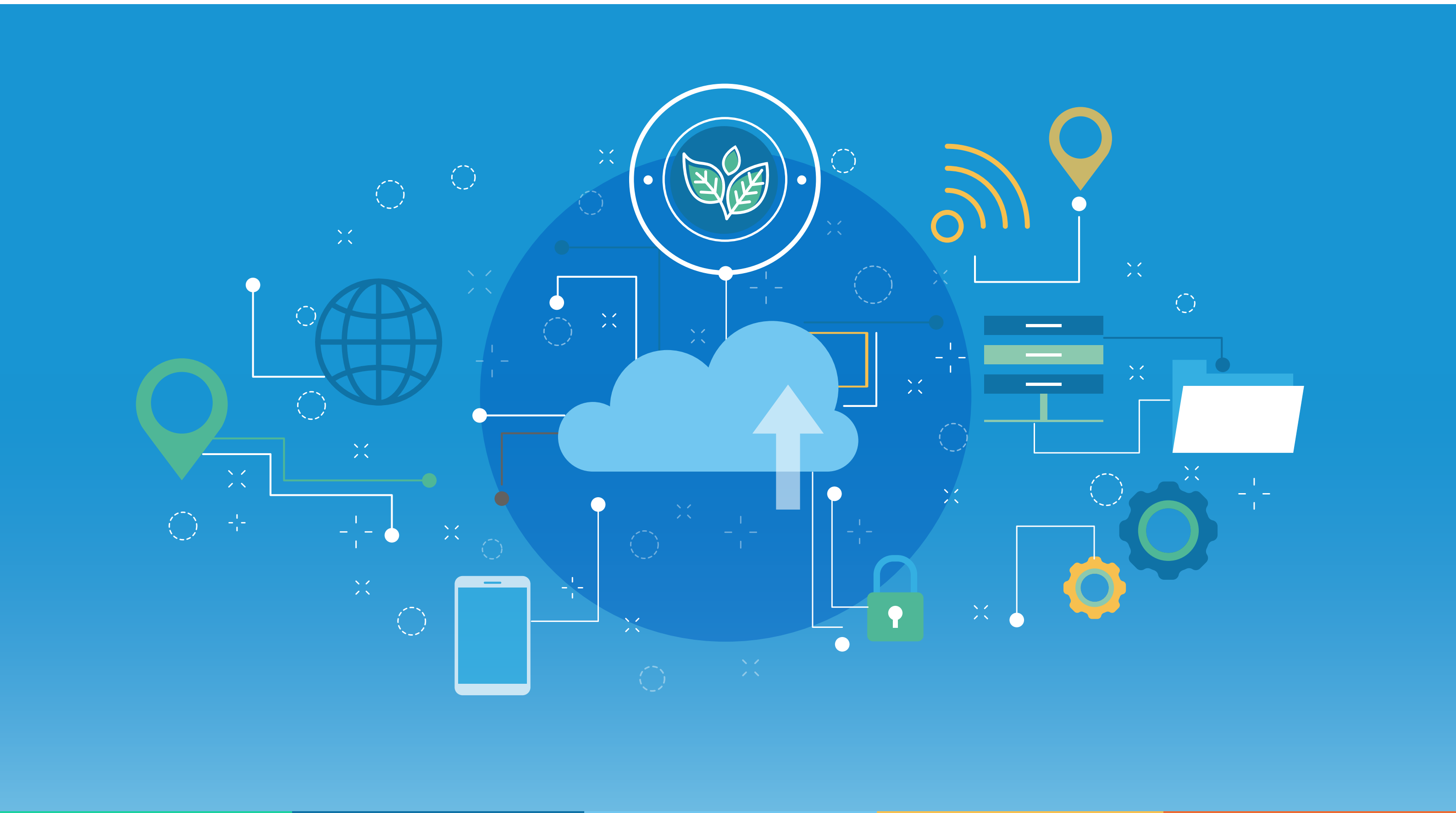
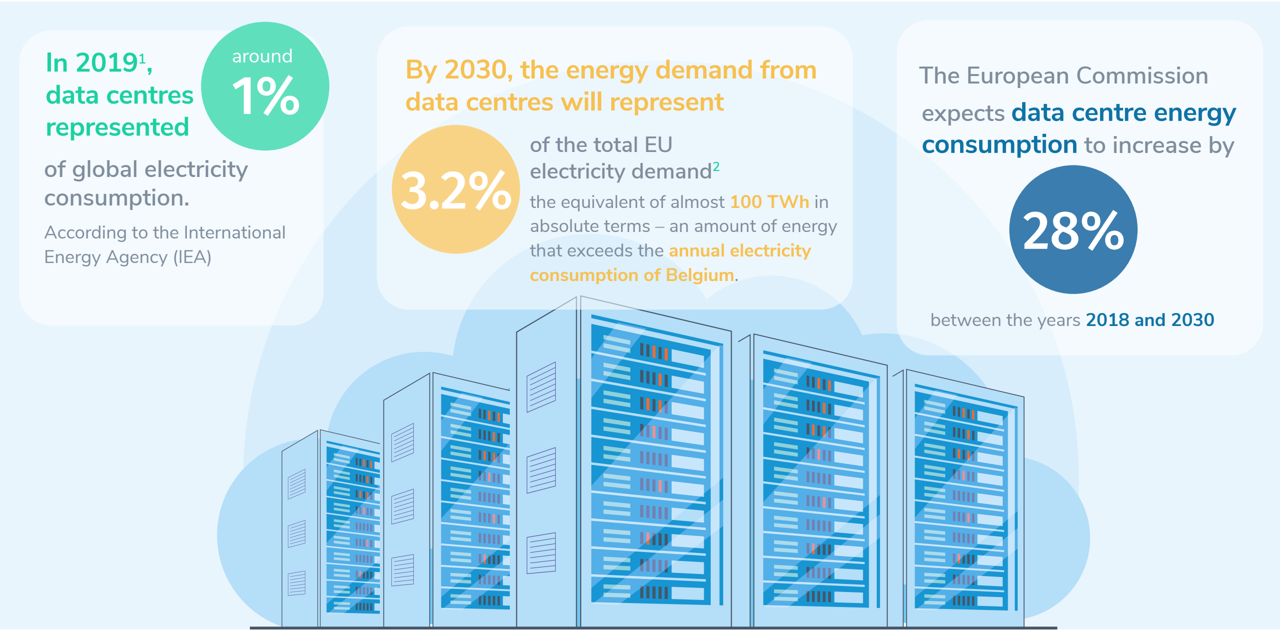
Climate neutral clouds?

What we need to achieve climate neutral, highly energy efficient and sustainable data centres in Europe by 2030



Introduction

Data centres are the factories of the digital world. As digital services move into the cloud, their energy consumption moves into data centres and their environmental impact becomes invisible to users.



In spite of significant improvements in the energy efficiency of data centre infrastructure in recent years, a booming expansion of digital services continues to drive the electricity demand up, and the growth has already more than offset the efficiency gains[3]. To achieve climate targets, it is important to bring this rising demand under control.

The European Commission has set a goal of achieving climate neutral, highly energy efficient and sustainable data centres by 2030. The existing policy aimed at managing data centre energy demand, however, is underdeveloped and a comprehensive and widely accepted method for assessing energy efficiency of data centres is missing. Without an appropriate performance metric, we cannot monitor progress towards the target, let alone set requirements towards achieving it.

In this briefing we highlight the shortcomings of current metrics and call for action that will ensure the development of an assessment method and deliverance of climate neutral data centres by 2030, urging the European Commission:

1. to ensure the swift development and adoption of an energy efficiency indicator for data centres;
2. to develop a definition of what constitutes a sustainable data centre, to allow evaluation of the 2030 goal;
3. to issue targets and legislative measures that ensure compliance with minimum energy performance requirements for data centres, and transparency in energy efficiency reporting;
4. to issue an ambitious standardisation request to European standardisers in support of the regulation, leading the way on standard-setting for data centres internationally.

The data centre stack

To understand the problem of assessing energy efficiency of data centres, it is necessary to be familiar with the structure of data centre operations.

A data centre is essentially a large collection of interconnected computers, gathered under one roof. They are digital infrastructure nodes where data is processed and stored before it is communicated via networks to the user. Data centre operations are, however, far from simple.

The complex structure of data centres

The technical structure of a data centre is often described as a stack, meaning that the physical and digital structures that make up the data centre are organised in layers, each building upon the services from the layers below.

An end-user application, such as an email client, exists on a supporting software infrastructure, including an operative system. These are different layers of code that, in turn, are hosted on ICT equipment, or hardware, that can be categorised into servers, storage and network. The ICT equipment needs a stable electricity supply, dust-free air and a cooling system to remove heat produced when it is working. These functions are called the data centre infrastructure. Finally, the ICT equipment and data centre infrastructure need to be put in an appropriate location, with a high-capacity electricity supply, and support systems including for example security.

Adding complexity to the picture, the different layers of the stack are often managed by different actors within the data centre industry. For instance, a single physical 'colocation' data centre, a common business model in the industry, will house ICT equipment from multiple clients who, in turn, provide so-called 'managed services' to many of their own clients.

Conversely, a single application can be operating on ICT equipment split across multiple physical data centres. For instance, Apple provides applications to its users from enterprise data centres owned and operated by Apple. However, Apple also uses services that they contract from Google. Google in turn operates many data centres of their own but also uses colocation services that they contract from third parties, to house additional ICT equipment.

Another example is the geographical spread of Netflix. Here, a single application (Netflix video streaming) uses Amazon managed services, as well as works with internet service providers all over the world to install servers within their networks (similar to colocation data centres). It was estimated that, in 2019, Netflix had approximately 8,500 servers in 620 locations globally.

These complex webs of services make the assessment of energy demand and efficiency challenging. Getting a full grasp of the efficiency requires energy data from all layers, split across multiple companies and locations. In addition, one needs a methodology to apportion work and energy appropriately between several clients of the same service provider operating at each layer.

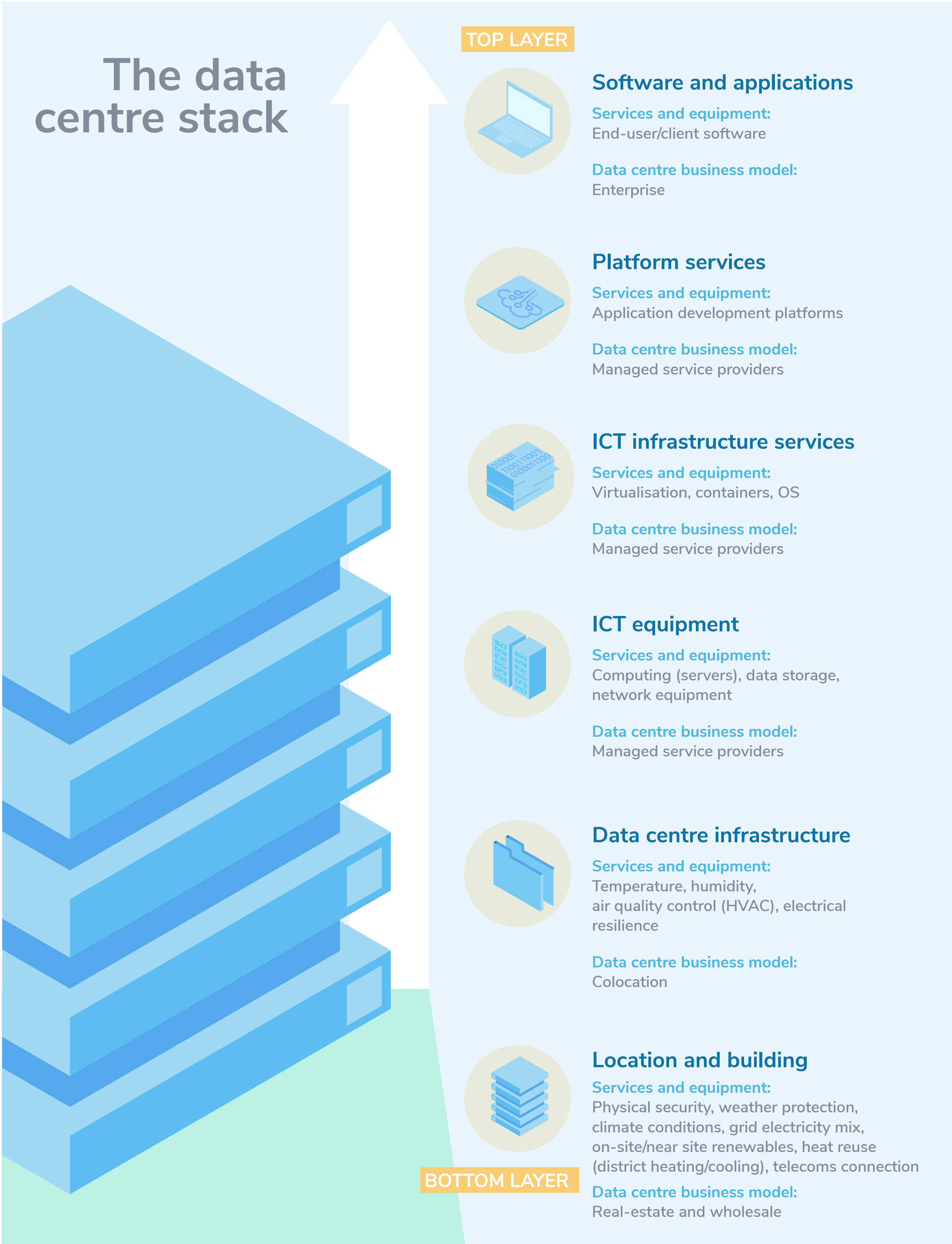


Figure 1 The data centre stack. Many layers, different actors

Measuring data centre energy performance

Assessing the full stack is – as we have seen in the previous section - a complicated exercise. Instead, companies often simplify the problem by focusing only on their own energy, and apply assessment methods, from the layer(s) for which they are directly responsible, or for which they have data. As a result, declaring energy performance of data centres is often focused on the lower layers of the stack.

Power Use Efficiency

The most widespread indicator of data centre energy performance is Power Use Efficiency (PUE). It is defined as the total data centre energy use over a year divided by the ICT equipment energy use for the same period. PUE is used to set efficiency limits for new data centre infrastructures and is often quoted by cloud and colocation data centres (but not necessarily with supporting data). It has also been used to monitor colocation data centres in the UK as part of the EU Emissions Trading Scheme. Moreover, Singapore, China and Amsterdam have PUE requirements for new data centres.

PUE defines useful energy as the energy used in ICT equipment, while the energy used in the two lower layers, building and data centre infrastructure, is viewed as losses. A problem with PUE, especially in view of its popularity, is that it implicitly assumes that an increase in energy supplied to the ICT equipment means an increase in data centre productivity, but this is not necessarily true. ICT equipment can, unfortunately quite often, consume energy without delivering any useful work.

Server and ICT efficiency

Other metrics have been developed in attempts to address efficiency of the layers above PUE. Server Energy Rating Tool (SERT) is used to assess computer server efficiency. Data Centre Compute Efficiency (DCCE) metric assesses server utilisation, while the Application Platform Energy Effectiveness (APEE) is designed to measure the efficiency of a single application running on a specific IT system. A general issue with these metrics is that their applicability to and representativeness of real data centre operations is limited, since these operations are very diverse. Besides, they do not allow comparison between different data centres.

Data centre productivity

A relatively promising metric for addressing the energy efficiency of the data centre as a whole is Data Centre Energy Productivity (DCEP). DCEP is defined as the useful work produced, divided by the total energy consumed in the data centre. It follows the generic energy efficiency formula for useful work output divided by the total energy input. It is the most intuitive way to assess a digital service provided by a data centre, and, in theory, it bypasses all the assessment methods for lower layers as long as the energy consumed can be apportioned correctly. As we have already mentioned above, this is not always easy. Another potential issue is linked to defining and measuring useful work and comparing it to other data centres.

An overview of the metrics mentioned above, as well as additional metrics relating to the building layer, are found in Table 1. A more detailed description of all metrics is found in the Annex.

	Full name	Layer	Description
PUE	Power Use Efficiency	Building and data centre infrastructure	Widespread, but does not account for ICT equipment and software efficiencies.
SERT	Server Energy Rating Tool	ICT equipment	Server performance, but not if performing useful tasks/work. Designed to be comparable, but does not represent real-life operation.
DCCE	Data Centre Compute Efficiency	ICT equipment	Server utilisation. In theory assesses useful work, but this definition is arbitrary and only used internally.
APEE	Application Platform Energy Effectiveness	ICT equipment and soft-ware layers	Efficiency of a single application running on a specific IT system. Only used in a controlled environment, not for an operational data centre.
DCEP	Data Centre Energy Productivity	All layers (in theory)	Useful work produced, divided by the total energy consumed in the data centre. Problem to define useful work and allocating energy use.
REF	Renewable Energy Factor	Building	Indicator of renewable energy integration, but not efficiency.
ERF	Energy Reuse Factor	Building	Indicator of energy reuse, but not energy efficiency inside the data centre.

Table 1 **Overview of sustainability metrics for data centres**

In general, energy efficiency assessment in data centres is stuck at the doorstep of the ICT equipment layer and has yet to move above it. The progress is painstakingly slow due to the lack of consensus and impetus to determine how to measure performance or service. It is simply not yet possible to assess end-user applications that deliver the actual useful work. In addition, there are no comparative metrics beyond data centre infrastructure.

Though research[4] has identified over 30 sustainability metrics for data centres, we are still lacking an assessment method that is comprehensive enough to capture the performance of the data centre as a whole, that gives a good enough representation of real data centre operations, and that can allow fair comparison between different data centres. Until this gap is filled, customers and policy makers alike will lack a fundamental tool to make informed decisions and set requirements.

Reaching climate neutral, highly energy efficient and sustainable data centres by 2030

As part of its Digital Strategy[5], the European Commission has set the goal of achieving climate neutral, highly energy efficient, and sustainable data centres by 2030. The three main policy initiatives at EU level to manage the rising energy demand from data centres to date are the EU Ecodesign requirements for servers, the EU Data centre Code of Conduct, and EU Green Public Procurement criteria for Data centres, the last two of which are voluntary initiatives. However, a recent report[6] concludes that the existing policy instruments are underdeveloped and work merely as a starting point or inspiration. In other words, European policymakers are currently lacking the means to achieving the goal of climate neutral data centres that should be fulfilled in less than 10 years. A uniform energy efficiency indicator for data centres is also recognised in the report as a key requirement for advancing policy for climate neutral data centres.

Given the formative and expanding phase that data centre industry is in, putting adequate measures in place is an urgent matter. We want to see a clearer commitment to reducing the environmental footprint of the digital sector.

We therefore urge the European Commission to develop an action plan to achieve climate neutral, highly energy efficient and sustainable data centres by 2030 without further delay. The planned actions should include:

1. **mobilising resources** to ensure the swift development and adoption of this much-needed energy efficiency indicator;
2. **developing a definition of what constitutes a sustainable data centre**, so that it will be possible to evaluate whether the 2030 goal was reached;
3. **setting ambitious targets and legislative measures** that ensure compliance with minimum energy performance requirements for data centres, and transparency in energy efficiency reporting;
4. **issuing ambitious standardisation requests to the European Standardisation Organisations** in support of the policies and regulations, with the co-benefit of Europe becoming the standard-setter for data centres globally.

Finally, while outside the scope of this briefing, reaching sustainable data centres also entails addressing many other relevant sustainability aspects, including circularity of hardware, water consumption, cooling, and the use of renewable energy. As the capacity of our digital infrastructure keeps increasing, it is also relevant to consider the future of this sector from the point of view of *sufficiency*, i.e. to not use more resources than required to meet our needs. These aspects must be considered holistically and alongside energy efficiency in the efforts recommended above.

In conclusion, we call on the Commission to move from what is currently at best wishful thinking to taking action that will ensure deliverance of climate neutral, highly energy efficient and sustainable data centres by 2030.

Annex
Data centre metrics

This annex presents, some of the most widespread sustainability metrics for data centres, categorised according to what layer they refer to.

Building layer metrics

At the building and location layer, sustainability assessment metrics show aspects relevant to energy system integration of data centres. Here we present two: Renewable Energy Factor (REF) is the ratio of the renewable energy owned and controlled by a data centre, and the total data centre energy consumption. The metric shows the level of renewable energy integration. Secondly: Energy Reuse Factor (ERF) is a measure of how well waste energy, primarily heat, from the data centre is put to use towards another activity, such as heating greenhouses, drying biomass or feeding thermal energy into district heating.

$$\text{REF} = \frac{\text{Renewable energy supplied to data centre}}{\text{Total data centre energy use}}$$

$$\text{ERF} = \frac{\text{Reuse energy}}{\text{Total data centre energy use}}$$

These metrics do not show the efficiency of data centre operations but merely give us information about energy system integration of data centres, as well as how well they contribute to a renewable-based and resource efficient energy system.

Data centre infrastructure

By far, the most widespread indicator of data centre energy performance is **Power Use Efficiency (PUE)**. It is defined as the total data centre energy use over a year divided by the ICT equipment energy use for the same period. In other words, PUE defines useful energy as the energy used in ICT equipment, while the energy used in the two lower layers, building and data centre infrastructure, is viewed as losses. The PUE metric calculates how many units of energy are used to deliver one unit of energy to the ICT layer with the ideal PUE value being 1.0.

$$\text{PUE} = \frac{\text{Total facility energy use over a year}}{\text{IT equipment energy use over a year}}$$

A development of PUE that addresses the carbon intensity of the data centre energy mix is **Carbon Usage Effectiveness (CUE)**, which is defined as the total CO_2 emissions associated with the data centre energy supply over a year, divided by the ICT equipment energy use.

$$\text{CUE} = \frac{\text{Total } CO_2 \text{ emissions caused by the Total Data Center Energy over a year}}{\text{IT equipment energy use over a year}}$$

Both these metrics, however, also have their flaws: they do not account for efficiency in the higher layers but simply assume that an increase in energy supplied to the ICT equipment means an increase in productivity of the data centre, which is not necessarily true.

ICT equipment and applications

Moving beyond the building and infrastructure layers is where the difficulty in defining and measuring useful work arises. Here we look at three metrics that attempt to address this is different ways.

Server Energy Rating Tool (SERT) is used to assess computer server efficiency. This is used for USA EPA ENERGY STAR and the Ecodesign regulations (EU) 2019/424. However, SERT can only measure the performance of the equipment, not how efficiently the work is being used. For example, a server can perform computational work, e.g. process a code, in a very efficient way, but if that code is not useful, the IT equipment will not be producing any useful service. Examples here are many: a search algorithm that looks for information in the same place several times is most certainly not an efficient code; servers "stranded" and forgotten in data centres do not provide any useful work even though they remain switched on and consume energy.

Furthermore, SERT is designed to be comparative and highly repeatable and therefore uses a synthetic benchmark that is not related to any particular type of application, or real-life measurements of servers actively in use in data centres.

Data centre compute efficiency (DCCE) is a metric to assess utilisation. However, the intention is to measure how much of the utilisation provides a useful service. For example, email servers primary use is email, but background services, such as frequent back-ups, will also use processing and storage resources. This is an internal metric for optimising operations, but companies define the useful work as they see fit.

Application platform energy effectiveness (APEE) is designed to measure the efficiency of a single application running on a specific IT system. It therefore accounts for every layer from the application to the ICT equipment, i.e. multiple software layers, and combined with PUE may also offer a rather complete assessment of the data centre in the specific case of the application being assessed. The intended use of APEE is for procurement of an IT system and, in that instance, it is more representative than, for example, SERT. However, APEE is not designed for measuring the efficiency of an operational data centre, as it uses an application-specific benchmark in a controlled environment. Furthermore, it is not clear how to apply this metric if the ICT equipment is running multiple applications, which is not unusual.

Assessing the whole data centre

In the sections above, we have presented some metrics that are used to assess parts of the data centre stack. We will now discuss a metric that attempts to take a holistic approach.

Data Centre Energy Productivity (DCEP) is defined as the useful work produced, divided by the total energy consumed in the data centre. It follows the generic energy efficiency formula for useful work output divided by the total energy input. It is the most intuitive way to assess a digital service provided by a data centre, and, in theory, this bypasses all the assessment methods for lower layers as long as the energy consumed can be apportioned correctly, but this is not always easy. Another potential issue is linked to defining and measuring useful work and comparing it to other data centres.

In simplified terms, DCEP defines useful work based on the number of tasks completed over a given assessment period, while also taking into account that some tasks are bigger and more important than others. Using this approach can, however, quickly become overwhelming if a large number of tasks are defined, along with individual weightings of importance. In addition, interpreting the result could be very difficult without intimate knowledge of every task. Nonetheless, taken in its simplest form (one task and no weightings), it provides a possible starting point to communicate efficiency for simple services, such as hours of video streamed, and to build from there, for example by including hours of user-generated video uploaded.

Notes and references

[1] International Energy Agency (2020) Data Centres and Data Transmission Networks URL: https://www.iea.org/reports/data-centres-and-data-transmission-networks

[2] European Commission (2020), Green and Digital: study shows technical and policy options to limit surge in energy consumption for cloud and data centres URL: https://ec.europa.eu/info/news/green-and-digital-study-shows-technical-and-policy-options-limit-surge-energy-consumption-cloud-and-data-centres-2020-nov-09_en

[3] Montevecchi, F., Stickler, T., Hintemann, R., Hinterholzer, S. (2020). Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market. Final Study Report. Vienna URL: https://digital-strategy.ec.europa.eu/en/library/energy-efficient-cloud-computing-technologies-and-policies-eco-friendly-cloud-market

[4] Vemula, Dr-Dinesh Reddy & Setz, Brian & Rao, G. & Gangadharan, G R & Aiello, Marco. (2017). Metrics for Sustainable Data Centers. IEEE Transactions on Sustainable Computing. URL: https://www.researchgate.net/publication/317560905_Metrics_for_Sustainable_Data_Centers

[5] European Commission Digital Strategy (2018) URL: https://ec.europa.eu/info/publications/EC-Digital-Strategy_en

[6] Montevecchi, F., Stickler, T., Hintemann, R., Hinterholzer, S. (2020). Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market. Final Study Report. Vienna URL: https://digital-strategy.ec.europa.eu/en/library/energy-efficient-cloud-computing-technologies-and-policies-eco-friendly-cloud-market

Authors

Author: **Carolina Koronen,** ECOS – Environmental Coalition on Standards

Edited by: **Kasia Koniecka,** ECOS – Environmental Coalition on Standards

ecos

Environmental Coalition on Standards

Mundo-b, the Brussels Sustainable House
Rue d'Edimbourg, 26
1050 Brussels, Belgium
+32 2 894 46 68
ecostandard.org

@ECOS_Standard ECOS-EU

In cooperation with:

coolproducts for a cool planet

coolproducts.eu

@CoolproductsEU