THE ENVIRONMENTAL FOOTPRINT OF DATA STORAGE OR WHAT CAN LIE BEHIND A GIGABYTE

Adrien Berthelot
University of Bristol
School of Computer Science
`adrien.berthelot@bristol.ac.uk`

Daniel Schien
University of Bristol
School of Computer Science
`daniel.schien@bristol.ac.uk`

What is the cost of a gigabyte? - Researchers working on the environmental impacts of information technologies (IT) often had to answer this question at least once. The first step in answering this question is generally to ask: what type of gigabyte (GB) are we talking about? Is it the cost of computing, transferring through networks or storing one gigabyte? This ambiguity, although seemingly trivial, nevertheless leads several scientific publications, even those from reputable academic journals, to produce misguided works [1]. In this paper, we will focus solely on the footprint associated with storing one gigabyte of data. We will explain why this issue is relevant in the current state of research on the environmental footprint of IT and why this simple question calls for complex answers. Finally, we will present our current research position on the footprint of storage as a service and its challenges.

1 Motivation

It seems, the amount of data generated and stored is growing rapidly. It is difficult to find data from verifiable sources, but estimates seem to agree on the order of magnitude, e.g. for the year 2025 we might globally store 175 ZB [2], 181 ZB [3] or up to 200 ZB [4] of data. One of the admitted starting point for this growth is the process of "digital transition" or digitalisation, i.e. the transition of IT from analogue to digital representation in storage; e.g. vinyl record to CD, which resulted in, more than a transfer, an explosion in data storage [5]. This significant and sustained growth in the total volume of stored data is accompanied by an increase of data processing (compute) and transfer (network) [6]. All of these contribute to the environmental footprint of digital, whether it is contribution to global warming [7] or other environmental impact like minerals and metals extraction [8]. However, compared to networking and compute, the storage footprint remains under-studied [9], even though data storage is a key function of any digital services.

This lack of research contributes to uncertainty and confusion about how to reduce the environmental impact of data storage. For example, to reduce the footprint of data storage, some publications will prefer HDD storage technology over SSD [9, 10], while others will prefer the opposite [11, 12]. This confusion extends also to information about the breakdown of the carbon footprint between embodied and operational. Naturally, as highlighted by [13], the operational footprint of storage depends mainly on the footprint of different used electricity mix, especially if the storage is distributed between different geographical location. Moreover the research field in Cloud Computing tends to focus mainly on issue like energy efficiency, therefore the focus is often put on operational impact. This confusion adds to the more general disorientation surrounding the measurement of the digital footprint [14] and can lead to potentially counterproductive incentives. For example, focusing on short-term reduction in data volume [15] or recommanding users to "send less emails" [16] has uncertain benefits in terms of environmental impacts.

2 Observation

One of the key factors for the uncertainties regarding the footprint of data storage can be summarised as follows: Behind the same generic requirement—storing one GB of data for one year—there is a wide variety of technical solutions. These technical solutions deploy sets of equipment with different environmental footprints, may consume electricity mixes with different footprints, and may be non-exclusive, i.e. a combination of technical solutions may be used at the same time for the same use-case. Traditionally, this heterogeneity of storage media is based on a logic known as hierarchical storage management (HSM) [17].

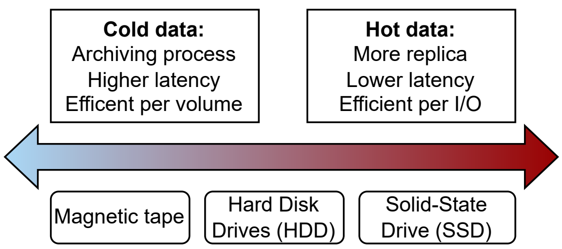

Figure 1: Placing the different technical solutions for data storage on a spectrum from cold(left) to hot(right) data

In this context, each type of data is stored in the most appropriate equipment according to the frequency of access to those data and its volume. Each piece of data can be placed on a cold-to-hot spectrum, where hot data are generally small volumes of data accessed frequently, and cold data are large volumes of data accessed infrequently. As described in Figure 1, different storage technologies are preferred according to data status on this spectrum, from magnetic tape storage [18] to NAND flash memory [11]. The characteristics of these technologies can be analysed in terms of efficiency per stored volume for cold data or efficiency per I/O operation for hot data. This is also reflected in their footprint, with greater energy efficiency for solutions at the hot end of the spectrum but a larger manufacturing footprint. The concept also extends beyond type of devices to choice of interface, e.g. SATA or NVMe [19], and protocol, e.g. RAID 1 or RAID 5 [20].

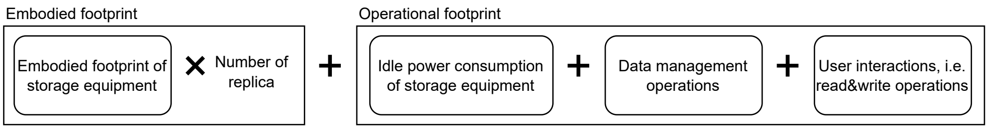

Figure 2: Summarizing the different embodied and operational costs linked to storing a defined volume of data, perimeter only includes the storage equipment in cloud storage context

In addition to the diversity of technical solutions, sometimes applied in combination, assessing the footprint of storage presents difficulties that are not yet fully addressed by existing methodological standards [21]. Whether we use methods such as carbon footprinting or life cycle assessment (LCA), attributing the embedded footprint as operational remains complex. As described in Figure 2, in order to estimate the footprint associated with storing a volume of data, e.g., one GB for one year, you will need to allocate several costs according to different allocation rules. The footprint of the storage servers must be allocated and taking into account replicas. But should this footprint be allocated proportionally to the volume of data stored or to dedicated I/O operations (in each case in relation to the total data stored on the server or I/O operations performed)? This choice among alternative allocation keys is critical because, as shown in Figure 1, the equipment involved is optimised for one of these two metrics. In addition, estimating the operational footprint also poses challenges with the inclusion of different types of electricity consumption, both non-dynamic, such as idle consumption and data maintenance operations [22], and dynamic, such as read-write operations. The attribution of the footprint of these non-dynamic consumption patterns and the fact that they may occur in different locations, i.e. with different electricity mixes, pose additional challenges to estimating the footprint.

Currently few, if none, end-to-end perspectives on data storage exist. The research field is very fragmented between energy efficiency in database management systems looking to optimise joules per query [23], Cloud storage management applying e.g. carbon aware computing to this process [22, 24], computer architecture [25] and information systems [26, 27]. This specialisation of research undoubtedly contributes to local efficiency gains but limits the overall understanding of the environmental footprint of storage and can contribute to confusing or counterproductive decisions, as storage solutions are dynamic and heterogeneous.

3 Position

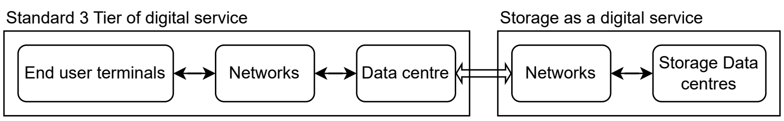

Figure 3: Describing storage a sub-digital service used by other standard digital services

In response to these gaps in the literature and the challenges of understanding the environmental impact of storage, we want to propose a new approach to this topic. We want to apply LCA, with its benefits [28], to storage services. But above all, we want to treat storage as a digital service in itself. As illustrated by Figure 3, we want to build on the already existing approach for LCA of digital services [29, 30, 31], as storage *as a service (StoaS)* is an underlying function of a vast number of digital services provided by specialised infrastructures. StoaS includes examples such as photo backup from smartphones all the way to storing particle collision data at CERN [32]. This approach will improve the quality of assessments and thus provide more reliable and accurate recommendations or environmental impact factors. This approach differs from others, including sectorial top-down models such as the sustainable web development model [33], or modelling an intensity of storage from too narrow bottom-up information, e.g. single disc energy intensity cloud carbon footprint [34] or BoaviztAPI [35]. We hope that with new results we will be able to better guide recommendations, build more comprehensive digital footprint models, and provide broader perspectives for research on reducing the storage footprint.

References

[1] Daniel Schien, Adrien Berthelot, Paul Shabajee, and Chris Preist. Global datacentre electricity consumption is not correlated with data demand: Responding to implausible assumptions and flawed modelling in castro et al. (2024). *Energy Policy*, 209:114947, 2026.

[2] Reinsel David, Gantz John, and Rydnin John. Data age 2025: The evolution of data to life-critical. Technical report, International Data Corporation (IDC), 2017. https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf.

[3] Statista. Total installed based of data storage capacity in the global datasphere from 2020 to 2025. https://www.statista.com/statistics/1185900/worldwide-datasphere-storage-capacity-installed-base/.

[4] Cybercrime Magazine. The world will store 200 zettabytes of data by 2025, 2020. https://cybersecurityventures.com/the-world-will-store-200-zettabytes-of-data-by-2025/.

[5] Martin Hilbert. Digital technology and social change: the digital transformation of society from a historical perspective. *Dialogues in Clinical Neuroscience*, 22(2):189–194, 2020.

[6] IEA. Global trends in digital and energy indicators, 2015-2022. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks#Energy.

[7] Charlotte Freitag, Mike Berners-Lee, Kelly Widdicks, Bran Knowles, Gordon S. Blair, and Adrian Friday. The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. *Patterns*, 2(9):100340, September 2021.

[8] Bordage, F., de Montenay, L., Benqassem, S., Delmas-Orgelet, J., Domon, F., Prunel, D., Vateau, C. et Lees Perasso, E. Le numérique en europe : une approche des impacts environnementaux par l'analyse du cycle de vie, December 2021.

[9] Sara Mcallister, Fiodar Kazhamiaka, Daniel S. Berger, Rodrigo Fonseca, Kali Frost, Aaron Ogus, Maneesh Sah, Ricardo Bianchini, George Amvrosiadis, Nathan Beckmann, and Gregory R. Ganger. A Call for Research on Storage Emissions. *SIGENERGY Energy Inform. Rev.*, 4(5):67–75, 2025.

[10] Baolin Li, Rohan Basu Roy, Daniel Wang, Siddharth Samsi, Vijay Gadepally, and Devesh Tiwari. Toward sustainable HPC: Carbon footprint estimation and environmental implications of HPC systems. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '23, pages 1–15. Association for Computing Machinery, 2023. Place: New York, NY, USA.

[11] Julie Sinistore, Mukunth Natarajan, and Eric Christensen. Comparative life cycle assessment of pure storage FlashArray//x70. ISO 14044-CONFORMANT REPORT Version 1 12/2/2022, WSP USA, 12 2022.

[12] Eric Burgener. How does the embodied carbon dioxide equivalent of flash compare to HDDs? Pure Storage - https://blog.purestorage.com/perspectives/how-does-the-embodied-carbon-dioxide-equivalent-of-flash-compare-to-hdds/, 09 2024.

[13] Swamit Tannu and Prashant J. Nair. The dirty secret of SSDs: Embodied carbon. *SIGENERGY Energy Inform. Rev.*, 3(3):4–9, 2025.

[14] Anne Pasek, Hunter Vaughan, and Nicole Starosielski. The world wide web of carbon: Toward a relational footprinting of information and communications technology's climate impacts. *Big Data & Society*, 10(1):20539517231158994, 2023.

[15] Schien Daniel, Shabajee Paul J S, Burak Akyol Huseyin, Benson Luke, and Katsenou Angeliki. Assessing the carbon reduction potential for video streaming from short-term coding changes. In *16th International Conference on Quality of Multimedia Experience (QoMEX)*, International Workshop on Quality of Multimedia Experience, QoMEx, United States, mar 2024. Institute of Electrical and Electronics Engineers (IEEE). https://qomex2024.itec.aau.at/.

[16] Luciano Rodrigues Viana, Mohamed Cheriet, Kim-Khoa Nguyen, Daria Marchenko, and Jean-François Boucher. Sending fewer emails will not save the planet! an approach to make environmental impacts of ICT tangible for canadian end users. *Sustainable Production and Consumption*, 34:453–466, 2022.

[17] Harry Katzan. Storage hierarchy systems. In *Proceedings of the May 18-20, 1971, spring joint computer conference*, AFIPS '71 (Spring), pages 325–336, New York, NY, USA, 1971. Association for Computing Machinery.

[18] M.A. Lantz, S. Furrer, M. Petermann, H. Rothuizen, S. Brach, L. Kronig, I. Iliadis, B. Weiss, E.R. Childers, and D. Pease. Magnetic Tape Storage Technology. *ACM Transactions on Storage*, 21(1), 2025.

[19] Charles Jaranilla, Hojin Shin, Seehwan Yoo, Seong-je Cho, and Jongmoo Choi. Tiered storage in modern key-value stores: Performance, storage-efficiency, and cost-efficiency considerations. In *2024 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 151–158. IEEE, 2024.

[20] Ralf Süß and Yannik Süß. Storage and virtualization. In *IT Infrastructure: Security and Resilience Solutions*, pages 107–160. Springer, 2024.

[21] Julia Meyer, Alexis Perez, NegaOctet, Resilio, Kleis, and Hubblo. Product category rule (PCR) for cloud services and data centre IT hosting services. See section 12.3, https://librairie.ademe.fr/industrie-et-production-durable/6105-7697-product-category-rule-pcr-for-cloud-services-and-data-centre-it-hosting-services.html.

[22] Varsha Rao and Andrew A. Chien. Understanding the operational carbon footprint of storage reliability and management. *SIGENERGY Energy Inform. Rev.*, 4(5):180–187, April 2025.

[23] Hemasri Sai Lella, Rajrupa Chattaraj, Sridhar Chimalakonda, and Manasa Kurra. Towards comprehending energy consumption of database management systems-a tool and empirical study. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*, pages 272–281, 2024.

[24] Morgan Séguéla, Riad Mokadem, and Jean-Marc Pierson. Dynamic Energy and Expenditure Aware Data Replication Strategy. In *IEEE International Conference on Cloud Computing Technical Program (CLOUD 2022)*, Barcelona, Spain, July 2022. IEEE.

[25] Jaylen Wang, Daniel S. Berger, Fiodar Kazhamiaka, Celine Irvene, Chaojie Zhang, Esha Choukse, Kali Frost, Rodrigo Fonseca, Brijesh Warrier, Chetan Bansal, Jonathan Stern, Ricardo Bianchini, and Akshitha Sriraman. Designing cloud servers for lower carbon. In *2024 ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA)*, pages 452–470, 2024.

[26] Keyi Zhong, Tom Jackson, Andrew West, and Georgina Cosma. Building a sustainable knowledge management system from dark data in industrial maintenance. In Lorna Uden and I-Hsien Ting, editors, *Knowledge Management in Organisations*, pages 263–274. Springer Nature Switzerland, 2024.

[27] Victoria Woo, Shuai Yuan, Melanie Ng, Jun Woo Bak, and Duanyuan Xu. Digibyte: Promoting awareness and habit change for sustainable digital file decluttering. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, pages 1–8. Association for Computing Machinery, 2025.

[28] Hannes Bluhm, Daniela Wohlschlager, Johanna Pohl, Severin Beucker, Jan Bieser, Daniel Schien, Kelly Widdicks, Adrian Friday, and Gordon Shaw Blair. Understanding digitalization's environmental impact: why lca is essential for informed decision-making. *npj Climate Action*, 4(1):41, 2025.

[29] Adrien Berthelot, Eddy Caron, Romain de Laage, Laurent Lefèvre, and Alexis Nicolas. Fine-grained methodology to assess environmental impact of a set of digital services. https://hal.science/hal-04928998.

[30] Bryan Lopez Londoño, Shoaib Azizi, and Göran Finnveden. Incorporation of software in the life cycle assessment of an ICT service: A case study of an ICT service for energy efficiency in the transport sector. *Journal of Industrial Ecology*, 28(6):1965–1978, 2024. https://onlinelibrary.wiley.com/doi/pdf/10.1111/jiec.13570.

[31] Adrien Berthelot, Eddy Caron, Mathilde Jay, and Laurent Lefèvre. Understanding the environmental impact of generative AI services. *Communications of the ACM*, Special Issue on Sustainability and Computing:1–14, 2025. https://hal.science/hal-04920612.

[32] ATLAS Collaboration et al. The environmental impact, carbon emissions and sustainability of computing in the atlas experiment. *arXiv preprint arXiv:2505.08530*, 2025.

[33] Wholegrain Digital, Mightybytes, Footsprint, EcoPing, and Green Web Foundation. Estimating digital emissions: What is the sustainable web design model? https://sustainablewebdesign.org/estimating-digital-emissions.

[34] Cloud carbon footprint. https://www.cloudcarbonfootprint.org/docs/methodology#storage.

[35] Thibault Simon, David Ekchajzer, Adrien Berthelot, Eric Fourboul, Samuel Rince, and Romain Rouvoy. BoaviztAPI: a bottom-up model to assess the environmental impacts of cloud services. In *HotCarbon'24. Workshop on Sustainable Computer Systems*, Santa Cruz, United States, July 2024.