Carbon emissions and long-term digital preservation

Matthew Addis
Arkivum
26 Mar 2024

arkivum
Bringing archived data to life

Scope

Environmental Sustainability

- Carbon Footprint
- Water consumption
- Noise pollution
- Wildlife disruption

https://doi.org/10.1038/s41545-021-00101-w

npj | Clean Water www.nature.com/npjcleanwater

PERSPECTIVE OPEN

Data centre water consumption

David Mytton[1]✉

The information communication technology sector will experience huge growth over the coming years, with 29.3 billion devices expected online by 2030, up from 18.4 billion in 2018. To reliably support the online services used by these billions of users, data centres have been built around the world to provide the millions of servers they contain with access to power, cooling and internet connectivity. Whilst the energy consumption of these facilities regularly receives mainstream and academic coverage, analysis of their water consumption is scarce. Data centres consume water directly for cooling, in some cases 57% sourced from potable water, and indirectly through the water requirements of non-renewable electricity generation. Although in the USA, data centre water consumption (1.7 billion litres/day) is small compared to total water consumption (1218 billion litres/day), there are issues of transparency with less than a third of data centre operators measuring water consumption. This paper examines the water consumption of data centres, the measurement of that consumption, highlights the lack of data available to assess water efficiency, and discusses and where the industry is going in attempts to reduce future consumption.

npj Clean Water (2021)4:11 ; https://doi.org/10.1038/s41545-021-00101-w

INTRODUCTION

The information communication technology (ICT) sector is expecting huge growth over the coming years. By 2023, 5.3 billion people will have internet access, up from 3.9 billion in 2015[1]. By then, 29.3 billion devices will be connected to the internet (up from 18.4 billion in 2018), with access speeds doubling between 2018 and 2023 to a global average of 110 Mbps[1]. More people having faster access to online services means internet traffic will double by 2022[2].

To reliably serve these billions of users, internet properties rely on millions of dedicated computers called servers. These servers are located in data centres, which provide reliable power, cooling and internet access. Around 40% of servers are in small data centres[3] such as cabinets in an office side room, but newer facilities are increasingly "hyperscale" warehouses, hundreds of thousands of square meters in size, and run by the big three cloud vendors (Amazon Web Services, Google Cloud Platform, Microsoft Azure)[4].

The energy consumption of data centres regularly receives attention in both the academic and mainstream press. Despite the ICT sector being responsible for some of the largest purchases of renewable energy[5], there remains considerable uncertainty about total data centre energy consumption. Estimates for 2018 range from 200[6] to 500 TWh[7]. Some extreme analyses even suggest energy consumption could quadruple by 2030[8], whereas other estimates show energy growth plateauing[6]. Regardless of the precise number, data centre energy is an important topic of public interest. However, it is just one aspect of the environmental footprint of ICT. A less well understood factor is water consumption.

Crucial for industry and agriculture, the availability and quality of water is a growing global concern[9]. Projections suggest that water demand will increase by 55% between 2000 and 2050 due to growth from manufacturing (+400%), thermal power generation (+140%) and domestic use (+130%)[10]. ICT is another sector contributing to that demand.

In Fiscal Year 2018 (FY18), Google reported 15.8 billion litres of water consumption, up from 11.4 billion litres in FY17[11]. Similarly with Microsoft who reported using 3.6 billion litres in FY18, up from 1.9 billion litres in FY17[12] (Fig. 1). Offices make up some of this total, but data centres also use water.

This paper examines the water consumption of data centres, how that consumption is measured by the ICT sector, and considers where the industry is going in attempts to reduce future water consumption.

DATA CENTRE WATER USE

Total water consumption in the USA in 2015 was 1218 billion litres per day, of which thermoelectric power used 503 billion litres, irrigation used 446 billion litres and 147 billion litres per day went to supply 87% of the US population with potable water[13].

Data centres consume water across two main categories: indirectly through electricity generation (traditionally thermoelectric power) and directly through cooling. In 2014, a total of 626 billion litres of water use was attributable to US data centres[4]. This is a small proportion in the context of such high national figures, however, data centres compete with other users for access to local resources. A medium-sized data centre (15 megawatts (MW)) uses as much water as three average-sized hospitals, or more than two 18-hole golf courses[14]. Some progress has been made with using recycled and non-potable water, but from the limited figures available[15] some data centre operators are drawing more than half of their water from potable sources (Fig. 2). This has been the source of considerable controversy in areas of water stress and highlights the importance of understanding how data centres use water.

This section considers these two categories of data centre water consumption.

Water use in electricity generation

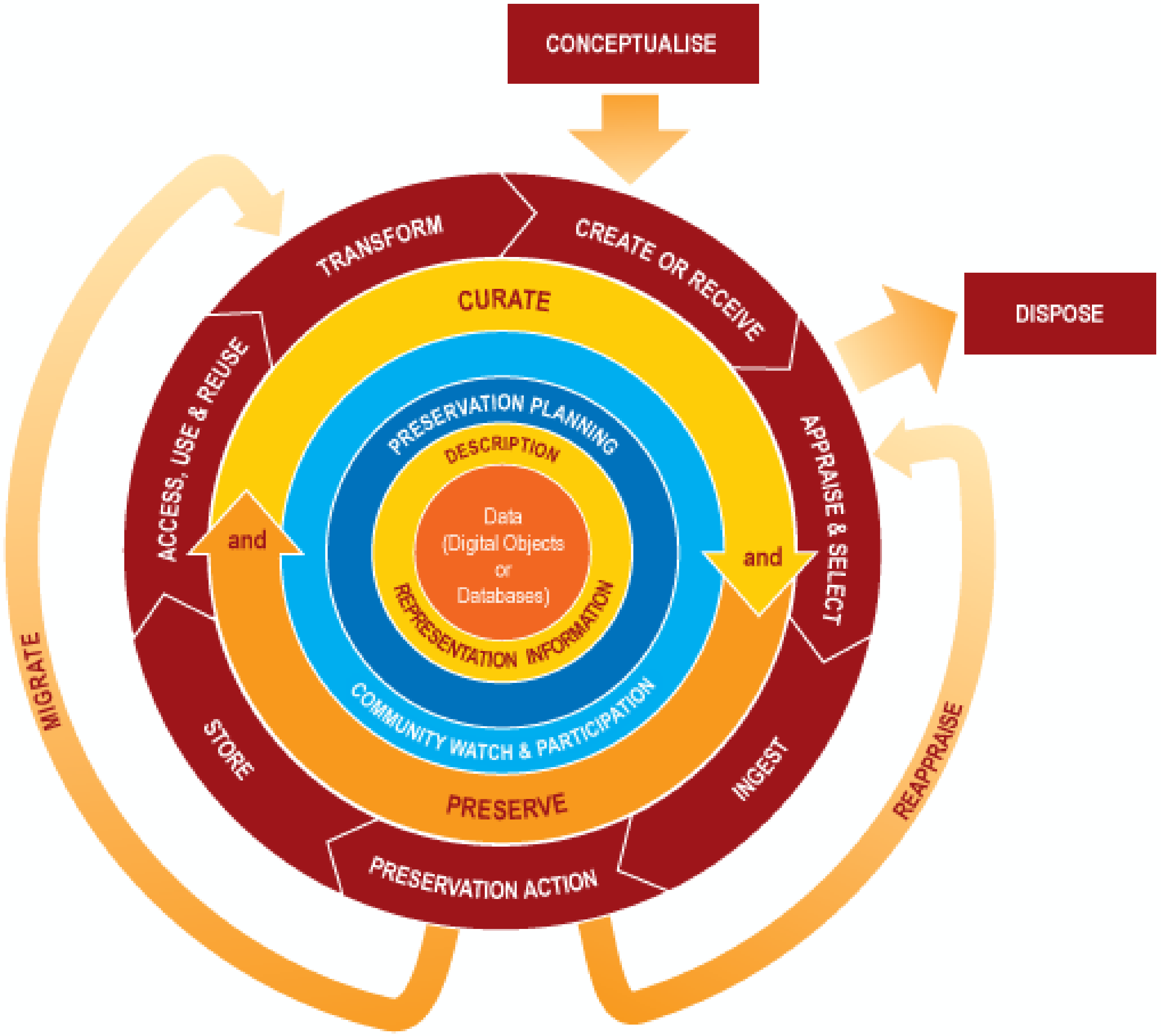
Water requirements are measured based on withdrawal or consumption. Consumption refers to water lost (usually through evaporation), whereas water withdrawal refers to water taken from a source such as natural surface water, underground water,

[1]Centre for Environmental Policy, Imperial College London, London, UK. ✉email: david@davidmytton.co.uk

Content Lifecycle

https://www.dcc.ac.uk/guidance/curation-lifecycle-model

- Create / Digitise
- Appraise and Select
- Ingest
- Preservation
- Storage
- Access
- Distribution / Transfer
- Use / Reuse



https://www.carbontrust.com/our-work-and-impact/guides-reports-and-tools/carbon-impact-of-video-streaming
https://www.iea.org/commentaries/the-carbon-footprint-of-streaming-video-fact-checking-the-headlines

Sources of Carbon Emissions

- Energy (power, cooling)
- ICT equipment (servers, storage, networking)
- Data Centres (buildings, equipment)
- People (staff, contractors)
- Travel (commuting, transport)

https://www.google.co.uk/about/datacenters/gallery/

LTDP in the Cloud

- Hyperscaler cloud infrastructures
- Published sustainability information
- Commitments to net zero
- Follow GHG Protocol
- Data available for actual emissions

https://sustainability.aboutamazon.co.uk/environment/the-cloud

https://cloud.google.com/sustainability

https://azure.microsoft.com/en-us/explore/global-infrastructure/sustainability

This Talk

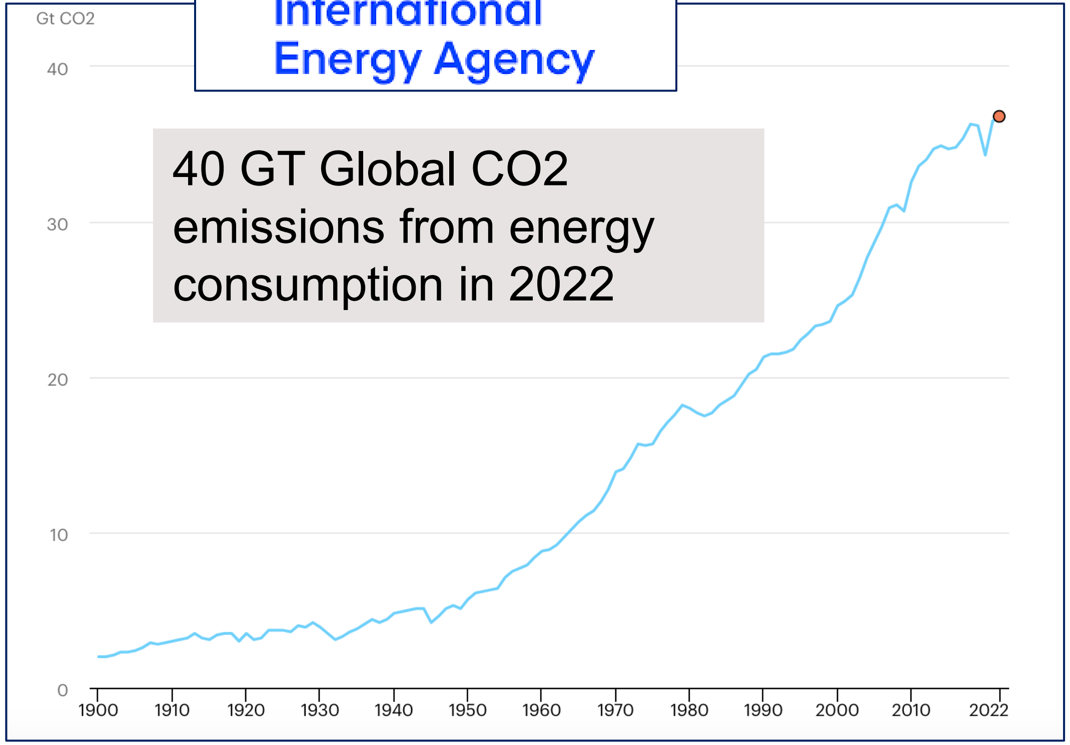
- LTDP services running in the Cloud
  - GHG Scope 3
- Carbon Footprint
  - Energy Use
  - Embodied footprint of ICT equipment
- Quantified emissions
  - kgCO2eq
  - Real world digital preservation activities

Carbon Emissions From LTDP: Energy Usage

Global Energy Consumption by Data Centres

Data centres, cryptocurrencies, and artificial intelligence (AI) consumed about 460 TWh of electricity worldwide in 2022, almost 2% of total global electricity demand.

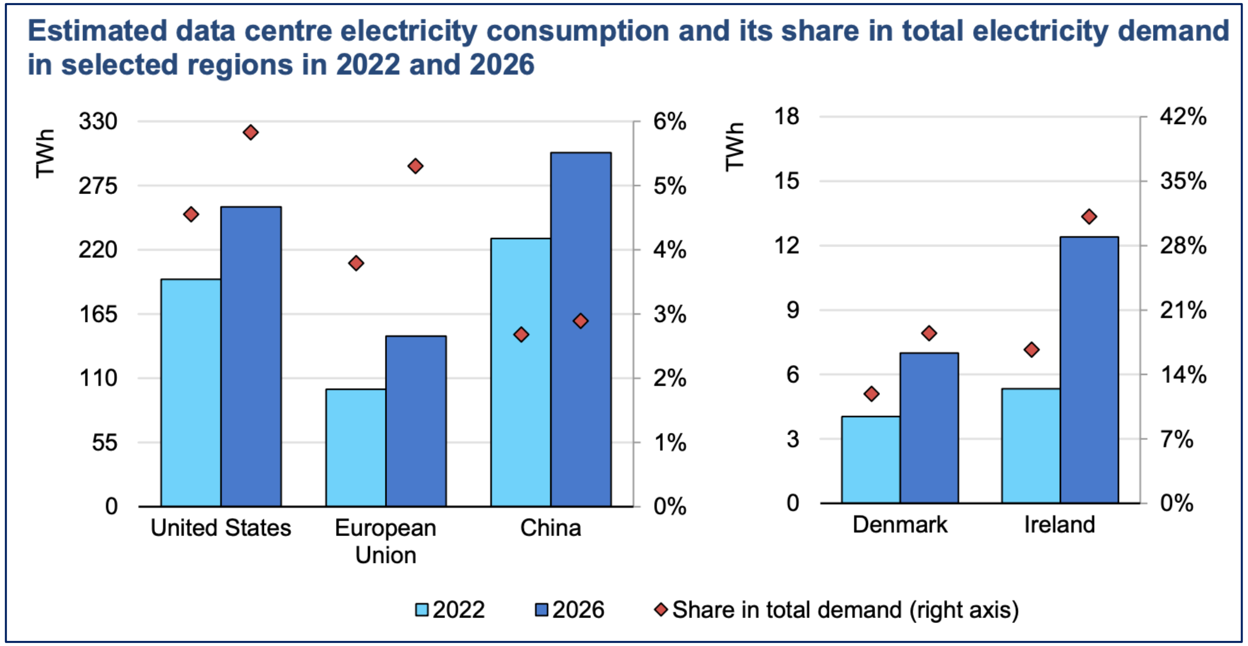
International Energy Agency



40 GT Global CO2 emissions from energy consumption in 2022

https://www.iea.org/reports/co2-emissions-in-2022

Data centres' total electricity consumption could reach more than 1000 TWh in 2026. This demand is roughly equivalent to the electricity consumption of Japan



https://www.iea.org/reports/electricity-2024

Misdirected Attention

Global energy consumption results in large CO_2 emissions

Global data centres use lots of energy

The cloud uses big data centres

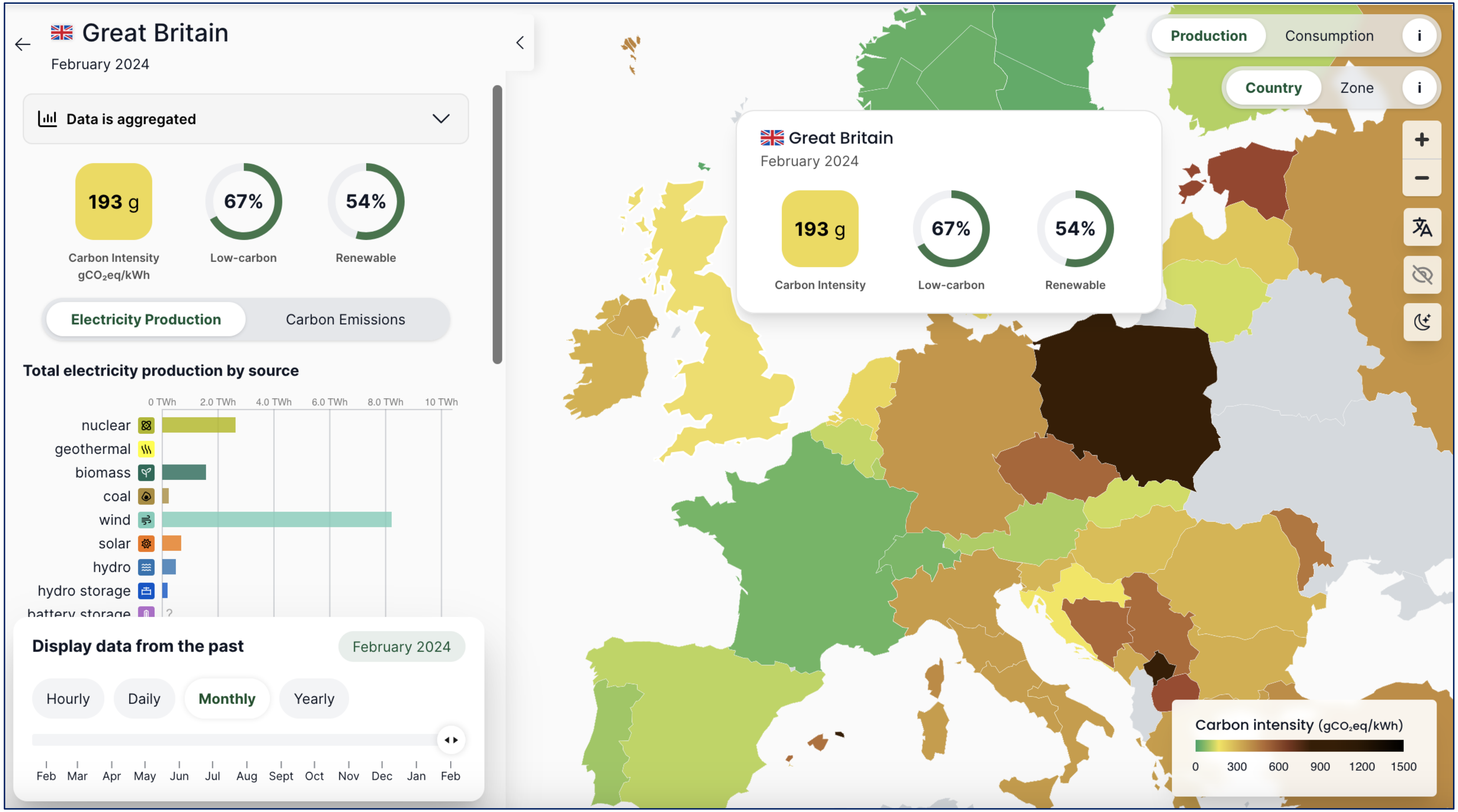
Digital Preservation is often done in the cloud

$\Rightarrow$

We need to worry (big time) about

CO_2 emissions from energy consumption by LTDP in the cloud

Carbon Intensity of Electricity over Time and by Location

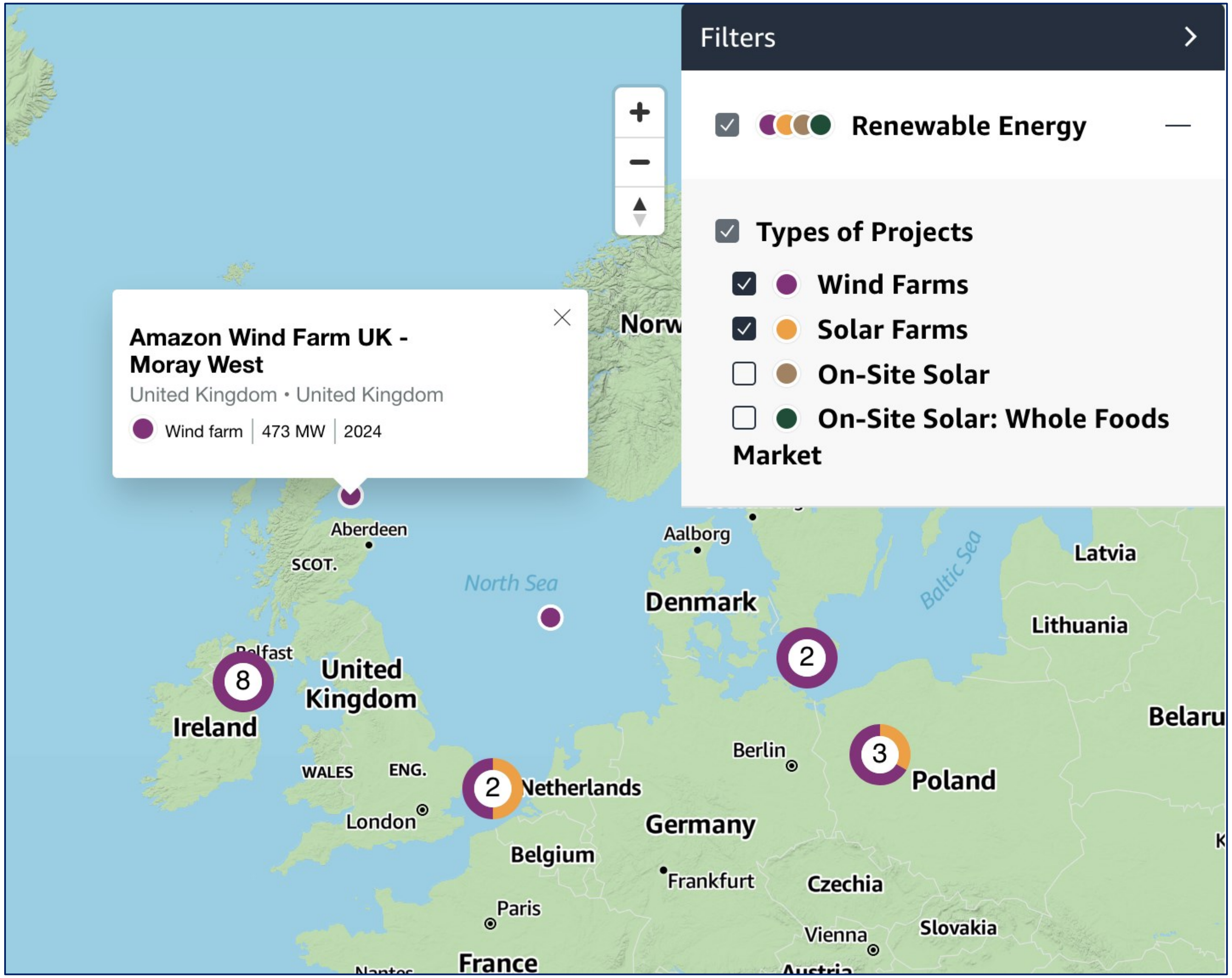


https://app.electricitymaps.com/map

Example: AWS Renewable Energy Initiatives

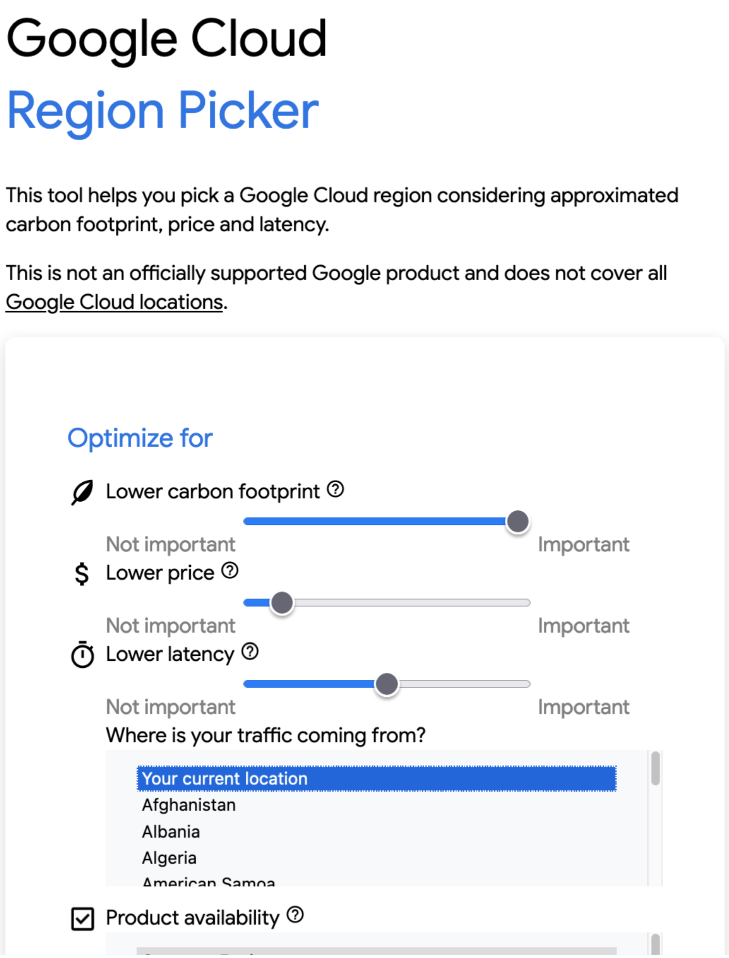
https://www.aboutamazon.co.uk/amazon-engie

- AWS has more than 500 wind and solar projects globally
- Once operational, they are expected to generate more than 77,000 GWh of clean energy each year
- AWS target is 100% renewable energy by 2025

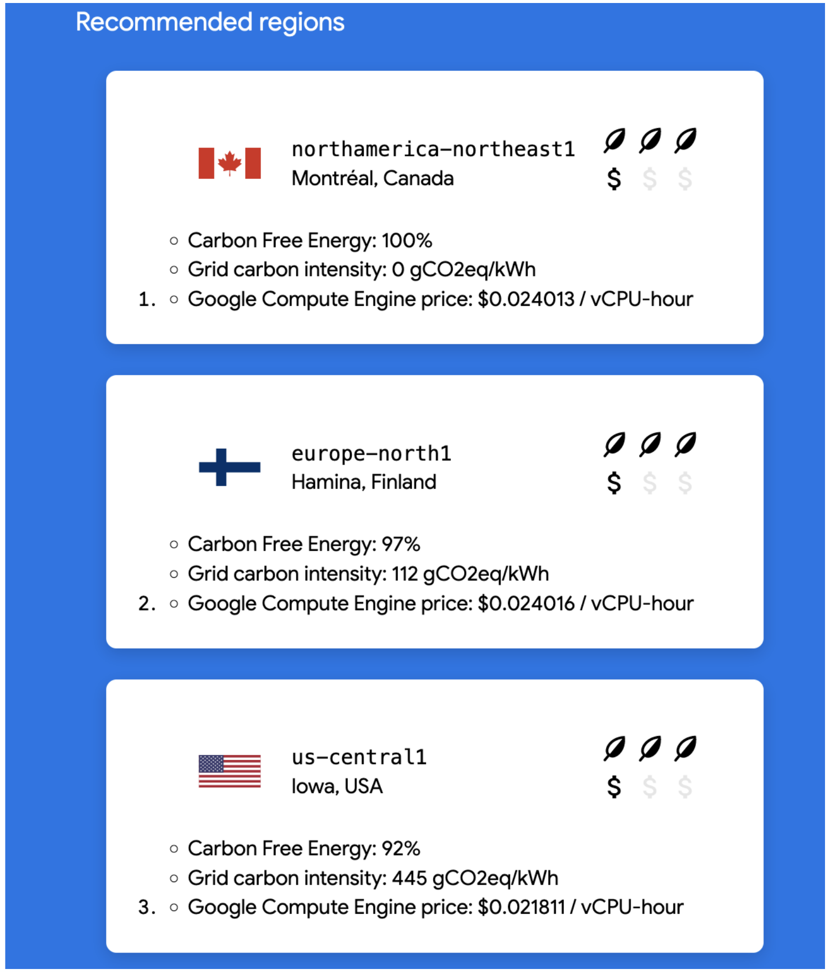


https://sustainability.aboutamazon.com/climate-solutions/carbon-free-energy?energyType=Wind+farm%2CSolar+farm

Example: Google Cloud Platform (GCP) Carbon Net Zero



https://cloud.withgoogle.com/region-picker/



https://www.google.com/about/datacenters/gallery/#hamina-exterior-landscape

Arkivum Measurement of LTDP Carbon Emissions

- Get resource consumption and carbon emissions from cloud provider reports
  - CPU resource consumption over 5 months (core-hours)
  - Storage consumption over 5 months (GB-months)
  - Gross emissions over 5 months per resource type (kgCO2 eq)
- Calculate metrics
  - kgCO2 eq per core-hour for compute
  - kgCO2 eq per TB-year for storage
- Measure resource consumption for specific preservation workflows (storage, compute)
  - Large files, small files, inside bagit bags, big ingests, lots of small ingests
  - File format identification, checksum generation, metadata extraction, replication etc.
  - Additional processing using Archivematica on-demand
- Calculate carbon emissions
  - kgCO2 eq per TB of data ingested for different scenarios

https://doi.org/10.6084/m9.figshare.20653101

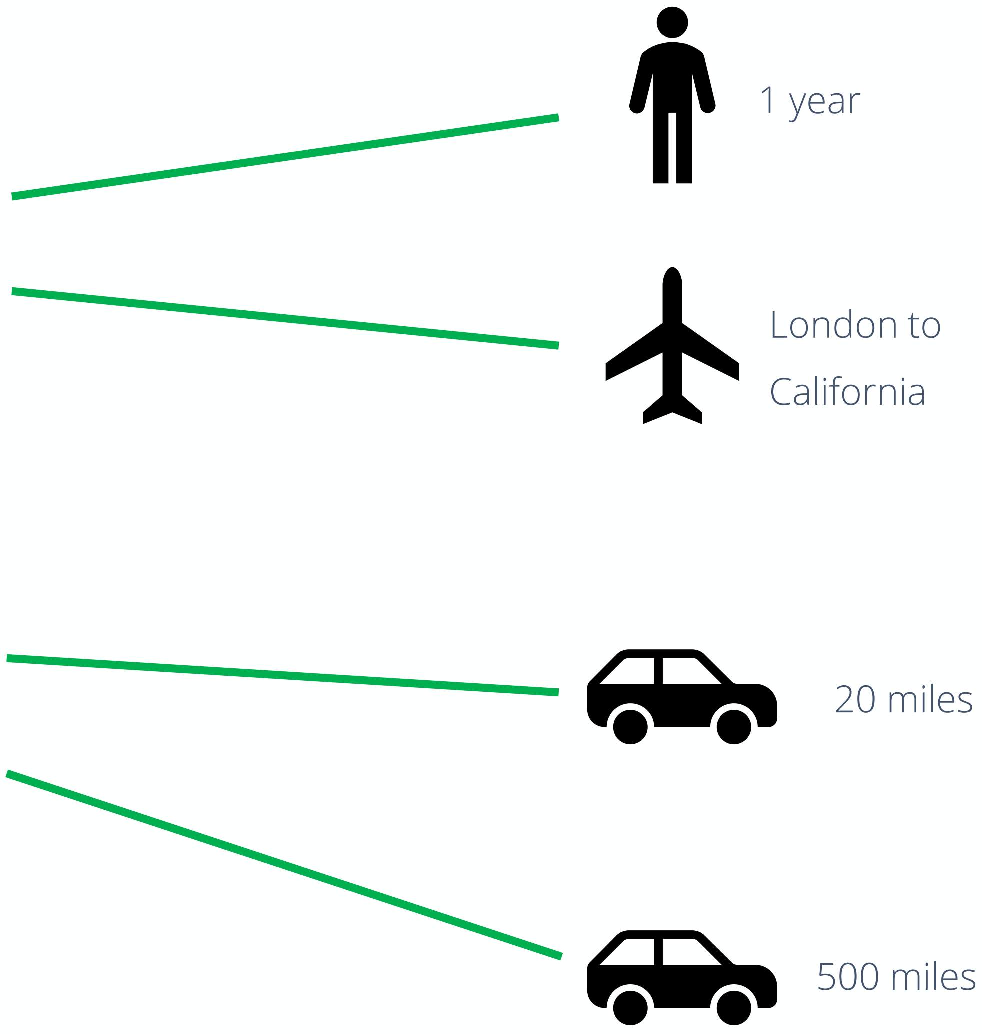
Arkivum: LTDP in the Cloud - Gross Carbon Emissions From Energy Consumption

Large image Datasets

1 PB data stored for 1 year	7800 kgCO2 eq
1 PB ingest of large image files	1600 kgCO2 eq

Large collections of office files

1M office files stored for 1 year	5.5 kgCO2 eq
Ingest of 1M office files.	140 kgCO2 eq



The net carbon emissions were zero

Summary: Carbon Emissions from LTDP Energy Consumption in the Cloud

- Already net zero (depending on your choice of cloud provider and location)
- Ambitious and rapid advances from all the major cloud providers on use of carbon free energy

- But not an excuse to be wasteful!
- Energy may be net-zero but the embodied footprint of the servers isn't…

Jevons paradox

https://en.wikipedia.org/wiki/Jevons_paradox

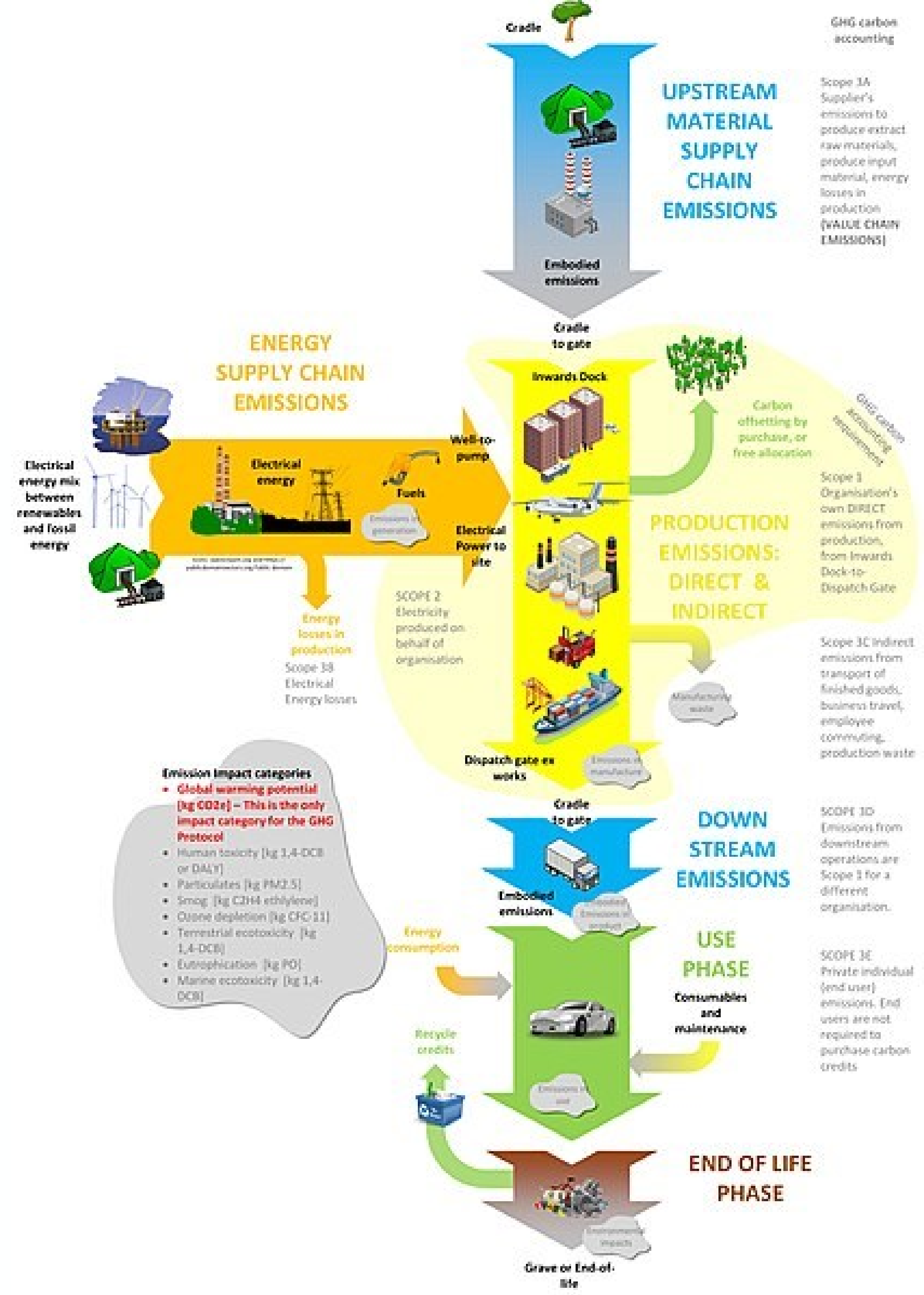
By Unknown author - Popular Science Monthly Volume 11, Public Domain, https://commons.wikimedia.org/w/index.php?curid=11022925

Carbon Emissions From LTDP: Embodied Footprint of ICT

Embodied Footprint of ICT in the Cloud

- “The cloud is just someone else’s computers”
- Raw materials, manufacturing, transportation, use, disposal, recycling
- The use stage (carbon footprint from energy used in the cloud) is a small part of the ICT lifecycle

- Life Cycle Assessment (LCA)
- ISO 14040
- Cradle to Grave



https://en.wikipedia.org/wiki/Life-cycle_assessment#/media/File:Life_cycle_analysis_and_GHG_carbon_accounting.jpg

LCA Numbers for ICT Servers

- Not published by Cloud Providers
- Very few examples from hardware manufactures
- Common approach is to use data from Dell servers and adapt it estimating cloud footprint
- Reality is that we don't know what AWS or GCP are actually using for their hardware and what it's embodied footprint is!

https://corporate.delltechnologies.com/content/dam/digitalassets/active/en/unauth/data-sheets/products/servers/lca_poweredge_r740.pdf

Life Cycle Assessment of Dell PowerEdge R740

Report produced June, 2019

From design to end-of-life and everything in between, we work to improve the environmental impact of the products you purchase. As part of that process, we estimate the specific impacts throughout the lifecycle. The lifecycle phases included in a LCA are illustrated in figure 1.

Figure 1: 'Cradle to grave' Life Cycle Assessment phases

LCA Definition

'A life cycle assessment is the compilation and evaluation of the inputs, outputs and potential environmental impacts of a product system throughout its life cycle' – *ISO 14040: 2006, sec 3.2.*

Dell PowerEdge R740

The product selected for this LCA is the Dell R740 server and represents that of a general-purpose rack server which provides computing services capable of handling very demanding workloads and applications, such as data warehouses, ecommerce, AI/Machine Learning, and high-performance computing (HPC). The server configuration modelled in this LCA represents that of a high-end configuration (see table 1).

Table 1: Assumptions

Assumptions	
Lifetime of product	4 Years
Use location	EU & USA
Memory	x12 32GB DIMM's
Storage	x1 400GB SSD x8 3.4TB SSD's
Processor	x2 Intel Xeon 140W CPU's
Platform	2U, 2-socket platform

Results Summary

The impact assessment results within this study include but are not limited to; global warming potential (GWP), ozone layer depletion potential and eutrophication potential. The results discussed in this LCA focus on the GWP impact category as it is considered the most robust and widely used impact category. Climate change is also referred to as GWP or the 'carbon footprint'. A detailed view of the carbon footprint is shown in figure 1. The major fraction of the impact (approximately 98%) derives from the manufacturing and use phase of the Dell R740. Transportation and end of life management has a less relevant contribution to the overall impact of the Dell R740 server.

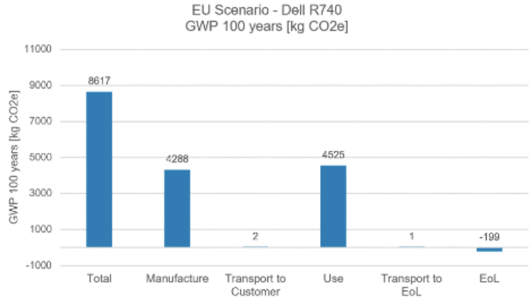


Figure 1: Contribution of the different stages of the lifecycle to the GWP of the Dell R740 (EU)

Key Findings:

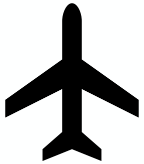
- *The use phase contributes to approx. 50% of the total life cycle global warming potential of the sever.*
- *The manufacturing stage contributes to approx. 50% of the product carbon footprint.*
- *Electronic components in the manufacturing stage have the largest environmental impact of all modules and are dominated by the x8 3.4TB SSD's. The manufacture of storage devices is complex and both energy and resource intensive.*
- *The majority of the SSD impact of the 3.84TB SSD's comes from the NAND flash chips. Results indicate that the die/package ratio of these chips significantly influences the GWP.*
- *The study scenarios assume three different die/package ratios of 30%, 60% and 80%. Overall manufacturing impacts of the server are reduced by ~40% if a die/package ratio of 30% is assumed for the SSD's.*
- *The two materials that are influenced by the different die/package ratios are the wafer manufacturing and gold.*
- *Recycling resulted in a net reduction of 200 kg CO2-equivalents. This represents a reduction of the total impact by around 1.8%.*
- *The largest net gains that come from recycling the Dell R740 server come from the recycling of gold (~84%), followed by steel (~10%).*

Dell R740 Server (General Purpose Rack Server)

- Manufacturing footprint ~4200 kgCO2eq
- Mostly comes from SSDs (~80%)
- Transport is negligible
- Recycling saves ~2%

- 4 year lifetime =>

1 tonne CO2eq per year



London to New York

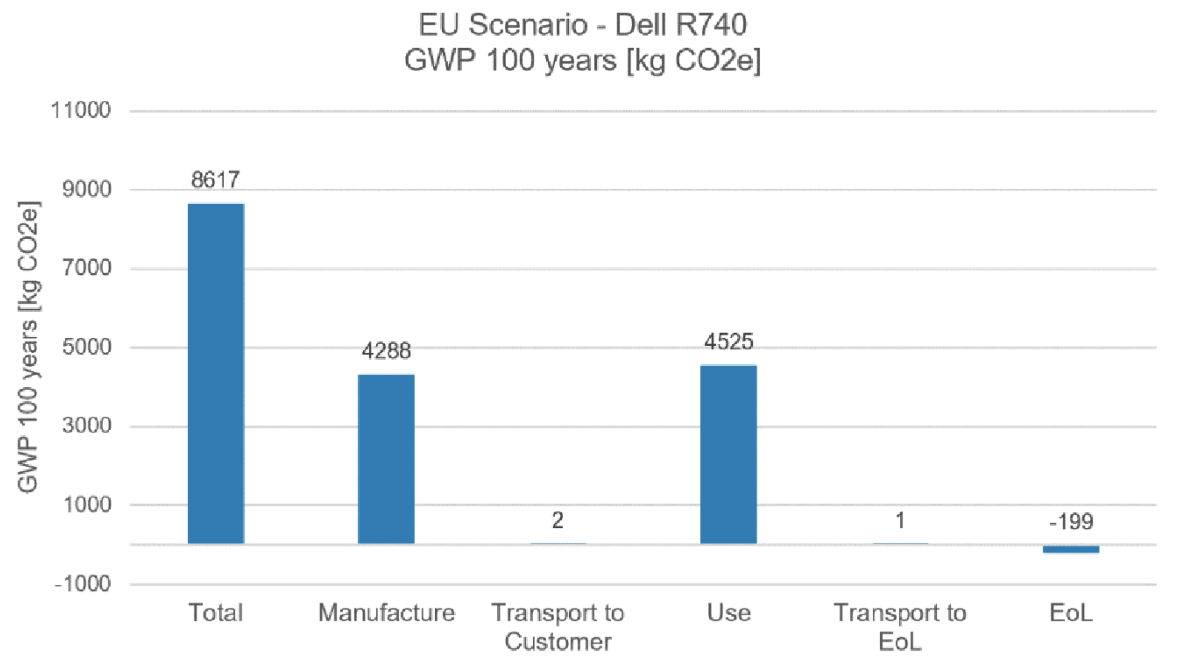


Figure 1: Contribution of the different stages of the lifecycle to the GWP of the Dell R740 (EU)

Embodied Footprint for Long Term Storage

- Information not published by cloud providers
- Little information on storage system composition
  - SSD, HDD, Data Tape
- Archival storage is a special case
  - Infrequent access
  - Large data volumes
  - Lower energy consumption
- Embodied footprint depends on media type and size
  - SSDs are worse than HDDs!
  - 1 HDD capacity could be anywhere between 1–20TB

The Dirty Secret of SSDs: Embodied Carbon

Swamit Tannu
University of Wisconsin-Madison
swamit@cs.wisc.edu

Prashant J Nair
University of British Columbia
prashantnair@ece.ubc.ca

Abstract

Scalable Solid-State Drives (SSDs) have revolutionized the way we store and access our data across datacenters and hand-held devices. Unfortunately, scaling technology can have a significant environmental impact. Across the globe, most semiconductor manufacturing use electricity that is generated from coal and natural gas. For instance, manufacturing a Gigabyte of Flash emits 0.16 Kg CO_2 and is a significant fraction of the total carbon emission in the system. We estimate that manufacturing storage devices has resulted in 20 million metric tonnes of CO_2 emissions in 2021 alone.

To better understand this concern, this paper compares the sustainability trade-offs between Hard Disk Drives (HDDs) and SSDs and recommends methodologies to estimate the embodied carbon costs of the storage system. In this paper, we outline four possible strategies to make storage systems sustainable. First, this paper recommends directions that help select the right medium of storage (SSD vs HDD). Second, this paper proposes lifetime extension techniques for SSDs. Third, this paper advocates for effective and efficient recycling and reuse of high-density multi-level cell-based SSDs. Fourth, specifically for hand-held devices, this paper recommends leveraging elasticity in cloud storage.

1 Introduction

Manufacturing, operating, transporting, and recycling computing systems, directly and indirectly, emit carbon dioxide (CO_2) and other greenhouse gases. As computing systems scale, their greenhouse contributions significantly impact global warming. This is highlighted by the pervasiveness of computing via hand-held devices, such as smartphones and tablets, and web services built around them. Moreover, digital data creation and consumption across the globe is snow bowling. As a result, carbon emissions due to personal devices, data centers, and networking infrastructure (known as the information and Communication Technologies (ICT) sector) are increasing rapidly. Today, about 2% of the total carbon emissions

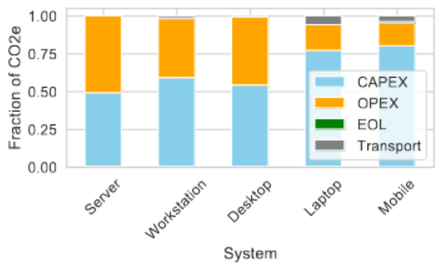


Figure 1: Breakdown of CO2e in Manufacturing (CAPEX) Operations (OPEX), Transport, and End of Life (EOL) phases.

are estimated due to computing and networking devices combined [22, 23], and it is estimated to double in the next decade. For example, the average household in the US has five to ten devices connected to the internet [30, 31]. We estimate that manufacturing and operating these devices for a year emits 2000 Kg CO_2 – equivalent to CO_2 emissions from driving a car for 5000 miles [20].

Most of the carbon emissions are because of the "conventional" electricity [6] that is used in the manufacturing and operation of computing systems [25]. For example, running and cooling the computing and networking hardware consumes significant electricity. If this electricity is generated from conventional carbon-intensive sources such as coal, natural gas, and crude oil, it will contribute to global warming. In contrast, electricity generated from renewable sources such as wind, solar, nuclear, and hydroelectric have a significantly small Global Warming Potential (GWP). Unfortunately, irrespective of whether they are hand-held devices or server nodes, manufacturing hardware and/or operating them require a significant amount of electricity – often from carbon-intensive conventional sources.

https://arxiv.org/pdf/2207.10793.pdf

Embodied Footprint of Servers and Storage

- Storage
  - Storage Embodied Factor: kgCO2eq per GB
  - HDD lifetime 4-6 years
  - Hard Drive 5 kgCO2eq per TB per year
- Servers
  - Cloud server lifetime 4-6 years
  - Cloud server utilization 50 – 65%
  - 1 core-hour ~ 0.5 gCO2eq
- Data Tape Libraries and media
  - Deep archive LTO 1 kgCO2eq per TB per year

https://arxiv.org/pdf/2207.10793.pdf

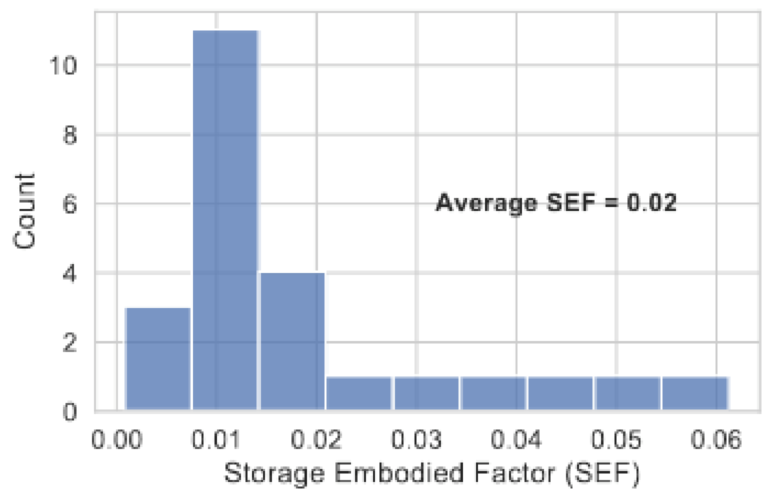


Figure 6: Distribution of estimated storage embodied factor for 24 Hard Drive Disks (HDDs).

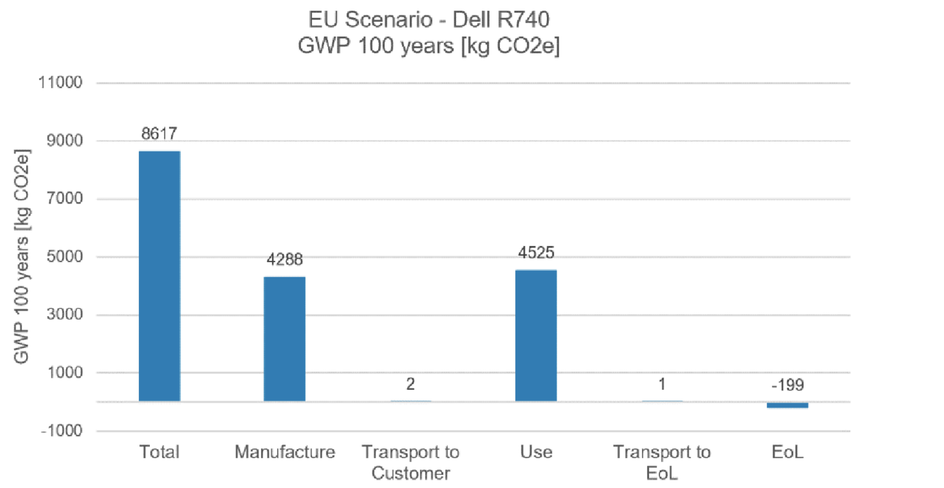


Figure 1: Contribution of the different stages of the lifecycle to the GWP of the Dell R740 (EU)

https://corporate.delltechnologies.com/content/dam/digitalassets/active/en/unauth/data-sheets/products/servers/lca_poweredge_r740.pdf

CALCULATING THE CARBON FOOTPRINT OF DIGITAL PRESERVATION
https://drive.google.com/file/d/1kMzU9cL975sR_1JwiQ-Rq8kRTp76bl_z/view

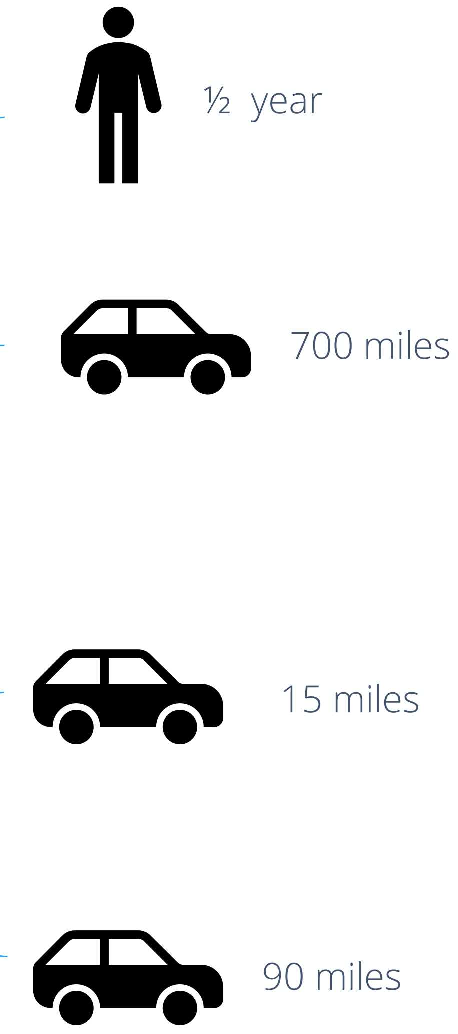
Carbon Emissions From ICT Embodied Footprint

Large Astronomy Research Datasets

	Embodied Footprint
1 PB data stored for 1 year	4000 kgCO2 eq
1 PB ingest of large image files	200 kgCO2 eq

Large collections of office files

	Embodied Footprint
1M office files stored for 1 year	4 kgCO2 eq
Ingest of 1M office files.	25 kgCO2 eq



Carbon Emissions from LTDP

Carbon Emissions from LTDP

Large image Datasets

	Gross Emissions from Energy Consumption	Estimated Embodied Footprint
1 PB data stored for 1 year	7800 kgCO2 eq	4000 kgCO2 eq
1 PB ingest of large image files	1600 kgCO2 eq	200 kgCO2 eq

Large collections of office files

	Gross Emissions from Energy Consumption	Estimated Embodied Footprint
1M office files stored for 1 year	5.5 kgCO2 eq	4 kgCO2 eq
Ingest of 1M office files.	140 kgCO2 eq	25 kgCO2 eq

The net carbon emissions from energy use are zero, the embodied footprint isn't!

Other Measurements and Calculations

	Date	Carbon Footprint from Energy Consumption	Embodied Carbon Footprint	Reference
Virginia Tech University, US	2022	Yes		https://osf.io/caub7
Digital Heritage Network, Netherlands	2022	Yes	Yes	https://doi.org/10.5281/zenodo.6341483
CSC, Finland	2023	Yes	Yes	https://drive.google.com/file/d/1kMzU9cL975sR_1JwiQ-Rq8kRTp76bl_z/view
The National Archives, UK	2023	Yes		https://www.nationalarchives.gov.uk/archives-sector/digital-services-and-carbon-emissions-in-the-heritage-sector-some-preliminary-findings/

LTDP carbon emissions into an organizational context

	Emissions (tonnes of CO2)			
Institution	2017-18	2018-19	2019-20	2020-21
National Trust			863,838	605,751
British Film Institute				19,684
Imperial War Museum	24,639	22,763	20,605	11,278
Natural History Museum	11,196	11,139	10,616	11,258
Royal Botanic Gardens, Kew	8,993	7,717	9,284	6,994
British Library	10,000	9,000	7,500	6,000
British Museum	8,516	7,080	7,164	5,861
National Gallery		5,762	5,391	4,716
English Heritage			3,993	3,591
Science Museum		3,548	3,563	2,541
National Maritime Museum		4,659		2,186
Historic Royal Palaces			4,605	1,993
National Library of Scotland	1,226	995	984	777

https://www.nationalarchives.gov.uk/archives-sector/digital-services-and-carbon-emissions-in-the-heritage-sector-some-preliminary-findings/

Carbon Emissions from LTDP On Premise

On-premise LDTP

- Server utilization
  - On prem — 15%
  - Cloud — 65%
- Data Centre efficiency (PUE)
  - On premise — 1.2 – 1.5
  - Cloud — 1.1
- Energy Mix
  - On prem — depends on local energy mix, e.g. 50%
  - Cloud — rapidly heading to 100% renewable
- Embodied ICT
  - On prem — inefficiencies in transport, recycling and disposal
  - Cloud — efficiencies of scale, extended lifetimes

LTDP Outside of the Cloud

If you do LTDP using your own facilities:

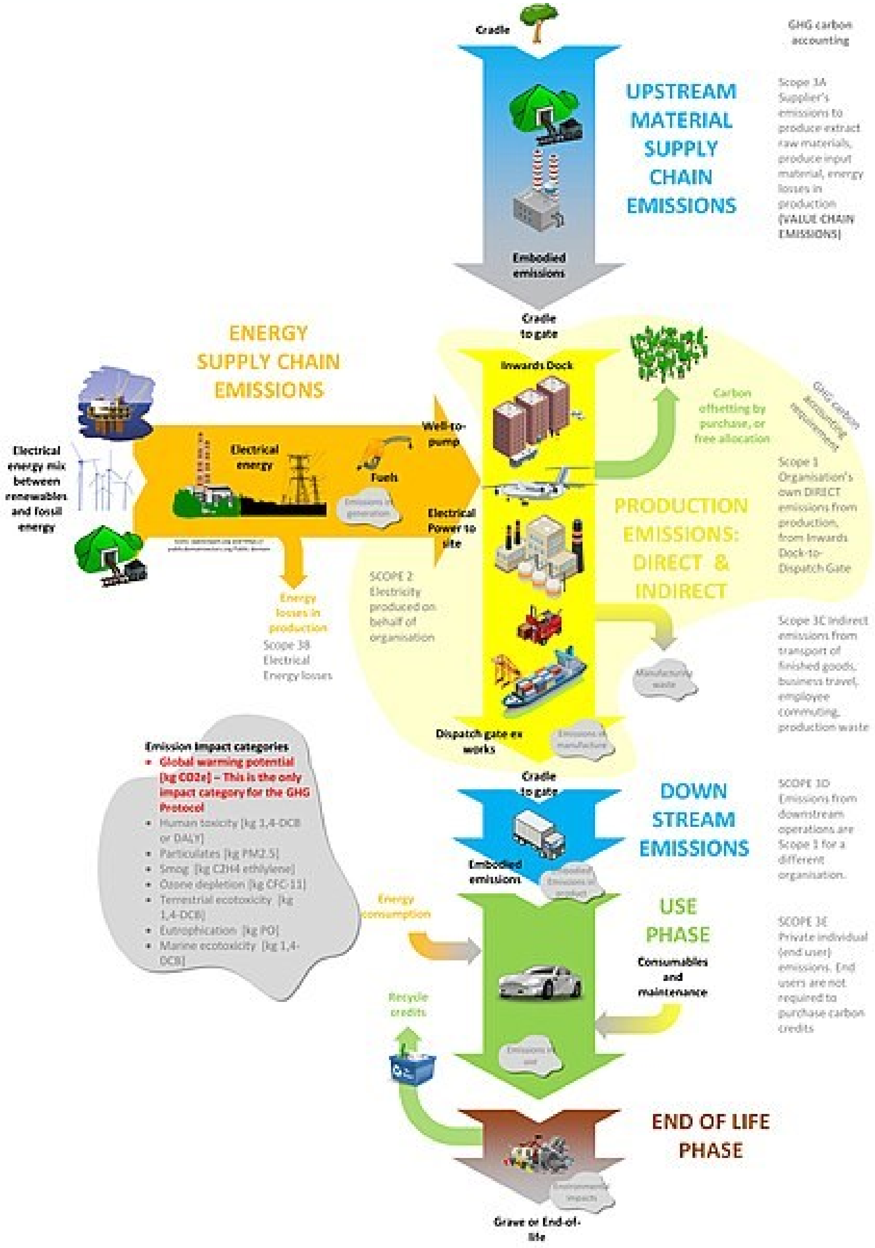
- Is your energy mix 100% renewable?
- How does your energy efficiency compare to AWS, Google, Azure?
- Are your hardware utilization levels as high as they could be?
- Are you getting the longest lifetime possible out of servers?
- Do you have a circular approach to hardware recycling?

Summary

Message 1: Worry about embodied footprint

- Carbon footprint from energy usage can be zero, but embodied footprint isn't
- ICT servers and storage have a significant embodied carbon footprint

- Keep less
- Do less
- Be efficient
- Share resources



Message 2: Carbon Footprint Of LTDP Can Vary Hugely!

- Carbon footprint of LTDP depends on:
  - Type and volume of data
  - Processing that gets applied
  - Where the processing takes place
  - When processing takes place
  - What tools and systems are used
  - How the data is stored and how often it is accessed

Measure your footprint

$\Longrightarrow$ Understand what makes the most contributions

$\Longrightarrow$ Make targeted reductions

arkivum
Bringing archived data to life

QUESTIONS?

Matthew Addis

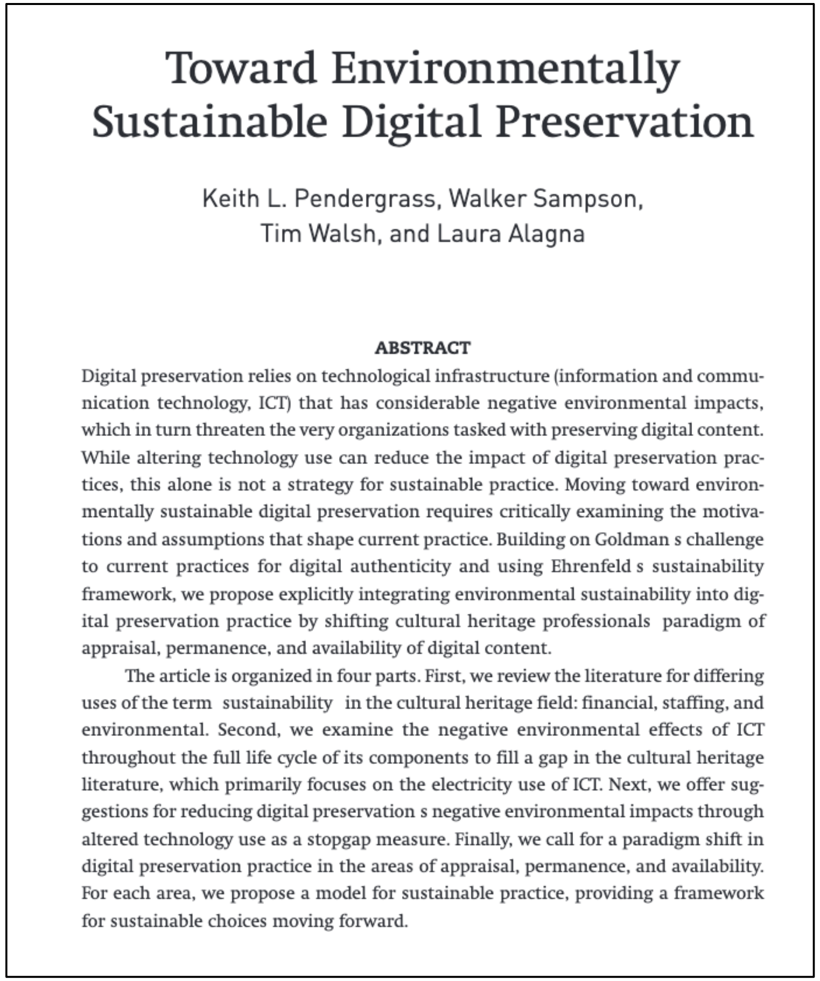
matthew.addis@arkivum.com

https://orcid.org/0000-0002-3837-2526

Resources

2019: Toward Environmentally Sustainable Digital Preservation

- Pendergrass, Keith L., Walker Sampson, Tim Walsh, and Laura Alagna. 2019. "Toward Environmentally Sustainable Digital Preservation." The American Archivist 82 (1): 165–206
- DPC Webinar April 2020: "Enacting Environmentally Sustainable Preservation"



Toward Environmentally Sustainable Digital Preservation

Keith L. Pendergrass, Walker Sampson, Tim Walsh, and Laura Alagna

ABSTRACT

Digital preservation relies on technological infrastructure (information and communication technology, ICT) that has considerable negative environmental impacts, which in turn threaten the very organizations tasked with preserving digital content. While altering technology use can reduce the impact of digital preservation practices, this alone is not a strategy for sustainable practice. Moving toward environmentally sustainable digital preservation requires critically examining the motivations and assumptions that shape current practice. Building on Goldman s challenge to current practices for digital authenticity and using Ehrenfeld s sustainability framework, we propose explicitly integrating environmental sustainability into digital preservation practice by shifting cultural heritage professionals paradigm of appraisal, permanence, and availability of digital content.

The article is organized in four parts. First, we review the literature for differing uses of the term sustainability in the cultural heritage field: financial, staffing, and environmental. Second, we examine the negative environmental effects of ICT throughout the full life cycle of its components to fill a gap in the cultural heritage literature, which primarily focuses on the electricity use of ICT. Next, we offer suggestions for reducing digital preservation s negative environmental impacts through altered technology use as a stopgap measure. Finally, we call for a paradigm shift in digital preservation practice in the areas of appraisal, permanence, and availability. For each area, we propose a model for sustainable practice, providing a framework for sustainable choices moving forward.

https://dash.harvard.edu/handle/1/40741399

DPC: Environmentally Sustainable Digital Preservation

- Blogs
- Reports
- Webinars
- Presentations

Home > Digital Preservation > Discover Good Practice > Environmentally Sustainable Digital Preservation

Environmentally Sustainable Digital Preservation

Environmental Sustainability

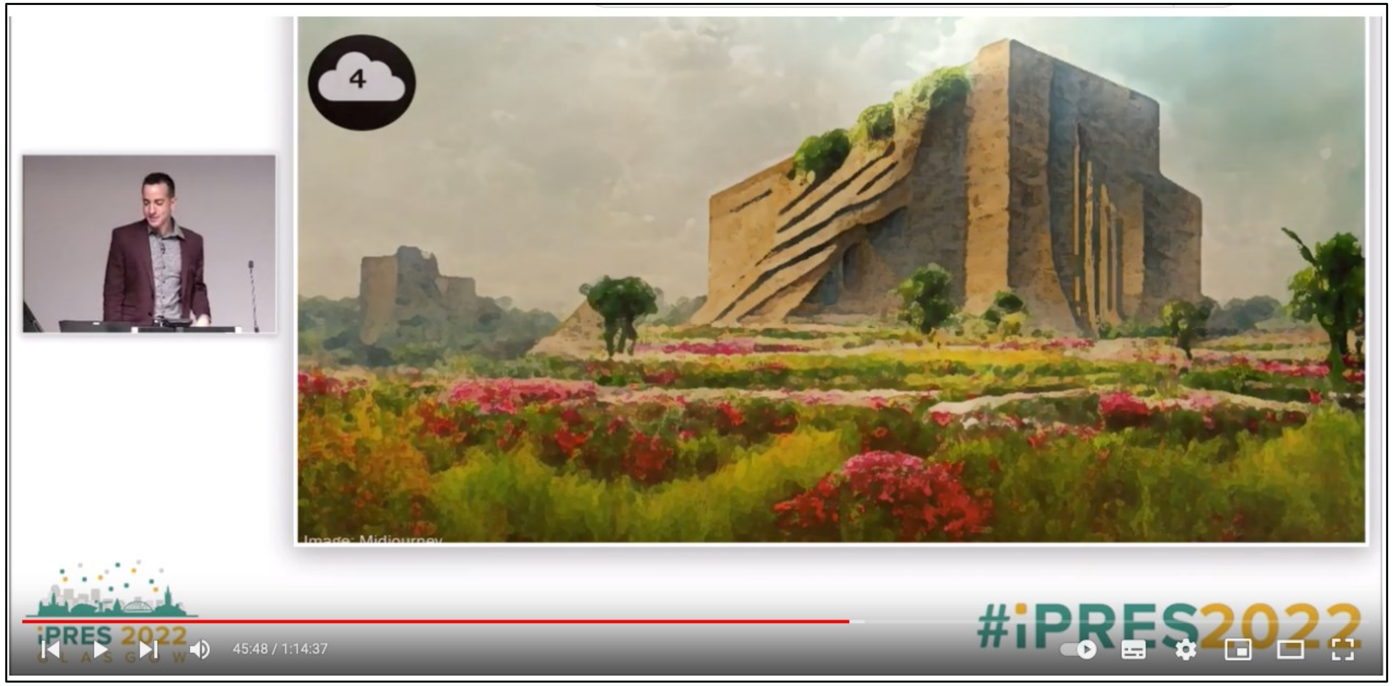
Digital preservation good practice is not solely about how successfully we preserve the bits and enable access to them, it must also take into account the broader context in which our work sits, and the wider responsibilities we have to society and the environment. Simply put, there is no point in preserving the bits if there is no one left to read and understand them. As a community we must therefore balance risks to the digital content that we hold not only against the financial cost but also the **cost to the environment**. We must consider how we reduce the environmental impact of our work, whilst continuing to maintain our valuable digital content for future generations. This is a challenging balancing act and we must work together as a community to evolve digital preservation good practice to minimise the environmental impact of our actions.

Environmentally sustainable digital preservation is not a new topic for the DPC and the wider digital preservation community, but it is certainly one that is growing in urgency. The DPC first addressed this topic in 2010 with an article in our newsletter from William Kilbride entitled **'Here comes the tide'** and William's involvement in a panel discussion at the iPRES conference **'How Green is Digital Preservation'**. In more recent years, other voices in the community have joined in this call to action and now we have a more substantial volume of content on this topic scattered across the DPC website.

https://www.dpconline.org/digipres/discover-good-practice/environmentally-sustainable-digital-preservation

iPRES 2022: Environmental Sustainability Sessions

- Green Goes with Anything: Decreasing Environmental Impact of Digital Libraries at Virginia Tech
- Seeking Sustainability: Developing a Modern Distributed Digital Preservation System, Penn State University Libraries
- The C02 Emissions of Storage and use of Digital Objects and Data. Exploring Climate Actions, Dutch National Archives / Digital Heritage Network
- After the Cloud: Rethinking Data Ecologies through Anthropology & Speculative Fiction, Steven Gonzalez Monserrate.



https://youtu.be/pFCqgmLgqzg

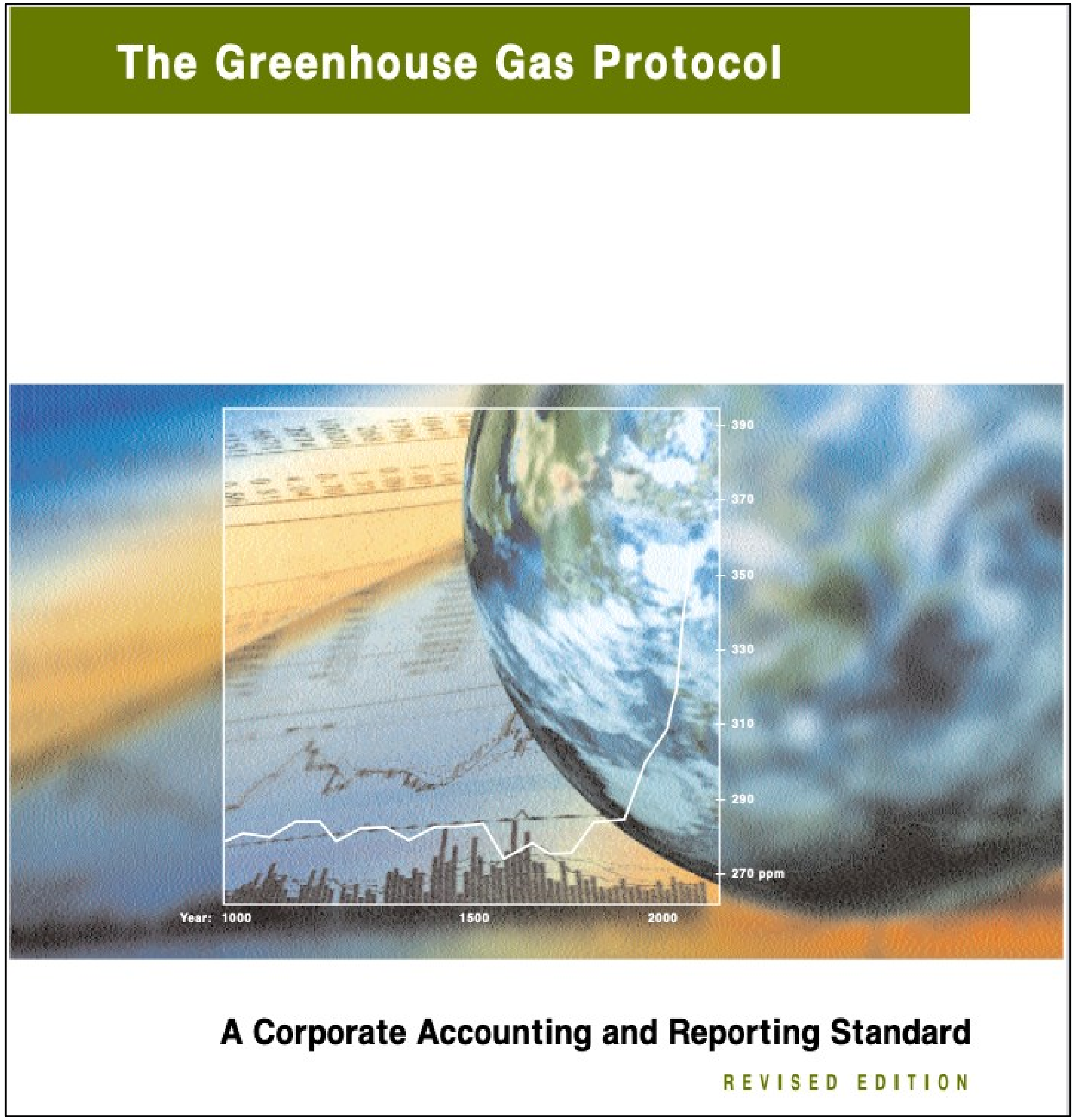
https://osf.io/caub7 (Virginia Tech)
https://osf.io/v9ub8/ (Penn State)
https://osf.io/7cbmd/ (Dutch National Archives)

iPRES 2023: Environmental Sustainability Sessions

- Mikko Tiainen and colleagues from CSC on Calculating the Carbon Footprint of Digital Preservation – A Case Study
  - https://drive.google.com/file/d/1kMzU9cL975sR_1JwiQ-Rq8kRTp76bl_z/view
- University of Illinois on The Curricular Asset Warehouse At The University Of Illinois: A Digital Archive’s Sustainability Case Study.
  - https://drive.google.com/file/d/1Ov2Q5X0f7QcnL0oJ0EpiMmjeAgdDpq8Q/view
- Tipping Point: Have we gone past the point where we can handle the Digital Preservation Deluge?
  - https://www.ideals.illinois.edu/items/128305

Frameworks and Standards

- GHG Protocol supplies the world's most widely used greenhouse gas accounting standards.
- Structured framework for thinking about emissions in supply chains

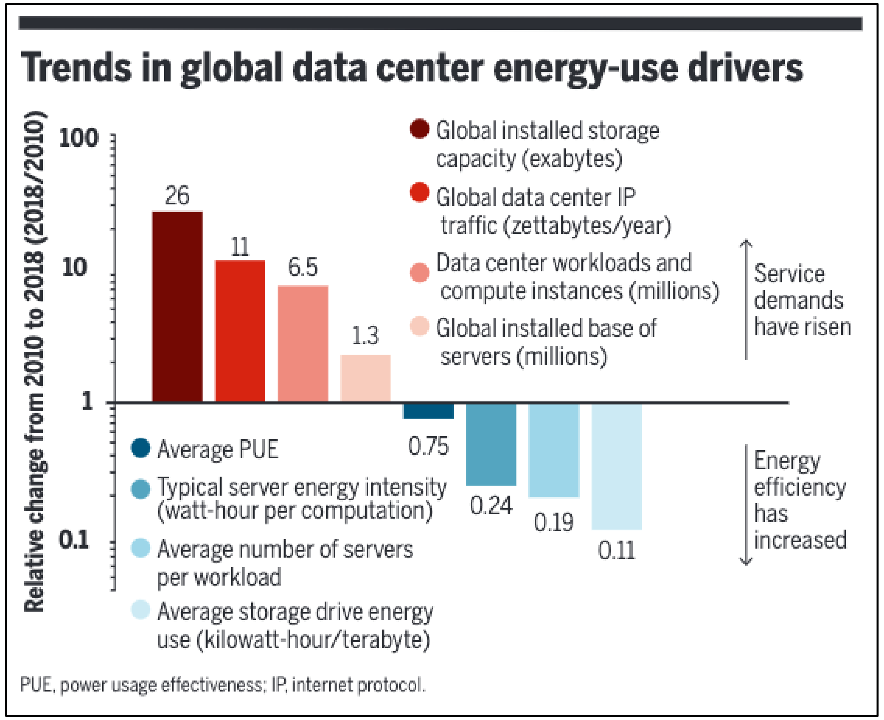


GREENHOUSE GAS PROTOCOL

https://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf

Fact Check of Energy Consumption Trends in the Cloud

- Many claims about cloud carbon footprint and energy usage are out of date or use flawed projections



POLICY FORUM

ENERGY

Recalibrating global data center energy-use estimates

Growth in energy use has slowed owing to efficiency gains that smart policies can help maintain in the near term

By Eric Masanet[1,2], Arman Shehabi[3], Nuoa Lei[1], Sarah Smith[3], Jonathan Koomey[4]

Data centers represent the information backbone of an increasingly digitalized world. Demand for their services has been rising rapidly (1), and data-intensive technologies such as artificial intelligence, smart and connected energy systems, distributed manufacturing systems, and autonomous vehicles promise to increase demand further (2). Given that data centers are energy-intensive enterprises, estimated to account for around 1% of worldwide electricity use, these trends have clear implications for global energy demand and must be analyzed rigorously. Several oft-cited yet simplistic analyses claim that the energy used by the world's data centers has doubled over the past decade and that their energy demand for data center services rises rapidly, so too must their global energy use. But such extrapolations based on recent service demand growth indicators overlook strong countervailing energy efficiency trends that have occurred in parallel (see the first figure). Here, we integrate new data from different sources that have emerged recently and suggest more modest growth in global data center energy use (see the second figure). This provides policy-makers and energy analysts a recalibrated understanding of global data center energy use, its drivers, and near-term efficiency potential.

Assessing implications of growing demand for data centers requires robust understanding of the scale and drivers of global data center energy use that has eluded many policy-makers and energy analysts. The reason for this blind spot is a historical lack of "bottom-up" information

As demand for data centers rises, energy efficiency improvements to the IT devices and cooling systems they house can keep energy use in check.

Bottom-up analyses tend to best reflect this broad range of factors, generating the most credible historical and near-term energy-use estimates (7). Despite several recent national studies (8), the latest fully replicable bottom-up estimates of global data center energy use appeared nearly a decade ago. These estimates suggested that the worldwide energy use of data centers had grown from 153 terawatt-hours (TWh) in 2005 to between 203 and 273 TWh by 2010, totaling 1.1 to 1.5% of global electricity use (9).

Since 2010, however, the data center landscape has changed dramatically (see the first figure). By 2018, global data center workloads and compute instances had increased more than sixfold, whereas data center internet protocol (IP) traffic had increased by more than 10-fold (1). Data center storage capacity has also grown rapidly, increasing by an estimated factor of 25 over the same time period (1, 8). There has been a tendency among analysts to use such service demand trends to simply extrapolate earlier bottom-up energy values, leading to unreliable predictions of current and future global data center energy use (3–5). They might, for example, scale up previous bottom-up values (e.g., total data center energy use in 2010) on the basis of the growth rate of a service demand indicator (e.g., growth in global IP traffic from 2010 to 2020) to arrive at an estimate of future energy use (e.g., total data center energy use in 2020).

But since 2010, electricity use per computation of a typical volume server—the workhorse of the data center—has dropped by a factor of four, largely owing to processor-efficiency improvements and reductions in idle power (10). At the same time, the watts per terabyte of installed storage has dropped by an estimated factor of nine owing to storage-drive density and efficiency gains (8). Furthermore, growth in the number of servers has slowed considerably owing to a fivefold increase in the average number of compute instances hosted per server (owing to virtualization), alongside steady reductions in data center power usage effectiveness (PUE, the total amount

https://datacenters.lbl.gov/sites/default/files/Masanet_et_al_Science_2020.full_.pdf

Matthew Addis' Blog Posts, Webinars and Reports

- Is digital preservation bad for the environment?
  - https://www.dpconline.org/blog/is-digital-preservation-bad-for-the-environment
- iPRES 2022: Climate Change and Environmental Sustainability
  - https://www.dpconline.org/blog/ipres-2022-climate-change-and-environmental-sustainability
- Does net zero emissions from energy usage in the cloud mean carbon free digital preservation is on the horizon?
  - https://www.dpconline.org/blog/blog-matthew-addis-enviornmental-23
- Quantified Carbon Footprint of Long-Term Digital Preservation in the Cloud
  - https://doi.org/10.6084/m9.figshare.20653101
- What is the carbon footprint of large-scale global digital preservation?
  - https://www.dpconline.org/blog/blog-matthew-addis-ipres23
- Webinar Recording: Environmental Sustainability of Digital Preservation in the Cloud
  - https://arkivum.com/webinar-environmental-sustainability-of-digital-preservation-in-the-cloud/

Yet More Reading!

- The Climate Impact of ICT
  - https://www.gla.ac.uk/media/Media_848209_smxx.pdf
- The carbon footprint of servers
  - https://www.goclimate.com/blog/the-carbon-footprint-of-servers/
- Environmental Footprint of Data Centers in the US
  - https://iopscience.iop.org/article/10.1088/1748-9326/abfba1
- Digital Preservation's Impact on the Environment
  - https://www.dropbox.com/s/csdc0ije7rru2j6/ALA_EnvironmentallySustainablePreservation_Tadic_20220428.pptx
- Walking a tightrope across the gap of digital preservation and environmental sustainability
  - https://kia.pleio.nl/attachment/entity/931f65cb-2058-4fe9-a500-99bc53dfde40
- Chasing Carbon: The Elusive Environmental Footprint of Computing
  - https://discovery.ucl.ac.uk/id/eprint/10147559/1/Chasing_Carbon_The_Elusive_Environmental_Footprint_of_Computing.pdf
- Cloud carbon footprint: Do Amazon, Microsoft and Google have their head in the clouds?
  - https://www.carbone4.com/en/analysis-carbon-footprint-cloud
- Digital Services and carbon emissions in the heritage sector: some preliminary findings
  - https://www.nationalarchives.gov.uk/archives-sector/digital-services-and-carbon-emissions-in-the-heritage-sector-some-preliminary-findings/

https://www.carbonindependent.org/

Comparisons

280 gCO2eq	1 mile travelled in an average sized car
250 kgCO2eq	1 hour per passenger on an international flight
10 TonnesCO2eq	UK average carbon footprint per year per person
50 TonnesCO2eq	Lifetime carbon budget per person from 2020 to stay within 1.5C global temperature rise