The CO2 Emissions of Storage and Use of Digital Objects and Data

Exploring climate actions

|  |  |
| --- | --- |
| **Lotte Wijsman** | **Arie Groen** |
| *National Archives of the**Netherlands,**The Netherlands**lotte.wijsman@nationaalarchief.nl* | *National Library of the Netherlands,**The Netherlands**arie.groen@kb.nl* |
|  |  |

|  |  |
| --- | --- |
| **Tamara van Zwol** | **Robert Gillisse** |
| *Dutch Digital Heritage Network,**The Netherlands**tvzwol@beeldengeluid.nl* | *International Institute of Social History,**The Netherlands**robert.gillisse@iisg.nl* |

**Abstract – The storage and use of digital heritage objects produce carbon dioxide (CO2) emissions. Cultural heritage organizations can take several measures into consideration in order to diminish these CO2 emissions. However, how much CO2 do storage and use produce and what measures could have (the most) effect? We examined the CO2 impact and possible measures on the basis of a case study. We have focused our investigation on the impact of servers, infrastructure, cloud storage and use. But the story does not end there. We look ahead, beyond the case study and beyond boundaries, introducing a research agenda within the Dutch Digital Heritage Network (DDHN).**

**Keywords – carbon footprint, sustainability, storage, users, carbon dioxide emissions**

**Conference Topics – Environment**

# Introduction

Preserving digital objects for the public contributes, like many human activities, to carbon dioxide (CO2) emissions and consequently has an impact on the environment. The Dutch digital heritage community is (becoming) conscious of the subject and wishes to examine the facts. What is the environmental impact of the storage and use of collections? What measures can be taken to lessen the CO2 impact? And what other issues are still untouched and are waiting for further investigation?

This paper provides insight into certain measures that can be taken when it comes to storage and use, based on a CO2 impact case study of the Delpher platform [1]. The published results and recommendations based on the case study [2] proved only to be the start of exploring climate actions for heritage organizations. To gain more insights, we asked the community to help us expand the research agenda to their specific needs and wishes. But first we will discuss the findings and recommendations based on the Delpher case study.

The Delpher case study was executed by the company PHI Factory and the Green IT expert group within the Dutch Digital Heritage Network. In Delpher you can search and find millions of digitized texts from Dutch newspapers, books, and other published works. These documents come from the collections of various Dutch scientific institutions, libraries, and heritage institutions. Delpher is developed and managed by the National Library of The Netherlands (KB). We have examined storage and data use in this case study, focusing on the CO2 impact of servers, the server environment/infrastructure, cloud storage, and the end use: searching through the files on the platform and downloading files. The study was based on the GreenHouse Gas Protocol [3]. PHI Factory used the guidelines from 'The GreenHouse Gas Protocol' to measure the CO2 footprint.

 This paper presents our findings in those four areas.

# Servers

Servers provide the computing power and storage required to store and make digital collections available for users. These servers are the main cause of CO2 emissions. This is due to both the electricity consumption and the indirect CO2 emissions from the production of the servers.

Creating digital compartments in the servers, like the KB has done for the data on Delpher, ensures that the capacity of these servers can be used more efficiently. This can be done by means of virtual machines or containers. The KB's servers consume now 242 MWh (or: 242,000 kWh) annually, which is equivalent to the electricity consumption of 98 average Dutch households in a year.

# Server environment

The location/environment of the servers has a major influence on the total of CO2 emissions. If data is stored locally, on the level of one institution, there is a good chance that actions facilitating the servers, such as cooling them, consumes as much or even more energy than the servers themselves. To reduce the CO2 impact of the infrastructure around the servers you can think about sharing servers with multiple organizations to use them most effectively. By moving the servers from the KB local location to a more efficient, external colocation data center, as in the case of Delpher, considerable savings can be made on electricity costs: saving annually the amount of 151 MWh. Because many servers are located here, facility systems such as cooling can do their job much more effectively. Therefore, this method is not only more sustainable, but also more economic.

You can also opt for more green energy, like the KB has done. Green energy is any energy type that is generated from natural resources, such as sunlight, wind or water. Because green energy is generated from a renewable source, the CO2 emissions are a whole lot lower than in the case of energy from fossil sources. The annual carbon footprint of Delpher's servers is less than 4 tons of CO2 equivalents per year, which equals 4 hot air balloons of 200 m2 (the size of a soccer field) filled with CO2.

# Cloud storage

With cloud storage, the data and computing power of many companies is divided over servers. This makes for very efficient use of (the capacity of) the servers since every available space is being occupied. The advantage of storage in a cloud environment is that the type of providers behind it (e.g. Microsoft and Amazon) are at the forefront of the development of facility systems and the use of containers to make the capacity of their servers as efficient as possible. Naturally, cultural heritage organizations have to consider if they are willing to store their data in a large datacenter under the control of such a provider in perhaps a different country, under different rules and regulations. Because Delpher concerns itself with national Dutch cultural heritage, it has been decided to store the data at a Dutch colocation and not via an international cloud provider.

# Data use

Retrieving files from a digital collection, loading webpages and using the search index causes CO2 emissions. In the case of Delpher a large part of the digital collection will not be downloaded by a user, but searched, which has only a limited impact. Still, there are ways to even diminish this impact. This could be done by e.g. offering lower resolution versions of the digital object files. In addition, to make it even more effective, you can also limit the user features on the website so that fewer files have to be searched in the data store. For example if you do not offer 'search all' as a standard option, but let users indicate which specific material (newspapers, books or magazines) should be searched.

# Discussion

1. *Case study as a starting point*

The case study provided calculated information on storage and use issues and recommendations for organizations to consider. But of course, one case study alone means there are limitations to what you can investigate. Some research topics are still left on the shelf.

For example we have not calculated the CO2 emissions of digital preservation workflows like pre-ingest. We have not considered the carbon costs of the data center building (materials) or the specific carbon costs of multiple information objects storage.

However the case study proved to be a starting point for further research and raising awareness.

1. *Further research*

In order to determine the topics for further investigation we involved the heritage community. During two sessions in November 2021 and June 2022 on the theme of Green IT, heritage professionals from the network were asked about their experiences and wishes for further research.

Important to note is that the environmental impact of digital heritage was not yet included in the policy of most organizations. At the top of the wish list stood and stands therefore climate awareness. However, the heritage community recognizes the complexity of this topic, the sense of urgency and the demand for more knowledge. And certainly, some measures are being considered or already executed, like a stricter selection and avoidance of duplicate storage of digital objects, cleaning up digital data, switching to green power or relocation to other data centers. In order to help organizations raise awareness, the main results and recommendations were visualized in an infographic [4].

During the sessions there were discussions. Does it help to centralize storage? For example, placing audiovisual material with an organization that has the specific expertise and services to do so, instead of trying to solve everything in the local situation.

We should also think about the accessibility requirements. Must everything always be immediately available? Organizations store more and more data everyday but should climate considerations be a reason to make stricter selection choices. Those are some of the questions raised. We cannot answer for all of them, but we can keep open the discussion as a community.

The topics that were mentioned for the research agenda were eventually based round three themes: organizational impact (e.g. cost savings, shared data storage), technology and suppliers ( the impact of digitization of materials, sustainable hardware, supplier comparisons) and use (user behavior, the impact of using audiovisual materials online, access of DIP on demand)

The network group has now new members from the community involved and is planning for a research agenda prioritization. Like DEN in the Netherlands, who share knowledge by bringing together sources on this topic [5].

1. *Global outreach*

Some of the more technological outcomes and recommendations in the case study, especially cloud storage and a more efficient way of dealing with servers and server environment, could be applicable for cultural heritage organizations around the world. We would also highly recommend for heritage organizations of different shapes and sizes to execute their own research and calculate CO2 emissions, sharing their results with the international community. The more information we can gather together, the better the result will be for the community.

# Conclusions

With the findings from the case study and the aforementioned recommendations, cultural heritage institutions can start to examine the CO2 impact of their own digital collections and make choices for a climate-resilient future. Also, the case study stimulates further discussion about selection and deduplications of collections in and between cultural heritage institutes. The case study is only a starting point for further research. Hopefully in future we can join forces with other (international) initiatives.

# REFERENCES

[1] Delpher platform <https://www.delpher.nl/>

[2] De CO2-impact van opslag en gebruik van digitaal erfgoed, DDHN <https://doi.org/10.5281/zenodo.6341483>

[3] The Greenhouse Gas Protocol <https://ghgprotocol.org/>

[4] Praatplaat CO2-impact opslag en gebruik van digitaal erfgoed, DDHN <https://doi.org/10.5281/zenodo.6411668>

[5] Ecological sustainability in digital transformation, DEN <https://www.den.nl/actueel/ecologische-duurzaamheid>