Data Centres

What are data centres and how will they influence the future energy system?

Every year National Grid Electricity System Operator (ESO) produces the Future Energy Scenarios (FES). These scenarios explore a range of different, credible ways to decarbonise the energy system to deliver the UK's target to achieve net zero carbon emissions by 2050, including the potential impact of this on future energy supply and demand.

Data centres are very important to the digitisation of society and the establishment of a smart energy system, yet also have implications for energy demand and greenhouse gas emissions in Great Britain. Joe Andrews, one of the ESO's commercial analysts, considers the role data centres play in society today and how this could change in the future, and the implications for future energy demand ahead of the publication of FES 2022 in the summer.

For FES 2022 we have been talking to stakeholders to better understand the data centre landscape, improve our modelling and better understand what this may mean for electricity demand from the data centre sector. Stakeholder engagement has indicated significant data centre capacity growth can be expected over the coming few years, which we are incorporating into our Future Energy Scenario assessments. This stakeholder engagement and feedback has been extremely important, as there isn't currently a central authority in Great Britain monitoring the data centre landscape.

Data centres can have a significant impact on electricity consumption and, as the population's demand for data processing continues to increase, so too does electricity consumption. This increase in demand and consumption will need to be matched by sufficient renewable and low carbon electricity capacity supply in the coming years. There is also a need to better understand how data centre energy demand could vary at daily, seasonal, and regional levels, to understand how this demand could be met by a carbon free electricity system.

Many data centre owners, such as Google and Equinix, have demonstrated their awareness of implications for emissions by pledging to make their data centres carbon neutral by 2030 via The Climate Neutral Data Centre Pact. Several countries in the EU, such as Ireland and the Netherlands, have already begun to tackle supply-challenges from the growth in this sector. The Irish government is currently in the process of deciding whether to restrict future data centre building in order to meet targets for emissions and renewable energy.

This thought piece focuses on what we already know about the British data centre landscape, what we currently envision for the future, and what needs to be done next.

What is a data centre?

Data centres have been around in some form or another since the emergence of computing in the late 1940s/1950s. Yet, due to the lack of a centralised governing authority, there is no official classification for what constitutes a 'data centre'.

Data centres can be described as physical facilities that organisations use to accommodate their computing applications and data. These can be situated on the premises of an organisation or hosted by third parties external to the organisation itself. The data centres, as defined above, host the core IT infrastructure that supports all digital activity across communities, businesses, and the government.

There is no doubting that data centres are fundamental to our modern way of living and our dependence on them is set only to increase as our consumption of data continues to grow. Every time we send an email, buy something online, save something to 'the cloud', or play online video games – it is all made possible via

processing that goes on inside the data centre. Saving to 'the cloud', while not taking up storage on your own computer, requires the data storage provided by physical data centres at another location. Even old emails that we haven't yet deleted, whilst being saved, are taking up storage and ultimately leave a large digital footprint.

While there is no official classification, there is a general industry-wide consensus that there are four or five dominant types of data centre. Their categorisation typically depends on whether they are owned by one or many organisations, whether they share and manage their own infrastructure and components, how they fit into the topology of other data centres, etc. The key components of any data centre, however, include firewalls, storage systems, servers, routers, switches, and application-delivery controllers.

The dominant types of data centre include:

Enterprise data centres

- These data centres are built, owned, and operated by the organisation that uses the IT resource contained within the data centre itself.
- They are most often likely to be housed on-premises.

Colocation data centres

- In this type of data centre, an organisation rents space within a commercial data centre owned by a third party and located off their own premises. The colocation data centre owner hosts the infrastructure (the cooling, bandwidth, physical edifice, security, etc), whilst the organisation which uses it provides and manages the components (switches, firewalls, storage, servers, etc).
- This way of managing data has often proved more cost-effective than housing an on-site data centre, with many organisations opting to transition in recent times from enterprise to colocation facilities.

Cloud data centres

- Another form of off-premises commercial data centre. In these, data and applications are hosted by commercial cloud services providers such as Microsoft, IBM Cloud, Google, Amazon Web Services, etc.
- Unlike colocation facilities, organisations do not provide or manage components in the data centre.

Hyperscale data centres

- Hyperscale data centres are renowned for their sheer size – typically at least 10,000sq ft with more than 50,000 computing and storage servers.
- They are often owned and operated by the organisation which they support (Google, AWS and Microsoft all own several hyperscale data centres).
- Hyperscale computing is critical for cloud services and big data storage. Efficiency, ultra-high speeds and high fibre count networks typically go hand in hand with hyperscale facilities.
- They are the fastest-growing type of data centre in the UK.

Managed Services data centres

- These data centres are managed by a third party (or a managed services provider) on behalf of an organisation. The organisation leases the equipment and infrastructure instead of buying it.

What does the current GB landscape look like?

Data centres have long been recognised as entities that consume large quantities of energy and, despite modern sites having improved PUE (Power Usage Effectiveness) ratings, they are still believed to be

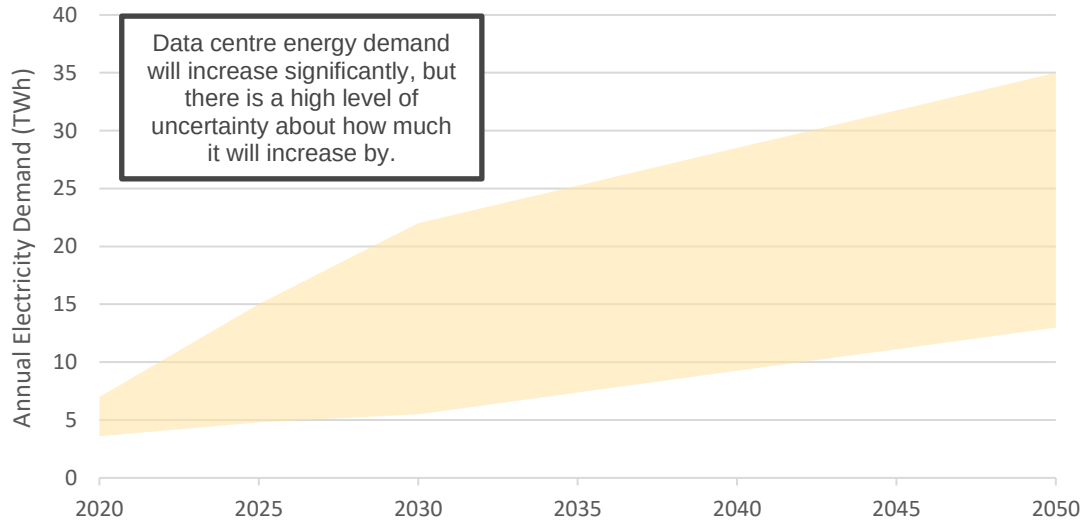
responsible for at least 1% of global electricity consumption.[1] Much of this is due to our hunger for instant-data processing; data centres are kept running 24/7 so our demands can be processed immediately. This can lead to up to 90% of energy consumed by a data centre being wasted whilst the systems are idling and waiting for the next surge of traffic, with only 10% used for actual heavy computation work. To avoid IT components becoming dangerously overheated, a lot of electricity is spent on keeping the facilities as temperature controlled, optimised environments – up to 40% of a facility's electricity consumption can go towards cooling the servers.

Research has shown us that the United Kingdom has the largest data centre capacity in Europe, with London being the continent's data centre hub. This is largely due to the size of the capital's financial centre, as well as the extensive size and reach of its fibre capacity, making it an attractive location to build. Consequently, there are between 400-600 known commercial data centres in Britain, split predominantly between colocation and hyperscale operators – accounting for 2.5% of the UK's electricity consumption. London, along with the other established data centre markets (Frankfurt, Amsterdam, Paris, Dublin), is now beginning to see data centre development running into physical limitations (Amsterdam has even announced a temporary development moratorium on new builds as land and power remain scarce across the state). This has given rise to the emergence of secondary markets such as Norway, Oslo, Berlin, and Reykjavik.

The future GB landscape:

Despite the limitations quoted above, the overall story for data centres in Britain over the coming decade will still be one of growth, with most of this growth being concentrated in London. Recent announcements and plans to build from data centre operators, in addition to market trends, have produced estimates of an almost doubling in colocation facility average power consumption between 2021-2030. This is accompanied by a potential more than ten-fold increase in hyperscale. This would amount to approximately just under 6% of the UK's total electricity consumption by 2030. Forecasting beyond 2030 towards 2050 is naturally more difficult, given the timeframe for un-foreseeable developments in data processing needs and solutions. The uncertainties around future data processing applications, the longevity of Moore's law[2], the future role of disruptive technologies, and the current lack of a government agency tasked with tracking the GB data centre landscape, will all play their part. Regardless of this, if we were to use existing market data and modelling projections, this could point to annual data centre electricity consumptions of between 3.6 TWh in 2020 to as much as 35 TWh by 2050.

Figure 1: Range of uncertainty of GB data centre electricity demand between today and 2050



[1] PUE (Power Usage Effectiveness) is a ratio that describes how efficiently a data centre uses energy. The lower the rating, the better the score.
[2] Moore's law refers to the observation that the number of transistors in a dense integrated circuit (IC) doubles every two years. This allows for greater efficiency and data centre capability. It is an observation rather than a law of physics.

What are the drivers for data centre growth?

This growth in data centres and their electricity consumption has many contributors. These drivers are ultimately all linked by the exponential year-on-year data consumption that they incite globally. In 2011, 1.8 ZB (zettabytes, 1.8 trillion gigabytes) of digital information were created. This reached 2.8 ZB in 2012, and by 2020 there was an estimated 44 ZB of data in the global datasphere.

One principal driver is an increasingly IT literate population, among an increasing UK population. Not only is the overall population rising, the percentage of the population connected to the internet is also on the up as we become an increasingly digital society. In 2005, 55% of the 60.3 million people living in the UK had household internet access. By 2010, the population had grown to 63.5 million but with 73% of households having access to the internet. The amount of internet users grew to 90% of the 67.1 million people living in the UK by 2018. This means the number of internet users consuming data in the UK nearly doubled in just 13 years between 2005-2018. Thus, the demand for additional data centres, required to process this increase in data consumption, is unsurprising.

As well as more people consuming data, the average person is consuming more data than they would have in the past. As we advance further into the digital era, people are increasingly abandoning their CD and DVD players in favour of online streaming services such as Netflix, Spotify, etc. These data-intensive services are accessed remotely over the internet, which requires the support of a data centre, and can be accessed at any time on a range of devices. We do not need to own a CD or have music stored on our phone to listen to it anymore, instead we can save this data to 'the cloud' and still have access to it anywhere, at any time, as long as we have an internet connection.

This change in societal behaviour is reflected in the growth of hyperscale data centres (often owned/leased by cloud service providers), which generally provide cloud and storage services. In Ireland, where cloud service providers such as Google host their European headquarters, hyperscale data centre average power consumption is expected to grow by up to 200% between 2021-2030.

Another key change in societal behaviour, albeit an initially forced change, is the growth of remote services to facilitate hybrid working because of the Covid-19 pandemic. There was a surge in demand for data services, such as cloud application delivery (e.g., Salesforce and Power BI), as well as an increased reliance on video calls, when many people were first instructed to work from home. Since then, the pandemic has fundamentally changed the way we view work. With many companies expressing support for the continuation of 'hybrid' working, we expect remote services required for those working from home will continue to increase demand of data centres.

The Internet of Things and smart technology

The development of data-generating and data-hungry technologies are arguably the greatest contributor to the ongoing explosion of digital information.

The Internet of Things (IoT) relates to the network of physical objects that are embedded with sensors, software, processing ability and connectivity which enables them to connect, collect and exchange data with other devices over the internet. Artificial intelligence has advanced the IoT further in recent times through machine learning - the study of computer algorithms which can automatically improve through experience and the use of data. The IoT has facilitated the development of a smart (Self-Monitoring, Analysis and Reporting Technology) energy system, allowing us to control and track home appliances like thermostats and lighting from our smartphones. However, it has simultaneously created a myriad of additional data points, all of which require the support of additional data processing (via a data centre).

This digitisation of existing processes is set only to rise, with the global Smart Home market revenue set to jump from £29 billion in 2017 to £152 billion by 2026.[3] Other industries outside the home, such as healthcare

[3] Global Smart Home Revenue Change, 2017-2026.

and transport, are also being transformed by the IoT (e.g., autonomous vehicles), and are all consequently generating more data.

The adoption of smart solutions across society is also growing the demand for 'edge' computing and thus the latest class of data centre – the edge data centre.[4] With a rapidly growing number of devices able to connect to each other via the IoT, more of them appear at the 'edge' of the digital network, producing more data at these endpoints and creating new challenges for the traditional network. This drives the demand for edge computing, which delivers computing capabilities to local endpoints of a network through which greater performance, lower latency and lower cost is obtained. This is achieved through small, locally distributed edge data centres, which can deliver cached content and cloud computing resources. The proximity between edge data centres and end users allows for rapid content delivery. This provides a greater user experience, whilst also relieving parts of the traditional network that can return to being focussed on less time-sensitive data and more big data analytics.

Many smart home devices such as smart bulbs or smart speakers require low-level processing work; requests are sent from the device to servers owned by the smart device operator in a data centre which could be some distance away. Through edge computing, such devices can receive and process a user's request entirely on the appliance itself.

Despite the network efficiency gains made by edge computing, overall data consumption and thus demand for data centres can still rise. This is because the greater user experience achieved by edge computing can ultimately increase its demand, accelerating the growth of the IoT and consequently the generation of more data – a phenomenon known as the Jevons paradox.[5]

What next?

There is still about a lot of uncertainty about the future of the GB data centre sector. Despite there currently being no central data centre governing agency, we have still acquired valuable insights into the current and future British data centre landscape through new industry stakeholders and market research.

We intend to build on our data sets and projections concerning data centres in FES 2022, refining the range given in Figure 1, covering a credible range of outcomes for growth across the four FES scenarios. We will continue to improve our modelling in this area for FES 2022 and beyond, and are keen to hear from stakeholders with insight in this area or who would like to discuss our forecasts.

We will improve the accuracy of our forecasts for new data centre capacity, and how much power consumption this new capacity will equate to. Understanding how long it will take data centre owners to build out their centres to maximum power consumption could be key to improving our forecasts of annual consumptions from 2030-2050. In addition to annual data, we plan to look at data centre usage at more granular, daily, seasonal, and regional levels. Understanding the impact of data centres at a more granular level will be important for predicting how they will affect energy supply issues during peak times, and therefore the balancing of a carbon free electricity system.

Stakeholder feedback has informed us of the relevance of future efficiency improvements when forecasting future data centre electricity consumption. A closer inspection of this is needed to understand the true impact these gains will potentially have. There is currently a significant gap in industry projections of how much efficiency gains will curtail future data centre electricity consumption. While some may point to efficiency and PUE[1] ratings showing signs of stagnation in recent years, along with some predictions of Moore's law ceasing to be relevant as early as 2025, others will point to innovations such as liquid cooling technology - which can potentially support ongoing energy efficiency improvements.

[4] Edge data centres are an emerging, relatively smaller class of data centre facility which are located close to the populations they serve. An edge data is typically one of many in a wider, complex network which are ultimately linked to a larger, central data centre.

[5] The Jevons paradox is most widely known in environmental economic circles. It occurs when government policy or technological progress increases the efficiency with which a resource is used, but the rate of consumption of such resource rises due to increasing demand.

Stakeholders further stated that much of the new data centre demand is coming from organisations wishing to transition away from their on-premises, enterprise facilities. Again, due to the lack of a centralised data centre agency, the scale of this is currently unreported. We therefore cannot currently be certain how much 'new' capacity demand is the result of the exponential growth in data consumption and how much is simply existing customers transitioning to a different type of data centre.

Uncertainties around efficiency gains, planned data centre capacity installations, and the unreported scale of enterprise to colocation transitions, all demonstrate the need for further research. FES 2022 will begin this process with improved, more detailed forecasts, while future FES publications will consider the GB data centre landscape in more granular detail. Understanding the data centre landscape will be crucial to the operation of the future energy system.

Get involved in the conversation

We'd love to hear your thoughts on Britain's data centre landscape and how these markets may develop in the future. Get in touch with us at box.fes@nationalgridso.com.