ENVIRONMENTAL CONSTRUCTION OF BIG DATA INITIATIVES

A Research Paper Submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia – Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Cameron S. Dorsch

Fall, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: ___________CSD________________ Date: 10/19/2020
Cameron Dorsch

Approved: Date:11/30/20
Richard Jacques

In 2006, Microsoft announced the purchase of seventy acres of land in Quincy, Washington, a dry, hilly town known for its fields of apples and potatoes. The land would go on to contain thousands of server modules for Microsoft's internet services, and despite officials' announcement that the new center would be "environmentally sensitive", relied of 40 diesel engines for its backup power, despite its proximity to a nearby elementary school. Despite receiving half the national industrial average electricity rate, when the location was faced with a $210,000 fine for over-estimating electricity use, it wastefully burned millions of watts of electricity to reduce the discrepancy, until the fines were reduced. As local farmer Randall Allred said to New York Times, "For a company of that size and nature, with all the "green" things they advertised to me, that was an insult." (Glanz, 2012, p. 1). Even when only considering their direct, localized effects, it is clear the expansion of Big Data and large-scale information communication technology (ICT) systems can come with risks, and that the identity of new digital technology initiatives is often deceptive to those who are directly affected by their materialization.

INTRODUCTION

Digital innovation stands at a crossroads between sustainability and waste, often being pitched as the cutting edge in sustainability, yet relying on power consumption habits which threaten long term net environmental impact. Since economic trends presented by industry groups such as Data Center Dynamics show the demand for data services has skyrocketed in recent years and shows no sign of slowing (Carlini, 2018, p. 1), technology companies are being called into question on how their infrastructure is affecting the environment, and more recently, how their data markets have led to higher order affects, such as those explored in life cycle assessments summarized by Pohl et al. Despite the increased focus of scholars on creating better models to gauge the net environmental impact of large-scale data technologies, most models only reaffirm the difficulties of consistently measuring higher order environmental impacts which could be associated with the technologies (Pohl et al, 2018, p. 5). The researchers, who reviewed all relevant academic work on these assessments for ICT since 2005 and analyzed twenty-five case studies, concluded that the net environmental impact of ICT remains relatively unknown. Additionally, their literature review revealed that significant research gaps exist in both the breadth of specific ICT products examined and the roll of certain higher order effects, such as user and behavioral effects.

Despite the uncertainty of their environmental impacts, data initiatives are widely praised for the aid in track sustainability efforts and overall consumption across the globe. In 2014, the United Nations identified Big Data as their number one tool to control pollution and fight climate change, establishing an advisory group to oversee its application (United Nations Independent Expert Advisory Group, 2014). As researchers Hazas, M. and Nathan, L. present in their book Digital Technology and Sustainability: *Engaging the Paradox*, sustainability and software development have been tied since the latter's birth, and most technology companies today actively aim to work under sustainable principles. However, the pro-environmental sentiments of technology companies are often at odds with the higher order effects of their products and services: increased digital consumerism, electricity usage, and ICT infrastructure (Hazas & Nathan, 2017, p. 13). The implementation of large-scale ICT technologies often has unforeseen material consequences, as explored by anthropologist Vonderau, A. in his analysis of the cultural and political effects of the establishment of a Facebook data center in a small Swedish city. These consequences often vary in scope and predictability, but ultimately, directly affect the lives of those localized to the technology's materialization.

While Vonderau was able to reasonably gauge the social and cultural impacts of the data center’s implementation due to the economic and political trends that followed, gauging the environmental impact proves a harder task, since a technology’s damage or change to the environment may take many years to manifest (Vonderau, 2019). Specific implementations such as Quincy, Washington can show the potential local dangers, but their true net environmental impact can only be gauged on a global scale. In addition to the challenges of measuring the long-term effects of a single case study, it is hard to measure the overall environmental impact without considering the technology as whole, in all its implementations. The increased deployment of large data centers is the manifestation of millions of online users who have contributed to the demand of data services, willing or unwillingly. This paper, synthesized from research gathered on the environmental assessment of data initiatives and the role of data initiatives in sustainability efforts, will analyze how the conflicting relationship between data initiatives and sustainability efforts affects large-scale ICT’s materialization and consequences to the individuals localized to its physical implementation. Using Social Construction of Technology (SCOT), the implementation of large-scale ICT will be analyzed in how it is influenced by three main groups: data scholars and academics, ICT service providers, and passive users (Bijker, Bönig, & Oost, 1984). How each of these groups may contribute to the overall environmental impact of large-scale data technologies will be considered, with the goal of identifying how a more responsible, unified identity for the technologies can be achieved, given their unknown environmental impact.

ACADEMIC EFFORTS OF ENVIRONMENTAL ASSESSMENT

Despite their inability to yield a specific estimate of all environmental impacts of large-scale ICT initiatives, academic models establish strong methodologies for identifying various sources of environmental impact and their seriousness (Pohl et al, 2018). To better understand how the various groups involved contribute to the environmental impacts, the most prevalent effects studied in the LCA of ICT initiatives will be considered within the framework of the three established groups. In addition to identifying how the environmental effects are influenced by each group, each effect's influence over the perception of ICT initiatives will be considered to aid in establishing the environmental identity of ICT initiatives, and if that identity lacks appropriate context given the technology's potential risks.

PREVALENT EFFECTS IN LIFE CYLCE ASSESSMENT

LCA is a methodology which aims to establish any environmental impact accrued over the entire creation and lifespan of a product or service. The first order effects of large-scale ICT can be considered using LCA much like for a traditional manufactured product. As researchers Pohl et al. described, "first order effects describe the effort required to provide an ICT-based service." The most impactful of these effects are the raw materials and electricity use associated with the service. In 2018, data centers accounted for approximately three percent of national annual electricity sales, or approximately 70 billion kWh (Shehabi, 2018). Additionally, the Department of Energy estimated data centers total water usage to surpass 660 billion liters per year in 2020, mostly as a result of indirect water used in the creation of electricity (Shehabi et al, 2016). As shown in a LCA of a large Chicago data center done by HSK, data centers can also accrue a significant carbon footprint just from their building materials and construction costs. The report considered the repurposing of an old building to avoid the carbon costs embodied in the construction and found that it resulted in an 88% decrease in carbon emissions. It was added regarding total impact that, "The elimination of the embodied carbon associated with building reuse is equivalent to the operational carbon of 200 server cabinets operating in a year," so the net impact of implementation is approximately one year's use for a small data center (Romero, 2020). These are all significant environmental hazards presented by large-scale ICT's continued growth and increased deployment but are relatively predictable and measurable.

On the other hand, higher order effects are much harder to integrate accurately into ICT's assessment, as these effects are all the intended and unintended effects of the services implementation and use (Pohl et al, 2018). Higher order effects can be further classified, however this varies between individual assessments, so a selection of prevalent effects will be reviewed. A shown in the figure, higher order effects have varying scopes of impact, and therefore vary in their potential net energy contribution (Horner, Shehabi, & Azevedo, 2016). For interpreting the most relevant higher order affects in the context of the construction of large-scale ICT, mainly the larger scope effects will be considered, because of their maximized potential impact.

Researchers Horner, Shehabi, and Azevedo also present a brief summary of these common effects, beginning with the basic efficiency and substitution effects. Efficiency effects

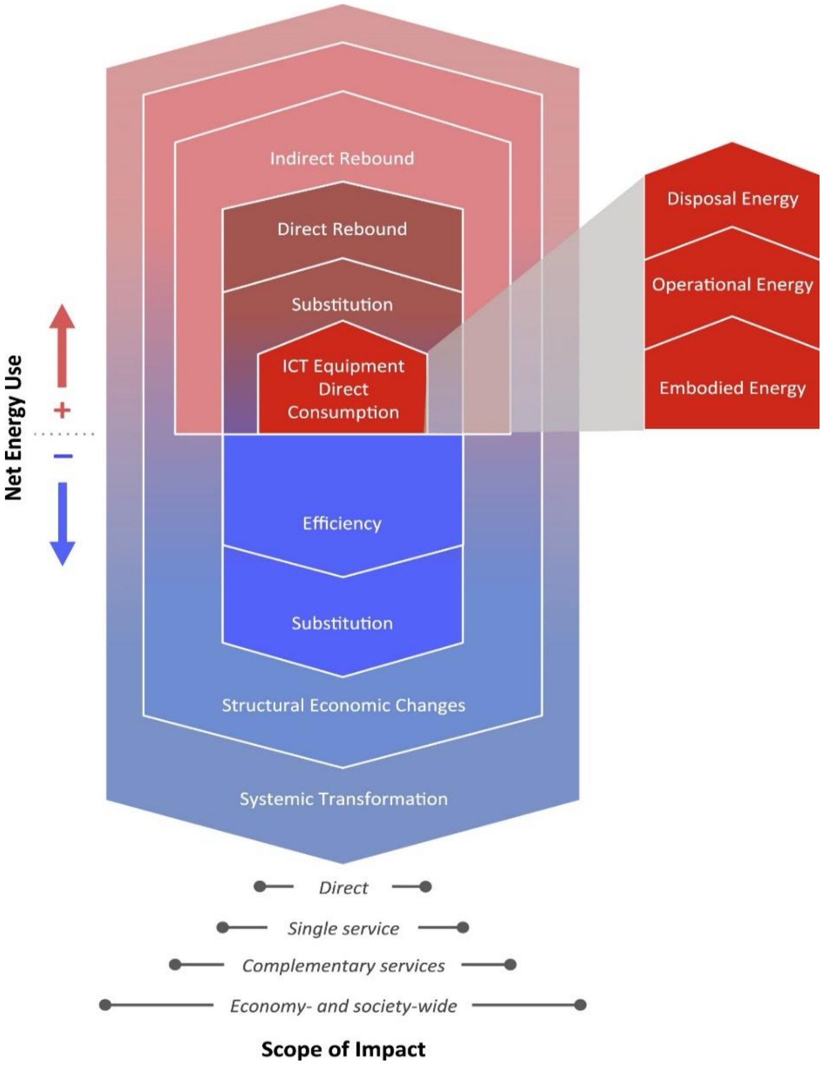

Figure 1: Effects of ICT. Graphical depiction of ICT effects order and scope. (Horner, Shehabi, & Azevedo, 2016)

are the increase in sustainability when existing systems are improved by a technology, and substitution effects are the change in environmental impact when a service is replaced by a technology. The latter could either result in a positive or negative impact, as shown in the figure. When the scope is expanded to a collection of services, efficiency improvements can lead to rebound effects. Author Azevedo, I., a researcher of rebound effects for ICT, explains that, "Efficiency measures that reduce energy service costs will free up resources that can be spent in the form of increased consumption—either of that same good or service or of other goods and services that require energy (and that have associated emissions)." When it is that same good or service, it is considered a direct rebound effect, and otherwise it is indirect (Azevedo, 2014). Structural economic effects simply expand this idea to the how the increase in efficiency as a result of ICT may affect economic sectors or industries. For example, Horner et al show how ICT has revamped the logistics industry and changed its carbon footprint; Freight sales have increased, and distribution centers modernized thanks to the implementation of ICT, increasing trucking and overall environmental impact. Finally, transformational effects are ways in which the technology may shift "human preferences and economic and social institutions" on a global scale (Horner, Shehabi, & Azevedo, 2016). The most obvious example of this is the societal change brought about by effortless teleworking capabilities and any associated environmental impact, but these effects are generally hard to predict and measure.

This range of effects will be used to show how the various groups may contribute to different levels of environmental impact associated with large-scale ICT. As the scope of an effect increases, our ability to measure it accurately inherently decreases due to the scale (Azevedo, 2014). Since the overall net impact is unknown, the maximum potential impact regarding each effect will be considered in order to cover all possible environmental risks. In order to understand how these effects may or may not play a role in the identity of large-scale ICT, popular environmental sentiments of Big Data will also be reviewed first to establish background.

BIG DATA'S ROLE AND IDENTITY IN SUSTAINABILITY

Large scale ICT's social and political identity regarding its environmental impacts is largely dominated by its praise as the new frontier in environmental protection efforts. In addition to the United Nations identifying Big Data as the newest tool for sustainability, it has launched several initiatives around the globe with the aid of data services. One of these is Pulse Lab Kampala, which established data services across diverse regions of Uganda to better gauge community sentiment about the country and its development. In addition to political efforts, many large ICT companies boast sustainability data initiatives, such as Amazon's open source sustainability data which offers various compiled environmental metrics from over the world, hosted on their website. Specific research by Dubey et al also explores the effectiveness of Big Data Prediction Analytics in increasing social and environmental sustainability of individual firms, concluding that there is a strong correlation between their use and sustainability, regardless of the firm's perceived flexibility to the technology (Dubey et al, 2017). The application of ICT to sustainability has had a measured impact in our ability to measure use and emissions and is widely perceived as an eco-friendly technology as a result. However, important hazards and unknowns in the environmental impact of Big Data are less often discussed, biasing the overall identity of large-scale ICT and its providers.

Pro-environmental bias contributes to the identity of many new digital technologies, especially since their materialization can be so disconnected from their users, such as with cloud services. Researcher Lucivero summarizes this disconnect in her academic paper Big Data Big Waste below.

> …using the term "cloud" to refer to computing and internet networks is a misleading metaphor. It suggests something impalpable, fluffy, untouchable, light, and transparent. This language strategically obscures the materiality of the infrastructure as well as its geographical presence and environmental impact. Cloud computing is in fact a highly tangible and touchable assemblage of material and heavy stuff. The material substrate of the cloud is made of cables, wires, servers, and shelves in buildings in every corner of the globe. Similarly, the language used to refer to data as an "unlimited and superabundant resource" implicitly suggests that data are virtual goods, always present, an ever increasing and never-ending resource, in contrast to other resources (such as oil, water, or land)… (Lucivero, 2019)

Lucivero establishes in her research not only how the materiality of Big Data can be hidden from its users, but that this characteristic of decentralized data services can be intentionally deployed to bias the perception of Big Data. Consequences of the implementation of ICT services, such as data centers in Quincy, Washington or Chicago, Illinois, are relatively unknown to users because one service is not tied directly to a single materialization. A decentralized ICT service is the result of every implementation the provider may have, and this significantly decreases perceived personal responsibility. Experimental research by psychologists Murtagh et al has also shown that the perceived automation increase with ICT can inhibit even simple actions for sustainability. When participants were led to believe the light system of a room may be automated or "smart", they were much less likely turn off the lights when the room was being left, despite the little effort required (Murtagh et al, 2015).

The disconnect users have from the physical impacts of ICT technologies only make it easier for users to perceive the technology as pro-environmental. While its physical consequences are hidden across the world in individual implementations, potential physical upsides are displayed in single data initiatives aimed at sustainability. Previously mentioned anthropologist Vonderau, who researched Sweden's transformation under Facebook's influence, also described the "invisibility" of ICT in his opening line, "Popular representations imagine the internet as being immaterial and fluid; hidden from the public eye are the industry and complex infrastructure securing the functionality of the World Wide Web, as well as this industry's social and environmental effects." (Vonderau, 2019, p. 1). Knowledge of these potential biases will aid in establishing responsibility for environmental impacts associated with large-scale ICT and help gauge awareness of the risks for each group in ICT's construction.

ENVIRONMENTAL MANIFESTATION OF BIG DATA'S CONSTRUCTION

As seen in the research presented by Lucivero and Horner et al, users of ICT technologies, especially those deployed on massive scale, often go uninformed with respect to the physical consequences of their consumption. Except for the tracking of user demand by ICT providers, ICT providers, users, and academics widely do not influence each other in the construction of Big Data, hurting overall awareness and education regarding the risks of the services. Each of these groups can significantly improve specific environmental impacts their work or behavior may have if communication is increased, bringing Big Data's social and political identity closer to its physical manifestations and real potential hazards. Using researched and established orders of environmental affect, each individual group's current involvement and ability to improve environmental awareness can be gauged in order to shape future implementation of large-scale ICT.

Providers have direct control over the simplest and most measurable effects of their services, the direct environmental consumption. However, as Horner et al found, “direct energy use is likely the simplest and least important ICT energy effect. The indirect energy effects are likely to be of much greater magnitude, owing to the breadth of the various mechanisms by which ICT services alter energy use.” (Horner, Shehabi, & Azevedo, 2016). Many providers also establish initiatives to directly improve the efficiency effects of their product. As established in Haza’s and Nathan’s book Digital Technology and Sustainability: *Engaging the Paradox,* this is generally one of two schools of thought for sustainable software engineering; if the product is created as efficiently and precisely as possible, its applications will provide the most environmental benefits possible (Hazas & Nathan, 2017). Despite their effort to improve the effects of their consumption and efficiency, providers rarely engage with higher orders of effect. As proposed by Lucivero in Big Data Big Waste, the providers can even strategically benefit off these negative environmental effects, such as the indirect rebound of digital consumerism (Lucivero, 2019). Without cooperation from the providers in gauging the magnitude and likelihood of high order effects, academics are unlikely to gain accurate predictions of the net environmental impacts of Big Data. This cooperation could help solidify the identity of large-scale ICT in the context of its environmental risks if the risks could be determined accurately. Aid in tracking rebound and structural effects could also benefit users in modifying behavior to reduce net environmental impact. As the automation research of Murtagh et al shows, individual users must be more informed in how their use patterns are contributing to environmental waste, especially for ICT products which are hidden from physical interaction. Only with the aid of the providers can it be identified how much change in user patterns as a result of the increase in large-scale in ICT contributes to the environmental impacts.

Users and academics are left relatively excluded from participation in the physical construction of the technology but provide context for understanding the physical impacts of their implementation. As predicted by McKinsey Global Institute in 2011, Big Data has revolutionized the economy and changed the way its users interact with the world (Manyika, 2011). Additionally, the individual user patterns of ICT determine its rebound effects greatly, as established by Pohl et al, and these user patterns are generally the hardest rebound effects to incorporate in assessment of ICT technologies. Since the LCA of large-scale ICT is inaccurate and requires further research, providers and users must work with academics in order to create better tools for gauging use patterns and their effects. Only once these effects are measured can providers inform users on how their consumption may manifest in the form of environmental impact. Increased communication between these groups is vital to improving the construction of Big Data initiatives. If all three groups can aid each other in establishing a more accurate idea of the physical manifestation of large-scale ICT, an identity for Big Data which is more productive to the reduction of net environmental impact can be established.

WORKS CITED

Azevedo, I. (2014). Consumer End-Use Energy Efficiency and Rebound Effects. *Annual Review of Environment and Resources, 39(1),* 393-418. doi:10.1146/annurev-environ-021913-153558

Bijker, W., Bönig, J., Oost, E. (1984). The social construction of technological artefacts. *Journal for Science Research:* 3, 39-51. Retrieved from https://cris.maastrichtuniversity.nl/en/publications/the-social-construction-of-technological-artefacts-problems-and-p

Carlini, S. (2018, December 12). Data center predictions for 2019. *Data Center Dynamics*. Retrieved from https://www.datacenterdynamics.com/

Dubey, R., Gunasekaran, A., Childe, S.J., Papadopoulus, T., Luo, Z., Wamba, S.F., Roubaud, D. (2017). Can big data and predictive analytics improve social and environmental sustainability? *Technological Forecasting and Social Change, 144(1),* 534-545. doi:10.1016/j.techfore.2017.06.020

Glanz, J. (2012, September 23). Data barns in a farm town, gobbling power and flexing muscle. *The New York Times.* Retrieved from https://www.nytimes.com/

Hazas, M., Nathan, L. (2017). *Digital Technology and Sustainability: Engaging the Paradox.* London, England: Routledge.

Hogan, M. (2018). Big data ecologies. *Ephemera: Theory and Politics in Organization, 18(3),* 631-657. Retrieved from http://ephemerajournal.org/

Horner, N.C., Shehabi, A., Azevedo, I. (2016). Known unknowns: indirect energy effects of information and communication technology. *Environmental Research Letters, 11(10),* 1.

Retrieved from https://iopscience-iop-org.proxy01.its.virginia.edu/article/10.1088/1748-9326/11/10/103001

Lucivero, F. (2019). Big data, big waste? A reflection on the environmental sustainability of big data initiatives. *Science & Engineering Ethics, 26*(2), 1009-1030. Retrieved from https://link-springer-com.proxy01.its.virginia.edu/article/10.1007/s11948-019-00171-7

Manyika, J. (2011). *Big Data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute.

Morley, J., Widdicks, K., Hazas, M. (2018). Digitalisation, energy and data demand: The impact of Internet traffic on overall and peak electricity consumption. *Energy Research & Social Science, 38(1),*128-137. doi:10.1016/j.erss.2018.01.018

Murtagh, N., Gatersleben, B., Cowen, L., Uzzell, D. (2015) Does perception of automation undermine pro-environmental behavior? Findings from three everyday settings. *Journal of Environmental Psychology, 42(1),* 139-148. doi:10.1016/j.jenvp.2015.04.002

Pohl, J., Hilty, H.M., Finkbeiner, M. (2018). How LCA contributes to the environmental assessment of higher order effects of ICT application: A review of different approaches. *Journal of Cleaner Productions,* 219, 698-712. doi:10.1016/j.jclepro.2019.02.018.

Romero, A. (2020, June 22). The cleanest data centers are ones that aren't built at all. *Server Farm*. Retrieved from https://www.serverfarmllc.com/

Shehabi, A., Smith, S.J., Masanet, E., Koomey, J. (2018). Data center growth in the United States: decoupling the demand for services from electricity use. *Environmental Research Letters*, 13(12), 1. Retrieved from https://iopscience-iop-org.proxy01.its.virginia.edu/article/10.1088/1748-9326/aaec9c#erlaaec9cs4

Shehabi, A., Smith, S.J., Sartor, D.A., Brown, R.E., Herrlin, M., … Lintner, W. (2016) United States data center energy usage report. *U.S. Department of Energy*. Retrieved from https://eta-publications.lbl.gov/

United Nations Independent Expert Advisory Group (IEAG) on a Data Revolution for Sustainable Development (2014). *A world that counts: mobilising the data revolution for sustainable development.* United Nations. Retrieved from https://www.undatarevolution.org/report/

United States Environmental Protection Agency. (2019). *Greenhouse Gas Equivalencies Calculator* (20) [Data Set]. EPA. Retrieved from https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator

Vonderau, A. (2019). Scaling the cloud: making state and infrastructure in Sweden. *Journal of Anthropology, 84(4),* 698-718. doi:10.1080/00141844.2018.1471513