techUK

techUK Data Centres Technical Committee

Briefing Note: Measuring Data Centre Performance

July 2017

A robust means of comparing data centre performance is one of the holy grails of the sector, and while we gradually get closer and closer to understanding how the performance of an individual data centre can be improved over time, developing indicators or metrics to compare the performance of different data centres against each other is very dangerous territory indeed. This is why.

What do we mean by data centre performance?

Data centre performance can mean lots of things. What we really want to know is how productive the data centre is in terms of inputs against outputs. This is why established metrics often use the term "effectiveness" instead of "efficiency". Inputs can mean anything from cost to raw materials. Operators measure performance to help them in strategic planning, for reporting and to identify where there is scope for investment or need for remedial action.

Performance metrics are not just financial and they vary in scope. A lot of data centre performance metrics are environmental (often referred to as "green KPIs") but each one has to be used with caution. The majority focus on energy, but environmental performance indicators include other aspects of sustainability such as asset disposal, water, energy re-use, life cycle impacts, and may even extend to biodiversity and responsible sourcing.

Because data centres are very energy intensive, energy is often used as an input rather than cost. So there are lots of energy related performance indicators. Some focus on how much peripheral energy the data centre needs as opposed to that needed for core functions (the most familiar indicator of this type for data centres is PUE – Power Usage Effectiveness). Some look at the type of energy purchased or generated, and some look at what happens to the energy coming out as heat. Others focus on the processing power of the servers in the data centre and their utilisation (how busy the server is, how close it is working to its capacity), compared to its total energy use or on the amount of data that is being transmitted in and out of the data centre – which is sometimes called "bits per Watt". Bits per Watt presents a really good example of why a simplistic approach to measuring data centre performance will deliver unreliable results – see box.

Box 1: Why not use Bits per Watt as a data centre KPI?

Data centres do different jobs. A single metric would favour some operations and penalise others. For example, high performance computing (HPC) involves very high quantities of data processing, very high utilisation of servers, high energy intensity (e.g. 30KW per cabinet) and high value but low volume output. Weather maps for instance use HPC because the size and complexity of the models and the sheer volume of data. However, the product is a map[1]. University research projects use this HPC and customers often have to wait for their turn[1] to access the computing resource. At the other end of the scale you might find an operation like Netflix, where there is storage but hardly any processing, but enormous quantities of content are delivered. So this data centre would have a high storage capacity, probably lower utilisation but very high levels of digital output as content is streamed.

This operation would perform very well against a bits per Watt metric, but the weather map would perform very badly. The metric would not give an indication of efficiency because the data centres are performing different functions.

Using metrics to compare sites – a word of warning

Performance metrics are not generally suitable for comparing one data centre with another and should never be used in isolation. This is largely because of the variety of operating models within the sector, the range of activities and the sheer complexity of the infrastructure, which brings together multiple disciplines like telecoms, IT, engineering and building control. PUE is the metric most frequently used as a single indicator of performance, but it would be unheplful if comparing sites designed for different levels of resilience, with different approaches to cooling or in different geographical regions.

So for instance a site with a very low PUE might have be very water intensive and therefore have a high WUE. This might not matter if the site were in Stockholm but it would mattter if that site were in Madrid. Or a site built for very high resilience with lots of redundancy built in would have a higher PUE than a site used for batch processing. Some hyperscale operations duplicate search facilities between two or even three data centres, and while individual facilities may have a very low PUE, observers argue that using two facilities to do the same job means that the PUE should be doubled. PUE is also a poor tool to compare new sites that are only partially filled with mature sites, because the fixed energy overhead at a new site is amortised across a smaller IT function until the facility fills up.

The data Centre conundrum: Infrastructure performance vs. IT performance

As mentioned above, data centres are complex environments that bring together multiple technologies. Best practice guidance aimed at improving energy stewardship in data centres (such as the EU Code of Conduct for Data Centres and the Green Grid Maturity Model) take account of this. There has been a tendency, however, to think that metrics designed for different parts of the data centre can be used interchangeably. They can't. In particular it is absolutely critical to make the distinction between infrastructure efficiency and IT efficiency. The former is the ability to deliver power to the compute (IT) elements. The latter is how effectively the compute elements do work. Some metrics, such as PUE, just cover infrastructure, and others, such as SERT, just cover IT. So we have to understand the relative scope of metrics, where and when they can be applied meaningfully and what they can – and cannot – tell us.

So PUE does not tell us anything useful about the efficiency of the computing operation in the data centre. SERT does not tell us anything about the efficiency of the infrastructure. Frustrated by this, observers often call for a single universal metric. However, this may not be productive. One of the most common data centre business models is the provision of serviced infrastructure within which third party customers install their own IT equipment. The data centre operator has no control over the IT function. That is why the best guidance materials take a more modular approach, addressing the different data centre functions individually. Any approach attemptingn to cover multiple functions will inevitably be very complicated.

Looking ahead

It's probably fair to say that metrics designed to measure data centre infrastructure efficiency are better developed than those measuring IT efficiency (i.e. we are better at measuring how efficiently we deliver energy to the computing function than how the computing function itself uses energy). Moreover, determining IT efficiency is going to become even harder as we move to cloud: measurement will become increasingly abstracted, with much higher dependence on software application architecture and efficiency of scheduling workloads. Even the concept of a 'server' and its utilisation as a proxy for useful workload is starting to erode as we move towards micro-services and even server-less computing[1].

[1] Physical servers were superseded by virtualisation which increased utilisation. Now 'Docker' containers make that virtualisation more efficient. Micro-services and server-less computing build on this and allow a workload to scale from a single container to an entire data centre and back down again with all customers and workloads sharing the server capacity which in thery enables 100% utilisation, dramatically improving efficiency.

What Makes a Good Metric?[2]

A common misconception is to confuse something that is measurable with something that is important. What we really need to do is try to make important things measurable. While we work on that, here are some of the things that we think metrics should be:

- **Widely Applicable:** Does the metric apply to most facilities? Can it be applied to different types of data centre operation? Can it work in different countries and regions?
- **Easy to measure:** Is it easy to apply the metric meaningfully? Can it be assessed relatively simply without adding cost or time burdens for the operator?
- **Can be used to compare facilities:** Can it be used to compare the performance of one facility with another in a meaningful way, even when facilities might have different functions?
- **Can be used to benchmark facilities over time:** Can the metric be used to track changes in performance in a single facility over time?
- **Application governed by a relevant standard:** Is there a standardised approach for applying and/or assessing performance against the metric?
- **Useful:** Does the metric actually tell us anything useful? A metric that rates data centres on the number of toilets they have might fulfil all the criteria above but not be useful at all.

In this context it is easier to understand why PUE is such a popular metric (or, perhaps more accurately, the "least worst" metric). It's relatively easy to measure, it's relatively easy to understand, its shortcomings are understood and it (mostly) drives the right behaviours.

So the Perfect Metric is....

Absent. The perfect indicator is simple, useful, widely applicable and allows comparisons between sites and benchmarking over time. Sadly, no perfect metric exists for data centres, despite sustained activity by international groups. Some create perverse incentives and all have disadvantages. These are illustrated in the following pages:

- Table 1 presents a précis some common performance metrics relating to energy efficiency
- Table 2 does the same for other sustainability related metrics used in the sector
- Figure 1 maps common metrics according to where they apply in terms of infrastructure and life cycle
- Annexe 1 provides a more complete list of standardised metrics, including those in development

[2] For a more detailed overview of what makes a good data centre metric see http://dcsg.bcs.org/sites/default/files/protected/DCSG_Metrics_-_DCD_-_December_2009.pdf

Table 1: Energy efficiency related performance metrics

Name	What does it do?	Pros and cons	Widely applicable?	Easy to use?	Can be used to compare sites?	Can benchmark over time?	Relevant standard?	Useful / relevant
CCF – Cooling capacity factor	It rates how efficiently cooling infrastructure is being used. High CCF suggests poor air flow management.	Not very well known so pros and cons largely unreported.	Yes	?	Not really	Yes	No	Yes
DCEP Data Centre Energy Productivity	Rather similar to DCIE but it uses a mixture of qualitative analysis and measurement. DCEP includes the "Bits per Watt" metric	Qualitative element means that it can be subjective. Very complicated. Different elements of DCEP have different pros and cons.	Not alw ays	No	Not really	Not alway s	No	No
DCIE Data Centre Infrastructur e Efficiency	Predates PUE and is its inverse. Expressed as a percentage, the higher the figure, the better. A data centre where all incoming energy is used by IT would be deemed 100% efficient.	A % measure of efficiency is more intuitive for people to understand, and in line with other efficiency measures like those for boilers. Can create perverse incentives for operators like PUE.	Yes	Yes	No	Yes	Yes (by inversion of PUE below)	Yes
DPPE Data Centre Performance per Energy	Developed in Japan, this holistic approach involves multiple KPIs presented in the form of a spider's web.	It is not clear whether the approach has been finalised or applied in practice.	Yes	No	No	Yes?	No	
ERF Energy Reuse Factor	This is a measure of how much of the waste heat from the data centre is re-used, for instance in district heating systems or other heat sinks.	Reuse of waste heat depends on demand and is easier to implement in new build than to retrofit. May be restricted to regions with scope for district heating systems.	Yes	Yes	Yes	Yes	Yes Future ISO/IEC 30134-6/EN 50600-4-6	Yes
PUE Power Usage Effectivenes s	The ratio of the total energy used by the data centre to the energy used by the IT. The lower the PUE the better.	Widely used and pros and cons understood, however it is often abused. "Gaming" to manipulate the PUE score favourably is common.	Yes	Yes	No	Yes	Yes ISO/IEC 31034-2/EN 50600-4-2	Yes
SCE Server Compute Efficiency	Developed by ISO 30134 as ITEE / ITEU, it looks at the efficiency of the server itself, broken down by equipment type using different subscripts.	If reflects the fact that different equipment behaves differently – e.g. small servers vs higher performance servers, storage on spinning disks as opposed to tape.	Yes	No	Serve rs only - not opera tional	No	Yes Future ISO/IEC 30134-4	See[3]
SERT Server Efficiency Rating Tool	Developed by SPEC, the Standard Performance Evaluation Corporation, this tool measures and evaluates the energy efficiency of servers.	There are many different types of servers which the tool accommodates but compromises are inevitable. However, this tool is widely respected.	Yes	No	Serve rs only - not opera tional	Yes	No- see note[4]	Yes
TGG PI Green Grid Performance Indicator	Combines PUE, IT Thermal Conformance (cooling efficiency) & IT Thermal Resilience (ditto, outside normal operations)	Complex and depends on detailed data collection. Recently developed so yet to be established.	Yes	No	No	Yes	No	?

[3] This is a server procurement KPI like ISO/IEC TR 21836 and ETSI EN 303 472

[4] SERT is a tool which is the basis of ISO/IEC TR 21836 and ETSI EN 303 472

Table 2: Other environmental / sustainability performance metrics relevant to data centres

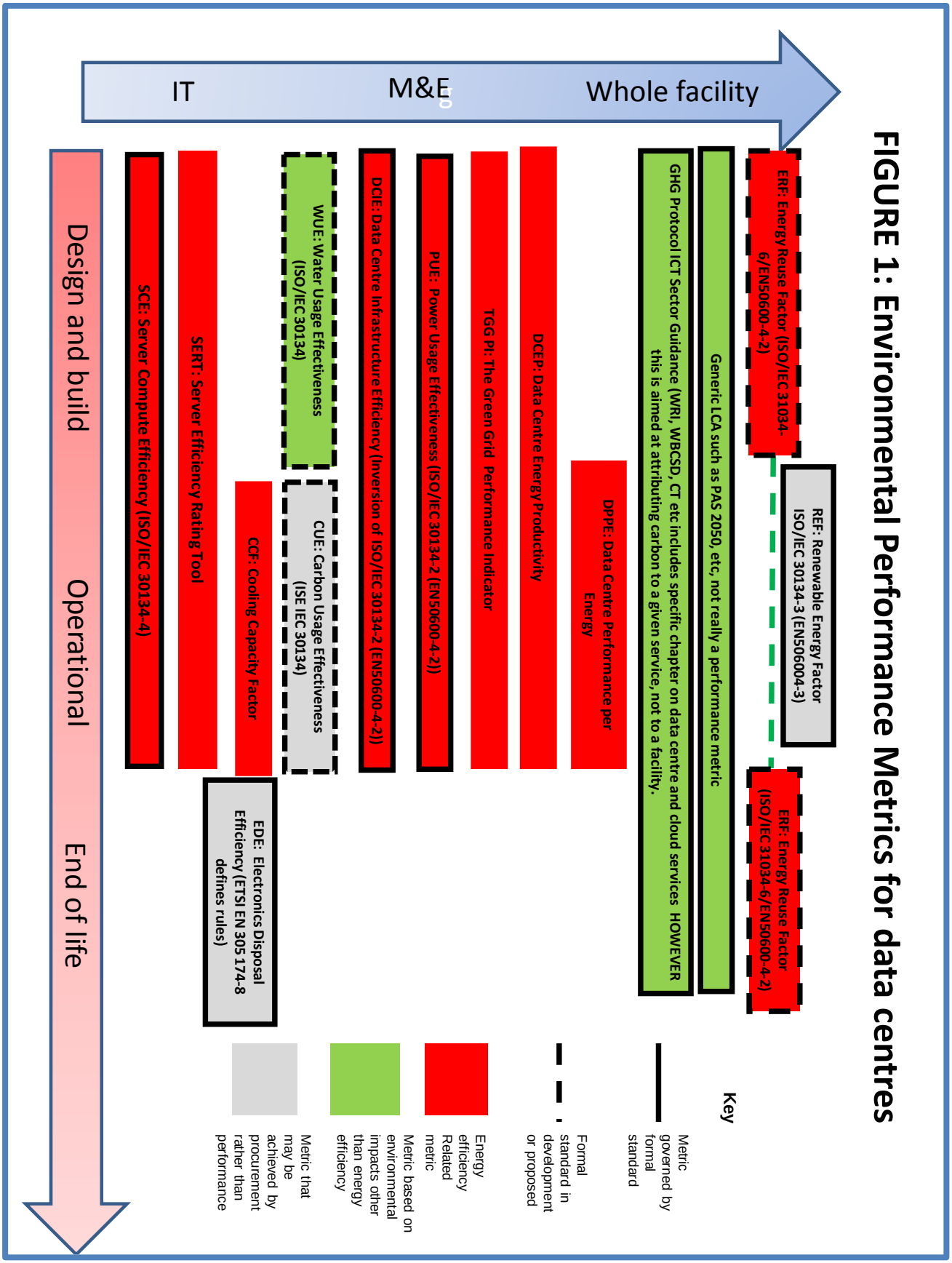
Name	Description	Pros and cons	Widely applicable?	Easy to use?	Can compare facilities?	Can benchmark over time?	Relevant standard?	Useful/ relevant?
CUE Carbon Usage Effectiveness	Similar in concept to PUE but measured in terms of carbon, so the outcome depends on the nature of the energy supply.	CUE could be improved purely through procurement decisions rather than through improved energy stewardship, so it is not necessarily a measure of energy management or efficiency.	Yes	Yes	No	Yes	In process: see note[5]	
EDE Electronics Disposal Efficiency	This is the percentage of IT and electrical equipment that, once decommissioned, is disposed of through formally recognised responsible entities. It was developed by the Green Grid to help operators ensure they were managing their old kit responsibly.	By law all electronic waste within the EU must be disposed of responsibly or recycled. In theory, therefore, 100% of electronic waste should be accounted for.	Yes	?	Yes?	Yes	Yes ETSI EN 305 174-8 defines rules - but not a KPI	Yes
REF Renewable Energy Factor	This is the proportion of renewable energy that is used within the data centre. This replaces the green energy coefficient because green energy is a meaningless term.	In some cases renewable energy is simply a purchasing decision and does not necessarily stimulate additional investment, nor should it be seen as a substitute for good energy stewardship.	Yes	Yes	Yes ?	Yes	Yes ISO/IEC 30134	?
LCA Life Cycle Assessment	A generic approach not specific to data centres that looks at all impacts across the life including manufacture, use and disposal. Likely to include many of the other metrics in this table.	LCA is intended to identify hotspots in a process so that efficiency measures can be targeted intelligently. Using LCA to compare products or services is extremely complex and problematic[6].	?	No	No	Yes	Yes lots	Yes within its intended scope
WUE Water Usage Effectiveness	Measures water use against IT function. Water reuse and recycling metrics are also used by some operators.	Water use is of very variable sensitivity depending on the location of the facility.	Yes		No	Yes	In process - see note[7]	Yes

[5] Possible future standard in ISO/IEC 30134 series

[6] See : our 2012 report: Evaluating the carbon impact of ICT: The Answer to Life, the Universe and Everything: http://www.techuk.org/insights/reports/item/459-evaluating-the-carbon-impact-of-ict-or-the-answer-to-life-the-universe-and-everything

[7] Possible future standard in ISO/IEC 30134 series

Figure 1
Map of environmental performance indicators for data centres



Annexe 1

Current List of standards governing data centre performance metrics (from official standards bodies)

ISO/IEC and CENELEC

- ISO/IEC 30134-2 (=EN 50600-4-2): PUE (and by inversion, DCIE)
- ISO/IEC 30134-3 (=EN 50600-4-3): REF
- Future ISO/IEC 30134-6 (ERF) – early stage development and then presumably EN 50600-4-6

ETSI

- ETSI EN 305 200-2-1: Global KPI for energy management of ICT sites (KPI_{EM})
- ETSI EN 305 200-3-1: Global KPI for energy management of ICT sites (KPI_{DCEM})

NB: None of the above are intended for comparison of data centres and this is clearly stated.

Existing (or in development) standards for server procurement:

- Future ISO/IEC 30134-4 (ITEEsv) – at final vote stage and then likely to become EN 50600-4-4
- Future ISO/IEC TR 21836 (SEEM – server energy efficiency metric) – only a technical report – based on SERT – not operational assessment
- Future ETSI EN 303 472 (server energy effieincy KPIs) – based on SERT – not operational assessment

NB: None of these server metrics can be used to compare facilities or benchmark operational trends

Existing (or in development) standards for server utilisation

- Future ISO/IEC 30134-5 (ITEUsv) – at final vote stage and then presumably EN 50600-4-5

Potential KPIs identified by ISO/IEC JTC1 SC39

- IS WUE
- IS CUE
- TR on application of 30134 series
- IS Water Reuse Factor (WRF)
- IS Resiliency Class or Type of Data centre
- IS Data centre cost expense (DCCX)
- IS Data centre cost effectiveness (DCCE)
- IS Cooling Effectiveness Ratio (CER)
- IS ITEE and ITEU networking
- IS ITEE and ITEU storage
- IS Seasonally adjusted Energy Efficiency Ratio (SEER)
- IS Coefficient of Performance (COP)
- TR Economic Output
- TR Resilience Risk Impact
- TR Data Centre Life Cycle Impact

IS - International Standard

TR – Technical Report

Note: Not all are environmentally focused but they indicate the range under consideration.

About techUK's Data Centres Technical Committee

techUK's Data Centres Technical Committee was originally established to advise techUK's Data Centres Council and provide expert technical input to policy responses, publications and other communications. However, the Committee's expertise is increasingly being sought during dialogue between industry and external stakeholders (predominantly government) for objective advice about the technical and market characteristics of the sector. Members collectively possess a wide spectrum of industry expertise; while the core focus is on engineering and technology, some cover operations and some are experts on business models and market trends, policy or legislation. The Committee includes external observers to ensure objectivity. Formal Terms of Reference provide governance for the group and a list of members is available.The Chairman is Professor Ian Bitterlin and the Vice Chair is Mark Acton.

Acknowledgements

We would like to thank the following for their technical and editorial contributions:

Mark Acton, CBRE
Ian Bitterlin, Critical Facilities
Mike Gilmore, e-Ready Building
Zahl Limbuwala, Romonet
Dave Smith, DataCentred
Andie Stephens, the Carbon Trust
Steve Strutt, IBM

Further information and 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

About techUK: techUK represents the companies and technologies that are defining today the world that we will live in tomorrow. The tech industry is creating jobs and growth across the UK. In 2015 the internet economy contributed 10% of the UK's GDP. 900 companies are members of techUK. Collectively they employ more than 800,000 people, about half of all tech sector jobs in the UK. These companies range from leading FTSE 100 companies to new innovative start-ups. The majority of our members are small and medium sized businesses.**www.techuk.org**