Future of data centre efficiency metrics

A review of progress so far and a suggested direction for further development
## Contents

1. Abstract 4

2. Summary 5
   2.1 Why do we want data centre efficiency metrics? 5
   2.2 Problems with achieving an overall efficiency metric 6
   2.3 Underlying growth in ICT services 6
   2.4 Efficiency or productivity metrics will not save energy 6
   2.5 Efficiency metrics do not equate to business value 7
   2.6 An alternative approach 8

3. The current state of metrics 9
   3.1 PUE 9
   3.2 Source energy adjusted PUE 9
   3.3 What about the IT equipment and software? 9
   3.4 Unit capacity based metrics 12
   3.5 An underlying problem 12
   3.1 Monitoring tools 14
   3.2 Gaming metrics and benchmark engineering 14

4. What do we want from metrics? 15
   4.1 Account the energy consumption of data centre services 15
   4.2 Understand the measure 15
   4.3 Understand the delivered carbon intensity to IT equipment 15
   4.4 Understand the level of energy re-use 16
   4.5 Support scope 3 disclosure 16
   4.6 Allow operators to compete directly on efficiency 16
   4.7 Traverse contract boundaries 16

5. Metrics for energy supply to the IT equipment 18
   5.1 Goals 18
   5.2 PUE 19
   5.3 Source energy 19
   5.4 Energy re-use 21
   5.5 Combining the metrics as IT delivery intensity 22

6. Accounting IT energy to delivered services 23
   6.1 Work or productivity measures 23
   6.2 Accounting the cost of services 23
   6.3 Cost and energy allocation traverses boundaries 24
   6.4 One rule – all cost and energy must be allocated 24
1. Abstract

This paper reviews the current state of data centre energy efficiency metrics, both those proposed and those already in use. The analysis of these metrics concludes that there are underlying, potentially insoluble problems in measuring the IT work or service output of a data centre in order to report efficiency. Despite a number of years of work by many skilled parties none of the current or proposed metrics is able to address this issue in a general form.

Further, even if these metrics were to be developed, many data centres are operated by more than one party and the areas of responsibility of these parties do not map well onto the parts of the data centre measured by each group of metrics. This mismatch would make it hard to obtain measurements in many data centres and even harder to assign responsibility or target the performance of each party when key components of the targeting metrics are not within their control.

The utility of functional data centre efficiency metrics in delivering the goals of reducing data centre or overall energy use is also examined and a number of significant issues are raised concerning their ability to deliver or support a market in more efficient data centre services.

Based on these underlying problems and recognising that there is a need for measures of data centre energy efficiency from operators, equipment vendors and regulators, each of whom wishes to see both the component and overall energy efficiency of data centres improved, an alternative suggestion is made.

The industry should implement a cost of services model to account for the energy and CO₂ consumed in delivery of each ICT service from a data centre. This approach utilises measurements, knowledge and processes which are currently available in the data centre industry and is based on well understood economic methods. The result is a metric which is both comparable and supports the development of an informed and effective market in efficient data centres and data centre services.

This alternate approach provides two major benefits over continuing to pursue overall data centre efficiency metrics; firstly, it is currently feasible and secondly it would support the goals of data centre metrics more effectively than overall efficiency metrics even if they were available.

Recommended actions

The BCS recommendation to the data centre industry, governments and regulators is:

- PUE\(^1\) should not be expanded to include onsite power generation, this would substantially devalue the metric
- Extend the PUE metric with the addition of carbon intensity\(^2\) of electrical supply metrics to produce a kg CO₂ per kWh delivered to the IT equipment metric
- Stop working on overall efficiency metrics unless a general and portable solution to the underlying problems is found
- Start implementing a common cost of services protocol to report the energy and CO₂ required at each stage of the data centre supply chain from IT equipment to delivered services

---

\(^1\) Power Usage Effectiveness – how effectively the data centre transfers power from the utility to the IT equipment

\(^2\) Such as CUE
2. Summary

PUE, a metric which describes how efficiently the power and cooling infrastructure of a data centre operates, has become established as the dominant measure of data centre efficiency. However, it only captures a relatively small part of the overall efficiency. This limited scope restricts the value of reporting PUE and can be quite misleading if it is misused to compare data centres. The PUE metric was not designed to consider other aspects such as the energy supply to the data centre or the efficiency of the supported IT equipment. Whilst PUE has been adopted relatively quickly by the industry, further progress towards metrics which capture more of the data centre activity has largely stalled.

There is still a desire for overall efficiency metrics, both for operators to track or report their own progress and for other parties to compare data centres. There is a broadly held view that it should be possible to measure the overall efficiency of a data centre by comparing the energy consumed with the “useful work” performed by the IT equipment.

Much of the current development effort for efficiency metrics is focussed on measures of this “useful work”, be this at the service output or IT equipment level. The concept is; given that the energy input in kWh can be measured, if the IT work performed can also be measured then the efficiency can be determined by comparing these values.

An alternative approach to directly measuring the work performed is to use a fixed measure of the IT device work capacity and then a measure of device utilisation to determine the overall work for an efficiency calculation. This is the basis of the DPPE metrics proposed by the Japanese GIPC.

During discussion of data centre metrics there is a subtle difference between the efficiency of an IT service and of the IT infrastructure which is often overlooked. The former includes the efficiency of the software and the application, the latter does not. Many of the proposed metrics, such as the Green Grid productivity proxies and DPPE fail to include the software and other externalities such as the energy consumption of the staff supporting a service. The proposed cost of services approach on the other hand is more holistic and includes these upper levels of the service delivery stack and externalities rather than just focusing on the hardware. This shortcoming is in itself a strong argument against proposals such as DPPE, as the objective for metrics is the efficiency of the IT service rather than further partial metrics. In this paper the full data centre stack is described to demonstrate the scope of coverage of the various metrics.

2.1 Why do we want data centre efficiency metrics?

Prior to discussion of the proposed metrics it is useful to review why we want metrics for the efficiency of data centres as this provides a set of deliverables against which we may assess the options.

Reduce overall energy consumption
The first major goal is to reduce overall energy consumption this may be viewed in both data centre specific and broader terms as;
- Reduce data centre energy consumption
- Reduce total energy consumption through the intelligent use of data centre services

Understand the (in) efficiency of data centres
In the context of the data centre the major goal is to understand the efficiency (or inefficiency dependent upon your viewpoint) of the data centre and where it may be improved. This creates a requirement for metrics which can help:
- Understand which data centres are more efficient
- Understand the efficiency of each part of a data centre
- Set and enforce targets for acceptable efficiency
2.2 Problems with achieving an overall efficiency metric
Many of the current proposals for productivity and utilisation metrics have been tested on a small scale and are able to measure or report values. However, there are a number of unsolved issues with this class of metric.

To be accurate enough to properly set targets for or compare data centre efficiency, they require a number of weighting factors to collapse the multiple dimensions of “IT work” or “utilisation” into a single value to compare with the input kWh. These weighting factors are user or service specific and therefore are no longer sufficiently general to be useful for targeting or comparison.

Many of the current proposals seek to measure the IT equipment activity and not the delivered services. This creates a specific weakness as they do not measure the actual output of the data centre and they are easy to “benchmark engineer”. This has already been seen in PUE measurements with many operators publishing claimed PUE figures which are clearly determined in a manner intended to optimise the score rather than provide comparable data.

Although not useful for external comparison or targeting, some of the proposed efficiency metrics may be effective as framework tools to operators and service providers who wish to develop internal, activity or company specific measures in the same way that monitoring tools allow the construction of service specific measures of performance or availability.

2.3 Underlying growth in ICT services
This paper does not directly discuss the underlying growth in demand for ICT services as this is a separate area of research. This growth does however impact the demand for data centre energy efficiency measures as the underlying growth in consumption of ICT services is driving growth in data centre capacity and energy consumption. This paper does address the issue of whether the proposed metrics are useful in reducing overall energy consumption where data centre ICT services are used to reduce carbon in a physical process.

2.4 Efficiency or productivity metrics will not save energy
The analysis determines that even if productivity or overall efficiency metrics existed, were practical to measure and produced strong measures which we could use to compare or target data centre efficiency, they would not deliver the goals of reducing data centre or overall energy consumption.

As a result of this it is recommended that governments and regulators should not seek metrics to describe overall efficiency to use as targets or regulatory measures for the data centre market as these would inevitably be a blunt instrument which would fail to deliver meaningful reductions in either data centre or overall energy consumption.

Efficient data centres drive further consumption
A related issue to the underlying growth in ICT services is the significant demand elasticity in the market for IT services. Improving the efficiency of data centre services makes them cheaper and unlocks further demand. Much of the evidence so far supports the view that the level of demand increase will increase the overall energy consumption of ICT services despite individual services being more efficient.

Efficiency metrics will not deliver a market
One key goal for governments and regulators is to support the creation of an informed market where the energy consumption or carbon of ICT services is a key factor in vendor selection.

Overall efficiency metrics will not only fail to deliver this market but are likely to obstruct and misinform it.

---

3 In the context of attempts to reduce energy consumption through improved energy efficiency this is also known as rebound effect or Jevons Paradox, [http://en.wikipedia.org/wiki/Jevons_paradox](http://en.wikipedia.org/wiki/Jevons_paradox)
A consumer of data centre services will not be able to use the “efficiency” score of the data
centre to evaluate the service specific efficiency or carbon due to three fundamental issues in
“reversing” these metrics:

- Without knowing the “work” required to deliver a unit of service, a “work / kWh” or “work
  / kg CO₂” metric is useless to the service consumer
- In many cases the services will not be directly comparable in terms of the data centre
  “work” required. For example, a service consumer comparing a webmail service such as
  Gmail with hosted Microsoft Exchange mailboxes will be comparing very different
  quantities of “work” per user month in addition to different functionality
- Multiple services are delivered from most data centres, each at a different level of
  efficiency which will vary with user load, time, workload on other platforms etc. There is
  no practical way for the service consumer to separate the efficiency of their service from
  the remainder in the facility

**Efficiency metrics will not support overall carbon reduction**
For the same reasons as stated above, overall efficiency metrics will be of no use in determining
the overall energy or carbon impact of an ICT service. In order to effectively reduce overall
energy we need to be able to compare on a case by case basis the ICT energy consumption
with the energy mitigation elsewhere due to the use of the ICT system, for example the kWh
used by a logistics system versus the diesel saved in the vehicle fleet.

**Efficiency metrics fail in a multi party supply chain**
Each of the metrics proposed so far requires at least two measurement points in the data centre
to determine efficiency. The common issue is that these measurement points do not map onto
the areas of control in many data centres. As the data centre market becomes more specialised
it is increasingly common to have several parties involved in delivering the ICT service. If there
is no useful metric which matches the area of control of each party it is hard to develop any
clear view of responsibility or apply targets.

**Efficiency metrics do not support international activity**
Overall efficiency scores do not allow us to deal with the import or export of services between
regulatory zones. If a data centre operator or service provider outside a regulatory zone
provides a service into it then there is no unit of work to which an “intensity factor” can be
applied. IT services are not delivered in tonnes or any other easily measurable or taxable unit.

This means that there can be a direct pressure to displace and a specific market distortion, not
from less to more efficient data centres, but from within to outside the regulated zone. Where a
regulatory zone has already implemented other measures such as increasing the proportion of
renewable energy in their grid this displacement is likely to carry the double penalty of reduced
data centre efficiency and higher supply carbon intensity. Equally, if services are exported from
a data centre in the region there is no method of returning the local carbon tax for the exported
portion of services

**2.5 Efficiency metrics do not equate to business value**
One of the major tasks within organisations is to demonstrate the value to the organisation of
the ICT services delivered from the data centre. A key step which is often overlooked is that
metrics of the efficiency of delivery of IT output would require a further metric converting the IT
output to received business value. This paper avoids this conversion by instead proposing that
the cost of service delivery be determined and leaving the consumer of the service to determine
whether the delivered value exceeds the cost.
2.6 An alternative approach
These structural issues with data centre efficiency metrics lead to the conclusion that they will not be useful in reducing overall or ICT specific energy consumption. Therefore an alternative is necessary to support the improvement of ICT cost and energy efficiency at the same time as reducing overall energy. It is proposed that a cost of services model is used as this approach is already used in many other sectors, is well understood and has been shown to deal well with complex multi-party supply chains such as those seen in the ICT market.

This approach:

- Allows us to use the knowledge, tools and measurements already in place in data centres to rapidly produce a common set of measures which can be implemented at a much lower cost than most metrics proposals.
- Supports the development of a market in ICT services based on the energy consumed to deliver the service which allows;
- Supports a later move toward full supply chain (scope 3 type) disclosure of an organisation’s carbon. This is not supported at all by the current proposed metrics.
- Allows data centre operators to compare their energy of delivery of services with others in their sector or the general market whilst also considering other factors which influence the service such as features, performance or availability.

The cost of services approach is discussed in two parts, from the energy supply to the kWh delivered to the IT equipment and then the conversion of IT kWh to delivered services.

Energy supply and data centre infrastructure metrics
The Green Grid PUE metric and the proposed CUE metric are both discussed as key components of this cost of services approach in the context of metrics and measurements of the supply of kWh to the IT equipment in the data centre.
3. The current state of metrics

This section discusses the metrics currently in use and those proposed. The data centre stack is introduced to provide context for discussion of what is being measured and what is excluded by each metric. The proposed approaches to measuring IT work are introduced and the underlying problems identified.

Progress so far
The data centre industry has spent several years and considerable effort in developing data centre metrics but in short, we have PUE and nothing much else. The last DCSG review of data centre energy efficiency metrics in 2008 proposed that there would be both difficulties in development of metrics above the PUE level and issues with the utility of any such metric. The recommendation was made that the industry should instead work toward metrics and methods which could “calculate a fair and reasonable approximation of the total environmental and financial cost of the service provision from the data centre”.

3.1 PUE
PUE, or Power Usage (in)Effectiveness is the only metric which has been broadly accepted, it describes the (in)efficiency with which the data centre delivers electrical energy to the IT equipment. DCIE, Data Centre Infrastructure Effectiveness is the same metric with the fraction inverted.

\[
PUE = \frac{Total\ Utility\ Energy}{Delivered\ IT\ Energy} = \frac{1}{DCIE}
\]

As an example; if a data centre has a PUE of 2 then for every 2 kWh of power purchased from the utility only 1 kWh is delivered to the IT equipment, the DCIE would be 50%.

PUE has been successful for three major reasons;
- PUE is unit-less
- PUE is easy to explain
- PUE is relatively easy to measure.

3.2 Source energy adjusted PUE
One current area of discussion is whether to extend PUE to cope with data centres which take in energy in forms other than electricity. This includes both facilities with onsite power generating capacity which use a fuel to create electricity and facilities which take in some other energy consuming resource such as a landlord’s chilled water supply or municipal steam feed. It is fairly clear that a data centre which takes landlord’s chilled water instead of running its own chillers should have to account for a share of the landlord’s chiller energy consumption in its PUE, but the value of including onsite generating facilities is questioned.

3.3 What about the IT equipment and software?
PUE and DCIE only cover the data centre plant section of the stack. To expand the scope of metrics and capture more of the data centre activities the goal for most groups developing metrics has been to move up the data centre stack and try to capture the “IT work” or “Useful work” delivered by the IT equipment or software platforms. A range of proposals have been made of greatly varying complexity and data gathering overhead.

Despite much effort there has been no significant progress in finding portable, measurable and effective metrics for anything above PUE. There are a series of underlying reasons why the
metrics higher in the stack are likely to remain out of reach and not generally applicable or transferable between data centres or organisations.

**Measuring work**
To move up the stack it is necessary to measure the work delivered by the IT equipment or software. Unfortunately this is quite subjective and two different organisations are unlikely to agree on precisely how to measure it. Worse, two different platforms within the same data centre may not be measurable in the same way.

**A single hard disk**
As an example of this problem we can consider a single hard disk drive being used in a data centre.

In the first use case the hard disk is acting as a backup device for a user's machine. The disk is inactive and hopefully powered off for most of the day and only working for a short period where the backup data is transferred. In this case we are concerned primarily with the total amount of data we can store on the disk (Capacity) and on how reliably that data is stored, measured by the MTTDL\(^4\) of the disk or disk system.

As a second case we might connect the disk to a streaming media server. In this case the primary copy of the content is elsewhere and we are not concerned by the MTTDL. Whilst we are interested in the capacity the primary constraint to the media server will be the sustained data transfer rate (SDTR) of the disk as once this is reached the server cannot deliver any more streams to users.

In a third case we could use the disk in a database server. Here the reliability of the disk is again of importance to us but our primary concern becomes not how quickly the disk can transfer long contiguous sections of data but how quickly it can read and write small sections of data with fairly random distribution on the disk surface. In this case it is common to not even format the entire disk as the major constraint is the seek time and more disks will be added to obtain IOPS\(^5\) performance before capacity is reached.

This range of measures of work for a single hard disk drive in a single data centre illustrates the problem which all metrics seeking to measure work must address. This is illustrated by the Venn diagram in Figure 3.

---

\(^4\) Mean Time To Data Loss
\(^5\) Input / Output Operations per Second
Complex measures of work
As the industry has explored the range of “work” that IT devices perform in the data centre and that many of them perform these different kinds of work simultaneously, a number of increasingly complex and data intensive measurement schemes have been put forward to capture everything down to individual HTTP requests on web servers to record the work performed.

The further we go into this assessment of “output”, “value” and “factors” the more apparent it becomes that whilst we may be able to present commonly applicable tools and mechanisms the specific values and weighting (or fudge) factors will always have to vary between organisations and frequently within organisations and data centres. This means that any “efficiency” metric derived from them is unlikely to be usefully comparable between different organisations or across the data centre market.

Measuring utilisation
Another approach which some parties are evaluating as an alternative to trying to measure the work delivered is to determine a specified “efficiency” for a given device under reference conditions and then to use the measured utilisation to determine the work delivered. This has the advantage that we can try to agree on one set of reference conditions to measure the efficiency of a server, such as a composite benchmark which does not depend on how we actually plan to use it but represents a reasonable average profile for the market. This is a common benchmarking approach and well understood.

The approach then, is to take the efficiency of the device and multiply this by the achieved utilisation, this avoids having to specifically measure the work performed and the discussion of which aspects of the work have which relative values. An example of this approach is shown below;

\[ Efficiency = DCIE \times IT\ Benchmark\ Efficiency \times IT\ Utilisation \]

We can alternatively express this in units to understand what is happening;

\[ Efficiency = \frac{IT\ kWh}{Utility\ kWh} \times \frac{IT\ benchmark\ work}{IT\ kWh\ at\ benchmark\ load} \times \frac{IT\ actual\ work}{IT\ benchmark\ work} \]

When we cancel through the units in the fractions we see that this is indeed a measure of the IT work per consumed energy.

\[ Efficiency = \frac{IT\ actual\ work}{Utility\ kWh} \]

Unfortunately the problem with this approach is that utilisation is subject to the same issues as measuring work. Even if we stick to using a very simple proxy such as processor load for server utilisation this approach will break down.

As an example of why this approach will break down, is as follows. In heavily virtualised environments it is not uncommon for the maximum achievable processor utilisation to be in the region of 30%. This is due to the number of virtual machines running on a single physical processor and the time it takes the processor to “context switch” between the different virtual machines. We cannot extract any more work by adding more virtual machines, the processor is choked, so should the operator be allowed to reset their maximum utilisation to 30% to calculate their efficiency? If so, should this threshold be determined on a per chassis basis? Should the operator report the maximum utilisation threshold they measured their efficiency with?
3.4 Unit capacity based metrics

Another approach has been to produce generic “capacities” for entire classes of IT equipment to eliminate the large quantity of asset and metering data required for useful work or utilisation metrics. The concept was that a capacity could be assigned to each class of IT device in the data centre and then be divided by the device power consumption to achieve a measure of efficiency. Unfortunately this approach failed to capture the diversity of the IT equipment and also failed to include whether the IT equipment was being utilised. This fell into the same trap as measuring data centres based on PUE, i.e. that of assuming that all IT electrical load is good and ignoring whether the server is doing any work at all. The opportunities to game such simplistic metrics and the perverse incentives created completely rule out this approach as doing more harm than good.

3.5 An underlying problem

Metrics based on measuring the work or utilisation which requires units of work must address an underlying problem. At the level of PUE we have one dimensional input (kWh) and one dimensional output (kWh) whereas as soon as we move up into the IT equipment or software we have multiple dimensions of output which we must somehow collapse or reconcile into a single ‘super’ unit to achieve an efficiency rating against our one dimensional input.

![Image](Figure 4 Dimensions of IT capacity or utilisation)

As shown in Figure 4 at a simple level we have at least three major types of equipment which would require separate measurement and weighting factors to reduce the measurement to a single value for comparison with the energy consumption kWh or kg CO₂.

\[ \text{Efficiency} = \frac{\text{Compute} \cdot \alpha + \text{Network} \cdot \beta + \text{Storage} \cdot \gamma}{\text{kWh}} \]

Unfortunately as already described we cannot simply measure the Network or Storage “work” undertaken. A firewall performing packet inspection is not equivalent work to that of a simple top of rack switch blindly forwarding packets. Figure 5 below shows some of the dimensions for storage as discussed above.

![Image](Figure 5 Dimensions of storage work or utilisation)

6 An example was Compute Units Per Second (CUPS) proposed as a productivity proxy

7 “gaming” a metric refers to deliberately choosing to optimise the achieved score on the metric at the expense of the intended target by exploiting the measurement process. In this case choosing to deploy a larger number of servers of lower compute capacity would improve the score on the metric whilst not improving overall efficiency.
To address this we need to expand our definition of work as shown in Figure 6 and therefore include further sub weighting factors for each area.

\[
\text{Efficiency} = \frac{\alpha (\text{Processor} \cdot \delta + \text{Memory} \cdot \varepsilon + I/O \cdot \zeta) + \beta (\text{Bytes} \cdot \eta + \text{Distance} \cdot \theta + \text{Inspections} \cdot \iota) + \gamma (\text{Capacity} \cdot \kappa + \text{SDTR} \cdot \lambda + \text{IOPS} \cdot \mu + \text{Reliability} \cdot \nu)}{kWh}
\]

Further, as described above in the example of the hard disk, we need to consider that each service platform in the data centre is likely to require a different set of weighting factors as the aspects of “capacity” and “utilisation” may be quite different for each device. This gives us;

\[
\text{Efficiency} = \frac{\sum_{\text{platforms}} \left( \alpha_n (\text{Processor} \cdot \delta_n + \text{Memory} \cdot \varepsilon_n + I/O \cdot \zeta_n) + \beta_n (\text{Bytes} \cdot \eta_n + \text{Distance} \cdot \theta_n + \text{Inspections} \cdot \iota_n) + \gamma_n (\text{Capacity} \cdot \kappa_n + \text{SDTR} \cdot \lambda_n + \text{IOPS} \cdot \mu_n + \text{Reliability} \cdot \nu_n) \right)}{kWh}
\]

Once these factors are taken into account to deliver an effective metric the weightings will inevitably be user specific and subjective, meaning that the metric will have little value for comparison of data centre operators, with this value further reducing as the nature of their business diverges. Even if one user can determine the weighting factors, these may not even be time stable for that operator, meaning that the metric would even fail to track progress for one operator so rendering it useless.

*If we cannot measure IT work then we cannot measure efficiency.*
Simple metrics - fairness vs. complexity
A common suggestion is to have a “simple” metric which is understood to be somewhat flawed but which provides a basic measure of the “efficiency” of the IT equipment or overall data centre. This is an attractive idea but leads directly back to the problems already discussed;

Any “simple” metric of work or efficiency will unfairly advantage or disadvantage different types of operator leading to both resistance to and gaming of the metric.

This will inevitably lead to further tuning of the metric to compensate for the identified issues or exploits which simply returns us to the problems already discussed with measuring work or utilisation.

3.1 Monitoring tools
The difficulty of obtaining common measures of IT work or delivered service is not unique to efficiency metrics. The same issues are well illustrated in the architecture of monitoring tools. These tools do not attempt to claim that a single configuration is appropriate for every operator and service but instead provide a set of capabilities which must be configured to the target environment. It is well understood that moving a fully configured monitoring platform from one data centre to another would not be useful, this is equally true of metrics at the IT level.

3.2 Gaming metrics and benchmark engineering
There is a significant risk of these complex metrics functioning as a diversion from the task of improving efficiency and reducing energy consumption. We have already seen that targeting a data centre operator on PUE can create a perverse incentive not to reduce IT energy consumption as this results in a worse PUE. The result of this is an incentive not to save energy at the IT equipment.

The more complex productivity or utilisation metrics present the same basic opportunity to game the metric but with many more methods of doing so. As the delivered services from the data centre are not measured an operator may improve their score simply by increasing their compute load by making the software less efficient. This may create a small increase the overall energy consumption but their score will improve. Where the weighting factors are defined or selected by the operator there are clear cases where selection of alternate factor sets would improve the metric score more cost effectively than implementing efficiency measures.

One important issue to accept and embrace in the construction of any metric or benchmark is that if it is successful then vendors and operators will inevitably be led to engineer their score. Where this “benchmark engineering” drives the market towards the improvements that the metric set out to deliver this is a positive outcome. The issue for all of the proposed productivity metrics is that there is substantial opportunity to improve the score of a component or a data centre through either careful measurement or specific efficiency reductions. Given the complexity of these metrics there can be little confidence that they will drive the requested behaviour.

8 This is already a recognised problem for PUE with many claims of “design PUE” or “partial PUE”
4. What do we want from metrics?

This section will review the specific deliverables for data centre efficiency metrics. For each deliverable the known issues or constraints will be discussed. These issues are not constrained to the technical measurement of data centre efficiency but include the contractual relationships and processes that must be accommodated.

4.1 Account the energy consumption of data centre services

It is important that the industry develops the ability to account the cost and energy consumption of data centre services. Applying an energy or environmental cost to each delivered service not only supports the development of a real market in energy efficient data centres but also directly supports market transformation activities.

A secondary advantage of accounting the energy of delivered services is that a service consumer making use of ICT to reduce energy consumption or cost in a physical process is able to make informed decisions to reach an overall optimum energy and cost for the process including the ICT contribution. Abstract “useful work” metrics will not be able to deliver this sort of decision support information.

Legal constraints

It is recognised that many countries do not allow data centre operators to charge their customers for the power consumed by the IT equipment, this results in an “all you can eat” sale of power and cooling capacity which does not include the energy consumption. Whilst this does present a useful incentive not to over provision power and cooling capacity it also removes the strong incentive for power management which is present where the data centre operator is permitted to charge for consumed energy. This is a matter for those states which prevent recharging for power to consider whether this constraint is in alignment with their environmental or energy security policy objectives.

4.2 Understand the measure

There some proposals to make the PUE metric much more complex by including other fuel sources in the PUE calculation to provide greater coverage. This generates a number of issues with the use and understanding of PUE. A large part of the success of PUE has been the simplicity with which it can be measured and understood. If the metric is made too complex we risk destroying any value it currently provides.

A user of the metrics should be able to easily understand the PUE of a given data centre, as a primarily electrical measure. Unlike other commercial buildings very little energy is used as heating or steam supply to large data centres. The major issue is onsite electricity generation which should not be included in the PUE and is better handled in a separate step.

4.3 Understand the delivered carbon intensity to IT equipment

We should be able to understand the carbon intensity of a delivered kWh to IT equipment in the measured data centre, specifically we should be able to multiply the PUE by the local grid carbon intensity to achieve a good approximation. In the cases where the data centre is not grid powered the carbon intensity of the onsite generation or mix should be published with the PUE data. To support this it is necessary to understand the utility electrical consumption of the data centre separately from any large scale onsite power generation activity.

This IT intensity metric represents a strong measure for environmentally aware procurement. If an operator builds in a coal fired state their data centre is unlikely to score well on a kg CO₂ / delivered IT kWh measure whereas the operator who built in a location with high renewable but higher power costs (such as Holland) is likely to achieve a much better delivered IT carbon intensity even if their data centre has a worse PUE. If these operators were simply compared on PUE the real difference in environmental impact would be masked.

---

9 An example of this legal restriction is France.
A key benefit of a delivered IT carbon intensity metric is the simplicity which promotes understanding and therefore use of the metric.

### 4.4 Understand the level of energy re-use
Many data centres are now exploring mechanisms to re-use some of the data centre energy for some other purpose. Whilst it is clearly desirable to re-use energy it is important that we are able to determine what the real savings are and provide a mechanism for the operator to report this improvement.

To achieve this, the operator should be able to publish, if they wish to, a reduction factor for a data centre which describes in a portable and effective manner the achieved energy saving value of any re-used energy from a data centre.

### 4.5 Support scope 3 disclosure
Metrics should support scope 3 disclosure of energy consumption or environmental impact. Specifically a customer consuming a data centre service should be able to declare the energy consumption or CO$_2$ of that service as part of their own environmental statement. Present metrics$^{10}$ do not support this and leave the CO$_2$ stranded every time a contract boundary is crossed.

### 4.6 Allow operators to compete directly on efficiency
Data centre operators should be able to compete with each other on the efficiency of their operations and services. Outsourcing providers should be able to compete with internal IT departments and data centres on specific and effective measures of energy consumption and cost.

### 4.7 Traverse contract boundaries
The most significant flaw with many of the metrics proposed so far is that they measure between technical boundary points in the data centre stack such as the IT energy consumed or IT work delivered. In practice most data centres are operated by more than one organisation (or by poorly connected divisions of a larger organisation). The division of control and therefore responsibility does not neatly map onto the measurement boundaries of the proposed metrics. This is a serious issue for the following two reasons:

**Measurement points**
First, in many cases an operator may not be able to measure their performance under a metric which has a measurement point outside their domain of control because that data may simply not be available to them. Further it may not be reasonable to make it available, as it may, in the case of a multi-tenanted data centre, require the disclosure of sensitive information about the operator or other tenants.

**Control**
Second, if we consider the case in Figure 7 of the service provider (shown in orange) which is a tenant in a Colocation data centre we see that both of the measurement boundaries are partly outside the operator’s control.

In order to measure or report the “efficiency” of the operator we would need to choose two measurement boundaries which describe the input and output of their activity. For this service provider in a colo data centre we would need to select an upper and lower measurement boundary; their efficiency would then be measured by the amount of upper boundary measure they delivered per unit of lower boundary measure consumed. Unfortunately the actual boundaries of the service providers control do not cleanly map onto the measurements.

---

$^{10}$ And some legislation such as the UK Carbon Reduction Commitment
The lower boundary would be either the data centre kWh or the IT kWh value. Whilst the IT kWh is mostly within the service provider’s control (assuming their customers do not specify the IT equipment) the achieved PUE of the data centre is split control with both the other tenants and the Colo operator, as all of these parties have an influence on what the facility can achieve. A good example of this is air flow management practices and temperature / humidity set points which are generally set to the worst customer contract constraint.\(^{11}\)

At the upper boundary, IT output end of their operation they are likely to be responsible for the operating system, some shared infrastructure and may operate the software specified by the client. This leaves the option of an incomplete measure of their activity at the IT equipment work boundary or a measure of IT useful work over which they may have little control.

We can see that this operator has no fully controlled boundary of measurement and it may therefore not be reasonable or productive to require them by target or legislation to meet any specified target “efficiency”.

Failure to address this boundary issue will render any metric useless in the data centre market. Multi party supply chains must be able to measure and report their performance in the same way as individual operators with full control.

\(^{11}\) It should be noted that only a small proportion of Colocation providers are taking significant steps in this area by providing air flow management education and components to customers and that this is so far at their own cost.
5. Metrics for energy supply to the IT equipment

The structural issues described above with data centre efficiency metrics at the IT device or software level lead to the conclusion that they will not be useful in reducing either overall or ICT specific energy consumption. Therefore an alternative approach is necessary. If we accept that we cannot realistically provide a common, portable, transferable and widely applicable metric above the kWh delivered to IT equipment then we need to ask what else can be done with metrics which will support the goals of understanding and improving cost and energy efficiency in the data centre?

In this section metrics between the energy source and the electrical kWh delivered to the IT equipment are discussed.

One of the key issues for metrics at the data centre infrastructure level such as PUE is how well they perform when integrated into broader measures. This is significant as new metrics which consider the overall energy efficiency of the energy supply system and data centre or the overall carbon intensity of the energy delivered to IT equipment should be able to build upon the existing metrics.

The next section will discuss how the metrics at the energy supply and data centre infrastructure level can provide a measurement foundation for implementing a cost of services model for the data centre and delivering per service or activity accounting.

The concept of source energy conversions for data centre metrics is considered in the context of whether this obstructs other key measures.

5.1 Goals

Any metric, protocol or structure created to improve ICT efficiency will need to deliver on one or more of the following basic needs;

- Measure the efficiency of part or all of the ICT system – PUE already does this for the data centre infrastructure
- Assist day to day optimisation decisions – as upgrades and changes are made in the data centre we should be able to determine both the local and the system level optimal decisions
- Support the determination of the per activity, platform or service energy cost of ICT to permit optimisation of the total ICT + non ICT energy or carbon
- Facilitate creation of demand side incentives to change ICT consumption behaviour based on the energy or carbon cost of the delivered service
5.2 PUE
The PUE (or DCIE) of the data centre should be measurable for all facilities, whether they are stand alone or part of a larger building and share electrical or mechanical plant\(^{11}\). Acceptable methodologies and compromises need to be agreed to allow for the measurement and reporting of a PUE figure in which we can have a reasonable (and commonly understood) level of confidence. Chasing the last 0.1% accuracy is not relevant or helpful here, the requirement is for a practical protocol which allows the majority or operators to determine and track their PUE for reasonable effort and cost at a reasonable accuracy (e.g. +/- 5%).

The PUE should be an electrical transfer efficiency value in order to stay portable and to be useful in the additional comparisons we wish to make in the next steps.

Where other energy sources are used directly in the data centre such as gas to produce humidification steam or heating water these should be converted to an approximate kWh equivalent (these are frequently referred to as primary energy). Where other energy consuming services are provided to the data centre such as shared chilled water feeds these should also be converted to a kWh equivalent.

Diagnostic metrics – Issues with PUE
It would be ideal if metrics provided insight into why the performance of a facility is good or bad and guidance for day to day decisions to improve that performance. Unfortunately PUE provides only a summary result without any insight into the cause of that result. PUE is a well understood and accepted metric but the industry is slowly coming to realise that it is not reversible\(^{iv}\), PUE is an output and should not be used as an input. Specifically your PUE describes what overall proportion of your utility power is delivered to IT equipment, it cannot assist with:

- What utility power draw a given IT load is responsible for, as PUE is not reversible
- What impact a change in IT power draw would have, we know PUE tends to get worse as the IT load falls but how much worse? How much impact does the weather have?
- What impact a change in mechanical or electrical infrastructure or installed IT devices will have

Once this non-reversibility is understood, either by theory or from experience of making PUE based predictions and then failing to deliver this leaves the industry with a substantial issue. How does an operator predict financial or environmental savings to justify a business case for change when they can only determine the impacts after the change?

An ideal metric would give us this additional diagnostic guidance. However, PUE is now established and understood. The additional complexity of a metric which provides this type of diagnostic information coupled with resistance to change in the market suggests that the industry should stay with PUE but be clear about what PUE should and should not be used for.

5.3 Source energy
There has been significant discussion of applying source energy conversions\(^{12}\) to data centre efficiency metrics, specifically extended PUE to include these conversions. For other types of building efficiency mechanisms have been developed which include source energy conversions and are structured such that the building should not receive a credit or penalty based on its utility provider.

For data centres this not applicable and may be specifically undesirable. Whilst an office building has little choice over its geographic location and the source energy of its utility supply due to the location of the company and its employees (or students, customers etc.) this is not true for the data centre. Data centres can be, and frequently are, sited in alternate states or countries in order to take advantage of tax benefits, varying utility power prices or the availability of renewable electricity generation.

\(^{12}\) These mechanisms convert other energy consumption such as gas or oil into electricity equivalents to provide a single value in kWh equivalent for energy consumption.
There are two major issues to the source energy approach:

- A major part of the success of PUE has been the simplicity of measurement and reporting. Proposing a complex new mechanism to measure a metric which is already in use will be ignored by much of the market leading to a devaluation of both the previous and “improved” version of the metric.

- By including energy sources such as gas and oil used for onsite electricity generation these conversions mean that the quoted kWh consumption of a data centre may or may not be from the grid electrical supply. This would obstruct the determination of the carbon intensity of a kWh delivered to the IT equipment from utility carbon intensity figures which is a key step in developing further metrics.

Where significant onsite power generation takes place this should be separately measured for efficiency and carbon intensity. The electrical output to the data centre is the measurement point which allows PUE to remain comparable.

**Source energy conversions**

Whilst onsite generation of power should not be included in the PUE metric there is an argument to provide a mechanism for the small proportion of data centres which consume significant amounts of energy in a form other than electricity. Examples of this may be those housed in an office building which use a shared building chilled water loop delivering cooling to both office and data centre space. It is also noted that many data centres use small amounts of gas for space heating and hot water which could be included in the source energy conversion.

Rather than exclude other types of energy and provide an unfair advantage to data centres using other energy inputs, there are available conversion mechanisms published by government bodies such as the EPA which allow an approximate conversion to kWh for chilled water, gas or other fuel consumption.

![Figure 9 IT delivery intensity](image)

Figure 9 provides an overview of how the energy consumption of the data centre may be broken up and accounted for. kWh of electricity are the input to the PUE metric with non electrical sources such as chilled water or gas for heating converted to kWh equivalent. This provides an overall kWh of electrical consumption for the PUE calculation. Onsite generation of electricity and any energy re-use from the data centre are considered separately.
Shared chilled water systems

Shared chilled water systems are a specific case that has to be considered as it is not uncommon for a data centre to share chiller plant with the remainder of a building. This is particularly common where a relatively small data centre is housed in a larger office building. This may also happen in reverse where the data centre chilled water supply is used for other activities in the building which are not part of the PUE calculation.

It is frequently possible, by measuring flow rate and temperature on the chilled water pipe work to determine the fraction of the cooling load the data centre is responsible for. This would allow an operator with effective metering to assign a suitable fraction of the total cooling plant electricity consumption to the data centre and complete the PUE equation.

If a source energy conversion is used in place of measurement it is important that this is set with a low estimate of the efficiency (CoP) of the cooling system to ensure that operators with inefficient data centre chilled water plant do not have the option of claiming that it is “shared” and reducing their reported PUE without improving their actual performance.

It should also be noted that data centres using shared building chilled water systems are likely to be less efficient as the water loop temperature required for comfort cooling is low enough to deliver substantial latent cooling for humidity control, this is specifically undesirable for most data centre equipment which requires sensible cooling. Modern designs and standards for data centres and IT equipment do not require close control cooling and operate at substantially higher supply air temperatures, meaning that any significant sized data centre should be more efficient using a dedicated cooling system.

Supply Intensity

Given a measurable and agreed electrical PUE we can turn to the supply carbon intensity of the data centre. This is expressed as the kg CO₂ per kWh of electricity. These figures are frequently made available by government bodies for electricity distribution grids. Where there is substantial onsite generation of electricity (such as combined cooling and power systems) the achieved kg CO₂ per kWh of the onsite system or onsite – utility mix should be published alongside the PUE.

\[
Supply\ Intensity = \frac{\left( Grid \frac{CO_2}{kWh} \times used \ grid\ kWh \right) - \left( Onsite \frac{CO_2}{kWh} \times used \ onsite\ kWh \right)}{used\ grid\ kWh + used\ onsite\ kWh}
\]

5.4 Energy re-use

The final element to capture is the energy re-use achieved by the facility. Clearly it would not be useful or reasonable to simply measure the kWh of heat delivered from a data centre to the recipient as this is not a reasonable or useful indicator of the other source energy which has been saved by the recipient of the re-used energy.

An agreed protocol is required to allow for the calculation of the energy which has not been used as a direct result of the re-used energy from the data centre. For example, if the data centre provides heat to a district heating system the re-use measure is the gas or oil that is not used in domestic boilers as a result of the system. From this reduction measure it would be possible to produce a value for the source carbon eliminated which could then be transformed into a carbon or energy reduction weighting for the data centre;

\[
ReUse\ Weighting \left( \frac{kg \ CO_2}{kWh} \right) = \frac{Annual\ ReUse\ CO_2}{Annual\ Total\ datacentre\ kWh}
\]

For onsite electricity generating systems there is a discussion over whether to factor any additional work they perform such as heating of other buildings into the supply intensity figure or leave this as an externality to be captured in the optional energy re-use metric. If the other work they perform is to be factored in their supply intensity then the same basic accounting rules for energy re-use need to be applied. Not all of the available heat output of a CHP system will be

\(^{13}\) Not gasoline
used across the full year in most instances, therefore only the other source energy eliminated should be subtracted from the intensity of the onsite generation. Specifically, we should not simply divide the total energy or CO$_2$ of the CHP plant fuel across the kWh of electricity and available heat that it produces.

5.5 Combining the metrics as IT delivery intensity

Figure 10 below shows how the kWh boundary around the PUE metric, energy supply, onsite generation and energy re-use work together to provide an extensible system of measurements which is as applicable to a simple grid fed data centre as one which uses onsite cogeneration and drives a district heating system.

Using this flow it is possible to produce a useful and meaningful statement about the efficiency of the data centre in the form of the kg CO$_2$/kWh delivered to the IT equipment. This encompasses the PUE, the supply intensity and optionally any source energy reduction of energy re-use.

$$IT\ Intensity \left( \frac{kg \ CO_2}{kWh} \right) = PUE \cdot \left[ \text{Source Intensity} \left( \frac{kg \ CO_2}{kWh} \right) - \text{ReUse Weighting} \left( \frac{kg \ CO_2}{kWh} \right) \right]$$

It should be noted that a measure of this type is currently proposed by the Green Grid in the form of the CUE or Carbon Usage Effectiveness$^vi$.

Errors

It is accepted that there will be small errors in this approach where energy consumption in the form of steam for humidification or oil for heating has been converted to kWh equivalents at the PUE calculation stage and is then multiplied by the utility carbon intensity. The important step is to minimise this error by excluding onsite generation capability.

These small errors are an acceptable part of the compromise between accuracy, ease of understanding, ease of measurement and cost of measurement which must be struck with these metrics if they are to be adopted and deliver any value. A “perfect” metric which is too complex for the market to understand and too complex or expensive to measure will have no transforming effect on the market.
6. Accounting IT energy to delivered services

Given an effective measure for the cost and carbon of a consumed kWh of electricity at the IT equipment, we can apply a series of relatively simple accounting rules to take this up the ICT delivery stack and apply it to the output services where it may be used, either as a demand side incentive to reduce energy consumption, or simply as information for more effective decision making. This section will discuss that process, what benefits it offers and how it can be achieved. The approach proposed is designed to both be flexible to suit the range of operators and facilities and not prescriptive, the methods of implementation are intended to be flexible dependent upon the level of maturity and capability in each data centre.

6.1 Work or productivity measures
Even if metrics for the overall efficiency of a data centre were available they would not provide any insight into how the energy consumption should be divided up across the IT equipment and services within the data centre or how a change in one area might impact other parts of the system.

Further, translation of the “productivity” or efficiency measure to the delivered business value is likely to be at least as difficult and operator dependent as the definition and measurement of the metric itself.

6.2 Accounting the cost of services
A cost of services model has a number of direct advantages over efficiency metrics. The process is already used in many other sectors, is well understood and has been shown to deal well with complex multi-party supply chains such as those seen in the ICT market.

Existing knowledge and metering
The cost of services approach allows us to use the knowledge, tools and measurements already in place in data centres. Combining these measures has the potential to rapidly produce a common set of measures which can be implemented across all types of data centre, IT service and operator at a much lower cost and effort than most of the efficiency or utilisation metrics proposals.

Supports the development of a market in efficient ICT services
Reporting the energy consumption of delivered ICT services allows for the development of a market in ICT services based on the energy or carbon efficiency which allows:

- Both business and personal IT service consumers to include the efficiency of delivery in their supplier selection criteria
- Service providers to compete directly with each other and with corporate IT departments on the efficiency of their service delivery
- Multi party supply chains to measure and report their performance in the same way as individual operators who run the entire data centre
- The inclusion of other energy consumption such as office space and staff used to create or support the service

Benchmark efficiency in the context of other factors
Cost of delivery allows data centre operators to benchmark their efficiency against others in their sector or the general market whilst also factoring in other factors which influence the
service such as features, performance or availability at whatever cost or value they apply to those factors.

6.3 Cost and energy allocation traverses boundaries
In contrast to the efficiency metrics which present potentially insoluble issues in a multi party supply chain, the process of allocating cost and energy up the data centre stack to the output services is inherently able to traverse the contract or responsibility boundaries which appear in the data centre. Contracts are, by their nature, structured to pass costs and to describe the responsibility of each party for the management and control of those costs.

Supply chain disclosure
There is a strong desire amongst many governments, NGOs and regulators to have a data centre efficiency metric which can be applied to the market as part of energy or carbon reduction. Unfortunately at the same time as requiring data centres to use less energy the same bodies are also suggesting that much of the energy savings in other sectors will be due to the use of ICT systems to drive other efficiencies.

Within the bounds of a single organisation we may be able to examine the overall energy consumption or carbon footprint of their physical activities and the ICT used to support them but this breaks down as soon as the ICT services are provided by another entity.

The reporting of the carbon of a delivered service allows the inclusion of the ICT supply chain carbon (scope 3\textsuperscript{14}) in the reported carbon of each company. A policy direction toward supply chain disclosure of ICT carbon would provide a strong incentive to structure service contracts to include energy as well as cost responsibilities\textsuperscript{15}.

6.4 One rule – all cost and energy must be allocated
To implement cost and energy allocation the only major rule is completeness. All cost and energy at one layer must be allocated to the next layer in the stack. The allocation regime need not be prescribed but the operator is not permitted to siphon off cost or energy into a bucket of “unallocated” consumption.

6.5 Allocating data centre energy to IT equipment
The IT equipment is supplied with energy by the data centre infrastructure for which measurements and metrics already exist. The completeness rule may be easily implemented by ensuring that all energy, cost and carbon for the data centre in the allocation period is distributed across the installed IT devices.

Allocation methods
Give the completeness rule the first step is to determine one or more methods of allocating the total data centre energy consumption or carbon to the installed IT devices.

The method of allocation should not be prescriptive, as the selection of method and accuracy of allocation may well change over time as the available data improve.

Three example methods are presented, each of which is possible with current technology in today’s data centres. The methods are presented below in Table 1.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
IT kWh & IT Equipment \\
\hline
Data Centre & Transformer \\
M&E Plant & \\
\hline
Energy Supply & \\
\hline
\end{tabular}
\caption{Allocating IT kWh}
\end{table}

\textsuperscript{14} See the Carbon Disclosure Project www.cdproject.net

\textsuperscript{15} “Scope 3: Other indirect emissions, such as the extraction and production of purchased materials and fuels, transport-related activities in vehicles not owned or controlled by the reporting entity, electricity-related activities (e.g. T&D losses) not covered in Scope 2, outsourced activities, waste disposal, etc.” see http://www.ghgprotocol.org

© 2010 BCS Data Centre Specialist Group http://dcsg.bcs.org Version 2.0.0 24 of 32
<table>
<thead>
<tr>
<th>Allocation Method</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per occupied rack</td>
<td>Provides a reasonable allocation of energy given that many data centres have relatively low per rack power limits</td>
</tr>
<tr>
<td></td>
<td>Does not require additional metering equipment to be installed</td>
</tr>
<tr>
<td></td>
<td>Reflects the fixed energy and cost overheads</td>
</tr>
<tr>
<td></td>
<td>Does not provide an incentive for power management</td>
</tr>
<tr>
<td>Per provisioned kW</td>
<td>Provides a reasonable allocation of energy which is more accurate than per occupied rack</td>
</tr>
<tr>
<td></td>
<td>Provides a direct incentive not to overprovision power and cooling capacity</td>
</tr>
<tr>
<td></td>
<td>Reflects the fixed energy and cost overheads</td>
</tr>
<tr>
<td></td>
<td>Does not create an incentive for power management</td>
</tr>
<tr>
<td>Per drawn kWh</td>
<td>Provides a reasonable allocation of energy, not significantly better than per provisioned kW</td>
</tr>
<tr>
<td></td>
<td>Does not reflect fixed energy and cost overheads</td>
</tr>
<tr>
<td></td>
<td>Does not provide an incentive not to overprovision capacity</td>
</tr>
<tr>
<td></td>
<td>Provides a direct incentive for power management</td>
</tr>
</tbody>
</table>

Table 1 IT energy allocation options

As shown above each of the methods has advantages and disadvantages, none are entirely accurate but they are reasonably fair. The per drawn kWh method is equivalent to multiplying the metered energy consumption of an IT device by the data centre PUE. Whilst the PUE metric is known to be non-reversible these methods are intended only to be useful estimates of the energy consumption. As the allocation models, metering and analytic capability matures more effective and accurate allocations will replace these estimates in many facilities.

Example allocation
An example allocation of data centre consumed energy to a set of installed IT service platforms is provided below. Table 2 shows the installed IT platforms and the recorded value for each of the allocation methods. The data centre energy is being allocated for one month.

<table>
<thead>
<tr>
<th>IT Equipment</th>
<th>Racks</th>
<th>Provisioned kW</th>
<th>Drawn kWh</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR platform</td>
<td>10</td>
<td>25</td>
<td>11,000</td>
</tr>
<tr>
<td>Storage platform</td>
<td>6</td>
<td>20</td>
<td>10,000</td>
</tr>
<tr>
<td>ecommerce platform</td>
<td>20</td>
<td>50</td>
<td>26,000</td>
</tr>
<tr>
<td>Network</td>
<td>10</td>
<td>30</td>
<td>16,000</td>
</tr>
<tr>
<td>Logistics platform</td>
<td>30</td>
<td>80</td>
<td>38,000</td>
</tr>
<tr>
<td>ERP platform</td>
<td>20</td>
<td>60</td>
<td>24,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>96</strong></td>
<td><strong>265</strong></td>
<td><strong>125,000</strong></td>
</tr>
</tbody>
</table>

Table 2 Installed IT platforms

The metering for the data centre reports that the total utility energy consumption was 250,000kWh over the month being allocated. This shows a PUE of 2 for the data centre over the metered period.
As shown in Table 3 there is some disparity between the allocated kWh to each platform between the proposed methods but the totals are consistent. The “true” allocable energy lies somewhere between the per provisioned kWh and per drawn kWh allocations as neither of these properly takes account of the fixed and variable energy overheads of the data centre.

Converting these consumed kWh values to kg CO₂ is trivial using a value for utility carbon intensity such as the 0.54 kg CO₂ per kWh of the UK grid.

<table>
<thead>
<tr>
<th>IT Equipment</th>
<th>Per rack</th>
<th>Per provisioned kW</th>
<th>Per drawn kWh</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR platform</td>
<td>26,042 kWh</td>
<td>23,585 kWh</td>
<td>21,368 kWh</td>
</tr>
<tr>
<td>Storage platform</td>
<td>15,625 kWh</td>
<td>18,868 kWh</td>
<td>19,943 kWh</td>
</tr>
<tr>
<td>ecommerce platform</td>
<td>52,083 kWh</td>
<td>47,170 kWh</td>
<td>53,419 kWh</td>
</tr>
<tr>
<td>Network</td>
<td>26,042 kWh</td>
<td>28,302 kWh</td>
<td>34,188 kWh</td>
</tr>
<tr>
<td>Logistics platform</td>
<td>78,125 kWh</td>
<td>75,472 kWh</td>
<td>74,074 kWh</td>
</tr>
<tr>
<td>ERP platform</td>
<td>52,083 kWh</td>
<td>56,604 kWh</td>
<td>47,009 kWh</td>
</tr>
<tr>
<td>Total</td>
<td>250,000 kWh</td>
<td>250,000 kWh</td>
<td>250,000 kWh</td>
</tr>
</tbody>
</table>

Table 3 Energy allocated to installed IT platforms

Issues for service providers
In a service provider data centre this means distributing the entire energy use and carbon of the data centre across the installed customers. It may be necessary to provide a short initial period over which the service provider operator is permitted to “sink” some of the energy after the first commissioning of the facility. Beyond this point sensible selection of infrastructure devices and modular deployment provide all the required flexibility and there is no further requirement for a safety valve.

6.6 Allocating IT equipment energy to software
The boundary between the IT equipment and the application software is defined by the transition from physical to logical systems, e.g. a physical server may be occupied, on the software side, by multiple operating system instances within a virtualisation environment.

At this boundary the full energy allocated to each piece of IT equipment, or groups of IT equipment where there are large, relatively homogeneous sets of equipment within the facility, should be applied to the software services delivered from the software layer. This may also be carried out at a group of IT devices level where the owner of the equipment is a single customer or department of the operator.

Figure 13 Allocating software kWh
Again, the completeness rule is the important element and it is largely up to the operator to determine the details of how they would divide up the consumed energy of a blade server chassis which houses hundreds of virtual servers.

**Shared infrastructure**
The energy consumption of any IT device which does not directly support an identifiable software activity, such as some network or storage devices should be cross-allocated to those devices which do.

**Example allocation**
Using the same example data centre we can now allocate the IT device energy to the software service platforms. In this simple example each platform had its own group of servers however both the network and storage platform are shared infrastructure. Specifically, the network and storage platforms do not provide any direct software service but support the other platforms by providing connectivity and data storage.

A method of allocation of the energy of the network and storage platform must be selected, in this case port count on the network and allocated raw space on the storage platform are used. It is up to the operator to determine on a per data centre, platform or service basis how best to represent the proportion due to each platform.

Using the per provisioned kW method the software level allocation is shown below in Table 5:

<table>
<thead>
<tr>
<th>Software</th>
<th>IT kWh</th>
<th>Share of network</th>
<th>Share of storage</th>
<th>Total kWh</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR platform</td>
<td>23,585</td>
<td>5,660</td>
<td>1,887</td>
<td>31,132 kWh</td>
</tr>
<tr>
<td>Storage platform</td>
<td>18,868</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecommerce platform</td>
<td>47,170</td>
<td>8,491</td>
<td>3,774</td>
<td>59,434 kWh</td>
</tr>
<tr>
<td>Network</td>
<td>28,302</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistics platform</td>
<td>75,472</td>
<td>8,491</td>
<td>5,660</td>
<td>89,623 kWh</td>
</tr>
<tr>
<td>ERP platform</td>
<td>56,604</td>
<td>5,660</td>
<td>7,547</td>
<td>69,811 kWh</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>250,000</strong></td>
<td></td>
<td></td>
<td><strong>250,000</strong></td>
</tr>
</tbody>
</table>

*Table 5 Energy allocated to software platforms*

**Software service energy implementation**
The software energy implementation follows the same pattern of complete forward accounting. The total energy use and cost should be allocated across the identifiable services or activities delivered from the data centre, be these user email accounts or large scale grid computing models of financial markets.

As for the software layer, infrastructure services which have no visible business service output such as authentication or backup should be cross allocated to services which are delivered to the service consumers.

![Figure 14 Allocating service kWh](image)

16 This is the actual disk capacity allocated not the usable capacity to account for different data redundancy schemes
6.7 Summary of the process
Whilst the overall process may seem complex, each step is localised, may be carried out by an independent team or organisation, and is relatively simple. Figure 15 shows a summary of the process.

- The existing incoming electrical energy measure for PUE is used, including any additional source energy equivalent such as gas or oil to provide the total kWh.
- The total kWh is allocated across the IT equipment based on the operators selected method and capability. The PUE and carbon intensity for a delivered IT kWh may also be calculated and reported at this point.
- The IT kWh for each IT device or platform are allocated to the software platforms.
- The software platform kWh are allocated to the services they deliver from the data centre.

**Accounting energy and cost in the data centre**

![Diagram of energy accounting in a data centre]

*Figure 15 Accounting energy in the data centre*
7. Benefits of the cost of services method

These rules are likely to be largely self tuning as any significant or structural unfairness is likely to be resisted by the next layer in the stack or the end consumer of the service. Unlike the complex overall efficiency metrics which may be easily gamed to improve the operators score, without any obvious disadvantaged party, gaming the allocation mechanism results in one or more service consumers receiving a larger allocation than they are responsible for which should create a restoring pressure.

These simple rules scale and flex to allow traversal of the departmental responsibility and contract boundaries which remain one of the most significant obstructions to improved data centre energy and cost efficiency.

7.1 Import and export of services

A major issue facing environmental policy in developed countries is the decoupling of economic growth from energy consumption or environmental impact in the form of greenhouse gas emissions. There are existing concerns that apparent reductions in the energy intensity of economies are in fact due to the exporting of energy and pollution intensive manufacturing processes to other countries and that the apparent improvement in GDP per Joule or kg CO$_2$ is simply due to importing goods instead of manufacturing them.

There are two basic instances where this type of issue will arise in ICT services both of which arise from the extreme portability of ICT services\(^\text{17}\). There are very few applications where the data centre needs to be in the same country as the consumer of the service;

1) Where a direct carbon tax or substantial regulatory burden is applied to energy consumption within states this is likely to drive “cloud” and outsourcing data centre operators to place new facilities outside this economic burden and import the delivered services over the Internet or private networks\(^\text{18}\). This will directly remove energy and greenhouse gas from both national totals and legislative control.

2) Where ICT is used to reduce the energy consumption or greenhouse gas emission of a physical process within a country if that service is delivered from outside the country then we will be unable to determine whether any net reduction has been achieved.

Clearly this presents a substantial issue, particularly if a state wishes to move to a carbon tax or reporting scheme based on consumption or scope 3 type allocation principles. ICT services will escape this taxation.

Intensity factors

The traditional approach to deal with this type of import or export issue is to apply a standard “intensity factor” of energy consumption or CO$_2$ emitted per unit of the commodity being sold or imported.

*The issue with applying intensity factors to ICT services is that there are no common measures of a unit of service.*

We do not have units to compare a satellite guided logistics system with a Hotmail account. The common fallback in this situation is to assume that the energy consumption of delivery is strongly related to the cost of the good or service which is quite true for many products. This is (fortunately) completely untrue for ICT services. Data centre derived services may only be considered energy intensive (where the supplier energy cost is the major part of the delivery cost) at the level of delivered power and space, the “IT kWh” on the stack diagram in Figure 7. Once the “value added” parts of the service such as management and the software are added, the energy cost becomes a small part of the price of the service.

\(^{17}\) Note that ICT services are more portable than manufacturing and that any displacement pressure is likely to be faster acting

\(^{18}\) There are already reports of this happening due to the UK Carbon Reduction Commitment
As a specific comparison we can compare the cost and energy of delivery of a commodity server from a provider as both Colocation, where the user purchases the server and the operator provides space, power and cooling and the same server, using the same energy if the user instead purchased a managed server which the operator owns and manages for them including the operating system and supporting the application software. This comparison is shown below in Table 6;

<table>
<thead>
<tr>
<th>Option</th>
<th>Monthly Energy</th>
<th>Monthly Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colo Server</td>
<td>250 kWh</td>
<td>Around £25</td>
</tr>
<tr>
<td>Managed Server</td>
<td>~ 250 kWh</td>
<td>Around £250 depending on service options</td>
</tr>
</tbody>
</table>

*Table 6 Cost versus energy consumption*

Whilst the operator will incur some additional energy cost in the personnel used to support the managed server this does not equate to 9 times the energy of operating the server as a single support engineer can manage hundreds of properly configured servers.

Clearly an intensity factor based on the price paid for an IT service would not a fair or reasonable approximation of the carbon of delivery and should not be used to tax or regulate the import, export or use of ICT services.
8. Further extension of the methods

These methods, which rely on simple accounting principles, are also relatively easy to extend to incorporate additional factors as this data becomes available to us.

8.1 Embodied carbon
At present there is very little data available on the embodied or disposal carbon of mechanical and electrical plant or IT devices. However, as this becomes available it may be easily incorporated into this scheme. The embodied carbon can be amortised over the lifetime of the equipment as we presently do for the cost and this amortised rate added to the energy supply carbon as the costs progress up the stack. The amortising M&E plant carbon would be added to the delivered IT kWh and the amortising IT device carbon added to the IT equipment carbon. The same completeness rules would apply and similar allocation regimes would function.

8.2 Support energy
One area of significant interest is the determination of the energy cost of the support and management functions, largely people and offices that develop and support data centre derived services. Again, these additional running or amortised energy or carbon functions can be added as the costs progress up the stack.
9. Document Information

Version History

<table>
<thead>
<tr>
<th>Version</th>
<th>Description</th>
<th>Version Updates</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0.0</td>
<td>Release</td>
<td>None</td>
<td>09 December 2010</td>
</tr>
<tr>
<td>1.1.5</td>
<td>Selected reviewers</td>
<td>Further review updates</td>
<td>08 December 2010</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Selected reviewers</td>
<td>Further content added to all sections, conclusion and summary replaced</td>
<td>29 November 2010</td>
</tr>
<tr>
<td>1.1.0</td>
<td>Internal review</td>
<td>Updated based on comments from Paul Latham and Steve Strutt</td>
<td>27 November 2010</td>
</tr>
<tr>
<td>1.0.0</td>
<td>DCSG members</td>
<td>None</td>
<td>22 October 2010</td>
</tr>
<tr>
<td>0.4.0</td>
<td>DCSG members</td>
<td>Added abstract and conclusion</td>
<td>22 September 2010</td>
</tr>
<tr>
<td>0.3.0</td>
<td>Selected reviewers</td>
<td>Paul Latham review comments</td>
<td>22 September 2010</td>
</tr>
<tr>
<td>0.2.0</td>
<td>Selected reviewers</td>
<td>Added 5.4.1</td>
<td>01 September 2010</td>
</tr>
<tr>
<td>0.1.1</td>
<td>Selected reviewers</td>
<td>Added 5.2.1</td>
<td>16 August 2010</td>
</tr>
<tr>
<td>0.1.0</td>
<td>Selected reviewers</td>
<td>Comments and changes cleared</td>
<td>06 August 2010</td>
</tr>
<tr>
<td>0.0.7</td>
<td>Internal review</td>
<td>Formatting updates</td>
<td>06 August 2010</td>
</tr>
<tr>
<td>0.0.6</td>
<td>Internal review</td>
<td>Additional content</td>
<td>06 August 2010</td>
</tr>
<tr>
<td>0.0.5</td>
<td>Internal review</td>
<td>Additional content</td>
<td>04 August 2010</td>
</tr>
<tr>
<td>0.0.4</td>
<td>Internal review</td>
<td>Paul Latham review comments</td>
<td>03 August 2010</td>
</tr>
<tr>
<td>0.0.3</td>
<td>Internal review</td>
<td>Additional content</td>
<td>02 August 2010</td>
</tr>
<tr>
<td>0.0.2</td>
<td>Internal review</td>
<td>Additional content</td>
<td>29 July 2010</td>
</tr>
<tr>
<td>0.0.1</td>
<td>Internal review</td>
<td>First version, no updates</td>
<td>28 July 2010</td>
</tr>
</tbody>
</table>

Release History

<table>
<thead>
<tr>
<th>Version</th>
<th>Description</th>
<th>Authoriser</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0.0</td>
<td>Public release</td>
<td>Liam Newcombe</td>
<td>09 December 2010</td>
</tr>
<tr>
<td>1.0.0</td>
<td>DCSG Members and selected reviewer release</td>
<td>Liam Newcombe</td>
<td>22 October 2010</td>
</tr>
</tbody>
</table>

---


iv See the “Data Centre Energy Efficiency Metrics” for a more detailed description of this problem

v For the UK this is published by DEFRA, [http://www.defra.gov.uk/environment/business/reporting/conversion-factors.htm](http://www.defra.gov.uk/environment/business/reporting/conversion-factors.htm), in the USA this is published by the EPA [http://www.epa.gov/cleanenergy/energy-resources/egrid/](http://www.epa.gov/cleanenergy/energy-resources/egrid/)


vii Price data supplied by Memset Ltd ([http://www.memset.com](http://www.memset.com))