

Coal Computing

How Companies
Misunderstand Their
Dirty Data Centers

Coal Computing: How Companies Misunderstand Their Dirty Data Centers

Ory Zik, VP of Analytics (ory.zik@luxresearchinc.com)

Avi Shapiro, Data Scientist (avi.shapiro@luxresearchinc.com)

Lux Research, Inc.

Despite their sophistication about data in other contexts, leading IT and computing giants in the U.S. use crude and outdated information to calculate the carbon footprint of their data centers, missing the real picture of their emissions by about 25%. Data centers comprise the fastest growing energy buying sector, and the companies that run them have the most advanced data analytics tools at their disposal, as well as high-minded public commitments to sustainability. They should lead rather than lag, by using more accurate data to report on their emissions – and to inform the actions they take to reduce them.

Executive Summary

Each year, the data centers that power social media, streaming video, cloud computing, and connected devices use 91 billion kilowatt-hours of electricity – enough to power New York City twice over – and their consumption is still growing rapidly. While companies like Google, Amazon, Facebook, and Apple profess sustainability goals and take measures to improve their energy profile, today they rely on the U.S. Environmental Protection Agency (EPA) Emissions & Generation Resource Integrated Database (eGRID) to estimate their emissions. However, eGRID divides the U.S. electricity grid into just 24 broad regions, and is updated only infrequently – the most recent information available is from 2012. Our analysis, using modern statistical tools now becoming available, shows that companies are miscalculating their emissions by about 25% on average.

The new Lux Grid Network Analysis (GNA) divides the grid into 134 regions, instead of just 24, providing more granular insight, and makes use of U.S. Energy Information Administration (EIA) data that is updated monthly, as opposed to three-year-old annual data. Applying the Lux GNA to U.S.-based data centers shows where operators are coming up short in their sustainability reporting:

- **Google misses the mark in four of out of its seven data centers.** Google uses eGRID to estimate its electricity emissions, but the Lux GNA model shows that four of Google's seven major U.S. data centers rely significantly more on coal than data reported by eGRID imply. As a result, Google's emissions are likely larger than estimated by 42,000 MT CO_{2e} per year – the equivalent of 8,500 additional SUVs on the road.
- **Amazon estimates are off in over 20 centers.** Amazon is less transparent than Google about how it calculates its emissions, but the Virginia electricity grid to which 23 of about 40 global Amazon cloud services data centers are connected uses about 43% electricity from coal based on the Lux GNA – not 35% as reported using eGRID.¹ As a result, Amazon's data centers are putting out 85,000 MT CO_{2e} per year more than inferred using eGRID – some 5,000 households' worth of emissions.

There is a stark contrast between the sophistication used by IT companies for monetized activities and their environmental impact analysis. While the former is done with location-based, real-time accuracy, the latter is disappointingly simplistic, based on coarse geographical estimations and with data three years out of date. As emissions become more critical to the planet, to regulators, and to the bottom line, it's time for companies to make a more rigorous reckoning of their usage and take action accordingly.

Data Centers are Big Energy Users – But the Impact Is Poorly Understood

As social media, streaming video, cloud computing, and ever-more-connected devices become a bigger part of our lives, the massive data centers that technology titans like Google, Amazon, and Facebook use to process and store the data are becoming a bigger part of our power grid. The [1,520 data centers in the U.S.](#) consume 91 billion kilowatt-hours of electricity per year, and the [National Resource Defense Council \(NRDC\) projects](#) they're on track to reach 140 billion kilowatt-hours by 2020. Data centers use enough electricity annually to power all the households in New York City for two years, and they pay more than \$10 billion for electricity each year, accounting for about 2.5% of total electricity revenues.

Data centers' thirst for power also makes them a major contributor to greenhouse gas (GHG) emissions, accounting for about 70 million metric tons of CO₂ per year. As global worries about climate change and the resulting policy response – from President Obama's Clean Power Plan to the climate change accords brokered at [COP21](#) – become increasingly serious, players like [Google](#), [Amazon](#), [Facebook](#), and more tout the efforts they're taking to make their data centers more sustainable. However, in a striking irony, the most data-savvy companies in the world are relying on surprisingly crude data about their energy sources in making these vital decisions about the climate impact of their activities.

The most data-savvy companies in the world are relying on surprisingly crude data to make these vital decisions about their climate impact.

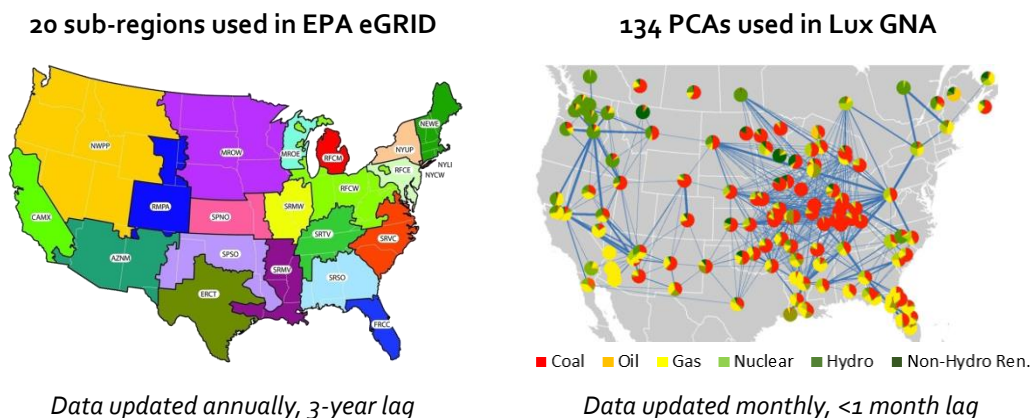
To be sure, the challenge isn't easy. Power generation is one of the most complex sectors of the economy, and [in the U.S. there are](#) nearly 20,000 power plants and 360,000 miles of transmission lines. What's more, once the electricity is put on to the grid, it can no longer effectively be traced back to its original source. However, the data source companies use today, the U.S. Environmental Protection Agency (EPA) [Emissions & Generation Resource Integrated Database \(eGRID\)](#) is not designed for the challenge of tackling this complex system. The eGRID construct dates to the 1990s and divides the U.S. electricity grid into just 20 broad sub-regions in the continental U.S. (24 total including Alaska and Hawaii). It also ignores electricity transfers with Canada and Mexico, and is updated only infrequently – the [latest version of eGRID](#) uses data from 2012. This white paper describes the new Lux Grid Network Analysis (GNA) model from Lux Research that allows companies to assess their energy usage and carbon footprint with much more granular and up-to-date information – and make better decisions about their environmental impact as a result. This GNA model is a part of Lux's comprehensive [Resource Intelligence for System Knowledge \(RISK\)](#) platform for analyzing resource usage.

Why Should Energy Buyers Care?

The question of eGRID's limitations might seem a bit academic, but the consequences are very real. The broad regions in eGRID obscure significant differences between the power sources and emissions levels for different sites within those regions. Meanwhile, the utility sector is currently undergoing huge changes, as both large-scale renewable installations and distributed generation from rooftop solar gain traction, meaning that outdated information is also likely to be misleading. For data center energy buyers and operators, unreliable data matters for:

- **Sustainability.** Theoretically, the most environmentally friendly solution is going totally "off grid" with renewable power. However, most renewable energy solutions are intermittent, and data centers depend on reliable power. Being connected to the grid and its volatile mix of energy sources means that companies need to have a good understanding of what they're buying in order to control the sustainability of their operations.

Fig. 1: Comparing the eGRID Framework Used Today with the New Lux Grid Network Analysis



- **Policymaking.** In the U.S. and around the world, public officials from the national to local levels are trying to make better energy decisions that promote the goals of public health, fighting climate change, and ensuring security and reliability of energy supply. Having a clear view of where the power is coming from – in particular for some of the largest users like data centers – is critical for choosing policies that will have the biggest impact on those goals.
- **Business decisions.** The shifting energy mix and consequent changes in the utilities business model lead to fluctuating contract prices and multiple distributed generation choices. In many markets, buyers can choose from a range of electricity providers, too, allowing customers to negotiate supply with local utilities as well as distributed generation vendors. However, doing so without accurate information leaves companies vulnerable to making costly decisions about energy sources.

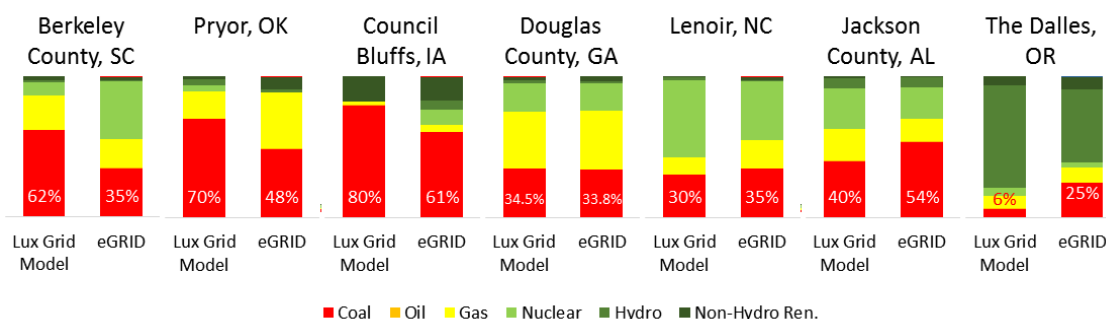
Better Data and Analytics Provide an Actionable Solution

To give energy users a more realistic view of their electricity sources and emissions than is available from eGRID, our team of scientists analyzed the U.S. grid empirically using network modeling (see Figure 1). We first identified the smallest entities that we could obtain data for, the Power Control Areas (PCAs), which allowed us to divide the contiguous U.S. and grid-connected regions of Canada and Mexico into 134 regions (compared to the 20 eGRID sub-regions that are used today). The U.S. Energy Information Administration (EIA) has opened up PCA information, aimed at allowing power stakeholders (energy buyers, utilities, distributed generation vendors, financial analysts, and policymakers) to better understand production and demand.

However, the EIA data alone does not allow users to know the electricity mix that they consume – users can know the mix of electricity produced in any given state, but this does not tell them the mix of electricity that they actually consume, since areas also import and export electricity. We used a spatiotemporal network topology to analyze the entire grid as a dynamic, equilibrium-based supply chain, using the PCAs' monthly generation data and estimating the true consumption mix using annual inter-PCA exchange data.

Overall, this new Lux GNA information allows a more than 80-fold increase in granularity compared to current solutions, by moving from 20 eGRID sub-regions to 134 PCAs, and monthly instead of annual data.

Fig. 2: Google Underestimates Its Dependence on Coal in Four Out of Seven Data Centers



Our analysis was recently [published in *Environmental Science and Technology*](#), where interested readers can find more detail on the methodology. In this white paper, we applied this new tool to the energy use of data centers to determine where the real picture differs from the conventional wisdom today.

Case Study 1: Google Underestimates Its Dependence on Coal

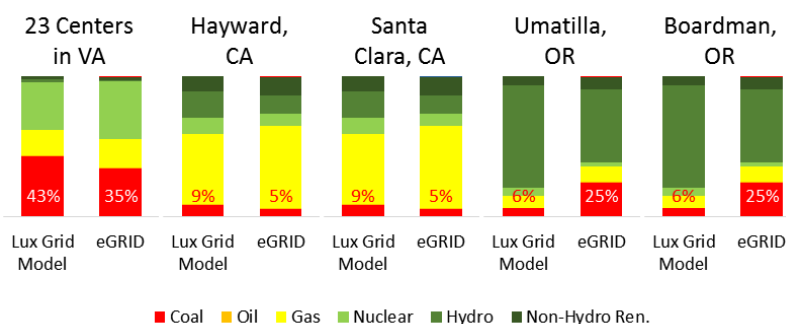
Google claims that including offsets, its net carbon footprint is zero, based on its tabulation of emission from its various energy sources, including electricity purchased from the grid. [According to Google Green](#), “For US locations (excluding those associated with green power purchase agreements), our grid renewables data are from the EPA’s eGRID.” Leaving aside [the contentious issue](#) of how well renewable energy credits (RECs) and other offsets do indeed offset emission, we looked to see what the improved Lux GNA tells us about Google's data center energy usage.

We found that eGRID significantly underestimates the amount of coal in the electricity consumed at four out of seven of Google’s U.S. data centers (see Figure 2). While our analysis does find that Google is using somewhat less coal than it estimates at three other sites, overall its emissions are likely significantly higher than it has projected. Using average values for the emissions from various types of generation and typical data center electricity usage, we estimate that the excess of actual emissions over Google's estimates from its seven data centers over the course of a year amounts to 42,000 metric tons of carbon dioxide equivalent (CO_{2e}) – the equivalent of putting 8,500 additional SUVs on the road.

Case Study 2: Amazon’s 23 Virginia Data Centers Use More Coal than Expected

Amazon is less transparent than Google about fuel consumption at data centers and how it's calculated. However, the Lux GNA reveals Amazon would likely miscalculate its carbon footprint where it relies on grid power, particularly in Virginia (see Figure 3). While the percentage grid coal usage in Virginia is only 8% higher using the Lux GNA than eGRID, 23 of Amazon's 28 U.S. data centers (a majority of its data centers worldwide) are in Virginia, making for a significant miscalculation of the amount of coal used. Again, we can estimate that the difference in emissions from the more accurate Lux GNA over eGRID amounts to 85,000 MT of CO_{2e} – equal to the emissions from around 5,000 households.

Fig. 3: Electricity Used in Amazon's 23 Virginia Data Centers Is Largely Coal-Generated



The U.S. Energy Market Is in Transition, Making Accurate Information Critical

The tech giants that operate data centers have all agreed that minimizing their carbon footprint is an important business objective, but, ironically, they haven't yet found a solid, data-driven way to do so. Making decisions about energy use based on accurate and up-to-date information is especially important given the changing U.S. electricity landscape. Investments in renewable energy and transmission are growing exponentially, while low prices are helping natural gas continue to displace coal, changing the composition of grid generation. Most importantly, the EPA's Clean Power Plan, and other changes coming globally as nations work to reduce greenhouse gas emission in line with their commitments made at the 2015 United Nations Climate Change Conference (COP21), will help provide more incentive for energy users to find cleaner ways to produce their power.

On-site distributed generation, large-scale off-site renewable energy power purchase agreements (PPAs) for utility-scale wind and solar, efficiency improvements, investment in offsets, and other solutions will all be options for data centers and other energy users. The right prioritization of what to do where begins with better analytics – with the tools now available, it's time for data centers to bring to their energy decisions the same data-driven rigor they use in the rest of their business.

About Lux Research

Lux Research is an independent research and advisory firm, providing strategic advice and ongoing intelligence on emerging technologies. Lux monitors emerging innovations and analyzes their market impact in its ongoing intelligence services, and provides customized advice and support to clients through consulting services. It also provides data analytics and corporate strategic intelligence in order to help put these innovations and market developments in context and help clients set their strategic priorities.

About Lux's RISK Platform

Resource scarcity, price volatility, and increased social and environmental pressures all threaten supply disruption and unplanned price spikes – but for most companies, the impact of resource dependence remains increasingly hard to evaluate. Lux Research's Resource Information for System Knowledge (RISK) platform guides companies through this treacherous landscape.

RISK provides the necessary support to help companies benchmark their global operations from the perspectives of profitability, resilience, and sustainability. Companies using the platform can determine where substitution strategies can be deployed, and foresee where breakthrough innovation can have the most impact. Lux's commitment is to allow organizations to improve their resource decisions by using the best science and up-to-date data. This approach is exemplified by the grid network analysis.

RISK and Data Centers

Organizations make decisions that involve complex value chains (in the case of data centers, largely the electricity grid) and complex resource interactions (for data centers, the interaction of distributed generation and the grid). Organizations need to think about resources outside the four walls of their facilities to mitigate resource cost, risk, and environmental impact. For data centers, only about 12% of their carbon emission is within the four walls of their facility; the rest of the losses (and emissions) come from resources procured from elsewhere. The analysis requires modern data and analytical tools like RISK.

¹ Amazon would not disclose the number of data centers they have, compelling us to present our best estimates. Additionally, they would not disclose what methodology they use to calculate their environmental footprint or whether eGRID underlies it.