


Stretched grid? Managing data center energy demand and grid capacity

David Mytton , Masaō Ashtine, Scot Wheeler and David Wallom

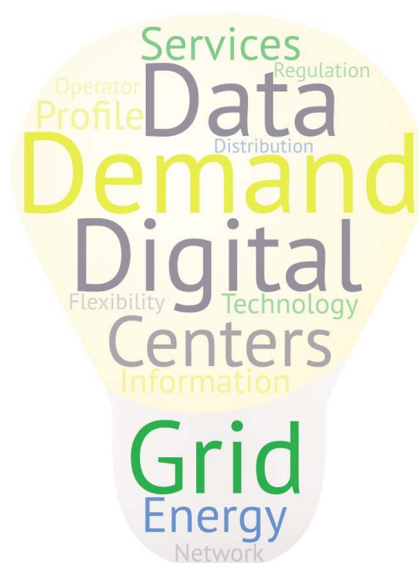
Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK.

*Corresponding address. E-mail: david@mytton.net

Abstract

Data centers are a crucial component of information technology systems because they provide a cost-effective, reliable, and secure location for the computer equipment that powers most, if not all, the services we use daily. However, their load-heavy utilization profile combined with a concentration of data centers within a few geographic areas is placing pressure on available electricity distribution network operator capacity, making it difficult to balance varying stakeholder needs. As more data centers are built, competing with other users like housing projects and electric vehicle (EV) charging, how should the energy system change to accommodate these evolving uses?

Graphical Abstract



Lay summary: Data centers are essential for digital services; however, the trend toward larger data centers also means larger power requirements. This can cause problems when they compete with other projects that also need access to electricity networks. Such conflicts are already arising in places like London, Dublin, and Virginia, USA, but it rarely makes sense for data centers to participate in demand response programs often used to deal with capacity constraints. Network operators, regulators, and data center owners need to work together to improve the frameworks used to prioritize projects and determine when and where to build new capacity.

Key words: data centers; grid; energy demand; digitalization; regulation; digital services

Article

Data centers are a crucial component of information technology (IT) systems because they provide a cost-effective, reliable, and secure location for the computer equipment that powers most,

if not all, the services we use daily—from video streaming and gaming to banking and email services. Although the definition of a data center can be broad—a facility containing IT equipment like servers and network devices—the trend is toward migration

Received: July 5, 2023. Revised: October 16, 2023. Accepted: October 17, 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

of workloads from older and smaller data centers toward large, “hyperscale” facilities operated by the biggest cloud providers such as Amazon, Google, and Microsoft. These facilities have at least 40 megawatts (MW) of power capacity, but some of the largest exceed 100 MW [1, 2]. Data centers also tend to be built in a small number of regions—68% of all data centers are in just 30 metro areas with Washington, DC/Northern Virginia, Shanghai, Beijing, Tokyo, and London accounting for 26% of the market [3].

In the UK, data centers have become highly concentrated in West London and the M4 corridor due to the proximity of fiber-optic cables that run undersea from North America making land-fall in Cornwall and then travelling to London along that route [4, 5]. This siting is a major reason why the London boroughs of Ealing, Hounslow, and Hillingdon represent 90% of applications for grid connections from data centers over the last 2 years [6]. Data centers are considered large energy users because of the high-density of IT equipment within. This load-heavy utilization is compounded by this concentration of data centers being within a small geographic area, placing pressure on available electricity distribution network operator (DNO) capacity when trying to balance varying stakeholder needs. The early signs of trouble appeared in a June 2022 article in *The Economist* [6], which revealed that battery storage systems and data centers were crowding out available DNO capacity. This quickly became mainstream news as the congestion was linked to multiyear delays to housing projects in West London, a particularly politicized issue [5, 7].

A freedom of information (FOI) request [8] revealed that the Greater London Authority had been informed of DNO capacity constraints by the city’s Housing and Land Infrastructure teams in March 2022. Due to this lack of capacity, Scottish and Southern Electricity Networks (SSEN), the regional DNO, provided local housing developers with a connection timeline as far out as 2030. The FOI response highlighted a high volume of requests for direct connections to the transmission network from data centers and battery installations, and SSEN further reported that capacity solutions were dependent on upgrades to the transmission network operated by National Grid Electricity Transmission not expected to be completed until 2029. The explanation by SSEN noted in the FOI response that data center connections were “unexpected” suggests a lack of communication and planning between data center owners, the regulators, the grid operators themselves, and local planning authorities, and it also highlights differences in construction timeframes. New data centers have historically taken up to 2 years to complete, but there are reports of faster buildouts completed in just 6–12 months [9]. In Virginia, USA, where there are 275 data centers with a combined power capacity of 1908 MW in 2022 [10], the local utility, Dominion, recently issued a warning that there was insufficient transmission capacity for new data centers [11]. Globally, more than 300 hyperscale data centers are planned for construction [12], rapidly adding to the number already in operation and outpacing the ability of DNOs to add sufficient capacity.

The demand profile is clearly growing, so how should this be managed? There are limited options for grid operators or regional authorities such as cities to choose from, ranging from developing grid flexibility products to legislative restrictions. After a temporary ban in 2019, Amsterdam, a large data center hub, has now imposed annual power quotas for new data center construction with a maximum budget of 670 megavolt amperes (MVA) until 2030 [13]—these regional impositions will affect the construction of hyperscale data centers. Ireland is another major data center hub, and EirGrid’s capacity planning statement for 2021–30 indicated that data centers could account for 23% of all electricity

demand in Ireland by 2030 [14], though this projection has a wide range of uncertainty. Based on these forecasts, the Irish Commission for Regulation of Utilities (CRU) published a direction to the system operators outlining new, stricter criteria for connections [15]. However, the use of speculative reservations was suggested as a reason for overprojecting capacity in Ireland—the Irish CRU reported that three-quarters of the capacity reserved was unused [16]. Similarly, Hillingdon Council in London rebutted the idea that housing was being delayed due to a lack of electricity grid capacity, stating that capacity is often reserved but not used [17].

Managing capacity is clearly an area where policy can be improved. Data center buildings are usually constructed with excess capacity to allow growth, anticipating future customer demand, but beginning operations well under maximum capacity. Grid capacity is reserved for this maximum, but may be unused, blocking other users. Reservations currently operate on a first-come, first-served basis with no cost incentive to only reserve what is needed. Microsoft has suggested altering the pricing mechanism to introduce reservation pricing that scales based on usage and has an expiry date [18]. As noted by the UK regulator, Ofgem, in 2020 [19], DNOs could also adopt locational pricing for grid connections based on regional demand akin to models used to price electricity in other areas of the world [20]. This could also flow back into pricing for data center customers [11]—there are already regional differences in cloud pricing that reflect the varying costs of providing services in each location. If proximity to existing fiber cables is important, that should be reflected in the price, allowing the data center (and fiber network) operator to balance the cost of additional electricity grid capacity with building more fiber connections. If demand for connectivity is high enough, data center operators can contribute to the improvements necessary, as is the case with substation upgrades by data center operators in West London [17]. Ofgem, having published its position in early 2022 [21], is now reviewing options to incentivize the right behavior at different times and locations. Despite solutions being proposed by both regulators and industry, including Ofgem reiterating charging reform as a priority in November 2022 [22], we have yet to see any definite progress with any of these approaches.

Aside from imposing capacity quotas, demand response is often discussed in the context of data centers because other large energy users, such as industrial refrigeration [23], can respond to requests to turndown events in return for lower energy or connection pricing agreements [24, 25], or revenue within network flexibility markets. These demand response actions can both help balance energy supply and demand at a national level (increasingly important with greater variable renewable generation) or alleviate network power constraints at a local level. Enabled by the digitalization of the energy system, network operators globally are looking to flexibility to manage the network in a smarter way, either for short-term deferral of network upgrades (where there is uncertainty in future demand) or improving utilization of network infrastructure long-term by reducing the peaks that drive network reinforcement. In the UK, this is primarily endorsed by Ofgem and the UK Government’s smart systems and flexibility plan [26] at a national level.

There are numerous pilot projects exploring flexibility at the distribution level such as EcoGrid 2.0 in Denmark, EnergyVille in Belgium [27] and the UK-based TRANSITION, Electricity Flexibility and Forecasting Systems, and FUSION (TEF) collaboration [28] and Project LEO (Local Energy Oxfordshire) [29]. Within the market trials run by Project LEO and TRANSITION, two categories of flexibility services were trialed: DSO Procured, where network

operators pay for flexibility to avoid the network going beyond its firm capacity during known periods of high demand or following fault conditions, and DSO Enabled, which includes peer-to-peer trading of contracted network capacity that may allow local consortia to lower their combined network capacity and therefore connection costs [30]. Another notable example of demand response flexibility is the recent world-first Demand Flexibility Service from National Grid ESO in the UK, which achieved more than 800 MWh in the first two events [31]. However, despite several pilot projects [32], flexibility markets at the distribution level are still very much in their infancy.

Theoretically, data centers could participate in these types of flexibility programs either through temporary computing load reduction or relaxing the set points for the thermal conditions maintained through air-conditioning systems. However, adoption has so far been limited—few data centers have implemented these mechanisms because they clash with customer service-level agreements. The type of workloads that are suitable for location migration—those that are not latency sensitive—also tend to involve large volumes of data, e.g. training machine learning models, running analytics queries on data warehouses, or batch processing such as encoding videos. Migrating large amounts of data between data centers far enough away to be subject to a different grid mix would take lengthy periods because of network transfer times. Providing such capacity is also expensive, particularly if stored in cloud computing services, which have prohibitively high network egress fees between cloud regions. Further, the ability to shift IT workloads between locations requires designing the capability into IT software systems with redundancy that is expensive to implement and difficult to maintain, something only large organizations like Google have so far managed to achieve for specific workloads [33].

As discussed in Microsoft's response to the Irish CRU directive [18], flexibility would not generally result in the data center "turning off." Instead, data centers would go "off grid," switching to on-site uninterruptible supplies such as diesel or gas generators. These are currently designed for emergency situations with batteries covering the data center load for a few seconds to a minute until generators are at speed and can take over. Aside from the risks associated with switching a live data center to a different source of energy, the carbon footprint of running fossil fuel-powered generators for extended periods of time runs counter to net zero goals [34]. Some data centers are now being built [35] with long-duration batteries designed to be charged whilst there is abundant clean energy, but this is not common due to the cost and experimental nature of such technologies. Microsoft has also just started a trial of providing grid services such as balancing via its on-site battery capacity [36].

At the heart of the issue is a mismatch of incentives. DNOs are under a reliability obligation to provide energy to existing customers, which means they have a conservative approach to network changes. In contrast, data center operators want to bring on new capacity quickly to address customer demand, and housing developers want to address the social and political imperative to build more. Considering the incentives involved is crucial to understanding the decisions made by the various stakeholders. Data center participation in demand response appears to be an intuitive solution to the overall problem, but only if the commercial incentives are ignored. Instead, data center operators are pushing for reform in other areas such as the pricing and management of capacity queues. Energy is a large part of the overall cost structure for data center operations, which impacts how decisions are made when choosing the location of new facilities,

so including infrastructure and not just the cost at the meter would better align incentives. The design of specific demand response services and how they could be codeveloped with data center and grid operators would help link system requirements to appropriate pricing mechanisms.

Looking ahead, a collaborative effort among data center operators, regulators, grid operators, and local planning authorities will be imperative to harmonize the diverse interests and accelerate the necessary infrastructural and regulatory reforms. This means altering regulations and market design to represent the desires of the various parties and their relative priority more accurately. The lessons learned from these reforms will be important for the future as electricity demand shifts from small numbers of large-scale users (e.g. data centers and battery storage) to large numbers of smaller users caused by the increase in electric vehicles and heat pumps. The grid must evolve at all levels to deal with this change.

Data availability

N/A

Conflict of interest

D.M. has a financial interest in StackPath, LLC, an edge computing company, and was engaged by the Uptime Institute as a Research Affiliate from December 2020 to November 2021. The authors declare no further interests.

Acknowledgements

N/A

Author Contributions

David Mytton (Conceptualization [lead], Investigation [lead], Writing—original draft [lead], Writing—review & editing [equal]), Masaō Ashtine (Writing—original draft [supporting], Writing—review & editing [equal]), Scot Wheeler (Writing—original draft [supporting], Writing—review & editing [equal]), David Wallom (Supervision [lead], Writing—original draft [supporting], Writing—review & editing [equal]).

References

1. Masanet, E. & Lei, N. How much energy do data centers really use? <https://www.agci.org/research-reviews/how-much-energy-do-data-centers-really-use> (2020). Accessed 5 Nov 2023.
2. Dawn-Hiscox, T. What is a hyperscale data center? Data Center Dynamics <https://www.datacenterdynamics.com/en/analysis/what-is-a-hyperscale-data-center/> (2022). Accessed 5 Nov 2023.
3. Synergy Research Group. Just 30 Metros Account for 68% of the Worldwide Colocation Market. <https://www.srgresearch.com/articles/just-30-metros-account-for-68-of-the-worldwide-colocation-market> (2022). Accessed 5 Nov 2023.
4. BBC News. 'Massive' transatlantic data cable landed on beach in Bude. BBC News 2021;
5. Hammond & Morris. West London faces new homes ban as electricity grid hits capacity. *Financial Times* 2022;
6. The Economist. Britain's overstretched electricity grid is delaying housing projects. *De Economist* 2022;
7. New homes may be delayed by power grid capacity. BBC News 2022;
8. Greater London Authority. FOI—electricity capacity in West London [Jun 2022] MGLA060622-2383. 2022;

9. Smolaks M. Data center costs set to rise and rise. *Uptime Institute Blog* 2023. <https://journal.uptimeinstitute.com/data-center-costs-set-to-rise-and-rise/>. Accessed 5 Nov 2023.
10. Olivo A. Northern Va. is the heart of the internet. Not everyone is happy about that. *Washington Post* 2023;
11. Judge P. Dominion energy admits it can't meet data center power demands in Virginia. *Data Center Dynamics* 2022. <https://www.datacenterdynamics.com/en/news/dominion-energy-admits-it-cant-meet-data-center-power-demands-in-virginia/>. Accessed 5 Nov 2023.
12. Synergy Research Group. Pipeline of over 300 new hyperscale data centers drives healthy growth forecasts. *Synergy Research Group* 2022. <https://www.srgresearch.com/articles/pipeline-of-over-300-new-hyperscale-data-centers-drives-healthy-growth-forecasts>. Accessed 5 Nov 2023.
13. Judge, P. Amsterdam resumes data center building, after a year's moratorium. <https://www.datacenterdynamics.com/en/news/amsterdam-resumes-data-center-building-after-years-moratorium/> (2020). Accessed 5 Nov 2023.
14. EirGrid. All-Island Generation Capacity Statement 2021–2030. <https://www.eirgridgroup.com/site-files/library/EirGrid/208281-All-Island-Generation-Capacity-Statement-LR13A.pdf> (2021). Accessed 5 Nov 2023.
15. Commission for Regulation of Utilities CRU Direction to the System Operators Related to Data Centre Grid Connection Processing. Commission for Regulation of Utilities, 2021
16. Ballard. Three-quarters of Ireland's data center power reserves left unused (updated). *Data Center Knowledge* 2022. <https://www.datacenterknowledge.com/energy/three-quarters-irelands-data-center-power-reserves-left-unused-updated>. Accessed 5 Nov 2023.
17. Donnelly, C. Datacentre sector hits back at claims that West London electricity grid capacity crunch is its fault | *Computer Weekly*. *ComputerWeekly.com* (2022).
18. Microsoft Microsoft Comments re: CRU proposed Direction to the System Operators related to Data Center grid connection (CRU/21/060). Microsoft, 2021
19. Ofgem. Electricity network access and forward-looking charging review: open letter on our shortlisted policy options. *Ofgem* 2020. <https://www.ofgem.gov.uk/publications/electricity-network-access-and-forward-looking-charging-review-open-letter-our-shortlisted-policy-options>. Accessed 5 Nov 2023.
20. UKERC. Exploring the implications of locational marginal pricing of electricity. *UKERC* 2022. <https://ukerc.ac.uk/news/locational-marginal-pricing/>. Accessed 5 Nov 2023.
21. Ofgem. Access and forward-looking charges significant code review—updates to our minded-to positions. *Ofgem* 2022. <https://www.ofgem.gov.uk/publications/access-and-forward-looking-charges-significant-code-review-updates-our-minded-positions>. Accessed 5 Nov 2023.
22. Ofgem. Open letter regarding prioritisation of electricity network charging reforms. *Ofgem* 2022. <https://www.ofgem.gov.uk/publications/open-letter-regarding-prioritisation-electricity-network-charging-reforms>. Accessed 5 Nov 2023.
23. Postnikov A, Albayati IM, Pearson S et al. Facilitating static firm frequency response with aggregated networks of commercial food refrigeration systems. *Appl Energy* 2019;**251**:113357
24. Hubbard J-J. Letter: boost grid capacity by building local markets. *Financial Times* 2022;
25. Cardoso CA, Torriti J, Lorincz M. Making demand side response happen: a review of barriers in commercial and public organisations. *Energy Res Soc Sci* 2020;**64**:101443
26. UK Government. Transitioning to a net zero energy system: smart systems and flexibility plan 2021. GOV.UK 2021. <https://www.gov.uk/government/publications/transitioning-to-a-net-zero-energy-system-smart-systems-and-flexibility-plan-2021>. Accessed 5 Nov 2023.
27. Wheeler S et al. Local energy markets: From concepts to reality. In: Shafie-khah M, Gafafrudi AS, (eds.), *Trading in Local Energy Markets and Energy Communities: Concepts, Structures and Technologies*. Springer International Publishing, 2023, 1–38
28. TEF. TEF. SSEN Transition . <https://ssen-transition.com/dso/tef/>. Accessed 5 Nov 2023.
29. Project LEO. Project LEO final report: a digest of key learnings. Project LEO . <https://project-leo.co.uk/reports/project-leo-final-report-a-digest-of-key-learnings/>. Accessed 5 Nov 2023.
30. Klyapovskiy S, You S, Michiorri A et al. Incorporating flexibility options into distribution grid reinforcement planning: a techno-economic framework approach. *Appl Energy* 2019;**254**: 113662
31. National Grid ESO. World-first Demand Flexibility Service exceeds expectations with businesses saving thousands of pounds while reducing carbon emissions. <https://www.nationalgrideso.com/news/world-first-demand-flexibility-service-exceeds-expectations-businesses-saving-thousands-pounds> (2023). Accessed 5 Nov 2023.
32. National Grid. EQUINOX (equitable novel flexibility exchange). National Grid 2022. <https://www.nationalgrid.co.uk/innovation/projects>. Accessed 5 Nov 2023.
33. Radovanovic A. Our data centers now work harder when the sun shines and wind blows. *Google* 2020. <https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows/>. Accessed 5 Nov 2023.
34. UK Government. Net zero strategy: build back greener. GOV.UK 2022. <https://www.gov.uk/government/publications/net-zero-strategy>. Accessed 5 Nov 2023.
35. Vigliarolo B. Google tests emergency battery backups for data-centers. *The Register* 2022;
36. Robinson D. Microsoft wants to export 'grid-interactive' Dublin DC setup. *The Register* 2023;