CALCULATING THE CARBON FOOTPRINT OF DIGITAL PRESERVATION

A Case Study

Mikko Tiainen
CSC – IT Center for Science
Finland
mikko.tiainen@csc.fi
https://orcid.org/0009-0000-8513-6262

Juha Lehtonen
CSC – IT Center for Science
Finland
juha.lehtonen@csc.fi
https://orcid.org/0000-0002-9916-5731

Heikki Helin
CSC – IT Center for Science
Finland
heikki.helin@csc.fi
https://orcid.org/0000-0003-4002-8203

Johan Kylander
CSC – IT Center for Science
Finland
johan.kylander@csc.fi
https://orcid.org/0000-0002-8084-8233

Abstract – Environmental sustainability is becoming an important factor in digital preservation. We have calculated the carbon footprint of our Finnish national digital preservation services, which we provide for cultural heritage and research sectors. We concentrate on the carbon footprint of manufacturing hardware and shipping the equipment to data centers, and the carbon footprint of the hardware service life and employees related to the services. Using data provided to us by the hardware manufacturers and other sources, we show that the majority of the emissions come from manufacturing and shipping of hardware, whereas the emissions created during the service life has a smaller role. As a whole, the annual carbon footprint of the services is smaller than the annual carbon footprint of three average Finns.

Keywords – sustainability, carbon footprint, data centers, hardware manufacturing, hardware service life

Conference Topics – Sustainability: Real and Imagined

I. INTRODUCTION

Our national digital preservation repository, funded by the Ministry of Education and Culture of Finland, provides services for preserving cultural heritage and research data [1]. Our concept includes two national services: (1) The Digital Preservation Service for Cultural Heritage (in production since 2015) preserves digital assets from the cultural heritage sector, represented by archives, libraries and museums, and (2) The Digital Preservation Service for Research Data (in production since 2019) preserves data from the research sector, represented by universities and other research institutes. Given the diversity of the user needs, the digital assets to be preserved make up a very heterogeneous whole while simultaneously requiring various and flexible solutions. Both of these services together are in this paper referred to as Digital Preservation Services (DPS). The technical solution behind the services is common for both cultural heritage data and research data.

The carbon footprint of an IT-service can typically be modeled by breaking the service down to its separate components. The hardware has a lifecycle carbon footprint starting from manufacturing the raw materials, transportation of the hardware, production usage, and lastly the disposal of the

iPRES 2023: The 19th International Conference on Digital Preservation, Champaign-Urbana, IL, US.

https://hdl.handle.net/2142/121104

hardware. On a data center level, data center power usage effectiveness (PUE) [2] is the driving factor together with hardware electricity usage when calculating the production usage carbon footprint. Enterprise level hardware vendors provide their own figures for the carbon footprint for their hardware.

In addition to the footprints mentioned above, the employee footprint includes emissions from offices, traveling, heating, waste management and so on. The employee footprint consists of carbon emissions resulting from the daily work of administrating, developing, and managing the DPS.

We calculate the total carbon footprint of our DPS in this paper. These calculations apply only to our current configuration and thus cannot be applied in general to any other DPS. They might however provide some general guidelines and insights for others.

For calculating the carbon footprint, our services can be divided into hardware, data centers, network, administration work, development work, and supporting ICT-services. The carbon footprint of constructing the data centers is not within the scope of our calculations: DPS's should in general be geographically distributed to several data centers. The density of data storage is now on a level where only a few server racks are needed to hold several petabytes of data. Therefore, our DPS does not need its own data centers and we utilize only a minor part of the existing data centers. The data centers thus facilitate many other IT-services in addition to our DPS.

The paper is divided as follows: In Chapter 2 we describe the hardware of our DPS, in Chapters 3 and 4 we show the carbon footprint of manufacturing and shipping the hardware and of the actual service life, in Chapter 5 we bind these findings together with some observations, and in Chapter 6, we conclude the paper with future work.

II. Hardware

Our DPS platform utilizes three separate data centers for storing preserved copies in order to reduce geographical risks. The available capacity of the DPS is currently 3.6 peta bytes per copy. Currently, the platform consists of the following hardware:

- 13 x HPE Proliant DL360 Gen10 frontend and validation servers (ingest)
- 10 x HPE Apollo 4510 Gen10 storage servers
- 4 x HPE Apollo 4200 Gen9 tape library front end servers
- 2 x HPE Apollo 4200 Gen10 tape library front end servers
- 2 x IBM TS4300 tape library with 7 IBM full height LTO-8 tape drives and 336 LTO-8 tapes
- 1 x IBM TS4300 tape library with 7 IBM full height LTO-9 tape drives and 231 LTO-9 tapes
- For tape drives 15 % duty cycle is estimated.

Our DPS platform also includes a dark archive storage for mitigating worst case disasters related to online storage copies. The dark archive can be divided into three components when making calculations about its footprint: (1) Dark archive copy manufacturing, (2) Dark archive copy logistics, and (3) Dark archive copy storage.

We are not required to have dedicated resources for dark archive logistics and storage as they are shared with multiple other customers. Logistics are organized into monthly transports to the dark archive.

The volume of the dark archive is the same as our DPS platform. Currently the dark archive consists of LTO-8 tapes stored in Pelican 1450 transport cases. The total number of these cases is 24, and the total number of dark archive LTO-8 media is 336.

III. Manufacturing and Shipping

The carbon footprints of hardware manufacturing and shipping (more accurately: raw materials, manufacturing, shipping, and disposal at the end of the life cycle) have been reported to us by the manufacturers. The calculations from both of the manufacturers are based on the Product Attribute to Impact Algorithm (PAIA) [3] and represent the status of the products in 2022. From these given calculations, Table 1 summarizes the carbon footprint of our DPS platform for hardware manufacturing and shipping.

The hardware components used in our DPS platform for ingesting and preserving contents can be divided into different roles. We can calculate the carbon footprint of the DPS platform based on the following roles: ingest, spinning disk storage, magnetic tape storage, and the dark archive. Fig. 1

depicts the relative size of the carbon footprint from manufacturing and shipping for each hardware role.

The carbon dioxide emissions for the last mile of transportation need to be calculated separately because the distances from the manufacturer sites to our data centers are different. The HPE servers are shipped to our data centers from within the EU while tape hardware is shipped from North America. The HPE servers are thus shipped into Finland by ground and sea transport whereas IBM tape hardware is transported via air. As an example, delivering a fully equipped IBM TS4300 tape library with seven full height LTO-8 tape drives to Finland has a logistics carbon footprint of 1449 kg CO_2ekv. In comparison, the logistics carbon footprint for ten HPE Apollo 4510 Gen10 servers is 675 kg CO_2ekv. These calculations are included in the sums in Table 1.

Table 1. The carbon footprints of manufacturing and shipping per unit.

Component	**Number of devices**	**Carbon footprint (kg CO_2ekv)**
HPE Proliant DL360 Gen10 (ingest)	13	14079
HPE Apollo 4510 Gen10 (spinning disk)	10	44890
HPE Apollo 4200 Gen9 (LTO-8, LTO-9)	4	9404
HPE Apollo 4200 Gen10 (LTO-8, LTO-9)	2	4704
IBM TS4300 tape library (LTO-8, LTO-9)	3	19116
IBM LTO-8 tape drives	7	665
IBM LTO-8 tape media (active, dark)	672	5020
IBM LTO-9 tape drives	7	665
IBM LTO-9 tape media	231	1726
Pelican 1450 transport case (dark)	24	210
Total DPS platform manufacturing CFP		**100479**

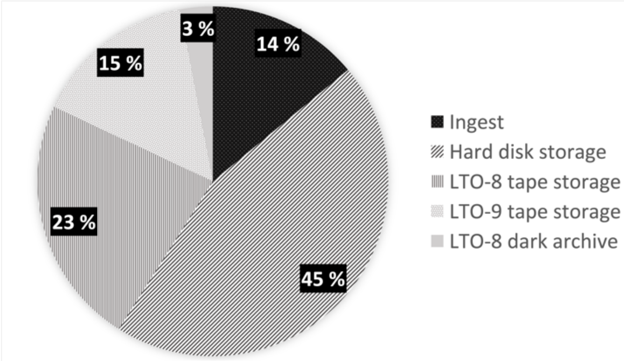

Figure 1. Carbon footprint division of manufacturing and shipping.

The calculation of the carbon footprint for the dark archive contains emissions resulting from manufacturing the LTO-8 media tapes and the Pelican 1450 transport cases. The exact carbon footprint of a case has not been provided to us, but we can estimate it by looking at the materials from which the case is constructed. A case weighs 2.5 kg and its raw material is polypropylene. Our figures are estimated from the carbon footprint of polypropylene pipe manufacturing [4] and they consist of producing polypropylene molecules and manufacturing the case. The total carbon footprint for manufacturing a Pelican 1450 case is estimated to be 8.4 kg CO_2ekv. This is an insignificant part of our whole carbon footprint.

IV. SERVICE LIFE

The carbon footprint of the hardware service life depends on data center Power Usage Effectiveness (PUE). Currently, our services are located in three separate data centers with different PUE values: (1) Data center A with a PUE value of 1.66; (2) Data center B with a PUE value of 2; and (3) Data center C with a PUE value of 1.2. The PUE value defines the energy efficiency of the data center. For example, a PUE value of 1.2 means that the data center requires 20% energy on top of the real power usage of the DPS platform. It can for example be cooling or lighting. The electricity production for the data centers is done with Finnish hydropower where the corresponding carbon dioxide emission is 24 kg CO_2ekv / MWh. This figure is based on information found in the carbon footprint calculation tool created by the Finnish Environment Institute [5].

Table 2 depicts the carbon footprint for each hardware component of our DPS during its service life. The calculations include the PUE of the data center where the components are located. We

Table 2. Service life carbon footprint.

Component	Number of devices	Service life (years)	Data Center	Data Center PUE	Annual electricity (kWh)	Service life carbon footprint kg CO_2ekv
HPE Proliant DL360 Gen10	13	5	A	1.66	1358	3517
HPE Apollo 4510 Gen10	10	5	A	1.66	1209	2408
HPE Apollo 4200 Gen9	2	5	B	2	3320	1594
HPE Apollo 4200 Gen9	2	5	C	1.2	3320	956
HPE Apollo 4200 Gen10	1	5	B	2	2812	675
HPE Apollo 4200 Gen10	1	5	C	1.2	2812	405
IBM TS4300 tape library	2	7	B	2	5472	1838
IBM LTO-8 tape drives	7	7	B	2	2711	911
IBM LTO-8 tape media	672	7	B	2	0	0
IBM TS4300 tape library	1	7	C	1.2	2736	552
IBM LTO-9 tape drives	7	7	C	1.2	2711	547
IBM LTO-9 tape media	231	7	C	1.2	0	0
	Summary of usage time carbon footprint					**13402**

assume in our calculations that servers with hard drives have a lifespan of five years while tape libraries and media have a lifespan of seven years. Fig. 2 shows the relative size of the carbon footprint of the service life for each role of the hardware: ingest, spinning disk storage and magnetic tape storage.

We have in close collaboration with our partner organizations (organizations that preserve their data in our DPS) defined common national preservation specifications, which in detail describe how digital assets should be prepared before ingesting them to the preservation service. This includes for example requirements for metadata and file formats. We put a lot of effort into automated validation of the submission information packages and their assets during the ingest phase: This includes for example virus checks, full metadata validation, file format validation and verification of checksums. Our service also performs continuous monitoring of integrity by calculating and verifying checksums. For all these operations, to mention only a few, we use the GlusterFS distributed file system[1], MongoDB databases[2], Python programming language, and various 3rd party open source components. Our software stack as a whole uses 100% open source solutions.

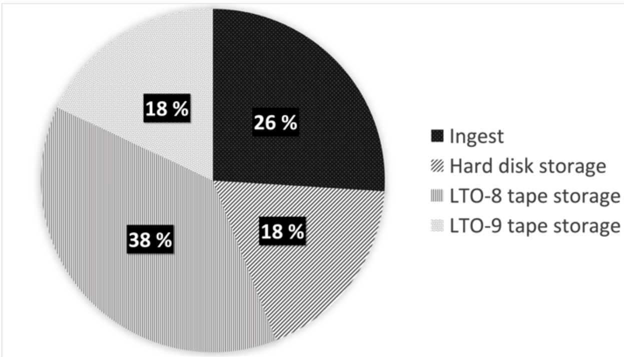

Figure 2. Service life carbon footprint division of hardware.

Our DPS have 17 experts working full time. The employee carbon footprint is calculated to have been 17.14 kg CO_2ekv in 2021, making our total annual carbon footprint for human resources in our services 292 kg CO_2ekv.

The carbon footprint of the dark archive is close to zero. We use one transport case per month, which makes the carbon footprint for the logistics around 4 kg CO_2ekv per year. Two years are needed to transfer all 3.6 peta bytes into the dark archive using LTO-8 tapes. The storage facility is located in a natural environment, shared with many other users, where

[1] https://www.gluster.org/

[2] https://www.mongodb.com/

external temperature and humidity control is not needed.[3]

V. Observations

Due to low carbon dioxide emissions of electricity production, the service life carbon footprint is only around 14% when compared to the manufacturing and shipping carbon footprint. This is shown in Fig. 3. This ratio is low even though we put a lot of computing resources into the validation of the submitted content during ingest. The majority of the carbon footprint is thus generated during manufacturing and shipping, and not during the actual service life of the hardware.

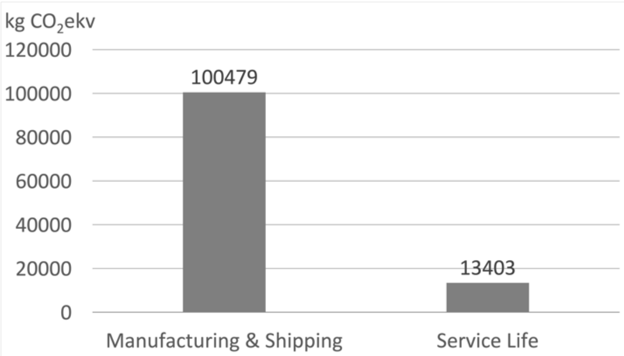

Figure 3. Manufacturing and shipping create a large carbon footprint compared to the service life.

When considering the storage areal density impact on the carbon footprint, the spinning disk areal density has the highest density and therefore its lifetime carbon footprint is not that far away from the footprint of tape environments. LTO-8, which has the lowest areal density, suffers from the fact that two modular tape libraries are needed to handle 3.6 peta bytes of storage.

Using electricity production with lower carbon dioxide emissions decreases the carbon footprint and reduces the impact that data center PUEs have on the total carbon footprint. Another major point of view that needs to be considered is however the total energy consumption during operation, regardless of the carbon footprint produced by it.

Table 2 shows the different life spans for the storage solutions. The annual carbon footprints for the different storage solutions with their differing life spans taken into account in the figures are shown in Table 3.

Table 3. The annual carbon footprint.

Component	**Annual carbon footprint kg CO_2ekv**
Ingest	3520
Spinning disk storage	9460
Magnetic tape storage (LTO-8)	4532
Magnetic tape storage (LTO-9)	3092
Dark Archive	273
Human resources	292
Total annual carbon footprint	**21169**

It can be noted that the dark archive with LTO-8 magnetic tapes has the lowest annual carbon footprint by far of all hardware components. Active tape environments suffer from tape servers that read and write the data, producing emissions in doing so.

The electricity production emissions play a role, if not a decisive one, in the total carbon footprint. Obviously, electricity production with low emissions should be prioritized.

As a collective result, our annual DPS carbon footprint is 21169 kg CO_2ekv. The Finnish Innovation Fund Sitra has calculated the average annual carbon footprint for a Finnish citizen in 2018, concluding that it is 10300 kg CO_2ekv [6]. The total carbon footprint of our DPS amounts to the carbon footprint of slightly less than three average Finnish citizens on an annual basis.

VI. Future Work

A few missing components from the calculation have been recognized: The results do not yet include carbon emissions of the optical network, data communication, or common support components for production and development. The carbon footprint relating to pre-ingest processing of digital content is also not within the scope of this paper.

The current calculations will become outdated when we increase the storage capacity or update the hardware. Carbon footprint calculations should be updated regularly whenever hardware infrastructure is changed or renewed.

Some possibilities to reduce carbon footprint are for example changing disk storage to other storage

[3] The PUE value is therefore effectively 1.

technology with a lower carbon footprint, favoring environment friendly technology and data centers, using emission free electricity, aiming for high areal density in storage media, and increasing the service life of hardware components in use.

By the end of 2023 the ingest and spinning disk components will be transferred from our site in southern Finland to Northern Finland. The data center cooling in the new site is implemented with open air free cooling which leads into an excellent PUE of 1.05. This means an annual reduction of 435 kg CO_2ekv to our carbon dioxide emissions.

We have not utilized Green Coding [7], but the possibility to reduce carbon footprint through efficient processing is something to consider in the future.

A large work is ahead for IT-infrastructure manufacturers. They have to learn to minimize their products manufacturing carbon footprint. One component in this would be extensive recycling of product materials. A second major change which will have a significant impact is the green energy transformation for the production phase of IT-hardware. This transformation has just started in Europe and the future is promising regarding this shift.

We as consumers must start to require and prioritize more environment friendly infrastructure. Hopefully, the digital preservation community and IT experts are able to find ways to influence this in a positive way.

REFERENCES

[1] Finnish National Digital Preservation Services. https://digitalpreservation.fi/en/

[2] Wikipedia: *Power Usage Effectiveness*. https://en.wikipedia.org/wiki/Power_usage_effectiveness

[3] Massachusetts Institute of Technology, Material Systems Laboratory: *Product Attribute to Impact Algorithm - PAIA*. https://msl.mit.edu/projects/paia/main.html

[4] Jeroen Wassernaar, *Polypropylene Materials for Sewerage & Drainage Pipes with Reduced Energy and Carbon Footprints*, Journal of Materials Science and Engineering B6 (11-12), 2016.

[5] Finnish Environment Institute: *Y-HIILARI Hiilijalanjälki -työkalu* [carbon footprint tool]. https://www.syke.fi/fi-FI/Tutkimus__kehittaminen/Kulutus_ja_tuotanto/Laskurit/YHiilari

[6] The Finnish Innovation Fund Sitra: *Carbon footprint of the average Finn*. https://www.sitra.fi/en/articles/carbon-footprint-average-finn/

[7] Green Software Foundation: *Principles of Green Software Engineering*. https://learn.greensoftware.foundation/