Summary

There's great work happening but there's still an *enormous* amount to do! (And I can't claim that this document is in any way an *exhaustive* list!)

And this is urgent work: Datasets are being published rapidly, and it's important to get the foundations in place so data publishers can incorporate these standards while they're preparing data.

Whilst it's urgent that we get started ASAP, selecting or designing technical standards for data and metadata will take years, and cannot be rushed. We're laying the foundations for the digital energy system that will last for decades. It's essential to get the foundations right.