Building and Environment 82 (2014) 151–159

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Assessing the environmental impact of data centres part 1: Background, energy use and metrics

Beth Whitehead [a,*,1], Deborah Andrews [a], Amip Shah [b], Graeme Maidment [a]

[a] Faculty of Engineering Science and The Built Environment, London South Bank University, 103 Borough Road, London, SE1 0AA, UK
[b] HP Labs, 1501 Page Mill Rd, Palo Alto, CA, 94304, USA

ARTICLE INFO

Article history:
Received 15 May 2014
Received in revised form
13 August 2014
Accepted 14 August 2014
Available online 23 August 2014

Keywords:
Data centres
Environmental impact
Carbon footprint
Life cycle assessment
Data centre metrics
Energy consumption

ABSTRACT

Data centres consume high levels of energy to power the IT equipment contained within them, and extract the heat they produce. Because of the industry's heavy reliance on power, data centre metrics have historically used operational efficiency as a proxy for sustainability. More recently the industry has begun to recognise that its focus needs to go beyond energy consumption, with the creation of metrics for issues such as carbon, water and compute efficiency. However, single-issue metrics often consider only the operational phase, omitting impacts from other issues, during other stages in a facility's lifetime. Further approaches exist to assess more holistically the impact of data centres, such as building environmental assessment methods, but none have the capacity to capture fully the interlinked nature of a system, where improvements in one area and to one impact, can adversely affect a totally different area and totally different impacts.

The following review of literature summarises the approach of the data centre industry to environmental impact, and provides direction for future research. Part 1 describes the energy consumption of the ICT industry and in particular data centres; current knowledge on the environmental impact of the industry; and how single-issue metrics have risen to prominence.

1. Introduction

Data centres contain IT equipment used for the processing and storage of data, and communications networking [1]. They are the backbone of IT networks across the globe [2,3] and include extensive supporting infrastructures required to power and cool the IT equipment. A data centre can be as simple as a single rack in a server closet or as complex as a large warehouse, typically having built-in redundancy for the avoidance of downtime.

Data centres are high energy consumers. In 2007 the ICT industry was estimated to account for 10% of total UK electricity consumption [4], and 2% of global anthropogenic CO_2 [5], approximately equal to the direct emissions of the aviation industry operation. The operation of data centres already accounts for around a quarter of these emissions [4], and is believed to have the fastest growing carbon footprint from across the whole ICT sector [5].

This energy consumption has drawn the attention of data centre owners and operators. Firstly because of the cost of energy bills, and more recently because of it's impact on the environment. However, exclusive consideration of energy consumption has meant that other impacts and stages in a data centres life cycle are not well understood.

This two-part literature review seeks to present the current energy consumption and environmental impact of the data centre industry, and how it is monitored, assessed and benchmarked, and concludes the need for a more holistic approach to the management of environmental impact in the future. It does not seek to establish ways in which to reduce the impact. The review aims to focus the industry on why it has approached environmental issues in the current manner, highlight the need for a change in approach, and suggest further research and work required to enable this. Part 1 describes the energy consumption of the ICT industry and in particular data centres; current knowledge on the environmental impact of the industry; how the industry benefits the environment; and how single-issue metrics

* Corresponding author. Tel.: +44 (0) 7739 019960.
E-mail addresses: bethfwhitehead@gmail.com, BethWhitehead@dc-oi.com (B. Whitehead).
[1] Permanent address: Operational Intelligence, 74 Kelvedon Close, Kingston upon Thames, Surrey, KT2 5LF, UK.

http://dx.doi.org/10.1016/j.buildenv.2014.08.021
0360-1323/

have risen to prominence. Part 2 builds on this foundation to describe the use of building environmental assessment methods and tools; and based on both parts of the review, concludes the need to apply life cycle thinking to assess the environmental impact of data centres.

2. Data centres

Data centres house servers, and networking and storage equipment, and are considered the central nervous system of the 21st century. They contain comprehensive mechanical and electrical infrastructures to support the energy intensive computing required to perform one or more of the following functions [6,7]:

- The physical housing of IT equipment such as computers, servers, switches, routers, data storage devices, racks, and related equipment.
- The storage, management, processing and exchange of digital data.
- The provision of application services or management for data processing, such as web hosting, internet, intranet and telecommunication.

Data centres vary in size from a single rack in a server closet to huge server farms with floor areas reaching 150,000 m^2. Some occupy floors within offices and others are steel sheds on dedicated sites like that shown in Fig. 1. Large facilities contain data halls as shown in Fig. 2, which contain racks of IT equipment; the remaining floor space houses power and cooling equipment. Typically, the extensive floor space required for the supporting infrastructures can be as much as two [9] to four (when there are no external services) [10] times greater than the data halls themselves. Tight controls on air quality mean that the data halls do not include windows, and in the UK are often built using a steel frame and concrete floor construction resulting in large, windowless boxes. They are high energy consumers, both for power and the extraction of the heat dissipated from the IT equipment, and although some have huge floor plates, they are incredibly low occupancy facilities.

Data centres are used by businesses, corporations, educational establishments and governments, to provide web hosting and the internet, the storage of company information, and the processing of business transactions. They can be public (accessible to all, such as those for Google searches) or private (for the storage of company information on network drives) and, based on the importance of continued access to the data, display varying levels of [12]:

Fig. 1. Facebook data centre [8].

Fig. 2. Inside a data hall [11].

- Reliability – probability that a component/system/data centre operates without failure over a set time period. Facilities can have the same availability, but a facility that has one outage per year is more reliable than a facility that has many failures lasting the same amount of time.
- Availability – the average time per time period (for example a year) that a component/system/data centre operates as designed, without downtime. For example 0.99999 availability is a facility that has a total yearly downtime of 315 s.
- Redundancy – the topology of supporting infrastructures that ensure a component/system/data centre remains available in the event of a failure.

Facilities are described using Tier classifications [13] – Tier I to Tier IV – which refer to the topology of the facility's supporting infrastructures (power and cooling), and reflect how the building performs under planned and unplanned outages. The ability of a data centre to continue to perform its function in the event of a problem is determined by the amount of redundancy (spare plant) built into the design. For example, two mains power feeds to a site would ensure continued operation if one feed is lost, because operations can be switched to the other. The amount of redundancy incorporated into a data centre is dependent on whether or not the business linked to the facility can continue relatively unharmed in the event of a fault.

The Tiers were established by the Uptime Institute (an industry research body), to provide a common language across which the availability and reliability (redundancy) of different facilities can be compared, and are described in Table 1.

3. ICT and the internet

Worldwide, the number of data centres is growing, in part, due to the increase in access to PCs (personal computers) and the internet. The global PC installed base (including laptops) is well documented, and has grown rapidly from 242 million in 1995 [15,16] to 592 million in 2002 [5] and 1 billion in 2009 [17]. Furthermore, projections up to 2004 [15,16] 2014 [17] and 2020 [5], and shown in Fig. 3, fit with a pattern of exponential growth suggested by these early figures; most of which will require access to the internet and networks supported by data centres [17].

Furthermore, at the end of 2012, 34.3% of the global population were internet users [18], a penetration that grew from less than 1% in 1995 [19], as shown in Fig. 4, and equated to a rise from 0.04% to 6% in less developed countries [15]. Between 2000 and 2012,

Table 1
Summary of Tier classifications [14].

Tier		Site level infrastructure topology
Tier I	Basic capacity	Site-wide shutdowns are required for maintenance or repair work. Capacity or distribution failures will impact the site.
Tier II	Redundant capacity components	Site-wide shutdowns for maintenance are still required. Capacity failures may impact the site. Distribution failures will impact the site.
Tier III	Concurrently maintainable	*Each and every* capacity component and distribution path in a site can be removed on a planned basis for maintenance or replacement without impacting operation. The site is still exposed to an equipment failure or operator error.
Tier IV	Fault tolerant	An individual equipment failure or distribution path interruption will not impact operations. A fault tolerant site is also concurrently maintainable.

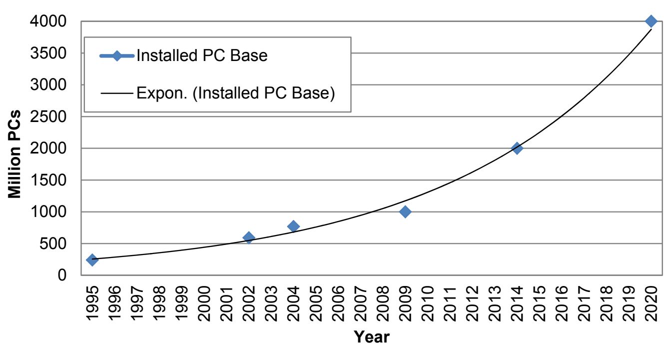

Fig. 3. Global PC installed base.

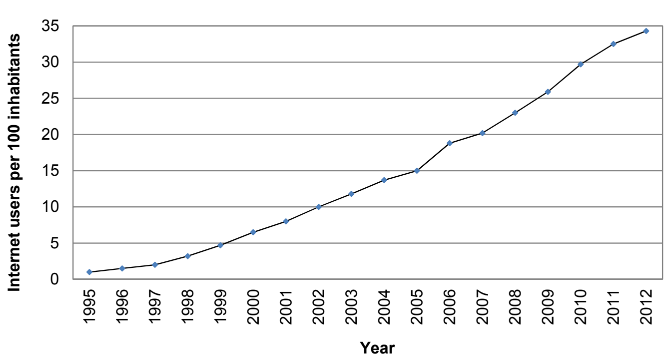

Fig. 4. Worldwide internet users [18,19].

because of rapid growth in parts of the developing world such as Africa, the Middle East and Latin America, access to the internet increased by 566% [18] (Fig. 5).

As global access to the internet increases and more businesses move their operations online, more ICT equipment is used and as a result more data centres are required [9]. In addition, as the internet has expanded, it is used in a more demanding manner. More powerful applications and faster speeds, and the online hosting of media and social networking sites [20] have therefore led to increases in the overall demand of data centres on power.

4. Operational energy use of ICT and data centres

Power demands in the UK have meant that the infrastructure is near breaking point, with demand set to outstrip supply by 2017 if a 'do nothing' approach is adopted for its generation [4]. Furthermore, in 2004 the UK 'became a net importer of energy' [4], making it politically vulnerable [3], and by 2016 the safety margin in the National Grid is expected to reach 4%, equivalent to a blackout risk of 1 in every 12 years [21]. The efficient use of energy by ICT, and in particular data centres, is therefore a key concern of the ICT industry.

4.1. Office and network equipment

The earliest research into the energy consumed by the ICT industry focused on IT components such as computers and peripherals (printers). Concern in the US was led by the Lawrence Berkeley National Laboratory (LBNL) as early as the 1980s, where research showed rising trends in energy use by personal computers and office equipment [22–24]. In 1983 LBNL estimated total average energy use by office equipment of 0.1 kWh/m^2/year, and projected levels would reach 0.4 kWh/m^2/year by 2011 [23].

In 1999, updated US research found office and network equipment consumed 74 TWh/year – 2% of total electricity or 3% when telecommunications equipment and electronics manufacturing were included [25]. By the start of the millennium, further research with improved data readings estimated higher levels of overall electricity usage of 97 TWh/year – 3% of total electricity [26] – showing an overall upward trend since research started.

4.2. The internet

Since the start of the internet, users have continued to increase as discussed in section 2.3. Alongside this growth in internet use there has been much speculation around the amount of electricity required to run it. At the end of the 1990s, Forbes published an article summarising a non-peer-reviewed report which assumed

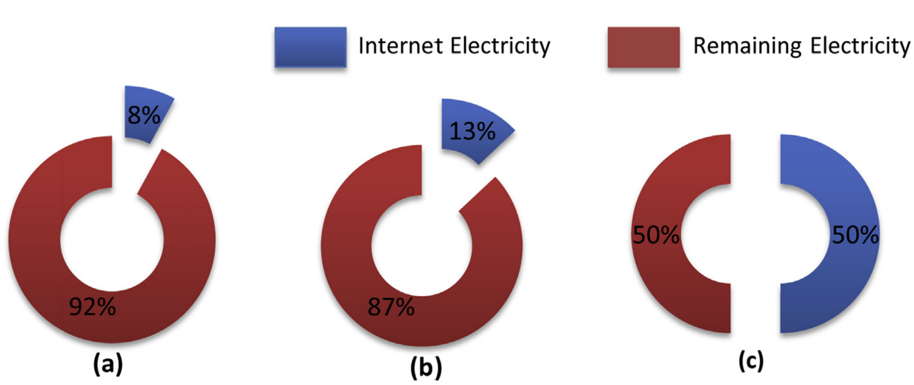

Fig. 5. Proportion of US electricity used for the internet (a), including embodied and operational impacts (b), and projected over next one to two decades (c).

the largest power draws of each equipment type (peak power without consideration of actual utilisation) and applied them universally [26]. The report claimed that 8% of all US electricity was used to power the internet, a figure which grew to 13% if embodied and operational energy for chips and computers was included, and 50% when projected over the ensuing one to two decades [27] as shown in Fig. 5.

These figures, however, have been widely rebuffed by research showing the article assumptions were overestimated almost eightfold [28–30]. This was later independently confirmed by the work of RAND [31] for the US DoE, in which four scenarios were built for the electricity requirements of a digital society from 2001–2021, and in which even the largest projected growth in digital devices was found to result in only a modest effect on electricity demands of 5.5% [31].

Nonetheless, the statistic persists and continues to be quoted today by journalists, and although it has been a catalyst for improved research, it shows the need to handle energy data with care.

4.3. Data centres

As access to IT and the internet has grown, so too has the number of data centres, and their consumption of electricity. There is little early evidence of peer-reviewed estimates on energy use of data centres [32,33], especially those that include infrastructure energy [34], but after incorrect estimates [28] made by Huber [27] were widely spread and then refuted, industry-wide research began to grow. Importantly in 2001, studies based on actual usage data, as opposed to assumptions, began to emerge [35,36] as shown in Fig. 6.

Early industry figures, based on overestimated facility assumptions and incorrect footprint areas, suggested data centre power demand (including infrastructure draw) could be as high as 2150–2690 W/m^2 [6] and frequently well over 1000 W/m^2 [35]. However, total computer room power densities based on actual usage data were approximated in 2001 at a much lower 355 W/m^2 [35], and emphasised the errors in assumptions made in previous studies. The data centre studied by Mitchell–Jackson et al. [35] was later investigated in more detail. From 2001 to 2002, the study of one internet data centre (IDC) found an increase in computer room floor area of 33%, yet due to energy efficiency measures, the same power density of 355 W/m^2 was displayed [36]. This shows that as early as 2000, efforts were being made to improve operational efficiency.

Later, the benchmarking of six data centres resulted in average densities in 2004 in the order of 538 W/m^2 [6], with the benchmarking of a further 22 data centres in 2006 yielding densities ranging from 54 to 1000 W/m^2 [37,38]. The figures, while sufficiently different, reflect the widely varying levels of efficiency that are displayed in facilities, and the upward trend in densities and the changes to technology over time.

Research into energy consumption of data centres was initially focused on power densities as described above. However, more recently this focus has shifted to total electricity consumption in areas across the globe, and is dominated by the work of Jonathon Koomey. In the period from 2000 to 2005, aggregate worldwide electricity use by servers doubled, largely due to increased numbers of cheap volume servers, and in part due to a small increase in power use per unit [33]. In 2000, annual data centre electricity consumption in western Europe (shown in Fig. 7) was 18.3 TWh, a figure that rose to 41.3 TWh in 2005, and assuming a 12% year-on-year growth was projected to reach 72.5 TWh by 2010 [34].

In 2007 the US EPA Energy Star programme compiled a Report to Congress which concluded that in 2006 US data centres and servers consumed 61 TWh of electricity – 1.5% of the country's overall demand and double that in 2000 [1] and equivalent to 33,672 kgCO_2e (calculated using conversion factors from the IEA 2013 edition of fuel combustion emissions [[39], page 110]).

Later, in 2011, Koomey [40] published a report for The New York Times which revisited the previous global and US projections from his 2008 paper [34]. The study found a slow-down in the growth of electricity because of efficiency improvements, the recession, and virtualisation, resulting in only a 56% growth in worldwide electricity consumption between 2005 and 2010, rather than a doubling as it did from 2000 to 2005 [40], as was projected by the Report to Congress [1]. As a result, global annual consumption grew from 70.8 TWh (37,382 kgCO_2e) in 2000 to 152.5 TWh (82,650 kgCO_2e) in 2005 and 238 TWh (125,900 kgCO_2e) in 2010 [39,40] as shown in Fig. 8.

Furthermore, if the 2010 consumption from the 2011 Koomey study [40] is extrapolated to 2015 (assuming the same 56% growth seen between 2005 and 2010, and split evenly between the five year period) a global consumption of 291 TWh is found in 2012 [34,40] and 371.1 TWh (197,500 kgCO_2e) by 2015 as shown in Fig. 8.

In addition to the research by Koomey [34,40], DatacenterDynamics (DCD) initiated a yearly census of the industry. In 2012, the DCD [41] Industry Census found a total global data centre power use of 322 TWh, 1.8% of global electricity use, and comparable to that suggested for 2011 by Koomey [40] of between 1.7 and 2.2%. Using a world average kgCO_2e per kWh of electricity generation this is equivalent to 171,630 kgCO_2e [39].

Moreover, the census found that the UK currently has the third highest global actual power demand of 2.70 GW in 2011 and 2.85 GW in 2012 [42] and 3.10 GW in 2013 [43], and which is forecast to reach 3.68 GW by 2016 [43] and shown in Fig. 9.

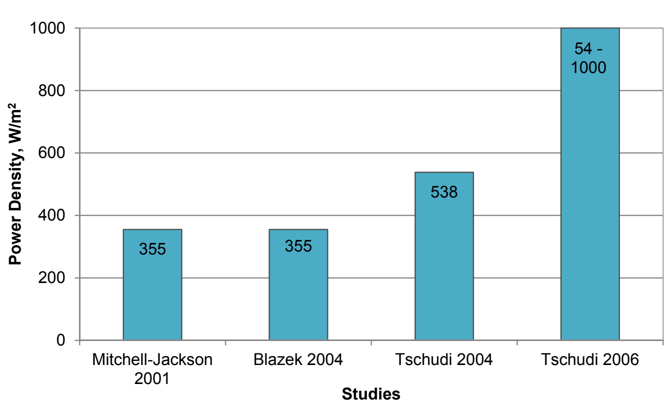

Fig. 6. Studies of data centre power densities.

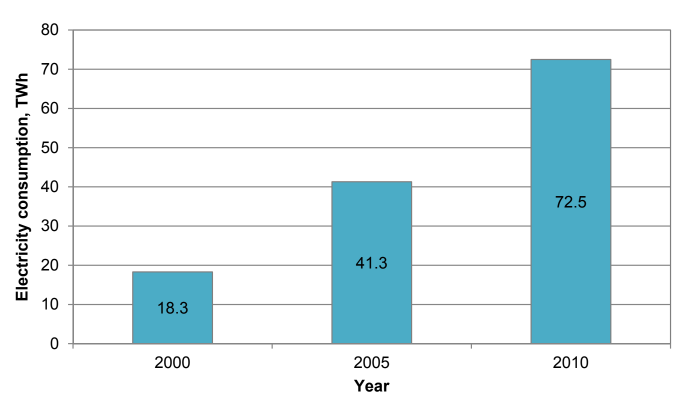

Fig. 7. Data centre electricity consumption in western Europe.

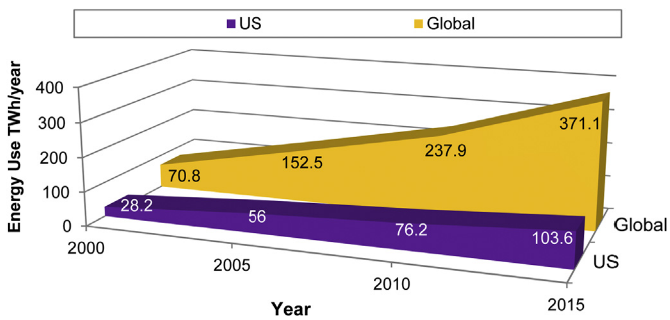

Fig. 8. Data centre energy use extrapolated to 2015 [40].

Assuming that the data centres are running 24 × 7, using the Carbon Trust conversion factors [44] for the UK electricity grid in 2013, this would have produced 10,536 $kgCO_2e$ in 2011, 11,122 $kgCO_2e$ in 2012, 12,097 $kgCO_2e$ in 2013 and will create 14,361 $kgCO_2e$ in 2016.

Furthermore, in 2011 the UK was estimated to have 7.59 million m^2 of dedicated data centre space – equivalent to 14 Pentagons – and a total maximum power consumption of 6.4 GW [45] or 24,975 $kgCO_2e$.

Throughout the energy literature reviewed in this section, early values for consumption across the industry were consistently varied as shown in Fig. 6. The differences reflect the difficulty in obtaining accurate data on energy use because of the lack of monitoring, the constant change in technologies, increasing densities, and changing approaches to calculating the consumption, and the reluctance of owners to share energy data [35,36].

Later results from Koomey [40] and DatacenterDynamics [43] are in the same order of magnitude and seem comparable. However, although there is a good correlation between the studies, the Koomey work recognises that further research is needed, based on actual energy use, and that while the DCD Census [41] is based on a sample of the industry, it is extrapolated to build the overall picture. It is clear therefore that both sets of results are exposed to uncertainties that need attention to improve their accuracy.

5. The impact of ICT and data centres on climate change

The growing power demand of data centres has led to a heightened awareness of their increasing impact on climate change from greenhouse gas (GHG) emissions. Comprehensive research into the impact of the whole ICT industry on climate change has been dominated by GeSI [5,46], but has been strengthened by recent research by Malmodin et al. [47].

In 2012 the growth in GHG emissions (embodied and operational) from the ICT industry was projected to rise at a faster rate than the total global footprint [46]. Note *footprint* and *emissions* are used interchangeably in this section, and refer to the GHG emissions including CO_2 and all GHGs converted to CO_2-equivalent (CO_2e). Of the three main sectors of the ICT industry, data centres are projected to have the fastest rate of growth at 7% per annum (p.a.) from 2011 to 0.29 $GtCO_2e$ in 2020 [46]. This growth echoes those found in an earlier study by GeSI [5] that suggested a rise in footprint of 7% p.a. from 2002 to 2020, but concluded a higher overall impact in 2020 due to an actual measured increase (rather than projected) of 9% p.a. from 2002 to 2011, and shown in Fig. 10.

In the 2008 study, GeSI estimated that ICT accounted for 2% of the global GHG footprint – a figure supported by research from Gartner [48] – and would grow to 2.8% by 2020. By 2011, however, research based on more accurate data and altered behaviours in the industry, meant only a 1.9% share of the total footprint was found [46].

A similar study by Malmodin et al. [47] compared its results to the first (Smart 2020) report [5], and found different percentage shares of the total global footprint, created by the use of old data in the 2008 GeSI [5] report, and therefore an elevated ICT carbon footprint. In the SMARTer 2020 report [46] these figures were updated, but the percentage share of the footprint by 2020 remained higher (2.3%) than in the Malmodin et al. [47] study (1.9%), suggesting these modelling differences remain.

Of the ICT impact, servers and cooling were found by Gartner [48] to account for 23% of carbon dioxide emissions. However, the GeSI report [5], which was more detailed and based on more accurate information, found a figure (including power systems and embodied impacts) closer to 14% shown in Fig. 11 below. Of this 14%, out-dated volume servers accounted for over a third of the impact, a figure reflected in the Koomey [34,40] studies. Although the difference in figures is big, it is clear that this is due to the omission of embodied impacts from PCs and monitors in the Gartner data [48].

6. ICT as a key enabler

Whilst the operation of ICT impacts negatively on climate change, ICT also has the potential to impact beneficially by enabling the reduction of emissions in other sectors. The potential of the internet (which cannot exist without data centres) to improve efficiency and reduce carbon emissions across sectors other than ICT was recognised as early as the 1990s when Romm et al. [49] noted the potential for the internet to turn retail buildings into websites and trucks into fibre optics.

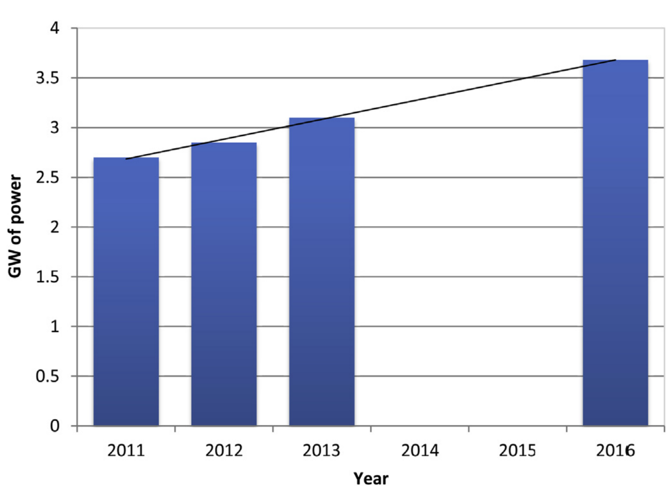

Fig. 9. UK actual power demand by data centres from 2011 to 2016.

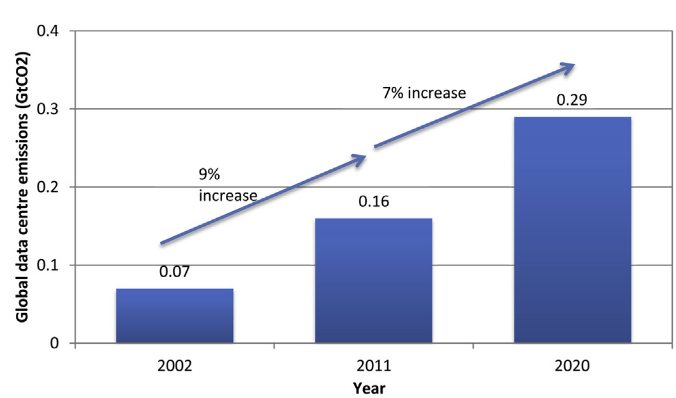

Fig. 10. Growth in data centre GHG emissions – 2002 to 2020.

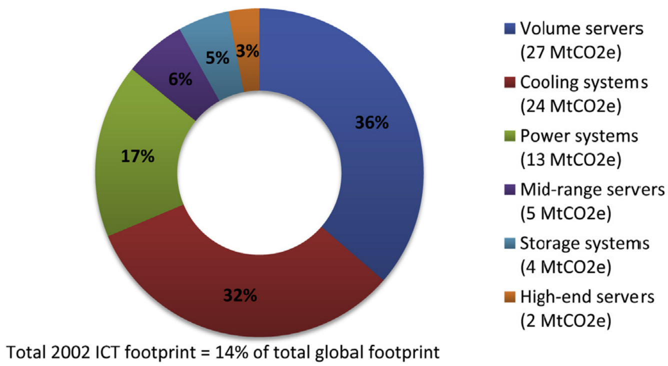

Fig. 11. Composition of global data centre footprint in 2002 [5].

Between 1997 and 1998, total US energy consumption increased by only 1%, yet the economy grew 4% year-on-year, and in 1998 GHG emissions rose by only 0.2%, the lowest since the recession year of 1991 [49]. These trends were largely attributed to increased low-energy production of computers and software (one third) and the knock-on gains in efficiency (two thirds) from their use [49].

This research by Romm et al. [49] is further backed by US EIA data [[50], Fig. 1.7], which illustrates the continued decrease in primary energy consumption per real dollar of GDP (gross domestic product) between 1973 and 2012. Alongside the growth in use of ICT discussed previously, the trend suggests that a growth in ICT occurs with a lowered increase in GHG emissions, which if continued could lead to an overall reduction in emissions.

6.1. Sources of savings from the use of ICT in other industries

The potential savings achieved by using ICT come from a number of areas: dematerialisation (the swapping of high carbon products for low carbon alternatives), energy, transport, buildings and industry [49,51], and can be achieved through smart grids, smart logistics, smart buildings and smart motor systems [5].

The reduced rate in increase of emissions discussed previously is largely influenced by the indirect effects of ICT, rather than dematerialisation. In 2008 GeSI [5] estimated that ICT could save up to 7.8 $GtCO_2e$ by 2020, a figure that is 5.5 times the projected impact of the ICT industry itself (1.4 $GtCO_2e$). However, in 2012 GeSI [46] increased this abatement potential to 9.1 $GtCO_2e$ by 2020, 7.2 times the projected industry impact of 1.3 $GtCO_2e$, which was reduced to take into account new technologies, improved data, and increased baseline emissions.

Building on the work of Romm et al. [49] and Turner et al. [51], GeSI [46] presented four main ways ('change levers') by which ICT can reduce emissions within other sectors: digitalisation and dematerialisation, improved data collection and communication, system integration, and optimisation of processes. Through the adoption of these methods, an abatement potential of 9.1 $GtCO_2e$ by 2020 is projected to come from improved efficiency within six sectors [46]:

- transport – ICT can improve logistics and create new technologies;
- agriculture and land-use – ICT can be used to control irrigation systems remotely;
- buildings – software can be used to more accurately design buildings in a way that reduces air conditioning and heating;
- manufacturing – software can be used to automate and optimise processes;
- power – software can be used to respond to dynamic changes in demand; and
- service and consumer – for example the move to online retail.

Although there are barriers to the realisation of the reduction potentials discussed in this section – such as poor economics and finance, lack of awareness and resistance to behavioural change – the potential of ICT to indirectly reduce climate change far outweighs the direct impact of data centres and the ICT industry. It is therefore clear that data centres will continue to grow in number, and reduction of their environmental impact will become increasingly important.

7. Data centre metrics

With the growing energy consumption discussed in section 4, efforts to improve the operational efficiency of data centres and the components they contain have been widespread. As a result, a number of common metrics and methods of assessment have been adopted to monitor and benchmark their performance.

Most metrics focus on the efficient use of individual resources during the operation of a data centre, and by this virtue can also help to reduce operational expenditure. Table 2 gives a summary of

Table 2
Commonly adopted metrics.

Metric		Equation
PUE	Power usage effectiveness [52–56,62]	$= \frac{\sum \text{Facility power}}{\sum \text{IT equipment power}} = \frac{P_{\text{mechanical}}+P_{\text{electrical}}+P_{\text{miscellaneous}}+P_{\text{IT}}}{P_{\text{IT}}}$
ERF	Energy reuse factor [62,64]	$= \frac{\text{Reuse energy outside of the data centre}}{\text{Total data centre source energy}}$
GEC	Green energy coefficient [62]	$= \frac{\text{Green energy used by the data centre}}{\text{Total data centre source energy}}$
SI-EER	Site infrastructure energy efficiency ratio [32]	Same as PUE
DCiE	Data centre infrastructure efficiency [53,54]	$= \frac{1}{\text{PUE}}$
DCeP	Data centre energy productivity [62,65]	$= \frac{\text{Useful work produced in the data centre}}{\text{Total data centre energy consumed producing this work}}$
ScE	Server compute efficiency	$= \frac{\text{No. of samples where server provides a primary service}}{\text{Total no. of samples over the time period}} \times 100$
A primary service is the main service provided by the server, for example the primary service of a mail server is to provide email [60]		
DCcE	Data centre compute efficiency [60]	$= \frac{\sum \text{ScE from all servers}}{\text{Total number of servers}}$
DPPE	Data centre performance per energy [59]	$= \frac{\text{IT equipment utilisation factor} \times \sum \text{IT equipment capacity}}{\sum \text{Data centre energy consumption} - \text{Green energy}}$
DC-EEP	Data centre energy efficiency and productivity [32]	$= \text{SI} - \text{EER} \times \text{IT productivity per embedded Watt}$
CUE	Carbon usage effectiveness [58,62,66]	$= \frac{CO_2 \text{ emitted } (kgCO_2e)}{\text{Unit of energy (kWh)}} \times \frac{\text{Total data centre energy}}{\text{IT equipment energy}}$
WUE	Water usage effectiveness (site) [66,67]	$= \frac{\text{Annual site water usage}}{\text{IT equipment energy}}$
$\text{WUE}_{\text{source}}$	Water usage effectiveness (source) [67]	$= \text{WUE} + \frac{\text{Annual source energy water usage}}{\text{IT equipment energy}}$
EDE	Electronics disposal efficiency [63]	$= \frac{\text{Weight of responsibly disposed of IT EEE}}{\text{Total weight of disposed of IT EEE}}$

some of the most common metrics used by the data centre industry.

Developed by The Green Grid [52–56], PUE (power usage effectiveness) is the most widely adopted metric, with reports from research analysts at Gartner that 80% of all new large data centres will have adopted the metric by 2015 [57]. The metric is used to measure the ratio of total power delivered to site to that used by the IT equipment, and is analogous to the miles per gallon metric for the fuel consumption of a car. It is dimensionless, and has an ideal value of 1.0 [58]; although Shiino [59] (2010) found actual values ranged from 1.25–3.75. Whilst Koomey [40] found values in the range of 1.36–3.6, from the EPA Energy Star programme's study of 61 data centres, and an average value of 1.92. In real terms this mean value of PUE shows that the average data centre consumes almost double the power required for just the IT equipment in power losses, cooling, lighting and other miscellaneous loads.

The PUE metric drives the need to minimise power used by anything other than IT. However, there are concerns that the metric does not consider the actual productivity or efficiency of the equipment [54,59]. As a result, a data centre in which no infrastructure upgrades are made actually achieves an improved PUE as the IT equipment ages and uses more power.

Recognising the need to consider more than just energy use, CUE and WUE consider operational carbon and water usage, following the same format as PUE. ScE and DCcE [60] consider the efficiency of the data centre compute infrastructure, and allow users to focus on operational efficiency, much like PUE. While DCeP is a productivity metric, which attempts to quantify the useful work performed by a data centre through a number of complex proxies [61,62] and is more advanced than the xUE family (PUE, CUE, WUE). EDE seeks to address the need for a metric to quantify the extent to which IT consumers are disposing of IT equipment responsibly at the end of their life [63]. Finally ERF quantifies the amount of energy reused outside of the data centre and GEC looks at the amount of renewable energy used. The remaining metrics are a variation on these themes.

7.1. Metric evaluation

There are a number of concerns with current data centre metrics. Most importantly, they are generally only concerned with the operational phase of data centres, and do not coherently account for other impacts that occur, for example, when the components are being manufactured (embodied impacts).

Although originally no consideration was made to the renewables content of the source electricity, the introduction of GEC (green energy coefficient) [62] allows for comparison of two sites with equivalent PUE values, one of which uses energy generated from renewable sources and the other of which relies on energy from coal-fired power stations. However, The Green Grid [62] currently only recognises three authorities across the world that issue green energy certificates that satisfy their requirements as proof of renewables content, in the EU, Japan and USA. It is also unclear whether purchase of these certificates is prohibitively expensive to facilities in countries where a high renewables content is almost a given. Nonetheless, the metric remedies the inability for global comparisons.

When The Green Grid metric WUE_{source} was released it introduced an important area of expansion that is omitted from most of the other industry metrics. Firstly, it considers water – not energy – and secondly, it considers the water used not only during the operation of the data centre (for example for humidification), but also during the production of the power that is used on site. Much like GEC now does for power usage. The metric, alongside CUE, recognises that impacts due to the existence of a data centre occur not only when it is in operation, but also in the production process, and acknowledges that environmental burden comes from more than just energy use.

This inclusion of impacts from the production phase is in line with life cycle thinking, in which the impact that a product or service has on the planet is assessed from the moment raw materials are extracted until eventual disposal of the product, while crucially considering more environmental impacts than simply energy and water use.

Without holistically considering the life cycle of a data centre for various environmental impacts and stages of the life cycle, it is hard to know how energy efficiency measures in the operational phase impact on other parts of the life cycle. For example, it is difficult to know whether the current drive to raise server inlet temperatures, and reduce or eliminate the need for mechanical cooling, could adversely affect the embodied impacts (impacts experienced pre- and post-operation) of servers due to a potential decrease in the time between technology refresh. In other words ‘pollution shift’ is difficult to gauge without relevant research.

Furthermore, the industry does not have a way to manage and assess when equipment should be replaced, based on an evaluation of the reduced efficiency and additional environmental impact from the replacement component against the energy savings that will be made as a result of the replacement [63]. The introduction of the EDE metric, which is concerned with the disposal of IT equipment, is incredibly important in this argument. Not only does the metric follow life cycle thinking, alongside the release of The Green Grid's guidelines to the application of life cycle assessment (LCA) to data centres [68], the introduction of EDE and similar metrics indicates that the industry is beginning to see the need to consider more than single-issue metrics with a life cycle approach, and recognises the need for benchmarking and tools to facilitate the change.

8. Conclusions

The review of literature in part 1 presents a clear picture of the rising access to ICT and it's increased energy consumption. However, there is still uncertainty in the true values of this consumption, and how it will continue to grow into the future. Historic information on installed equipment has changed, as more accurate information becomes available, meaning projections are dynamic and need revisiting on a yearly basis. Work is required to ensure this information is more accurate in the first instance. In addition, data centre asset management needs a better handle of the amount of servers that are installed, and consuming energy to idle, but are unused for compute. It is noted by Koomey [40] that the omission idle servers from the estimates could mean the industry is actually consuming more than currently known. Irrespective of model uncertainties, energy consumption by data centres continues to grow, and although ICT is an enabler for energy reduction in other industry, its consumption needs to be monitored.

In response to rising energy consumption, and with growing concerns about energy security and availability [3], environmental impact has become increasingly important for the data centre industry. One method adopted to monitor and benchmark this impact is data centre metrics.

Metrics, such as power usage effectiveness (PUE), focus on operational efficiency, using it as a proxy for sustainability. These metrics have gained in popularity because of the documented rise in operational energy consumption, but miss impacts that are embodied in the facility due to energy consumed and emissions created during the manufacturing and disposal of data centres and their components [69]. By only considering one issue – for instance energy, water or carbon – in one stage of the facility's lifetime, it is not possible to detect the effect that improving the efficiency of one issue has on another at any other point of the building's lifetime [69].

Although the industry's primary concern is the financial implications of this rising energy consumption, the industry is becoming increasingly aware of its environmental impact and its vulnerability from the uncertain future of its fuel supply. In response to the resulting rise in energy bills, and power infrastructures pushed to their limits, the industry has focused almost exclusively on energy efficiency as a proxy for sustainability. The metrics have instigated a change in behaviour for the industry to one with more concern for sustainability, however, 'pollution shift' cannot be accurately evaluated by them and tends to be considered by intuition.

There is currently little evidence of detailed research that considers the impact using a life cycle perspective. In order to ensure impacts are not going unnoticed because of the operational focus, it is imperative that more research is conducted into the interrelated nature of environmental impact. This research should look to provide greater information on the most environmentally impacting parts of the facility beyond operational consumption; and seek to determine whether a life cycle perspective is required. The work of GeSI [46] and Malmodin et al. [47] already points to the contribution ICT has to the global footprint of GHG emissions, and should be strengthened by more detailed research.

As single metrics are so widely adopted by the industry, future work from life cycle benchmarking should look to establish similar simple metrics for a wide range of impacts, to enable the broadest section of the industry to use and report against them. These could take a similar form to impact factors for electricity from different energy mixes, and global warming potential of different refrigerants, to allow designers and owners to understand their impact more holistically.

Finally it should be reiterated that this paper seeks to provide a picture of the current impact of the industry and methods used to monitor it. It does not include consideration of how the impact can be reduced.

Other options for assessing the impact of data centres are discussed in part 2 of this paper. This concluding part describes and critiques the use of building environmental assessment methods and tools; and based on both parts of the review, concludes the need to apply life cycle thinking to more holistically assess the environmental impact of data centres.

Acknowledgements

This paper has been developed with the financial support of the Engineering and Physical Sciences Research Council (EPSRC), EP/H50169X/1 and HP; who have also gone to great lengths to share their knowledge and expertise in the fields of data centres and sustainable IT. Additional support and review of work was provided by supervisors: Dr Robert Tozer, Dr Alan Dunn and Sophia Flucker.

References

[1] US EPA. Report to congress on server and data center energy efficiency. Public law 109–431; 2007. Available from: http://hightech.lbl.gov/documents/DATA_CENTERS/epa-datacenters.pdf [accessed 04.10.10].

[2] Covas MT, Silva CA, Dias LC. Multicriteria decision analysis for sustainable data centers location. Int Trans Oper Res 2012;20(3):269–99.

[3] Uddin M, Rahman AA. Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics. Renew Sustain Energy Rev 2012;16:4078–94.

[4] Global Action Plan. An inefficient truth; 2007. Available from: http://greenict.org.uk/sites/default/files/An%20Inefficient%20Truth%20-%20Executive%20Summary.pdf [accessed 03.08.10].

[5] GeSI. SMART 2020: enabling the low carbon economy in the information age; 2008. Available from: http://www.smart2020.org/_assets/files/02_Smart2020Report.pdf [accessed 01.07.10].

[6] Tschudi W, Xu T, Sartor D, Nordman B, Koomey J, Sezgen O. Energy efficient data centers. Final report to California Energy Commission; 2004. Available from:, http://escholarship.org/uc/item/1s93b4rd [accessed 16.01.11].

[7] LBNL. Definition of data center; 2010. Available from: http://hightech.lbl.gov/DCTraining/definitions.html [accessed 17.09.10].

[8] Intel Free Press. Outside facebook data center. Available from:; 2012. Creative commons license: https://www.flickr.com/photos/intelfreepress/6722296855. https://creativecommons.org/licenses/by/2.0/legalcode [accessed 09.07.14].

[9] Shah A, Krishnan N. Optimization of global data center thermal management workload for minimal environmental and economic burden. IEEE Trans Comp Packag Technol 2008;31(1):39–45.

[10] ISG. Data center case study. HP Enterprises; 2014. Available from: http://www.isgplc.com/datacenters/casestudy/default.asp#thebuilding [accessed 11.07.14].

[11] Dennis van Zuijlekom. End-of-life. Available from:; 2013. Creative Commons License: https://www.flickr.com/photos/dvanzuijlekom/8522770418. https://creativecommons.org/licenses/by/2.0/legalcode [accessed 09.07.14].

[12] Arno R, Friedl A, Gross P, Schuerger R. Reliability of example data center designs selected by tier classification. In: Tallahassee, FL, USA: proceedings of industrial and commercial power systems technical conference (I&CPS); 2010.

[13] Uptime Institute. Data center site infrastructure. Tier standard: topology; 2012. Available from: http://uptimeinstitute.com/publications [accessed 24.07.13].

[14] Uptime Institute. Data center site infrastructure. Tier standard: operational sustainability; 2013. Available from: http://uptimeinstitute.com/publications [accessed 24.07.13].

[15] Chinn MD, Fairlie RW. ICT use in the developing world: an analysis of differences in computer and internet penetration 2010. Rev Int Econ 2010;18(1):153–67.

[16] PRB. How many people have ever lived on the earth?; 2011. Available from: http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx [accessed 09.06.13].

[17] Stamford C. Gartner says more than 1 billion PCs in use worldwide and headed to 2 billion units by 2014; 2008. Available from: http://www.gartner.com/newsroom/id/703807 [accessed 23.06.08].

[18] IWS – Internet World Stats. Internet world stats usage and population statistics; 2013. Available from: http://www.internetworldstats.com/stats.htm [accessed 09.06.13].

[19] ITU. Measuring the information society; 2012. Available from: http://www.itu.int/ITU-D/ict/publications/idi/ [accessed 19.10.10].

[20] Leiner BM, Cerf VG, Clark DD, Kahn RE, Kleinrock L, Lynch DC, et al. A brief history of the internet. ACM SIGCOMM Comput Commun Rev 2009;39(5):22–31.

[21] Buchanan A. Will GB's lights stay on and will the gas keep flowing: a look at the next decade?; 2013. Available from: https://www.ofgem.gov.uk/ofgem-publications/64123/lecture-19th-february-2013.pdf [accessed 08.08.13].

[22] Harris J, Roturier J, Norford LK, Rabl A. Technical assessment: electronic office equipment. LBL-25558-Rev; 1988. Available from:, http://www.ntis.gov/search/index.aspx?frm_qry_Search=DE89013375 [accessed 06.10.10].

[23] Piette MA, Eto JH, Harris JP. Office equipment energy use and trends. LBL-31308. UC-350; 1991. Available from:, http://dx.doi.org/10.2172/7001015 [accessed 06.10.10].

[24] Koomey JG, Cramer M, Piette MA, Eto JH. Efficiency improvements in US office equipment: expected policy impacts and uncertainties. LBL-37383 UC-1600; 1995. Available from:, http://enduse.lbl.gov/info/LBNL-37383.pdf [accessed 06.10.10].

[25] Kawamoto K, Koomey JG, Nordman B, Brown RE, Piette MA, Ting M, et al. Electricity used by office equipment and network equipment in the U.S.: detailed report and appendices; 2001. Available from: http://enduse.lbl.gov/projects/infotech.html [accessed 07.10.10].

[26] Roth KW, Goldstein F, Kleinman J. Energy consumption by office and telecommunications equipment in commercial buildings. Volume I: energy consumption baseline. In: Asilomar, Pacific Grove, CA, USA: proceedings of 2002 ACEEE summer study on energy efficiency in buildings; 2002.

[27] Huber P.W. Dig more coal – the PCs are coming (updated: 31.05.99). Available from: http://www.forbes.com/forbes/1999/0531/6311070a_2.html [accessed 04.10.10].

[28] Koomey JG, Kawamoto K, Nordman B, Piette MA, Brown RE. Memorandum LBNL – 44698; 1999. Available from: http://enduse.lbl.gov/SharedData/IT/Forbescritique991209.pdf [accessed 04.10.10].

[29] Koomey JG, Chong H, Loh W, Nordman B, Blazek M. Network electricity use associated with wireless personal digital assistants; 2004. Available from: http://hightech.lbl.gov/documents/DATA_CENTERS/koomeyfinalPDAarticle.pdf [Accessed 04.10.10].

[30] Koomey JG. Sorry, wrong number – how to separate fact from fiction in the information age. IEEE Spectr 2003;40(6):11–2.

[31] Baer WS, Hassell S, Vollaard BA. Electricity requirements for a digital society; 2002. Available from: http://www.rand.org/pubs/monograph_reports/MR1617/index.html [accessed 04.10.10].

[32] Brill KG. Data center energy efficiency and productivity. Site infrastructure institute white paper; 2007. Available from: http://www.uptimeinstitute.org/symp_pdf/(TUI3004Chttp://www.uptimeinstitute.org/symp_pdf/(TUI3004C)DataCenterEnergyEfficiency.pdf [accessed 05.11.10].

[33] Koomey JG. Estimating total power consumption by servers in the US and the world; 2007. Available from: http://sites.amd.com/us/Documents/svrpwrusecompletefinal.pdf [accessed 08.10.10].

[34] Koomey JG. Worldwide electricity used in data centers. Environ Res Lett 2008;3:8.
[35] Mitchell-Jackson J, Koomey JG, Nordman B, Blazek M. Data center power requirements: measurements from silicon valley. Energy 2003;28(8):837–50.
[36] Blazek M, Chong H, Loh W, Koomey JG. Data centers revisited: assessment of the energy impact of retrofits and technology trends in high-density computing facility; 2004. Available from: http://hightech.lbl.gov/documents/DATA_CENTERS/blazekfinalpdf.pdf [accessed 07.10.10].
[37] Greenberg S, Mills E, Tschudi W, Rumsey P. Best practices for data centers; lessons learned from benchmarking 22 data centers. In: Asilomar, Pacific Grove, CA, USA: proceedings of 2006 ACEEE summer study on energy efficiency in buildings; 2006.
[38] Tschudi W, Mills E, Greenberg S, Rumsey P. Measuring and managing datacenter energy use. Findings – and resulting best practices – from a study use in 22 data centers. HPAC Engineering; March 2006. p. 45–51.
[39] IEA. CO_2 emissions from fuel combustion highlights. 2013 ed. 2013 Available from: http://www.iea.org/publications/freepublications/publication/CO2EmissionsFromFuelCombustionHighlights2013.pdf [accessed 14.07.14].
[40] Koomey JG. Growth in data center electricity use 2005 to 2010; 2011. Available from: http://www.analyticspress.com/datacenters.html [accessed 01.02.12].
[41] DCD. DCD Industry census 2012: energy. DatacenterDynamics Focus 2012;3(26):38–41.
[42] DCD. DCD Industry census 2012: UK energy. DatacenterDynamics Focus 2012;3(26):46–50.
[43] DCD. DCD Industry census 2013: UK figures. DatacenterDynamics Focus 2013;3(32):82–3.
[44] Carbon Trust. Conversion factors. Energy and carbon conversions 2013 update; 2013. Available from: http://www.carbontrust.com/media/18223/ctl153_conversion_factors.pdf [accessed 14.07.14].
[45] DCD. Special market report: data centers in the United Kingdom; 2011. Available from: http://www.datacenterdynamics.com/research/uk-market-report [Accessed 19.03.12].
[46] GeSI. SMARTer 2020: the role of ICT in driving a sustainable future; 2012. Available from: http://www.gesi.org/SMARTer2020 [accessed 22.03.13].
[47] Malmodin J, Bergmark P, Lunden D. The future carbon footprint of the ICT and E&M sectors. In: Zurich, Switzerland: proceedings of the first international conference on information and communication technologies for sustainability; 2013.
[48] Gartner. Green IT: a new industry shockwave. In: Orlando, Florida, USA: gartner presentation in: gartner symposium/ITxpo; 2007.
[49] Romm J, Rosenfeld A, Herrmann S. The internet economy and global warming. A scenario of the impact of e-commerce on energy and the environment. Version 1; 1999. Available from: http://www.p2pays.org/ref%5C04%5C03784/0378401.pdf [accessed 04.10.10].
[50] US EIA. Monthly energy review – February 2014. 1. Energy overview. Fig 1.7; 2014. Available from:, http://www.eia.gov/totalenergy/data/monthly/index.cfm#summary [accessed 22.03.14].
[51] Turner V, Bigliani R, Ingle C. Reducing greenhouse gases through intense use of information and communication technology: part 1. IDC White Paper; 2009. Available from:, http://download.intel.com/pressroom/archive/reference/IDCWP31R.pdf [accessed 13.10.10].
[52] The Green Grid. Green grid metrics: describing datacenter power efficiency. White Paper; 2007. Available from:, http://www.thegreengrid.org/~/media/WhitePapers/Green_Grid_Metrics_WP.ashx?lang=en [accessed 24.05.10].
[53] The Green Grid. Green grid data center power efficiency metrics: PUE and DCIE. White Paper #6; 2008. Available from:, http://www.thegreengrid.org/~/media/WhitePapers/White_Paper_6_-_PUE_and_DCiE_Eff_Metrics_30_December_2008.pdf?lang=en [accessed 01.02.12].
[54] The Green Grid. Usage and public reporting guidelines for the green grid's infrastructure metrics PUE/DCiE. White Paper #22; 2009. Available from:, http://www.thegreengrid.org/library-and-tools.aspx [accessed 12.12.10].
[55] The Green Grid. Recommendations for measuring and reporting overall data center efficiency. Version 1 – measuring PUE at dedicated data centers. Task force report; 2010. Available from:, http://www.thegreengrid.org/en/Global/Content/Reports/RecommendationsForMeasuringandReportingOverallDataCenterEfficiency [accessed 17.01.11].
[56] The Green Grid Global Task Force. Recommendations for measuring and reporting overall data center efficiency. Version 2 – measuring PUE for data centers; 2011. Available from: http://www.thegreengrid.org/en/Global/Content/Reports/RecommendationsForMeasuringandReportingOverallDataCenterEfficiencyVersion2 [accessed 19.05.11].
[57] Stamford C. Gartner says energy-related costs account for approximately 12 percent of overall data center expenditures; 2010. Available from: http://www.gartner.com/newsroom/id/1442113 [accessed 04.09.13].
[58] The Green Grid. Carbon usage effectiveness (CUE): a green grid data center sustainability metric. White Paper #32; 2010. Available from:, http://www.thegreengrid.org/library-and-tools.aspx [accessed 12.12.10].
[59] Shiino T. Japan's approach to reducing greenhouse gas emissions from data centers; 2010. Available from: http://www.nri.co.jp/english/opinion/papers/2010/pdf/np2010157.pdf [accessed 17.01.11].
[60] The Green Grid. The green grid data center compute efficiency metric: DCcE. White Paper #34; 2010. Available from:, http://www.thegreengrid.org/library-and-tools [accessed 02.02.11].
[61] The Green Grid. Proxy proposals for measuring data center productivity. White Paper #17; 2009. Available from:, http://www.thegreengrid.org/library-and-tools.aspx [accessed 12.12.10].
[62] The Green Grid. Harmonizing global metrics for data center energy efficiency. Global taskforce reaches agreement regarding data center productivity. March 13, 2014; 2014. Available from: http://www.thegreengrid.org/library-and-tools.aspx [accessed 22.03.14].
[63] The Green Grid. Electronics disposal efficiency (EDE): an IT recycling metric for enterprises and data centers. White Paper #53; 2013. Available from:, http://www.thegreengrid.org/library-and-tools.aspx [accessed 07.03.13].
[64] The Green Grid. ERE: a metric for measuring the benefit of reuse energy from a data center. White Paper #29; 2010. Available from:, http://www.thegreengrid.org/library-and-tools.aspx [accessed 12.12.10].
[65] The Green Grid. A framework for data center energy productivity. White Paper #13; 2008. Available from:, http://www.thegreengrid.org/library-and-tools.aspx [accessed 12.12.10].
[66] The Green Grid. The green grid introduces data center sustainability metrics. Porter Novelli Press Release; 2010. Available from: http://eon.businesswire.com/news/eon/20101202005473/en/data-center%3B-energy-efficiency%3B-carbon%3B-power-usage [accessed 02.12.10].
[67] The Green Grid. Water usage effectiveness (WUE): a green grid data center sustainability metric. White Paper #35; 2011. Available from:, http://www.thegreengrid.org/library-and-tools [accessed 31.01.12].
[68] The Green Grid. Data centre life cycle assessment guidelines. White Paper #45, v2; 2012. Available from:, http://www.thegreengrid.org/library-and-tools.aspx [accessed 20.11.12].
[69] Shah A, Bash C, Sharma R, Christian T, Watson BJ, Patel C. The environmental footprint of data centers. In: San Francisco, California, USA: proceedings of IPACK2009(2009); 2009.