techUK

Data Centre Energy Demand: The Viking Helmet

October 2021

Why do we overstate data centre energy demand?

We all know that data centres are electro-intensive and that the demand for digital services is increasing rapidly. This creates an energy challenge for our sector – how to accommodate this explosion in data without a parallel explosion in data centre energy consumption. At the moment the sector is able to keep energy demand surprisingly flat [i] due to improvements in infrastructure management, processor efficiency, technologies like virtualisation, and the trend to outsource and consolidate inefficient on-premises IT.

Nevertheless, data centre energy consumption has always been a controversial subject. Some claim that sector energy use is spiralling out of control and data centres have been the subject of a long catalogue of exaggerated and sensationalist claims about energy consumption. The result is that both current estimates for sector energy use and predictions for future demand vary wildly, and that the numbers that get picked up and promulgated tend to be at the top end of the scale.

So why does this happen? Think of the horned Viking helmet. There is absolutely no evidence that the helmets worn by Vikings ever had horns; the horned helmet was the creation of 19th century Wagnerian opera. Yet who can imagine a Viking helmet without horns? Maybe someone who has watched Vikings [ii] or who listens to Saga Thing [iii]. But for most of us it's very hard to deconstruct that myth, probably because the real helmets, if Vikings ever wore them, were much less striking. This problem is captured by Brandolini's Law [iv], which states that it takes ten times more energy to debunk myths than to create them.

The classic example is the statement made by Gartner in 2007 that the ICT sector's emissions were equivalent to those of the airline industry. While it made a legitimate point about the growth of ICT and its under-reported energy consumption, the comparison does not bear detailed scrutiny for multiple reasons [v]. Moreover the aviation sector is a fraction of the size of the ICT sector, so it is like saying a cow and a hamster produce the same amount of manure and that we really need to do something about the cow. In the intervening fifteen years the debate should have moved on, yet we still see regular coverage shouting about ICT and airlines, except that recently the term "ICT" has sometimes magically been replaced by "data centres".

But why does energy use get exaggerated in the first place? And by whom? We've thought about this in the light of the claims that have been made and have come up with the following suggestions. This is not a comprehensive list, and it is based on observations rather than formal study, but it's worth bearing these points in mind when you see the next attention-grabbing headline about sector energy consumption.

The core problem, however, is that inaccurate and sensationalist claims distract us from the more important dialogue underway about data centre energy use, which is essentially how our electro-intensive sector can contribute meaningfully to our collective net zero commitments whilst accommodating sufficient growth to ensure adequate capacity so that the UK continues to be a world leader in digital services.

This boils down to an examination of how we optimise efficiency, how we accelerate the consolidation and/or migration of inefficient legacy on-premise activity, how we improve transparency and do a better job of evaluating the carbon productivity of our data centre estate, both directly and indirectly (don't forget data centres underpin all those smart, ICT-enabled energy saving and dematerialised alternatives), how operators can fund additional renewable generation, act as anchor customers and testbeds for new technologies like fuel cells and battery storage and finally how we can leverage the physical infrastructure itself to enable a more dynamic, distributed electricity grid.

Where Does It All Go Wrong? Summary

Lack of Data

Industry data that is Inadequate or incomplete encourages guesswork.

Outdated sources

Old data doesn't reflect current performance and leads to big errors.

Long Horizons

Technology changes too fast to allow long range predictions.

Definitions

Definitions vary so we aren't comparing like with like

Making Headlines

The urge to exaggerate to attract attention is strong: sensationalism sells!

Magic Numbers

We love a simple number but aren't interested in the caveats it comes with.

"Propheteering"

The bigger the problem, the more compelling the solution!

Regulatory Evangelism

It's easier to justify policy intervention when the issue is big and important

Reality Check!

Calculations and models always need a sense check but not everyone does it.

Schoolboy Errors

Schoolboy errors are much more common than you would expect.

Laziness

It's tempting to round up the usual suspects without checking the facts.

Matrioshkas

Layered business models an encourage duplicate reporting.

Extrapolating

It's fun to extrapolate but first we need to be sure it's an established trend.

Misinterpreting trends

We shouldn't assume that a given trend will deliver a particular outcome.

Partial picture

Drawing conclusions from only part of the picture can be risky

Lure of Scenarios

The tendency to confuse scenarios and predictions is very common.

Why Do We Overstate Data Centre Energy use?
More detail on our suggestions.

Lack of robust data

This is obviously an exaggeration but the core problem is that data centre energy is generally not reported systematically by operators or recorded formally as a proportion of national consumption figures, which means that analysts have to resort to modelling, accompanied by the inevitable assumptions. The UK is unusual, perhaps unique, because commercial operators in the UK report energy consumption that is measured and audited through the Climate Change Agreement (see more below). However, even in the UK the data is not captured in a meaningful way in national energy statistics and moreover the CCA data is incomplete. It does not include enterprise sites, whether purpose-built or on-premise, so misses a significant proportion of activity.

Ageing Sources

If you are studying medieval Icelandic literature then very old sources are invaluable. However, when it comes to research on digital technologies, Granny doesn't know best. Data centres are underpinned by rapidly evolving technology which means that basing calculations and projections on old data can quickly lead to order of magnitude errors. For instance, the amount of energy it takes to process a given amount of data has reduced by around seven orders of magnitude over the last three decades (this relates to Moore's Law), and the energy intensity of the fixed communications network also appears to halve roughly every two years [vi].

Long Horizons

The further ahead people predict, the less reliable those predictions are. This is because technology is still evolving fast, so the same rules apply to projecting ahead as they do to using ageing sources. However it is fun to look at some of the predictions that were made about ICT and data centres in the past and compare them with what has actually happened. For instance a 2009 paper sponsored by the Japanese government projected that by 2030 the internet routers in Japan would consume the entire 2005 electricity grid capacity [vii]. In 1999 Huber and Mills predicted that in ten years ICT would consume 50% of US power. There are many more, but all the results tend to suggest that a more cautious approach would have been advisable.

Inconsistent definitions

One of the biggest stumbling blocks when we try to evaluate the energy impact of data centres is that we don't define data centres consistently. Should we only include larger facilities, say those with resilient power supplies, and set minimum thresholds in terms of space and power consumption as we do for the CCA [viii], or should we include all those small, on-premise data centres and server rooms? If the latter, where do we stop – cupboards, closets and individual units? And then how we know where these are, how many there are or how much energy they use, as this is generally not recorded or reported.

HEADLINES!

We all love exciting news, especially journalists. Nobody really wants to know that data centre energy use is growing incrementally, or that the sector is relatively carbon productive compared to most other economic activity. What everyone wants to hear is a big shocking number or an exciting statement like "ICT will consume over a fifth of our energy by 2030" (which was erroneously reported in Nature in 2018 [ix] and by the Guardian in 2017 although the latter cited 2025 as the relevant date [x]) or that ICT has the same carbon footprint as airlines" or that "downloading a video uses the same energy as the country of Burundi". It doesn't matter that none of these are true, or that some are patently nonsense: they have stuck, like the horns on that Viking helmet.

Magic Numbers

We love a simple number, shorn of all those boring caveats, but there is no such thing as data without context, especially when it relates to data centre energy consumption. When publishing their numbers, analysts and researchers usually provide detailed explanations of the scope of their studies, the assumptions and modelling approaches and usually a list of caveats. However, when these numbers are quoted by third parties, they often lack those accompanying but all-important qualifications. If two estimates vary, there will be a reason: they may be based on different models, they will have made different assumptions, the baseline datasets may differ and the scope may vary. For instance the CCA data only includes commercial (third party – colocation) data centres, so the 3.6TWh a year reported for the UK's commercial sector [xi] will under-represent total consumption.

"Propheteering"

Propheteering is not a real word but we think it neatly captures a growing trend to dress up sales pitches as climate concern. While most proposals do involve genuine products and services, there is a tendency to start off by exaggerating the problem, perhaps to instil a sense of urgency or to increase the apparent size and benefit of the deliverables: a service or solution that will save 20% of 10TWh is obviously going to be more exciting and newsworthy than one that saves 20% of 10KWh. One such pitch started by claiming that there were eight million data centres, which is patently untrue unless you count individual servers as data centres.

Policy Evangelism

Non-commercial "propheteering" or doom-mongering is also popular with the more evangelistic brand of policy makers for whom apocalyptic predictions provide a much needed rationale for regulation and other legislative interventions. So when something promises to address an apocalyptic problem, check it out for commercial or political agendas. The truth is that climate change **IS** an apocalyptic problem, but we need to deploy ICT intelligently rather than vilify it.

Forgetting the reality check

Sometimes it is easy to get carried away with projections and charts and modelling and maths, but we have to remember to sanity check the results, especially if we have taken the highest figure from one set of studies, say energy intensity of the internet, and then multiplied it by the highest figure from another, say number of data centres. You can quickly end up in a very strange place, where we will all have to go and live in caves and eat bugs because data centres will have used up all the available power. But what is mathematically possible is not necessarily societally or economically possible. By multiplying two erroneous figures you also multiply the error margin! So the person who asserted confidently that UK data centres use 320TWh of power a year should have checked their claim against the UK's total electricity consumption, which at the time was 330TWh, and thought again.

Schoolboy Errors

We all make silly mistakes from time to time but they crop up with remarkable regularity here and always seem to end up exaggerating power demand. For instance the Shift Project Report [xii] confused bits with bytes – there is an eightfold difference – and had to issue a correction which was less widely publicised than the original. The uncorrected version is still being regularly quoted back to the industry. Other reports have multiplied power provisioning by 8760 to provide a figure in MWh, forgetting that data centres only use a fraction of provisioned power. Confusing energy and electricity is also surprisingly common. A country's electricity demand is usually only a fraction of its primary energy consumption, so it's important to use the right terms.

Laziness - Rounding up the Usual Suspects

It is surprising just how lazy people are when they look for information on data centre energy use, often going no further than the top two links from a Google search or resorting to the usual suspects – it's astonishing how often the 2006-7 Gartner analysis (see above) is still quoted. For the reasons already mentioned this comparison does not bear scrutiny, yet we are constantly being hectored about it. Laziness is also favoured by the fact that robust studies take time and so misinformation has plenty of opportunity to get embedded while those assessing its validity may take years to do the complex maths involved.

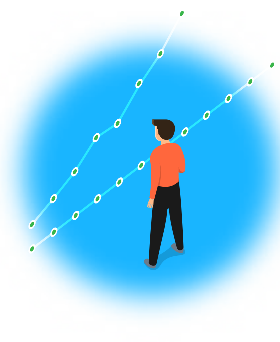
Matrioshkas – Double Counting

A single data centre supports multiple layers of customer activity, rather like a Russian doll or a sherry trifle [xiii]. The little doll in the middle is the colocation provider who leases space to customers who may use that space for their own servers and/or sell an IT service function to their own customers. So when trying to work out how much data centre space there is, we encounter an "I'm Spartacus" tendency in the industry: each company with presence in a data centre will view that presence as "their data centre". In the past this was occasionally exacerbated by operators engaging in what we called "MegaWatt Waving" [xiv]. This kind of over-reporting tends to be particularly problematic for industry surveys. A 2013 survey suggested that data centres consumed 20% of UK power by failing to de-duplicate responses before factoring up the results.

Extrapolating

Extrapolating is fun, but you have to be careful. Extrapolating trends is really problematic for industries where the underlying technology is changing rapidly so you have to be sure to go back far enough to reflect the underlying trend. There is a great explanation of this by Dr Jonathan G Koomey in his excellent analysis "Does not Compute: Facts and Fiction about Computing and the Environment" [xv] : He uses a wedding as an example: In 24 hours the bride has progressed from zero husbands to one husband so by extrapolation she will have around 50 husbands before the end of the next month.

Misapplying trends



It is easy to assume that two trends are interdependent but this is not always the case. For instance, assuming that an explosion in the demand for digital data will lead to a parallel explosion in data centre energy demand. This might seem logical – if we wanted more cheese, we would need more cows. But data is not cheese and the relationship is complex: yes the two trends are related, but more loosely than you might expect. Have a look at the IEA's chart of internet traffic and data centre energy use [xvi] where internet traffic is rising fast and data centre power is relatively flat. In fact the relationship is more likely to be, at least partially, cost-related: internet activity is price elastic and this acts as an indirect market control. If energy use spiralled in the same way as demand, then the cost of our online activity would increase and act as a constraint.

Using only part of the picture as a basis for the whole

Not seeing the whole picture can be a dangerous pitfall, as Arctic seafarers will have discovered, but is surprisingly common: we look at growth in one part of the sector, where it is most apparent, like hyperscale, and assume that the rest of the sector is growing at an equal rate, when in fact traditional on-premise and enterprise may be shrinking (which it is!). The problem for data centres is that growth is very visible, with large sites being built in tight clusters, but the parts that are shrinking are hidden on premises. So people are only aware of the growth, not what it is replacing.

We also have to remember that most studies only look at the problem from one angle - bottom up or top down, and surprisingly few industry-wide studies actually measure energy – they use modelling, along with a variety of accompanying assumptions.

Confusing Scenarios with Predictions

A zombie apocalypse is a scenario, not a prediction. A scenario is an imagined situation that may not be supported by existing data or evidence or previous events or practices. Scenarios are really important in risk management planning as this process has to accommodate low-likelihood high-impact events, such as coronal mass ejections, tsunamis and volcanic ash clouds. However, we run into problems when we use scenarios inappropriately, especially if we mistake them for predictions. As a result unrealistic, unlikely and unfeasible scenarios have all been used as a basis for declarations about data centre power. A frequently cited study [xvii] presented three different scenarios for power consumption depending on different rates of efficiency improvements in technology. The unrealistic 'worst case' scenario in this study was widely reported in media articles, despite the report authors highlighting other, more likely outcomes.

Is there anything we underestimate?

Er, yes, we probably underestimate – or at least under-report - the energy consumption of on-premises data centres and server rooms, for two reasons. Firstly, those in the public sector, of which there are many, are not obliged to report energy use, so we have no idea how much or how little they use. Secondly, on-premises data centres are unlikely to be run as business units or obliged to meet performance KPIs and so may be less incentivised to optimise energy use.

The only insight we have into this hidden world comes from the EURECA project which analysed around 350 on-premise public sector data centres in 2018 and reported average PUE [xviii] of around 5. So for each KWh used by the IT, there is a facility overhead of 4KWh. Compare that to PUE in the colocation sector which according to the CCA data is around 1.7, giving therefore a facility overhead of 0.7. This seems to indicate that the on-premise approach to computing is roughly six times less efficient than outsourcing, and that's just the facility infrastructure – we haven't started on the IT yet. When we do, we find that utilisation (how busy the servers are) and computational efficiency (how efficient the processors are, which tends to decline with age) were also low [xix].

Puzzling over these tantalising figures and factoring up in line with analyst data that suggests about half of computing activity is still on premise (let's hope this is wrong), we think the power being wasted every year within these environments could be measurable in TWh. This figure sounds so outlandish that Reality Check and Schoolboy Error alarm bells should be ringing. The problem is that nobody knows for sure because no systematic auditing or reporting has ever been conducted – which takes us back to the Available Data problem!

Since the UK's commercial sector collectively uses about 3.6TWh, all of which is measured, audited, reported and subject to efficiency KPIs, the on-premise energy consumption mystery is certainly something that merits further investigation. In the meantime, we alook forward to a more rational dialogue on data centre energy consumption.

Contacts

Emma Fryer
Associate Director, techUK
Tel: 01609 772 137
Mob: 07595 410 653

emma.fryer@techuk.org

Lucas Banach
Programme Assistant
Tel: 020 7331 2006

Lucas.banach@techuk.org

Further reading and links to relevant resources

Data Centre Energy Analysis: Past, Present and Future
Explanatory video by Eric Masanet, Professor and Mellichamp Chair in Sustainability Science for Emerging Technologies, UC Santa Barbara, 2021. https://www.youtube.com/watch?v=-o8j5zIM0iA

Electricity Intensity of Internet Data Transmission: Untangling the Estimates: Aslan, Mayers, Koomey and France; Journal of Industrial Ecology 2018: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3226029

Does Not Compute: Avoiding Pitfalls Assessing the Internet's Carbon and Energy Impacts:
Jon G Koomey and Eric Masanet, 2021: https://www.gwern.net/docs/cs/2021-koomey.pdf

Energy Tracking Report on Data Centres and Data Transmission Networks, International Energy Agency (IEA) 2020: https://www.iea.org/reports/data-centres-and-data-transmission-networks

The Carbon Impacts of Video Streaming: The Carbon Trust 2021
https://www.carbontrust.com/resources/carbon-impact-of-video-streaming

Relevant techUK Publications

Ten Myths About Data Centres (2019):
https://www.techuk.org/asset/67DBA646-0EC5-450C-B6F71E94DFDFD02B/

Does streaming really have a dirty secret? (2020)
https://www.techuk.org/asset/B9574F9E-0F76-4EAF-8CC61DAD056513C9/

Data Centres and Power: Fact and Fiction (2013)
https://www.techuk.org/asset/04020E61-9AD8-4C6E-BCF0E2C2A5B63F59/

Data Centre Energy Routemap (2019)
https://www.techuk.org/asset/502783FB-8F7B-44A3-B3931F2E2600A7A9/

The UK Data Centre Sector: The Most Important Industry You've Never Heard of (2020)
https://www.techuk.org/asset/3A709707-FAA0-495C-AEFF4D7A349C2D9A/

[i] See https://www.iea.org/reports/data-centres-and-data-transmission-networks, from the International Energy Agency, which explains the relationship between internet traffic, workloads and data centre energy consumption.

[ii] A recent historical TV drama series charting the Viking Age through the activity of Ragnar Lothbrok and sons.

[iii] A long running, entertaining and informative podcast reviewing the Icelandic sagas: https://sagathingpodcast.wordpress.com/.

[iv] Otherwise known as the Bullshit Asymmetry Principle, this is an internet adage based on proverbs about the relative speeds at which lies and truth travel, captured early on by Jonathan Swift as "falsehood flies and the truth comes limping after" and later by Churchill, among others.

[v] Firstly it compared emissions from aviation fuel with the life cycle emissions associated with ICT- a like-for-like comparison would have put aviation's emissions between 2 and 5 times those of ICT. Secondly the comparison confuses Scopes 1 and 2 (aeroplanes combust fuel but ICT emissions depend on the local electricity grid mix, so as grids decarbonise, ICT emissions reduce). Thirdly it doesn't allow for radiative forcing, which relates to the relative impact of emissions depending on where in the atmosphere they occur. Fourthly it does not mention the ability of ICT to help reduce emissions across the wider economy. Finally it does not reflect improvement potential: ICT has improved efficiency by around 7 orders of magnitude in the last 3 decades; aviation does not need a logarithmic scale to monitor efficiency improvements. So you can see why we hate this comparison. You can find a good analysis of it here: https://tekdeeps.com/no-the-internet-is-not-a-bigger-environmental-culprit-than-aviation/

[vi] See https://onlinelibrary.wiley.com/doi/full/10.1111/jiec.12630 We like to call this pattern of reduction Aslan's Law.

[vii] S. Namiki, T. Hasamaand H. Ishikawa, 2009

[viii] The Climate Change Agreement's definition for a data centre includes the criteria such as back up power supply, environmental controls for temperature and humidity and power supply of at least 200KW.

[ix] See https://www.nature.com/articles/d41586-018-06610-y

[x] https://www.theguardian.com/environment/2017/dec/11/tsunami-of-data-could-consume-fifth-global-electricity-by-2025

[xi] 2020 data from the Climate Change Agreement

[xii] https://theshiftproject.org/en/article/lean-ict-our-new-report/

[xiii] See Data Centre Business Models: The Sherry Trifle: https://www.techuk.org/asset/67F13798-15D7-4D87-A3AFA3536498375F/ and Data Centre Business Models and Services Infographic: https://www.techuk.org/asset/CF3DB785-E916-4DDA-9AC5B4BCE4E6A515/

[xiv] This refers to the now outdated practice where developers sometimes made competitive "mine's bigger than yours" type statements about power provisioning.

[xv] See: Recalibrating Data Centre Energy Use Estimates: https://www.science.org/doi/abs/10.1126/science.aba3758

[xvi] https://www.iea.org/reports/data-centres-and-data-transmission-networks

[xvii] Andrae (2015) On Global Electricity Usage of Communication Technology: Trends to 2030, Challenges, 6(1), 117-157; https://doi.org/10.3390/challe6010117

[xviii] PUE means Power Use Effectiveness and is the ratio of total power delivered to the facility to the power consumed by the IT within it. A high PUE is therefore undesirable, as it indicates a high energy overhead and low facility efficiency. PUE cannot go below 1 but the closer to 1 a facility can get, the more efficient it is considered to be. PUE is not a perfect metric: it is frequently and incorrectly used as a metric for overall data centre energy efficiency. PUE does not indicate overall data centres energy efficiency as it excludes the efficiency of the ICT that it houses. It is a useful trend analysis tool.

[xix] https://cordis.europa.eu/project/id/649972/results

About techUK

techUK is the UK's leading technology membership organisation, with more than 850 members spread across the UK. We are a network that enables our members to learn from each other and grow in a way which contributes to the country both socially and economically.

www.techuk.org

10 St Bride Street, London EC4A 4AD
techUK.org | @techUK | #techUK