

An Environmental Chargeback for Data Center and Cloud Computing Consumers

Edward Curry¹, Souleiman Hasan¹, Mark White², Hugh Melvin²

¹Digital Enterprise Research Institute (DERI),
IDA Business Park, Lower Dangan, Galway, Ireland
{ed.curry, souleiman.hasan}@deri.org

²Discipline of Information Technology, National University of Ireland, Galway,
University Road, Galway, Ireland
{m.whitel, hugh.melvin}@nuigalway.ie

Abstract. Government, business, and the general public increasingly agree that the polluter should pay. Carbon dioxide and environmental damage are considered viable chargeable commodities. The net effect of this for data center and cloud computing operators is that they should look to “chargeback” the environmental impacts of their services to the consuming end-users. An environmental chargeback model can have a positive effect on environmental impacts by linking consumers to the indirect impacts of their usage, facilitating clearer understanding of the impact of their actions. In this paper we motivate the need for environmental chargeback mechanisms. The environmental chargeback model is described including requirements, methodology for definition, and environmental impact allocation strategies. The paper details a proof-of-concept within an operational data center together with discussion on experiences gained and future research directions.

Keywords: Sustainability, Chargeback Model, Data Center, Energy Efficiency, Cloud Computing.

1 Introduction

Google estimates that to answer a single search requires 0.0003kWh of energy and generates the equivalent of about 0.2g of CO₂. A 1-minute YouTube stream requires 0.0002kWh of energy and generates approximately 0.1g of CO₂. A single Gmail user requires 2.2kWh every year, and generates 1.2kg of CO₂. In 2010 Google’s total carbon footprint was 1.46 million metric tons of CO₂¹. However, is Google solely responsible for these emissions or do the 1 billion users that consume Google’s services bear some responsibility? Do these users bear the responsibility equally (1.46 billion

¹ All figures published by Google at <http://www.google.com/green/bigpicture/> retrieved on 23rd April 2012. Figure are for data-center emissions only and do not include end-user footprint.

kg CO₂/1 billion users = 1.46 kg of CO₂ per user per year) or do some power users cause more emissions than occasional light users? Are the users aware of the environmental effects of their usage? Should consumers of data center-based cloud services be accountable for the emissions associated with their service usage?

Government, business, and the general public increasingly agree that the polluter should pay. Carbon dioxide and environmental damage are considered viable chargeable commodities. The net effect for data center and cloud computing operators is that they should look to “*chargeback*” the environmental impacts, in addition to the financial costs, of their services to the consuming end-users.

Chargebacks can have a positive effect on environmental impacts by linking consumers to the indirect impacts of their service usage, allowing them to understand the impact of their actions. In this paper we motivate the need for environmental chargeback mechanisms to inform consumers of their data center environmental impacts.

The rest of this paper is organized as follows: Section 2 discusses the need for environmental chargeback for data center consumers. The environmental chargeback model is introduced in section 3 with discussion on the requirements, a methodology for defining chargeback models, and environmental impact allocation. A proof-of-concept implementation of the model within an operational data center is presented in section 4. Section 5 outlines related work and section 6 concludes with proposals for future research directions.

2 The Need for an Environmental Chargeback for Consumers

In this section we examine the environmental impact of data centers and cloud computing. For the sake of brevity, discussion on environmental impacts is limited to impacts associated with electricity generation in the operational phase of a data center. The authors note that power consumption does not tell the full story of the impacts of data centers on the environment². However, we believe the approach purposed in this paper has the potential to be applied beyond power to include other impacts such as water, construction materials, facilities equipment, and IT equipment.

2.1 Data Centers and Cloud Computing Energy Impact Analysis

As corporate and home users move their IT services to the cloud, the growth of data center-based services is set to continue. Power consumption largely defines a data center’s environmental impact: The amount of power that a data center uses on a day-to-day basis determines how much irreplaceable fossil fuels it consumes and the

² In order to understand the full environmental burden a full Life Cycle Assessment (LCA) of the data center facilities and IT equipment is needed. Take for example Microsoft’s data center in Quincy, Washington that consumes 48 megawatts (enough power for 40,000 homes) of power. In addition to the concrete and steel used in the construction of the building the data center uses 4.8km of chillers piping, 965km of electrical wire, 92,900m² of drywall, and 1.5 metric tons of batteries for backup power. Each of these components has their own impact that must be analysed in detail.

quantity of carbon emissions for which it is responsible. In 2010 the total electricity use by data centers was estimated between 1.1% and 1.5% of all electricity use for the world. For the US, that number was between 1.7% and 2.2% [1].

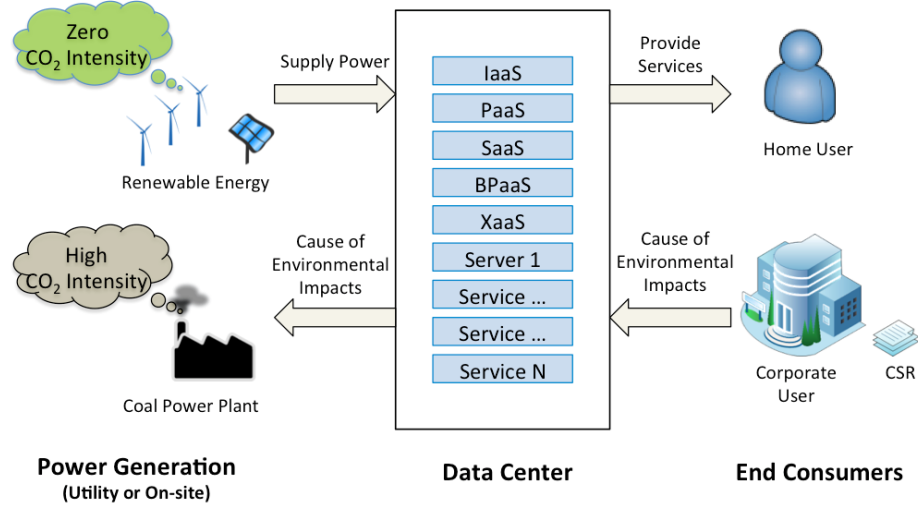


Fig. 1. Data Center Service Supply Chain

Within the supply chain of data center services, illustrated in Figure 1, the main emission occurs at the power generation site. Location is a key factor for the CO₂ intensity of the power consumed by the data center. A gas or coal fired power utility creates much more CO₂ than a hydro- or wind-power utility. For this reason, many new data centers (e.g. Google's) are located near low-cost and environmentally friendly power sources.

Data centers generate heat and must be cooled; the required equipment can be a significant consumer of power. Geographic location is also a key factor in cooling; a data center in a cool climate such as Ireland requires less cooling power than a data center in a warmer climate such as Mexico.

Based on these factors (and many others) the resulting CO₂ footprint of a data center can vary significantly, thus the execution of similar workloads in two different data centers can also vary.

2.2 Empowering the Consumer with Environmental Information

The principle of '*the polluter pays*' is gaining widespread acceptance among governments, business, and the general public. The end-users of data center services and their needs for IT services are the ultimate reason for the existence of the data center. However, very little information flows to the service consumer about the environmental impacts associated with their service execution. The result is that consumers are not well informed of the environmental consequences of their service usage, and thus have little opportunity to change their behavior to be more ecologically sound.

The challenge is how to tie environmental impacts back to the point of usage, so that the consumer can be better informed of their contribution to data center activity. The objective is to promote the reduction of environmental impacts by:

- **Raising Consumer Awareness of Environmental Impacts:** Improving sustainability performance requires information on the use, flows and destinies of energy, water, and materials including waste, along with monetary information on environment-related costs, earnings, and savings. This type of information is critical if we are to understand the causal relationships between the various actions that can be taken, and their impact on sustainable performance. Increased clarification will lead to consumers making more informed choices when choosing the services they use and the data centers providing those services.
- **Induce Efficient Usage of Data Center-Based Resources:** Consumers are concerned about the environmental impacts of their actions and will make environmentally friendly choices where possible. Studies have shown that improving access to information on consumption can reduce the overall usage of a resource (i.e. paper [2], energy [3][4]). It is reasonable to assume that if appropriate usage information were available for data center services, it would reduce usage. Empowering end-users to make sustainable choices requires them to know the environmental impacts of their action at the point of consumption, so they can make informed choices. Could the service be scheduled (invoked) when more renewable power sources are available? Could it be invoked less often?
- **Embed Service Usage within Sustainable IT Practices:** Corporate IT departments concerned with sustainability may have a sustainable IT program with the objective of reducing the environmental footprint of IT [5]. A chargeback would allow environmental impacts of IT service usage to be embedded within business and decision-making processes. It would enable IT departments to consider environmental impacts within the full life-cycle [6] of their cloud computing strategies.

3 An Environmental Chargeback Model for Consumers

Carbon dioxide and environmental damage are becoming more accepted as chargeable commodities. Determining how much environmental impact is being caused by an service end-user within a data center would make it possible to levy charges based on the impacts occurred, thus linking consumer activity with the environmental cost of the IT supporting it.

Pay for Use has long been a cornerstone of many business models (i.e. telephony, water, & waste). Pay for use models can increase awareness of the costs of resource usage and promote more efficient and selective usage, resulting in less waste and lower costs. To this end we propose the use of a pay for use “chargeback” model for environmental impacts associated with data center services and cloud computing.

The purpose of the chargeback model is to allocate the environmental impacts of providing data center services to the service consumers: that is, making the consumer accountable for the environmental impacts of their service usage. Developing a chargeback model with billing based on actual resource usage, instead of resource

allocation or reservation, is a fundamental requirement to encourage users to have more sustainable behaviors. An effective chargeback model should have the following benefits:

- Correlate service utilization with service consumers or corporate departments.
- Provide visibility into service and associated resource utilization.
- Enable consumers to understand their data center environmental footprint.
- Add transparency to sustainability of outsourced enterprise IT.
- Encourage the use of green power with lower environmental footprint.

3.1 Model Requirements

A chargeback model should meet the following requirements:

- **Equitable:** The consumer is only charged for the impacts they cause. One consumer should not subsidize the impacts of another consumer.
- **Accurate & Auditable:** Charge for actual impacts accurately and fully, and maintain records to handle inquiries and disputes.
- **Understandable:** Charging process & methodology must be comprehensible to consumers.
- **Controllable & Predictable:** Consumers must have the ability to control or predict the cost of performing a particular activity.
- **Flexible & Adaptable:** Ability to handle multiple service types (i.e. PaaS, IaaS, SaaS) and dynamic cost models (i.e. include capital impacts, operational impacts, and intermittent availability of renewables that can vary over time or by region).
- **Scalable:** Can handle small- and large-scale services.
- **Economical:** The model itself must be relatively inexpensive to design, implement, deploy, and run, including data collection, processing, and reporting to consumers.

3.2 Model Definition Methodology

We propose the following methodology to define environmental chargeback models:

Step 1. Identify service and define environmental system boundary: Identify the target service. Define the system boundary for environmental impacts of the model. Determine what type of information is needed to inform consumers and decision makers. Define the functional units that will be used (environmental impacts, energy efficiency, life span, cost per use, etc.).

Step 2. Identify the billable items and, for each one, identify the smallest unit available as a service to consumers: The goal of this step is to find a unit of measurement that makes it easy to aggregate and store billing data. The unit should also be an easily understood *charging unit* to the consumer.

- *Billable Service Items:* Resources for which consumers will be charged. These will be part of the IT service catalog, and consumers will be able to purchase these

items. Examples of billable service items include, servers, virtual machines, storage, email, search, etc.

- *Atomic Service Units*: The smallest possible unit of measurement and collection, for a billable item, that will be used for billing purposes. The consumer bill will typically contain information on how many atomic units of a resource were used.

Step 3. Identify, analyze, and document relevant environmental impacts: Determine service resource use and associated environmental impacts. The data must be related to the functional unit defined in Step 1 and includes all data related to environmental impacts (e.g. CO₂) within the system boundaries.

Step 4. Define an environmental cost allocation strategy for each billable service item: After the environmental impacts have been identified, the billable service items have been identified, and the atomic service units have been defined, it is possible to build one or more environmental cost allocation strategies. Building an allocation strategy requires associating impacts to billable service items that are offered to service consumers. Each billable service item can have different allocation method that can use fixed, variable, or mixed charging. In order to maximize the benefits from cost allocation, it is necessary for allocation to reflect actual usage.

Step 5. Identify, integrate, and deploy the tools necessary to collect data and to calculate the environmental chargeback: A chargeback model implementation, illustrated in Figure 2, will typically require environmental data collection, DC resource utilization, service workload, chargeback calculation, and billing & reporting. The collection tools will vary based on the service and the data center.

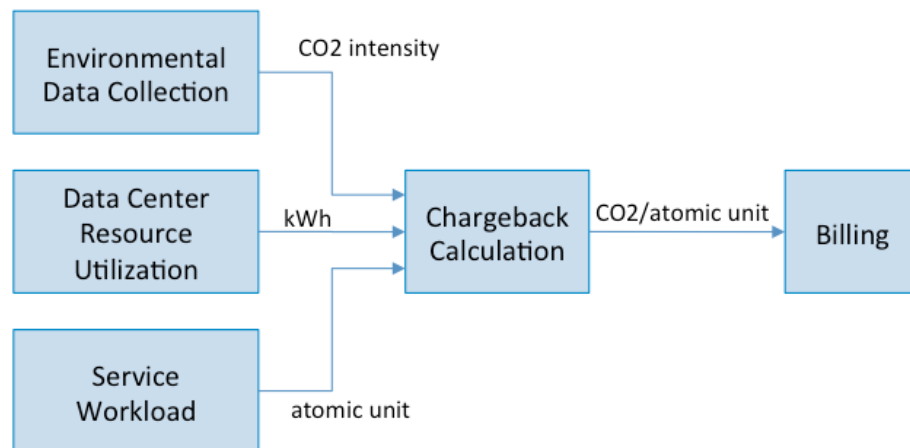


Fig. 2. Information workflow within environmental chargeback model

3.3 Allocating Environmental Impacts

To be able to determine the correct environmental impact allocation strategy it is necessary to know the direct and indirect costs of providing data center services. Sim-

ilar to financial costs, environmental costs can be broken into Capital (initial/setup) or Operational (ongoing/running).

- **Capital Impacts** include the impacts of building the data center facilities and the impacts of the associated IT (server, network, storage, racks, cabling, etc.) and facilities equipment (i.e. power and cooling infrastructure). For capital items, the impact needs to be amortized over the life of the item. Typically servers have a lifespan of 3 to 5 years, while data center facilities have a lifespan of 10 to 15 years. Capital impacts may also extend to software artifacts. For example, the cost of building a search index may be orders of magnitude more expensive than the costs of user searches against the index. Software artifacts could have a useful lifespan in days, week, or months. A chargeback model should reflect these costs in a fixed charge over the estimated useful life of the equipment/software.
- **Operational Impacts** include all environmental impacts for keeping the data center running. The primary operational impact is power generation and water for cooling. Operational impacts are more straightforward to allocate and can usually be allocated by usage, such as the energy costs of running a server.

Both of these types of impacts must be taken into account within a chargeback model. It is important to be pragmatic with respect to the complexity of the chargeback model and the environmental data available. Setting a realistic environmental system boundary for the chargeback is essential to keeping its implementation straightforward.

4 Proof of Concept Implementation

In order to validate the proposed approach for the chargeback model a proof of concept (PoC) has been realized for a service within the DERI data center. In this section we discuss how the PoC chargeback model was defined, and how the resulting information is presented to the consumer. The section ends with a discussion on insights gained from the PoC.

4.1 Model Definition

The chargeback model in the PoC was defined as:

Step 1: The target is a transaction-based data service. The environmental system boundary of the model will be the carbon dioxide emissions associated with power generation. The chargeback model will be applied to the carbon emissions that are produced as a result of using the data service. The functional units are CO₂ in grams (gCO₂), kilowatts (kW) and kilowatt-hours (kWh).

Step 2: *Billable Service Items:* User accounts for the data service. *Atomic Service Units:* Single data transaction executed on the service.

Step 3. The chargeback model was scoped to cover the emissions associated with power generation. The data service runs on 27 dedicated servers³, and the power supplied is from a mixture of fossil fuel power plants, and renewable energy sources (primarily wind). This results in a varying CO₂ intensity of the power based on the availability of renewable energy.

Step 4. Within the data service the computational workload of all transactions is similar, thus the energy cost between transactions is similar. This enabled us to treat transactions as equal from an impact allocation perspective. To quantify a single transaction we need to know 1) the energy consumed by the total service, 2) the CO₂ emissions of the energy, and 3) the total number of transactions. In order to establish this in real-time we utilize a sliding window-based calculation⁴. During the sliding-window the following data is gathered: *Total Service Energy*, *Number of Transactions Served*, and *CO₂ Intensity of Power Supply*. Impacts are allocated to the atomic service units according to the formula in equation 1.

$$\text{CO}_2 \text{ per Transaction} = \frac{\text{Total Service Energy} \times \text{CO}_2 \text{ Intensity}}{\text{Number of Transactions}} \quad (1)$$

Step 5. The implementation of the model leverages a number of existing monitoring infrastructures within the data center. These systems have been integrated together to deliver the chargeback model. The major sub-systems of the implementation are:

- **Real-Time Web Service for Power CO₂ Intensity**⁵ based on the energy source mix (renewables, gas, oil.) that is used by power utilities in Ireland.
- **Data Center Resource Energy Monitor:** Provides data on the power consumption of the data center hardware used by the data service. Power distribution units in the datacenter are equipped with electricity metering capability.
- **Data Service Workload Monitor:** Provides data on user workload of the service.
- **Chargeback Calculation:** The chargeback model is encoded as rules within a Complex Event Processing (CEP) [7] engine. The CEP engine constantly receives events from each of the above systems, allocates impacts in real-time, then forwards the charge to the billing system. Data interoperability between existing infrastructures is achieved within the Linked dataspace for Energy Intelligence (LEI).
- **Billing System:** Provides reporting on charges to consumers.

4.2 Chargeback in Action

Figure 3 illustrates a screenshot showing actual chargeback results from the PoC. This interface provides the service consumer with an overview of their service usage, together with information on the current cost of a service invocation. As the CO₂ inten-

³ Network and data storage devices were excluded due to insufficient metering.

⁴ A limitation of this approach is that it ignores transactions that may have been initiated prior to the start of the window and those that do not complete prior to the end of the window.

⁵ The Energy Research Center at UCD provides this service <http://erc.ucd.ie/realtimedata/>

sity varies with the proportion of renewable energy that is used for power generation, consumers are notified of the current proportion of renewables and forecasts for the predicted future availability. This can enable them to plan for greener use of the service. The interface details:

- Consumer Service Usage
 - Data transactions: Total number of transactions during billing period.
 - Cost per transaction: Average gCO₂ caused by single transaction during billing period.
- Current Service Cost
 - Current Cost of transaction in Wh, and a mean cost over the last 5 seconds.
 - Latest fuel mix for power generation.
 - Current cost of transaction in gCO₂.
- IT Infrastructure
 - Number of machines involved in service execution.
 - Total average acclimated power per machine during billing period.

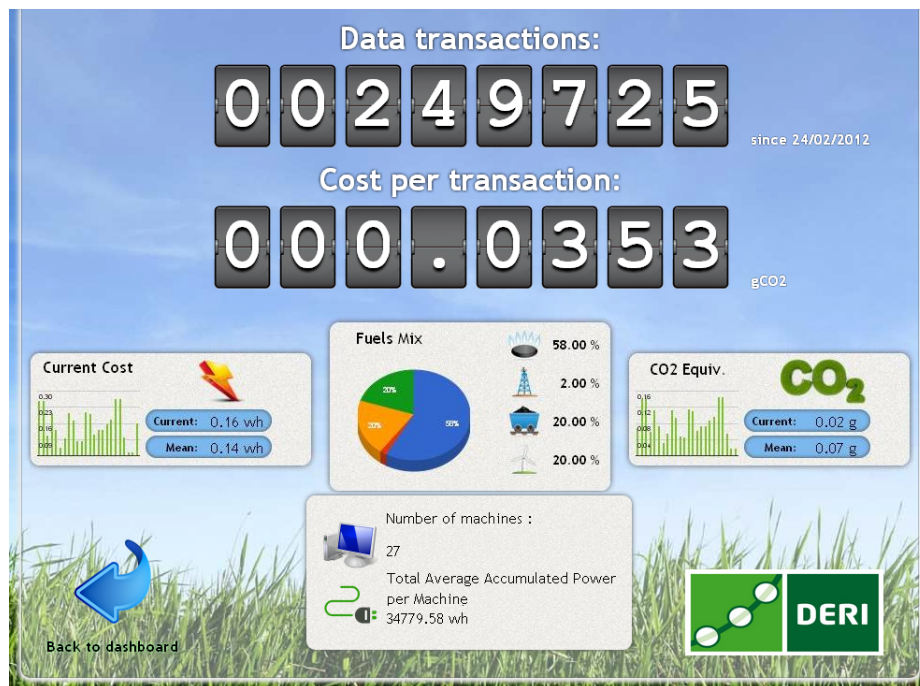


Fig 3. Chargeback results from DERI Energy

4.3 Experiences and Discussion

In the process of implementing the PoC, we gained a number of insights into chargeback models.

- **Metering and Monitoring:** The PoC piggy-backed on existing billing/monitoring infrastructure for the services. Where such infrastructure is already in place, the model implementation can be simplified by reusing it. Service-level monitoring will need to match the atomic service units identified. Where this is not available the service-monitoring infrastructure may need to be extended to support it.
- **Service & Infrastructure Complexity:** Defining the allocation strategy is dependent on the complexity of the service interaction model, the supporting IT infrastructure, and resource variation between atomic service units. The PoC had a straightforward IT deployment and atomic service units. However, if a service is delivered by complex infrastructure, that is shared and federated across multiple data centers, the implementation of the cost allocation mode will be more difficult.
- **Stakeholder Collaboration:** Co-operation between facilities and IT in the data center has been a challenge within the industry. Deploying chargeback models will require collaboration from more players, such as service managers and developers.
- **Security and Privacy:** While not a direct focus of this work, the authors acknowledge that security and privacy concerns can arise with chargeback models. These issues should be taken into consideration within the wider context of security and privacy for data centers and cloud computing.

5 Related Work

The model proposed within this paper is complementary to existing work on improving the sustainability of data centers and cloud computing. SLA@SOI [8] outlines an approach to service management by embedding Service-Level Agreement (SLA) aware infrastructure within services. It defines assessment criteria using a custom representation language. Energy can be considered as part of the SLA, however consumer-centric usage-based environmental reporting is not directly addressed.

The EU GAMES project [9] focuses on the improvement of IT service centers energy performance with respect to quality of service agreements during service composition. The energy efficiency improvement is based on a knowledge base of application level impacts on the IT service centers energy efficiency.

FIT4Green aims at contributing to ICT energy reducing efforts by creating an energy-aware layer of plug-ins for data center automation frameworks that will improve energy consumption. FIT4Green [10] looks to integrate all devices connected (including networking) with service delivery into optimization policies. ALL4Green, a follow-on to FIT4Green, aims to enable data centers, power suppliers, and end-users to communicate their expected supply and demand. This will allow ICT resources to be better allocated to provide requested services, while saving energy and reducing greenhouse gas (GHG) emissions.

5.1 Data Center Energy Efficiency Metrics

When assessing the financial health of a business, one should not look at one metric in isolation. The same is true for assessing the efficiency of a data center. In this section we will examine a number of key metrics defined by the Green Grid to understand the sustainability of a data center⁶.

Power Usage Effectiveness (PUE) is a measure of how efficiently a data center uses its power. PUE measures how much power the computing equipment consumes in contrast to cooling and other overheads. The reciprocal of PUE is *Data Center Infrastructure Efficiency (DCiE)*. Both PUE and DCiE metrics give an indication as to the use of power by supporting infrastructure of the data center. Ideally PUE would equal one, meaning no additional energy is consumed to support the IT load.

Water Usage Effectiveness (WUE) measures data center water usage to provide an assessment of the water used on-site for operation of the data center. This includes water used for humidification and water evaporated on-site for energy production or cooling of the data center and its support system. *Carbon Usage Effectiveness (CUE)* measures data center-level carbon emissions. CUE does not cover the emissions associated with the lifecycle of the equipment in the data center or the building itself.

The Data Center Productivity (DCP) framework is a collection of metrics that measure the consumption of a data center-related resource in terms of data center output. DCP looks to define what a data center accomplishes relative to what it consumes. *Data Center compute Efficiency (DCcE)* enables data center operators to determine the efficiency of compute resources. The metric makes it easier for data center operators to discover unused servers (both physical and virtual) and decommission or redeploy them.

All the above metrics focus on data center efficiency to help data center operators identify opportunities for efficiency improvements. While these metrics can inform the consumer that they are using an efficient data center, they do not inform the consumer of the cost of their service usage. They do not give consumers the information necessary to improve the sustainability of their behavior.

6 Conclusions and Future Work

In this paper we have motivated the need to link data center service and cloud consumers with the impacts of their service usage, allowing them to understand the environmental consequences of their actions. We proposed the need for environmental chargeback models including discussion on their requirements, definition methodology, and environmental impact allocation. A proof-of-concept implementation of the model within an operational data center is described together with experienced gained.

Future research directions will focus on a user evaluation to determine if chargeback models can effectively change user behavior and reduce the impact of services. For example, will users choose to use the service when more renewable energy is

⁶ White papers detailing all Green Grid metrics are available at <http://www.thegreengrid.org>

available, meaning less CO₂ emissions? Other avenues of investigation include challenges associated with deploying chargeback models within different data center environments (i.e. homogenous & heterogeneous), at large scale (i.e. warehouse size), and appropriate strategies for the allocation of capital environmental impacts.

Acknowledgements. We would like to acknowledge the constructive comments of reviewers. The work presented in this paper is funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion- 2).

References

1. Koomey, J.: Growth in Data center electricity use 2005 to 2010. Analytics Press, Oakland, CA (2011).
2. Medland, R.C.: Curbing paper wastage using flavoured feedback. Proceedings of OZCHI 2010 Design–Interaction–Participation. pp. 22 - 26. Association for Computing Machinery (ACM) Press (2010).
3. Abrahamse, W., Steg, L., Vlek, C., Rothengatter, T.: A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*. 25, 273-291 (2005).
4. Darby, S.: The effectiveness of feedback on energy consumption. A review for DEFRA of the literature on metering, billing and direct displays. Environmental Change Institute University of Oxford. 22, 1–21 (2006).
5. Donnellan, B., Sheridan, C., Curry, E.: A Capability Maturity Framework for Sustainable Information and Communication Technology. *IEEE IT Professional*. 13, 33-40 (2011).
6. Conway, G., Curry, E.: Managing Cloud Computing: A Life Cycle Approach. 2nd International Conference on Cloud Computing and Services Science (CLOSER 2012). , Porto (2012).
7. Hasan, S., Curry, E., Banduk, M., O’Riain, S.: Toward Situation Awareness for the Semantic Sensor Web: Complex Event Processing with Dynamic Linked Data Enrichment. 4th International Workshop on Semantic Sensor Networks 2011 (SSN11). pp. 69-81. , Bonn, Germany (2011).
8. Wieder, P., Butler, J.M., Theilmann, W., Yahyapour, R. eds: Service Level Agreements for Cloud Computing. Springer Berlin / Heidelberg (2011).
9. Bertoncini, M., Pernici, B., Salomie, I., Wesner, S.: GAMES: Green Active Management of Energy in IT Service Centers. In: Soffer, P., Proper, E., Aalst, W., Mylopoulos, J., Rosemann, M., Shaw, M.J., and Szyperski, C. (eds.) *Information Systems Evolution*. pp. 238-252. Springer Berlin Heidelberg (2011).
10. Basmadjian, R., Bunse, C., Georgiadou, V., Giuliani, G., Klingert, S., Lovasz, G., Majanen, M.: FIT4Green - Energy aware ICT Optimization Policies. COST Action IC0804 on Energy Efficiency in Large Scale Distributed Systems- 1st Year (2010).