THE HIDDEN ENVIRONMENTAL COST OF DATA

An email instead of a letter, online shopping instead of driving to a mall, a video conference instead of an in-person meeting. Are these activities as green as we think, or do they hide an environmental footprint?

CONTENT

Acknowledgements

This paper was written by the team at **Eticas** as part of a collaboration under **Ashoka's Tech and Humanity initiative.**

Project team: Ethics Oversight
Research Director: Dr. Gemma Galdon-Clavell, Founder of Eticas Foundation
Research Lead: Matteo Mastracci, Ethics and Technology Researcher at Eticas
Other Contributors: Emilia Paesano, Project Manager at Eticas, Patricia Vázquez, Head of Marketing and Communications at Eticas, Marta Vergara, Head of Storytelling and Funding at Eticas

The Hidden Environmental Cost of Data

Data has become an integral part of our daily lives, helping us to make better decisions and to improve our quality of life. However, the increasing amount of data that we generate and process requires an enormous volume of energy, which has negative social, economic, and environmental impacts. Data processes are all over. From sending emails and video streaming to browsing cookies, our reliance on data is driving up energy consumption at an alarming rate. The good news is that there are ways to mitigate this impact.

Environmental Data Pollution

Data pollution is a new and confusing concept and most existing studies do not address it sufficiently. This document will use the term "**environmental data pollution** (EDP)" to refer specifically to the carbon footprint of data and the potential environmental harm caused by data and data processing.

Environmental data pollution is the carbon footprint derived from data processes, from its generation, collection, processing, exchange, consumption, storage, etc. Concrete examples go from sending an email, to streaming a TV show, having a video conference, etc.

The ICT Industry Rising Footprint

A 2008 report by the Global eSustainability Initiative (GeSI) found that the **total footprint emissions from the ICT sector**, including personal computers (PCs) and peripherals, telecoms networks, devices and data centers, amounted to **830 metric tons of carbon dioxide** equivalent (MTCO2e) in 2007 corresponding to **2%** of the estimated **total emissions from human activity** released that year. The overall total figure, as of today, is likely to be much bigger and, a recent study (Freitag et al., 2022) came to the conclusion that the global **emissions** from the **ICT sector** are as high as **2.1% – 3.9%**.

A groundbreaking study conducted by the European Commission from 2020[1], pointed out that energy consumption of data centers is set to increase in the coming years. Between 2010 and 2018, the energy consumption of data centers in the EU28[2] increased from 53.9 TWh/a to 76.8 TWh/a which accounted for 2.7% of the electricity demand in the EU28 in 2018. At the same time, the EC study, after noting that a precise method for measuring CO2 emissions is far from being achieved, estimates that European data centers could produce between 0.4% and 0.6% of the entire EU greenhouse gas (GHG) emissions (EC, 2020: 57).

[1] See, European Commission. (2020): https://digital-strategy.ec.europa.eu/en/library/energy-efficient-cloud-computing-technologies-and-policies-eco-friendly-cloud-market

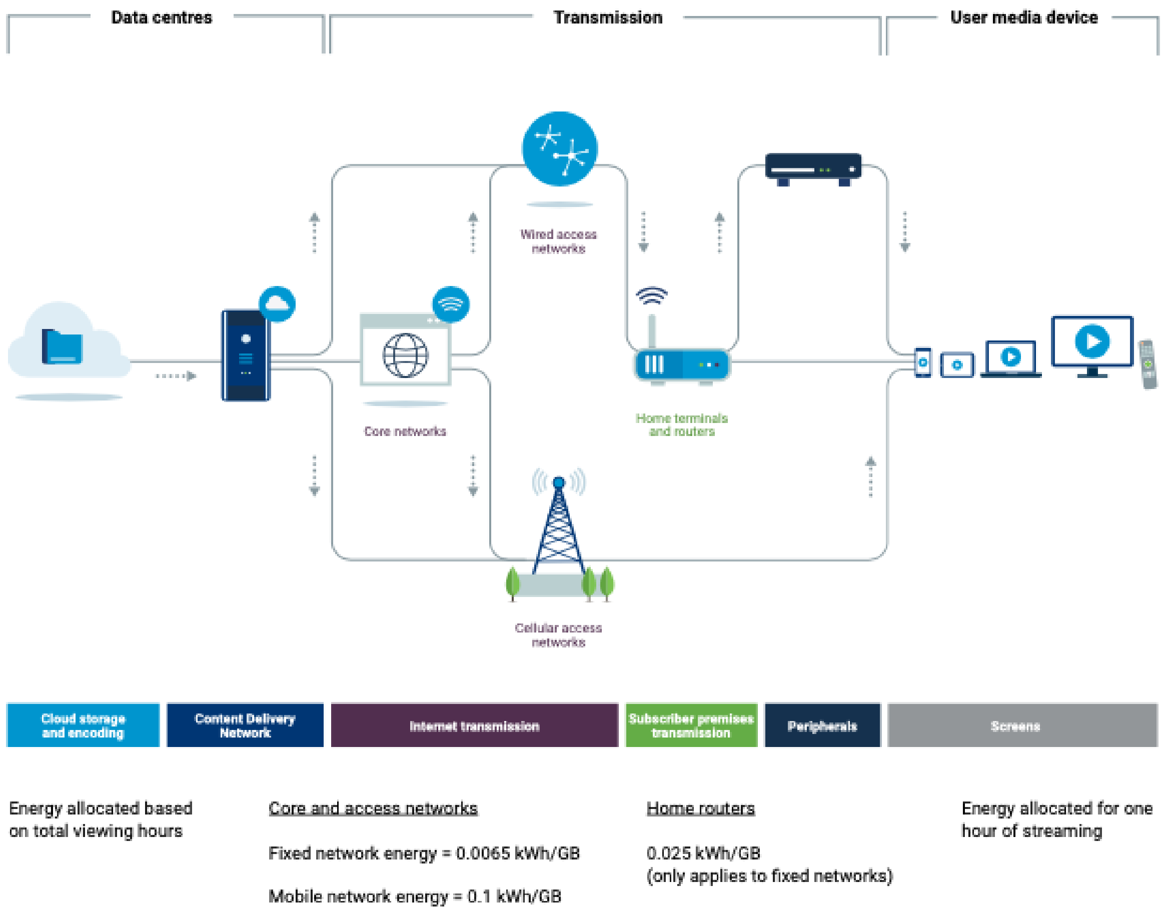
[2] EU-28 is the abbreviation of European Union (EU) members before the exit from the union by the United Kingdom.

1.1 Measuring EDP, but How?

Estimating the amount of carbon emissions that data processes produce is a complicated matter. One of the reasons is that **each data process is made up of innumerable and imperceptible steps** (which together lead to the data life cycle) and for each of the phases, it might be necessary to **resort to as many measurement techniques**. This may be why most studies prefer to focus on environmental measurement for specific data processes and not on a general purpose measurement. What follows is a list of data processes methods that we have unearthed and which we believe are as interesting as the most reliable in terms of their methodological approach.

Video Streaming

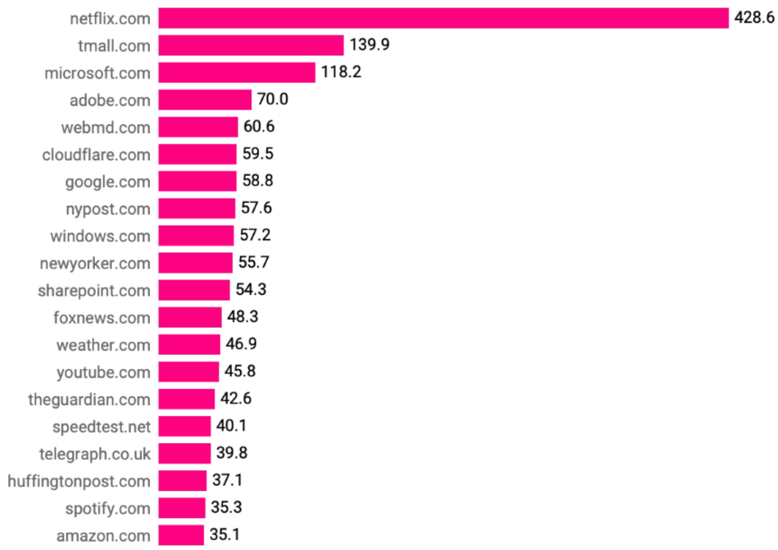
The carbon emissions of video streaming services have been the subject of a study by **Carbon Trust** that relied on what they called the **conventional approach**. Taking into account the different components of the video streaming life cycle, the study concluded that the estimated **European average carbon footprint** corresponded to around **55 gCO2e per hour** of video streaming for the year 2020, where the component of the user's devices accounted for the largest share of emissions in the video streaming footprint. Indeed, the carbon footprint of the end-user devices accounted for 401 MtCO2e, followed by networks (198 MtCO2e) and data centers (141 MtCO2e).



Graphic 1: Data life cycle of the video streaming process. Source: The Carbon Trust. (2021).

Browsing Cookies

Cookies can also have an impact on the environment, further increasing the production of carbon emissions, as demonstrated by Carbolytics[3]. Carbolytics research examined and analyzed the carbon emissions produced by the total number of cookies belonging to the top one million websites. The survey identified more than 21 million cookies for every single visit to all of these websites, belonging to more than 1200 different companies, which translates into an average of 197 trillion cookies per month. It was also found that the median cookie size (byte length of its name and value) was 35 bytes. The study, in the end, estimated that the **total number of carbon emissions** for the cookies from the top one million websites amounted to **11,442 metric tons of CO_2 per month** with a lower bound of 1,400 and an upper bound of 17,100 metric tons of CO_2 per month. This means that the carbon footprint of cookies from top websites in **one year** would be approximately **138.000 metric tons of CO_2**. Cookies' environmental impact is something that deserves to be brought to attention beyond the numbers in absolute terms, as it highlights a deeper problem of the existing ecosystems of data: the **breach of individuals' privacy which causes an environmental hazard**. Most interestingly, Carbolytics found (see **Graphic 5** below) that the site www.netflix.com turned out to be the top site for carbon emissions per cookie with 428.6 metric tons of CO_2 per month.



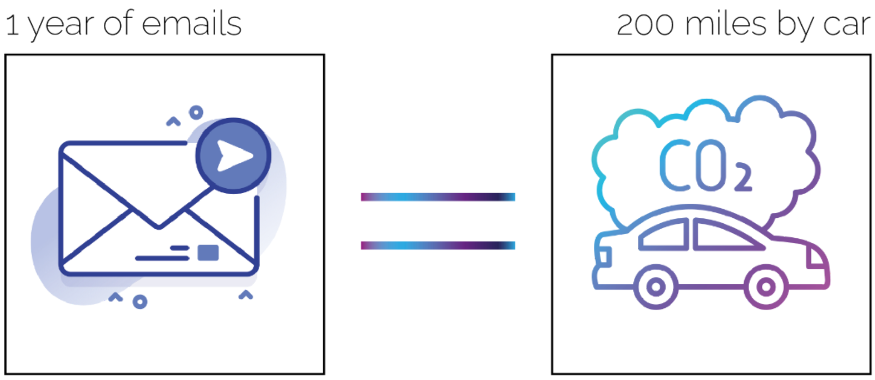
Graphic 2: Top 20 sites by emission from cookies. Source: Carbolytics. (2022). An analysis of the carbon costs of online tracking.

Sending Emails

In his work "**How Bad Are Bananas?**" (2010), Berners-Lee argues that an average spam email has a footprint of **0.3g CO_2e**, a proper email has a carbon footprint of **4g CO_2e** and emails with long and tiresome attachments have a carbon footprint of **50g CO_2e**. It mainly depends on the size and other characteristics not too dissimilar to those for sending postal mail. The environmental impact and the energy consumption are mainly related to the power that both data centers and personal computers are using for

[3] See, https://carbolytics.org/

sending, filtering, and receiving emails. The author also clarifies that a year of incoming emails adds up to 135 kg (300 lbs.) CO2e which corresponds to over 1 percent of the 10-ton lifestyle and equivalent to driving 200 miles in an average car.



Graphic 3: Emails vs Cars. Source: Own elaboration.

The book is based on a study by McAfee which found out based on an estimated 62 trillion spam emails delivered in 2008, that spam's annual energy use amounts to 33 billion kilowatt hours (KWh), or 33 terawatt hours (TWh), a total of **0.3 grams of CO2-e** has been the average for a single spam email where the **biggest amount of energy consumption** (52 percent) comes from **end-users deleting spam and searching for legitimate email** (false positives). On their side, OVO Energy, a Bristol-based trading company found that for **every fewer thank-you emails sent per day** by a British adult, there would be a **saving of around 16,433 tonnes of carbon per year**, equivalent to something like **81,152 flights to Madrid** for a middle class passenger.

Bitcoin and Crypto Mining

Among all the data processes, Bitcoin is the industry with the **highest degree of risk** for the environment and carbon footprint emissions. According to the **Cambridge Bitcoin Electricity Consumption Index** (CBECI)[4] the global electricity demand of Bitcoin miners in 2022 (electricity load) reached 11.55 gigawatts (GW). The total **yearly electricity** consumption of the **Bitcoin network**, in the same year, peaked 101.29 terawatt-hours (TWh). The trend of Bitcoin electricity consumption is constantly growing, **and the more electricity is consumed, the more carbon emissions are released.**

One of the latest studies on the subject, De Vries et al. (2022) estimated that the Bitcoin network could be responsible for around **65.4 megatonnes of CO2 per year**, **which is comparable to country-level emissions in Greece** (Digiconomist, 2022). In its latest figures from their Bitcoin Energy Consumption Index[5], Digiconomist found that the annual consumption of Bitcoin leads to 72.65 Mt CO2 of carbon emissions (comparable to the carbon footprint of Turkmenistan.), 130.25 TWh of electrical energy (comparable to the

[4] See: https://ccaf.io/cbeci/index
[5] See,
https://digiconomist.net/bitcoin-energy-consumption#:~:text=The%20carbon%20footprint%20per%20VISA%20transaction%20is%20only%200.45%20grams%20CO2eq.&text=The%20number%20of%20VISA%20transactions,on%20average%20(1395.13%20kWh).

power consumption of Argentina) and 37.82 kt of electronic waste (comparable to the small IT equipment waste of the Netherlands). Another notorious study (Stoll et al., 2018) came to different conclusions about Bitcoin's annual carbon emissions, which given its November 2018 annual electricity consumption of 48.2 TWh, that would be **between 21.5 and 53.6 megatonnes of CO2 per year.**

AI, ML and Deep Learning Models Training

Also **training an AI model**, in that case a Natural Language Processing (NLP) one, can be **environmentally costly** as shown by a 2019 research paper (Strubell et al., 2019). In fact, the researchers by training four different AI models (Transformer, NAS, ELMo and BERT) demonstrated that the overall cost of training in terms of CO2 emissions and cloud computing cost was not entirely negligible. In the case of training of the NAS AI model, the carbon footprint emissions (626,155) was nearly comparable to the carbon dioxide emission of five American cars during their lifetime.

Model	Hardware	Power (W)	Hours	kWh·PUE	CO_2e	Cloud compute cost
$\text{Transformer}_{base}$	P100x8	1415.78	12	27	26	$41–$140
Transformer_{big}	P100x8	1515.43	84	201	192	$289–$981
ELMo	P100x3	517.66	336	275	262	$433–$1472
BERT_{base}	V100x64	12,041.51	79	1507	1438	$3751–$12,571
BERT_{base}	TPUv2x16	—	96	—	—	$2074–$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	$942,973–$3,201,722
NAS	TPUv2x1	—	32,623	—	—	$44,055–$146,848
GPT-2	TPUv3x32	—	168	—	—	$12,902–$43,008

Table 1: *Estimated cost of training a model in terms of CO2 emissions.* Source: Strubell, E., Ganesh, A., & McCallum, A. (2019)

Similarly to what happens with AI models, **training a machine learning (ML) model** can also have a significant environmental impact. Lacoste et al. (2019) demonstrated that the presence of GPU cloud servers located in North America can lead to very different carbon emissions, from 20g CO2eq/kWh in Quebec to 736.6g CO2eq/kWh in Iowa. Finally, **deep learning**[6] **model training** can also lead to a massive amount of computational time needed and, thus, to carbon emissions as illustrated by the study conducted by Schwartz et al. (2019). The study supports **Green AI**, which is more environmentally friendly and inclusive AI research, as opposed to **Red AI** as an artificial intelligence research method that seeks to achieve cutting-edge results in terms of accuracy through the use of enormous computing power.

Video Conferencing

In a blog post published in November 2020[7], David Mytton, researcher of sustainable computing at the University of Oxford, attempted to **debunk the myth that video**

[6] As clarified by IBM, "deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to "learn" from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy." Available at: https://www.ibm.com/cloud/learn/deep-learning

[7] See, https://davidmytton.blog/zoom-video-conferencing-energy-and-emissions/

conferencing tools are much greener than face-to-face meetings. Mytton began by asking a series of basic questions:

> "In the context of video conferencing, it is safe to assume that arranging a Zoom call is better than flying business class from London to New York, but is it better than both participants walking to a cafe in the city they both already live in? Maybe someone was going to drive their EV to the office. What about if you have many participants all over the world? How about if some connect via 4G vs others on a laptop?"

After realizing that Zoom has never made it clear the exact amount of carbon emissions from its online video conferencing systems, he started calculating an estimation data of the electricity consumed by the platform. Zoom offers different **bandwidth** requirements for 1:1 call and group video calling, The numbers were later converted into gigabytes as per Aslan et al. (2018) who identified the total amount of electricity intensity of fixed-line internet data transmission. It follows that a 1 hour 1:1 call could generate 1.08 – 3.24GB of network traffic using 0.0162 – 0.0486 kWh of electricity, while a group call of six people around 4.86 – 14.85GB of traffic and use 0.0729 – 0.22275 kWh of electricity. Mytton found that for the UK, a 1:1 call HD 1080p of 1 hour between two people would require 0.25358 kgCO2 per kWh with a corresponding CO2 emissions of 0.012 kgCO2 using the 2019 UK greenhouse gas conversion factors[8]. Similarly, in the US, for the same 1:1 call it would be around 0.28839 kgCO2 per kWh.

Assumptions About EDP

Negative assumptions:

- There is **no single method** to measure environmental data pollution, and this is mainly due to the diversity of each data process.
- Each of the individual methods used for a specific data process tends to be **closed and self-referential** and does not refer or communicate with the methods used for other types of data processes.
- Overall, **time and place** are two of the most important and problematic factors of the data lifecycle that can have an impact on the carbon footprint calculation.
- **The more steps** there are in the individual data lifecycle, **the more difficult** would be calculating the environmental impact of that data.

Positive assumptions:

- All methods show that, albeit the uncertainties exposed above, **environmental data pollution can be measured** at least by an approximated estimate.
- The best **two measuring methods** are the **conventional approach** by Carbon Trust, that relies on a blended approach, teaming up with some of the best expertises on the field, and the **5 building blocks method** by De Nederlandsche Bank, which for each phase of the data life cycle is based on some of the best indices in the sector.

[8] See, https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2019

1.2 Lowering Polluted Data: Mitigating Measures

We propose data minimization as a **general** mitigation measure and a good practice according to the GDPR standards, which is applicable to any data process regardless of the type and location, and three **particular** measures that are only suitable for certain data processes.

Data Minimization

This mitigation measure is based on a simple principle: **the less data is collected, the lower the risk of producing environmental data pollution**. Article 5(1)(c) of the European GDPR, "Data minimization", says that personal data shall be:

> "**adequate, relevant and limited** to what is **necessary** in relation to the **purposes** for which they are processed ('data minimization')"

Modular Process, Less AI Training

A particular method to mitigate the carbon footprint resulting from training an AI model, is the **AI modular building block** developed by Elevait[9], in which all the AI models are connected to each other, so the AI training will be only one. The modular process aims to simplify the AI training by lowering the amount of time and energy required for training different models within the same process.

PoS Replacing PoW for Cryptos

Another particular method concerns the field of cryptocurrencies and questions the energy efficiency of their current protocol, the **Proof-of-work** (PoW). To reduce the environmental impact of the Bitcoin network, it was proposed to switch to the **Proof-of-Stake** (PoS) protocol. Last September, **Ethereum**, the world's second-largest cryptocurrency, switched its protocol from PoW to PoS saving **99.95%** of their energy consumption[10].

Spam Filtering

As a mitigation measure specifically designed for emailing, spam filtering could be an effective way to counter the carbon emissions of email spam as noted by the McAfee & ICF International research (2009). The study found that **spam filtering saves 135 TWh of electricity per year**. However, it would be more desirable to combat spam at the source, as was the case with US-based web hosting provider **McColo**. In this case, this major web hosting provider and source of online spam, was taken offline in late 2008 by its upstream Internet Service Provider (ISP), followed by approximately **70% drop of total spam volume** resulting in **energy savings equivalent to 2.2 million off-road passenger vehicles**.

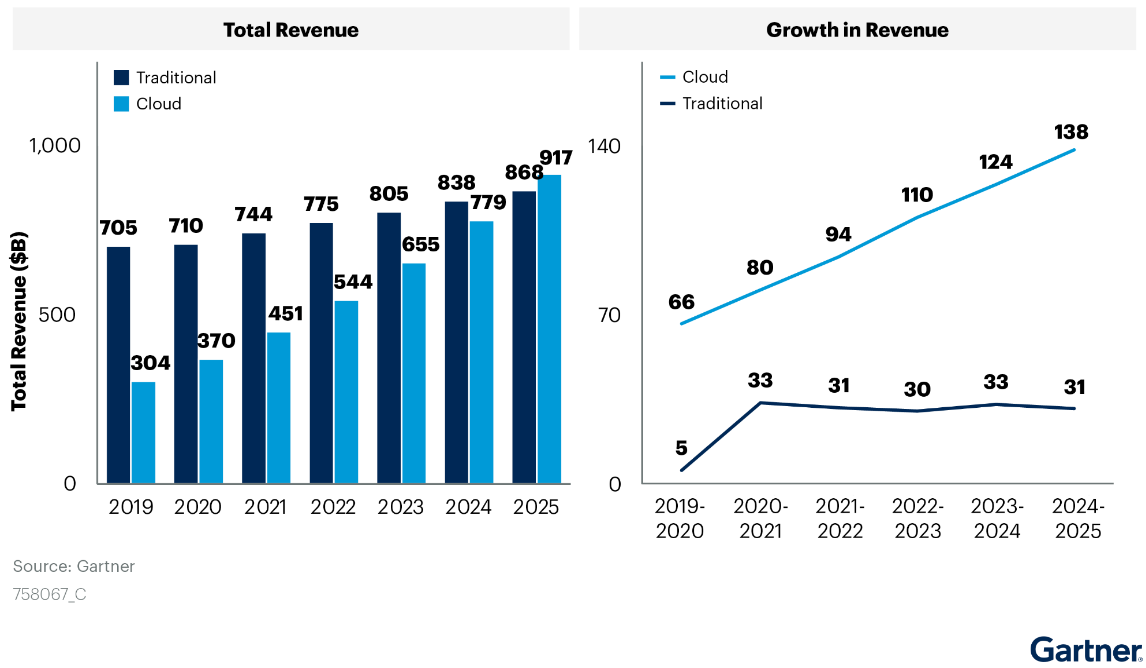
---

[9] See: https://www.elevait.de/
[10] See: https://ethereum.org/en/upgrades/merge

Data Centers: How They Impact our Daily Lives

As the global demand for data continues to rise exponentially, data centers are rapidly expanding around the world. While data centers can offer numerous advantages, such as the potential for stimulating investments and job creation, they can also have some disadvantages. For example, data centers can have substantial impacts on the environment, such as through their energy consumption, water footprint and the potential for greenhouse gas emissions. To gain a complete understanding of data centers and their significance in modern society, it is crucial to consider the challenges, as well as the benefits and drawbacks, they present.

The data center market is quickly growing. The storage and processing of data and its subsequent need for access explains why data centers are proliferating and will continue to do so. Beyond the current cloud service delivery models, there is a latent trend in the cloud computing market to migrate all types of cloud applications to offer them as services. Some voices point to the danger of this paradigm, called **XaaS** (Everything as a Service), as more and more resources will be needed and data will have more environmental impact. Gartner (2022) forecasts that by 2025 business spending on **new cloud computing-based technologies will exceed business spending in the traditional cloud computing market**. new technologies, driving the continuous shift to the cloud, as illustrated in the figure below:



Graphic 5: Sizing Cloud Shift, Worldwide, 2019 – 2025. Source: Gartner (2022)

This will cause associated a growth of **consumption of electricity and GHG emissions that it generates**, an urgent problem to tackle, if we take into account that nowadays **the largest element of this kind of pollution within the technology industry are data**

centers, with an annual consumption estimation that exceeds 200 billion kWh (Xinyuan, 2022).

2.1 Environmental Impact of DCs: Electricity, Emissions and Water Footprint

In 2019, it was estimated that **DCs worldwide used more than 2% of the world's electricity** and generated the same volume of carbon emissions as the global airline industry (TNW, 2020). Despite the enormous energy expense and associated pollution of data centers, the development of increasingly efficient technologies means that the energy efficiency of the devices is increasing. However, this increase is slow and the effects of the data-intensive demand of the technologies, on the contrary, continues to grow exponentially. This double effect generates a growing concern about the still uncertain general environmental impacts of the sector in the coming decades. (IEA, 2022).

Energy Consumption of DCs

According to the International Energy Agency (IEA, 2022), data centers and data transmission networks **are responsible for nearly 1% of total energy-related GHG emissions**, which is about 205 TWh, and their use of electricity in 2021 was 220-320 TWh2, or **about 0.9-1.3% of global final electricity demand**, excluding energy used for cryptocurrency mining, which was 100-140 TWh. If we look at the impact of the growth of DCs in some of the most industrialized countries, the figures for energy use are huge: in the US, electricity generation is the second largest consumer of water and the main cause of GHG emissions (Siddik et al., 2021); Denmark alone is expected to host several large-scale DCs, whose demand for electricity in 2040 may reach 33% of 2017 national electricity consumption (Petrović et al., 2020); In Ireland, the DC consumption of electricity increased by 144% between 2015 and 2020 (Burke-Kennedy, 2022) according to an analysis from the Central Statistics Office (CSO) and in the country have an average floor area of 20,000 sq.m, and an average site area of approximately 11 hectares (O'Leary, 2022). Also in China the trend is the same: according to the China Academy of Information and Communications Technology (CAICT), by the end of 2021, China's data centers had a combined total of more than five million standard Rack Units (U) (Xinyuan, 2022).

Among all the functionalities that can cause large energy demands in DCs, consumption related to **computing power (43%)** and **cooling systems (32%) stand out**. Their need for constant cooling means that **DCs are buildings between 10-40 times more energy intensive than a typical office building** and this intensity of energy is completely linked to its energy consumption (ICF, 2009).

The relationship between refrigeration and GHG emission is very high: **40% of the total energy consumed by the DC is dedicated to cooling.** Unfortunately, **the most widely used system, air cooling, is also the most inefficient,** although its maintenance is relatively simple and inexpensive (Zhang et al, 2021). This means that the forecast for the future is in any case of growth in energy demand and therefore in greenhouse gas emissions, although this scenario is difficult to draw.

Water Footprint of DCs

Closely related to the large amount of electricity used to cool the DCs is the **huge water footprint** of these infrastructures, which spend large amounts of water to function. The water footprint can be associated mainly with two factors. The first one is the amount of water consumed indirectly that, as a primary source, is needed to generate the electricity used by a DC. The second one is the water footprint associated with the water used within the infrastructure, normally used for cooling through cooling towers. Water consumption can become very significant: data centers that have 15 MW of IT capacity can consume between 80 and 130 million gallons per year and it is estimated that the typical water consumption of a DC is 1.8 liters (0.46 gallons) per kWh (Shehabi et al, 2016). Siddik et al (2021) made estimates at the US level of what the water footprint was in 2018, in which they estimated an average of **1 MWh of energy consumption by a data center requires 7.1 m3 of water,** although it is confirmed that in general, the calculation of a water impact footprint is difficult to estimate and a better understanding of the global water footprint of DCs is necessary to understand its real scope. As noted by Siddik et al. (2021) in the US, only power plants with generating capacity greater than 100 MWh must report water consumption.

The water footprint can also cause significant impacts on the environment and on water itself, as demonstrated by a rather controversial case in the US panorama. Following the acquisition of an aging coal-fired power plant located in Finger Lakes, New York state by a private equity firm, Atlas Holdings LLC, the plant was converted into a bitcoin mining company named Greenidge Generation Holdings Inc., which continues to arouse much controversy today (Gopalakrishnan, 2022). One of these is the alleged **huge water footprint** that the plant is causing by the dumping of tens of millions of gallons of hot water into the glacial Seneca Lake (De Chant, 2021) which has led many residents to denounce how the temperature of the lake has significantly increased (Morgenson, 2021). More recently, in a much-anticipated decision, New York's Department of Environmental Conservation has denied the air permit requests to the company claiming that following the previous permit released back in 2016 and 2019, the greenhouse gas emissions from the plant have increased dramatically. The case of the Greenidge Generation plant and its impact on nearby Seneca Lake is one of many that should draw attention to the adverse effects that the mismanagement of data centers can entail on the environment.

Data Brokerage: Uncovering a World of Data

Personal data is the real engine of a bigger machine, the personal **data ecosystem**. This data can include information such as browsing history, location data, purchasing behavior, and social media interactions. Data brokers are one of the main parts of this complex machinery and they play a significant role in the personal data ecosystem, as they facilitate the flow of personal data between various parties. Data brokers often acquire this data from a variety of sources, including social media companies, retail websites, and mobile app developers. However, **data brokers often operate in the shadows** and the vast majority of data owners remain unaware about what is happening in the digital world with their personal data. This has raised concerns about the privacy and security of individuals' personal data, the potential for abuse and misuse of this information, and the environmental impact of the big amounts of data used.

Data users buy data, especially big data, because **owning the data** can give them a massive **competitive advantage** in the market that translates into **significant economic return**. Data users, for instance, can include law enforcement agencies (LEAs) that purchase location data (Morrison, 2021), banks that can collect data to analyze clients' incomes and expenditures (Ostapchenya, 2020), private investigators and lawyers buying cell phone records in seek of evidence against cheating spouse in divorce cases (Matwyshyn, 2009) and many others.

Brokering Data, Who are the Brokers?

Data is at the heart of data ecosystems and data brokers are the main actors within it for one simple reason: they **maximize** the **usefulness and value** of each single piece of **data**. But one question still remains open, who the data brokers really are and what they do in practice. According to European and US regulation, data brokers can be defined in a more comprehensive manner in this term:

> ***Data brokers** are some key players who **may or** may **not aggregate data and** whose **main purpose** is to **sell data** collected from other sources to data users and without the data owners being aware of this transaction.*

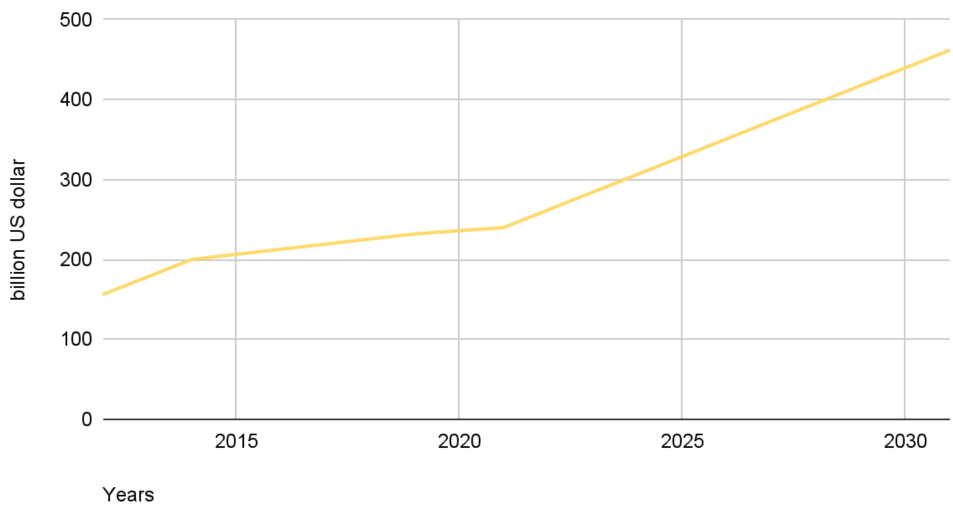
The **Vermont** legislature has passed a law[11] requiring **data brokers to be registered** in a list as of February 2019 (Melendez, 2019; Melendez & Pasternack, 2019). At that moment, the list is collecting names of about **121 data brokers** operating in the United States, further confirming that the U.S. market is the main and fastest growing market globally. Indeed, although there are numerous data brokers active also in Europe, the European data broker landscape is not comparable to the U.S. market in terms of market size (Open Society, 2016:13). Some of the biggest names in the data brokerage industry like **Acxiom, LexisNexis, Nielsen, Equifax, CoreLogic, Verisk, Oracle and Epsilon** all have headquarters in the United States (Sherman, 2021). Acxiom, which is the largest data broker in both size and revenue by far, advertises the selling of entire U.S. households

[11] Vt. H.764 (Act 171), https://legislature.vermont.gov/bill/status/2018/H.764

clusters based on behaviors and attitudes through its **Personicx**[12] LexisNexis in February 2021 signed a multi-million dollar agreement (Biddle, 2021) to sell personal data to the US Immigration and Customs Enforcement (ICE) and so forth.

Growing Data Broker Industry

Data brokers not only conclude countless personal data transactions almost per minute, but they grow in terms of new market slices at an astonishing rate. Not surprisingly, the data broker market is **constantly expanding** and in full growth. In **2012** alone, the brokerage market reached **US$ 156 billion** in net revenues (Harcourt, 2016:198), in **2014** the same broker industry has likely peaked at **US$ 200 billion** in annual revenues (Harcourt, 2016) for then reaching **US$ 232.634 billion** of revenues in **2019** (Blueweave, 2021). The overall growth of the data brokers' market does not seem to know any declines as in **2021** it was valued at **US$ 240.3 billion**, it is estimated to **grow at a CAGR of 6.8% from 2022 to 2031** and is expected to reach **US$ 462.4 billion** by the end of **2031** (TMR, 2022).



Graphic 9: Title: Data brokers expected growth in annual revenue. Source: own production.

As the market projections already clearly reveal, the data brokerage industry will continue growing at an impressive rate that remains difficult to predict with certainty today. **Data accountability**, however, does not seem to be sufficient yet and a major problem to be solved in the near future.

[12] See: https://www.personicx.co.uk/

Conclusion

There are a number of challenges associated with the concept of environmental data pollution. For one, there is the issue of how to accurately measure the pollution levels of different types of data due to the **lack of coherent and uniform standards** in the **measurement of carbon emissions**. Also, it is also true that using **renewable energy** to power data processes **does not necessarily mean** that **environmental data pollution will disappear.** This is because renewable energy is only a part of the whole data life cycle, and there are many other stages in the life cycle of data where environmental pollution can occur. For example, even if renewable energy is used to power data processes, the production and disposal of electronic devices used for data processing can still result in environmental pollution.

Regarding **data centers**, as more and more companies rely on the internet and digital technology to conduct their operations and serve their customers, their demand has increased significantly. In the first half of 2022, the data center industry was worth approximately $24 billion in new deals. At the same time, the rising data center industry involves both **pros and cons**. On one hand, modern data centers are essential to the functioning of many digital businesses and can contribute to increased investment and job creation. On the other hand, data centers also have a number of drawbacks, including some significant environmental and socioeconomic impacts. Data centers can generate, among other things, substantial amounts of waste and greenhouse gas emissions, which contribute to climate change.

In the **short term**, a good advocacy strategy, at least in the EU, would be to push for the **inclusion** of the information and communication technology (ICT) industry **within** the EU Emissions Trading System (**EU ETS**). Currently, the ICT industry is completely left out from the EU ETS, even though the EU has warned that the industry is one of the fastest growing for electricity consumption. Including the ICT industry in the EU ETS would allow for the creation of a market-based mechanism to incentivize the reduction of greenhouse gas emissions from this sector. It is possible that, hypothetically, the inclusion of the ICT industry in the EU ETS could create a **spillover effect outside the EU**, affecting other jurisdictions. This is because the EU ETS is a large and influential system, and any changes to it could have broader implications. For example, if the inclusion of the ICT industry in the EU ETS leads to significant reductions in greenhouse gas emissions in the near future, it could encourage other jurisdictions to adopt similar approaches.

References

Andress, J., & Leary, M. (2017). In *Building a practical information security program*. essay, Elsevier, Syngress.

Aslan, J., Mayers, K., Koomey, J. G., & France, C. (2017). Electricity intensity of internet data transmission: Untangling the estimates. *Journal of Industrial Ecology*, 22(4), 785–798. https://doi.org/10.1111/jiec.12630

Berners-Lee, M. (2010). *How bad are bananas?: The Carbon Footprint of Everything*. Profile Books

Biddle, S. (2021, April 2). *LexisNexis to provide giant database of personal information to ice*. The Intercept. Retrieved December 8, 2022, from https://theintercept.com/2021/04/02/ice-database-surveillance-lexisnexis/

BlueWeave. (2021). Global Data Broker Market Size, Share, Opportunities, COVID-19 Impact, And Trends By Data Type (Consumer Data, Business Data), By End-User Industry (BFSI, Retail, Automotive, Construction, Others), And By Geography - Forecasts From 2021 To 2026. https://www.knowledge-sourcing.com/report/global-data-broker-market

Burke-Kennedy, E. (2022, January 20). *Electricity consumed by data centres jumps by 144% in five years*. The Irish Times. Retrieved December 11, 2022, from https://www.irishtimes.com/business/energy-and-resources/electricity-consumed-by-data-centres-jumps-by-144-in-five-years-1.4781365

Carbolytics. (2022). *An analysis of the carbon costs of online tracking*. https://carbolytics.org/report.html

De Chant, T. (2021, July 6). *Bitcoin power plant making part of Glacial Lake 'feel like a hot tub,' residents say [updated]*. Ars Technica. Retrieved December 11, 2022, from https://arstechnica.com/tech-policy/2021/07/bitcoin-power-plant-is-turning-a-12000-year-old-glacial-lake-into-a-hot-tub/

de Vries, A., Gallersdörfer, U., Klaaßen, L., & Stoll, C. (2022). Revisiting bitcoin 's carbon footprint. *Joule*, 6(3), 498–502. https://doi.org/10.1016/j.joule.2022.02.005

Digiconomist. (2022, February 25). Bitcoin less "green" than ever before. *Digiconomist*. https://digiconomist.net/bitcoin-less-green-than-ever-before/

Edelman, G. (2022, July 21). Don't Look Now, but Congress Might Pass an Actually Good Privacy Bill. *WIRED*. https://www.wired.com/story/american-data-privacy-protection-act-adppa/

European Commission. (2020). *Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market*. https://digital-strategy.ec.europa.eu/en/library/energy-efficient-cloud-computing-technologies-and-policies-eco-friendly-cloud-market

Ferré-sadurní, L., & Ashford, G. (2022, November 23). *New York enacts 2-year ban on some crypto-mining operations*. The New York Times. Retrieved December 12, 2022, from https://www.nytimes.com/2022/11/22/nyregion/crypto-mining-ban-hochul.html

Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G. S., & Friday, A. (2022). The real climate and transformative impact of ICT: A Critique of estimates, trends, and regulations. *Patterns*, 3(8), 100576. https://doi.org/10.1016/j.patter.2022.100576

Gartner. (2022, February 9). *Gartner says more than half of enterprise IT spending in key market segments will shift to the cloud by 2025*. Gartner. Retrieved December 11, 2022, from https://www.gartner.com/en/newsroom/press-releases/2022-02-09-gartner-says-more-than-half-of-enterprise-it-spending

Global eSustainability Initiative (GeSI). (2008). *SMART 2020: Enabling the low carbon economy in the information age*. https://gesi.org/research/smart-2020-enabling-the-low-carbon-economy-in-the-information-age

Gopalakrishnan, S. (2022, September 2). *Bitcoin mining billed as 'green' has dirty coal ash problem*. Energy News Network. Retrieved December 11, 2022, from https://energynews.us/2022/08/25/in-the-finger-lakes-a-bitcoin-mining-plant-billed-as-green-has-a-dirty-coal-ash-problem/

Gschossmann,I., van der Kraaij, A., Benoit, P., & Rocher, E. (2022). *Mining the environment – is climate risk priced into crypto-assets?*. European Central Bank (ECB). https://www.ecb.europa.eu/pub/financial-stability/macroprudential-bulletin/html/ecb.mpbu202207_3~d9614ea8e6.en.html#:~:text=With%20current%20yearly%20GHG%20emissions,2030%20(120.9%20million%20tCO2).

Harcourt, B. E. (2015). *Exposed: Desire and disobedience in the Digital age*. Harvard University Press.

ICF International. (2009). Opportunities for Combined Heat and Power in Data Centers. https://www.energy.gov/sites/prod/files/2013/11/f4/chp_data_centers.pdf

International Energy Agency. (2022). Data Centres and Data Transmission Networks. IEA. https://www.iea.org/reports/data-centres-and-data-transmission-networks

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the Carbon Emissions of Machine Learning. https://arxiv.org/abs/1910.09700

Matwyshyn, A. M. (2009). *Harboring data: Information security, Law, and the Corporation*. Stanford Law Books.

McAfee & ICF International. (2009). *The Carbon Footprint of Email Spam Report*. https://www.siskinds.com/wp-content/uploads/carbonfootprint_12pg_web_rev_na-1.pdf

McCook, H. (2015), An Order-of-Magnitude Estimate of the Relative Sustainability of the Bitcoin Network. https://www.academia.edu/7666373/An_Order-of-Magnitude_Estimate_of_the_Relative_Sustainability_of_the_Bitcoin_Network_-_3rd_Edition

McCook, H. (2018). The Cost & Sustainability of Bitcoin. https://www.academia.edu/37178295/The_Cost_and_Sustainability_of_Bitcoin_August_2018_

Melendez, S., & Pasternack, A. (2019, February 3). *The data brokers quietly buying and selling your personal information*. Retrieved December 8, 2022, from https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information

Melendez, S. (2019, February 3). *Over 120 data brokers revealed under landmark Vermont Law*. Fast Company. Retrieved December 8, 2022, from https://www.fastcompany.com/90302036/over-120-data-brokers-inch-out-of-the-shadows-under-landmark-vermont-law

Morgenson, G. (2021, July 5). *A bitcoin business is polluting Seneca Lake, say critics. here's how*. NBCNews.com. Retrieved December 11, 2022, from https://www.nbcnews.com/science/environment/some-locals-say-bitcoin-mining-operation-ruining-one-finger-lakes-n1272938

Morrison, S. (2021, July 31). *Here's how police can get your data - even if you aren't suspected of a crime*. Vox. Retrieved December 8, 2022, from https://www.vox.com/recode/22565926/police-law-enforcement-data-warrant

Mytton, D. (2020, November 16). Zoom, video conferencing, energy, and emissions. *Davidmytton*. https://davidmytton.blog/zoom-video-conferencing-energy-and-emissions/

O'Leary, K. (2022, January 6). *Renewed calls for less data centres in South Dublin over 'blackout' fears*. DublinLive. Retrieved December 11, 2022, from https://www.dublinlive.ie/news/dublin-news/renewed-calls-less-data-centres-22665875

Open Society Foundations. (November, 2016). *Data Brokers in an Open Society*. https://www.opensocietyfoundations.org/publications/data-brokers-open-society

Organisation for Economic Co-operation and Development (OECD). (2013), "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value", *OECD Digital Economy Papers*, No. 220, OECD Publishing, Paris, https://doi.org/10.1787/5k486qtxldmq-en.

Ostapchenya, D. (2021, June 11). *The role of Big Data in banking : How do modern banks use big data?* Finextra Research. Retrieved December 8, 2022, from https://www.finextra.com/blogposting/20446/the-role-of-big-data-in-banking--how-do-modern-banks-use-big-data

Petrović, S., Colangelo, A., Balyk, O., Delmastro, C., Gargiulo, M., Simonsen, M. B., & Karlsson, K. (2020). The role of data centres in the future Danish Energy System. *Energy*, 194, 116928. https://doi.org/10.1016/j.energy.2020.116928

Schwartz, R., Dodge, J., Smith, Noah A., & Etzioni, O. (2019). Green AI. https://arxiv.org/pdf/1907.10597.pdf

Shehabi, A., Smith, S., Sartor, D., Brown, R., Herrlin, M., Koomey, J., Masanet, E., Horner, N., Azevedo, I., & Lintner, W. (2016). United States Data Center Energy Usage Report. https://doi.org/10.2172/1372902

Sherman, J. (2021). *Data brokers and sensitive data on U.S. individuals: Threats to American civil rights, national security, and democracy*. Duke Sanford. https://techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2021/08/Data-Brokers-and-Sensitive-Data-on-US-Individuals-Sherman-2021.pdf

Siddik, M. A., Shehabi, A., & Marston, L. (2021). The environmental footprint of data centers in the United States. *Environmental Research Letters*, 16(6), 064017. https://doi.org/10.1088/1748-9326/abfba1

Stoll, C., Klaaßen, L., & Gallersdörfer U. (2018). *The Carbon Footprint of Bitcoin*. MIT Center for Energy and Environmental Policy Research (CEEPR). https://ceepr.mit.edu/wp-content/uploads/2021/09/2018-018.pdf

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/p19-1355

The Carbon Trust. (2021). *Carbon impact of video streaming*. https://www.carbontrust.com/resources/carbon-impact-of-video-streaming

Transparency Market Research (TMR). (2022). Data Brokers Market [Data Type: Unstructured Data, Structured Data, Custom Structure Data] - Global Industry Analysis, Size, Share, Growth, Trends, and Forecast, 2022-2031. https://www.transparencymarketresearch.com/data-brokers-market.html

eticas

info@eticas.tech

+34 936 005 400

Mir Geribert, 8, 3rd
08014, Barcelona