The CO_2 impact of digital heritage storage and use

Using the Delpher platform as a case study

31 December 2021

dutch digital heritage network

Imprint

Research and text: PHI Factory

Process support: Green IT network group, with the cooperation of experts at the KB (National Library of the Netherlands).

Translation: Vertaalbureau Noorderlicht BV

This is a publication of the Dutch Digital Heritage Network, December 2021.
More information is available at: netwerkdigitaalerfgoed.nl
Comments are welcome via: info@netwerkdigitaalerfgoed.nl

Contents

Summary

The Dutch Digital Heritage Network is made up of several institutions in the areas of culture, heritage, education and research. Within this network and among members of the Green IT working group created within it, the desire arose to know the environmental impact of the storage and use of collections from these organisations. Because of their digital data management, the organisations contribute to the impact of energy consumption on the planet. So how significant is the ecological footprint exactly, and how can it be made as small as possible?

This was investigated with a case study: the Delpher platform, managed by the KB, the National Library of the Netherlands. In Delpher, millions of digitised texts from Dutch newspapers, books and magazines can be searched. In particular, the Delpher study focuses on the use of the servers that run the platform. The use of this platform was also examined, in order to map indirect greenhouse gas emissions from the transmission of data. For this purpose, the guidelines from ['The Green House Gas Protocol'] have been used.

The total Delpher carbon footprint in 2021 was 53 tonnes of CO_2 equivalents. This is equivalent to:

- the CO_2 uptake of 2,650 trees in 1 year;
- 371 flights from Amsterdam to Paris;
- the emissions of more than 5 households (including all indirect emissions of, for example, purchased items).

And looks like:

- 53 balloons of 200 m^2 filled with CO_2.

Servers provide the computing power and storage needed to host and make available a digital collection. These servers are the core of CO_2 emissions. This is due both to the electricity consumption and to the indirect CO_2 emissions from the manufacturing of the servers. By purchasing green electricity, Delpher's impact is limited.

The location of the servers has a major influence on the total electricity consumption. If data is stored locally, it is likely that facilitating the servers (including cooling them) will consume as much or even more energy than the servers themselves.

The decision was made to move the Delpher digital collection from the KB to a governmental data centre: the Centre for Infrastructure and Exploitation, run by the Tax Authorities (ODC B/CIE). This move is expected to save annually:

- 196,000 kWh (as many as 79 average Dutch households in a year);
- 109 tonnes of CO_2 (based on Dutch grey electricity).

This is equivalent to the amount that 5,500 trees take out of the air in a year. The question why these savings are calculated based on grey electricity will be elaborated later in the report.

This energy saving was achieved by moving the Delpher data to the ODC B/CIE data centre where various companies host their servers. This is due to the choice of more efficient hardware and the energy efficiency that a large data centre provides in terms of facilities.

In addition to electricity consumption, the **hardware** of the servers has a hidden CO_2 impact. This impact is divided into three groups of hardware: the servers that provide the computing power for Delpher, the storage modules that hold the data for this computing power, and the hardware where the digital collection is stored.

One situation in which a great deal of data is transmitted all at once is in the **migration** of digital heritage. Heritage institutions that want to store their data elsewhere must take this additional energy consumption into account. At Delpher, this has resulted in around 31 tonnes of CO_2 emissions.

In addition to the calculations of emissions from servers, hardware and data migrations, this project also looked at the **use** of Delpher itself. It has been investigated what the energy consumption is when someone carries out a search on the Delpher platform or opens or downloads a file.

Data usage when opening newspapers and watching (instructional) videos requires the most data traffic. It was noted that of all media, object pages (where an image of, for example, a newspaper page can be viewed in detail) were responsible for the largest share of data consumption. In order to reduce data usage, the KB could offer the object pages in lower resolution on the platform or give the user more options to replace wide searches ('search everything') with more targeted searches.

Other measures that were not investigated within the scope of the case study, but might contribute to limiting the carbon footprint, are:

- Preventing and removing duplicates: parts of a collection may be stored in duplicate. Removing this duplication prevents the digital collection from becoming unnecessarily large.
- Making agreements between heritage institutions to prevent duplication.

The findings in this case study show that the final CO_2 impact depends on many different factors. The selected storage of digital heritage (servers, the hardware and the location of the storage) in particular determines the size of the CO_2 emissions. In this context, various measures can be taken by heritage institutions to reduce this impact.

1. Introduction

Organisations emit a large amount of CO_2 through their operations. Just think of the energy consumption of the building and the transport of staff and all goods. However, that is not all. Organisations have even greater impact on global CO_2 emissions than they think, namely in all the materials they use in business operations. This impact is also referred to as the 'hidden impact'. It concerns the CO_2 emissions that are produced throughout the entire life cycle of a product (from the extraction of raw materials, manufacturing and transport, to use and post-use processing).

To provide insight into the hidden impact, PHI Factory has developed the PHI Impact Calculator. This tool uses carefully selected standards and key figures (based on scientific data and life cycle analyses of materials) to calculate the hidden impact of an organisation's operations. In order to identify the sustainable potential of digital heritage, it is important to know what the CO_2 impact of the storage and use of digital heritage is, and what measures can be taken to limit this impact.

1.1. The assignment

The Dutch Digital Heritage Network is made up of several organisations in the fields of culture, heritage, education and research. Within this network and the Green IT working group, the desire arose to know more about measuring the CO_2 impact of the storage and use of collections from these organisations. This is because they contribute to the impact of energy consumption on Earth through their digital data management. What is the ecological footprint and how can they make it as small as possible? This was investigated with a case study: the Delpher platform, managed by the KB, the National Library of the Netherlands. In Delpher, millions of digitised texts from Dutch newspapers, books and magazines can be searched. The case study can serve as a best practice for the heritage field in terms of awareness, behaviour and the substantiation with numerical data of the CO_2 impact.
These figures lay the foundation for possible follow-up steps or measures.

Project objectives:

- raising awareness about sustainable goals and storage among heritage institutions within the network;
- providing substantiation with numerical data for mapping the footprint of a heritage organisation by means of a concrete case.

1.2. Scope

The Delpher platform was taken as the starting point for this case study. We looked at:

- electricity consumption servers;
- electricity consumption infrastructure (such as cooling);
- hidden impact of hardware;
- data transmission;
- use of Delpher.

The servers on which Delpher runs were located at the KB (National Library) for a long time, but have now been relocated to the Centre for Infrastructure and Exploitation, a governmental data centre run by the Tax Authorities (ODC B/CIE). A copy of the collection is also stored at the Netherlands Institute for Sound and Vision (NIBG) as a backup. The relocation to the ODC B/CIE provides great insight into the various factors that play a role in the carbon footprint and is therefore the focus of considerable attention.

In most cases, a carbon footprint is determined for one organisation. Because Delpher comprises the digital collection for multiple organisations, the carbon footprint is spread across these organisations. Unfortunately, it was not possible to determine the share per heritage organisation on the basis of user data. This means that the results in this report are always reported at full Delpher level. Since the KB developed Delpher and runs the platform, the results can be interpreted as part of the KB footprint.

To calculate the footprint, we considered data from 2021 as much as possible. Where necessary and possible, these have been extrapolated to a full year, in order to show the different numbers in perspective.

1.3. About Delpher

Delpher is a digital platform that provides access to digitised texts from 2 million newspapers, 12 million magazine pages and more than 900,000 books. These documents have been digitised in such a way that it is possible to search with search terms, after which Delpher displays the most relevant texts. The total number of digitised pages is 120 million, representing 1.1 petabytes of data (1.1 million gigabytes).

Delpher is an initiative of several knowledge organisations: The universities of Groningen, Leiden, Utrecht and Amsterdam, and the KB (National Library of the Netherlands). For years, the KB, together with the NIBG in Hilversum, was the location where a copy of Delpher's digital collection was stored on tape. Up to and including 2021, the Delpher platform itself was also hosted on the KB servers. In addition, part of the collection (400 terabytes) was stored on hard disk at the KB, making it available to users faster than via tape.

At the time of writing this report, the KB was in the process of transferring date from Delpher to a large government data centre: the ODC B/CIE. This relocation has had a major effect on the electricity consumption of the servers and thus provides good insight into the impact of storage and use of the sustainable provision of a digital collection.

Delpher can be accessed via www.delpher.nl.

1.4. Terminology and context

This report uses various terms related to CO_2 impact and electricity consumption. In order to interpret these terms correctly, the most important ones are explained below.

1.4.1. Electricity consumption

Electricity consumption is measured in kWh (kiloWatt hours). This is the amount of electricity that a device with a power of 1 kilowatt consumes in 1 hour. A power of 1 kilowatt is equivalent to 1,000 Watts. For comparison: a classic light bulb has 40 to 60 kilowatts of power, a vacuum cleaner 900, a kettle around 2 kilowatts. A kettle therefore consumes 1 kWh of electricity in half an hour. 1000 kWh is equivalent to 1 MWh (MegaWatt hour).

The KB servers consume 242 MWh (or: 242,000 kWh) annually, which is equivalent to the electricity consumption of 98 average Dutch households in a year[i].

1.4.2. CO_2 emissions

Several greenhouse gases are causing global warming. For comparison purposes, all greenhouse gases are converted into CO_2, the most important greenhouse gas. The carbon footprint (also called CO_2 footprint or ecological footprint) of organisations is expressed in kg CO_2.

One kg of CO_2 is released when:

- driving about 5 km by car;
- consuming approximately 2 kWh of grey electricity;
- producing 50 grams of minced meat.

One tonne of CO_2 (=1000 kg):

- is emitted per passenger on an 8-hour flight;
- is released when burning 320 litres of diesel;
- is the amount of CO_2 absorbed by 50 trees in a year;

The average Dutch person is responsible for about 10 tonnes of CO_2 per year[ii]. The total emissions of the Netherlands in 2020 amounted to 165 megatonnes (165 million tonnes) of CO_2.

In this report, the terms CO_2, CO_2 footprint, carbon footprint and CO_2 impact are used as synonyms for readability.

2. Calculation method

Calculating an organisation's carbon footprint requires an understanding of all activities that (indirectly) cause greenhouse gas emissions. This not only considers emissions during the activity of the organisation itself, but also the emissions indirectly resulting from those actions. This can be traced back in particular to the purchase and use of products. The entire life cycles of these products is considered.

Life cycle assessments (LCAs) are used to map the environmental impact of products throughout the chain: from raw material extraction, manufacturing, transport and use to end-of-use. In each phase, greenhouse gases can be emitted, and it differs per product which phase contributes most. For instance, the emission of petrol is highest during use because that is when the greenhouse gases are released into the air. However, if you look at a kilo of beef, there are many emissions earlier in the chain: when ruminating, a cow emits methane, which is a strong greenhouse gas.

For ICT products, the environmental impact varies greatly from product to product. In the case of products such as batteries and circuit boards, many emissions arise during the extraction of the raw materials, while others have relatively high emissions during use due to high electricity consumption.

PHI Factory determines the carbon footprint of companies by linking emission figures to the products used and purchased by organisations. Figure 1 shows an elaboration of this.

The emission factors are determined based on scientific data in relation to the unit, according to the 'Technical Guidance for Calculating Scope 3 Emissions' of the 'Green House Gas Protocol'. The scientific sources used include the International Environmental Product Declaration System[iii] and the Idemat[iv] tool that was created by the Delft University of Technology[v].

Each organization can decide for itself which production groups responsible for indirect CO_2 emissions it includes in the calculation. A comparison between different organisations can therefore create a distorted image. PHI Factory is collecting more and more data from other organisations in the database to make comparisons better and better.

The Delpher case study focuses in particular on the use of the servers on which the platform runs. The use of this platform was also examined, in order to map indirect greenhouse gas emissions from the transmission of data.

Figure 1: Calculation method CO_2 equivalents for products and services in the facilities domain (PHI Factory, 2019)

2.1. Direct and indirect emissions

PHI Factory uses the guidelines from 'The Green House Gas Protocol' to measure the carbon footprint. A distinction is made between three different scopes:

- **Scope 1:** Direct CO_2 emissions. This is caused by own resources that are owned and managed by the organisation. This includes the consumption of gas and cooling systems during the use of the building.
- **Scope 2:** Indirect CO_2 emissions. Energy and heat purchased. These emissions take place outside the organisation but can be directly attributed to it.
- **Scope 3:** Indirect CO_2 emissions. Emissions that are caused by chain partners for the purpose of business operations, such as the extraction, manufacturing and transportation of purchased products.

The full carbon footprint mapped in this case study falls under scope 2 and 3 (from the perspective of the KB).

Scope 2 includes the electricity used by the KB. The remaining emissions, such as the hidden impact of the hardware, the electricity consumption at the ODC B/CIE and the NIBG and the use of Delpher, are all scope 3 emissions. For the sake of clarity, no distinction will be made between the different scopes in the further discussion of the results.

2.2. Calculation unit

The calculation unit is CO_2 equivalents in kg or tonnes. One tonne is 1.000 kg CO_2 and is equivalent to:

the CO_2 uptake of 50 trees in 1 year, or

seven flights from Amsterdam to Paris.

One tonne of CO_2 looks like:

a hot air balloon of 200 m^2 filled with CO_2. This is a balloon the size of a football field.

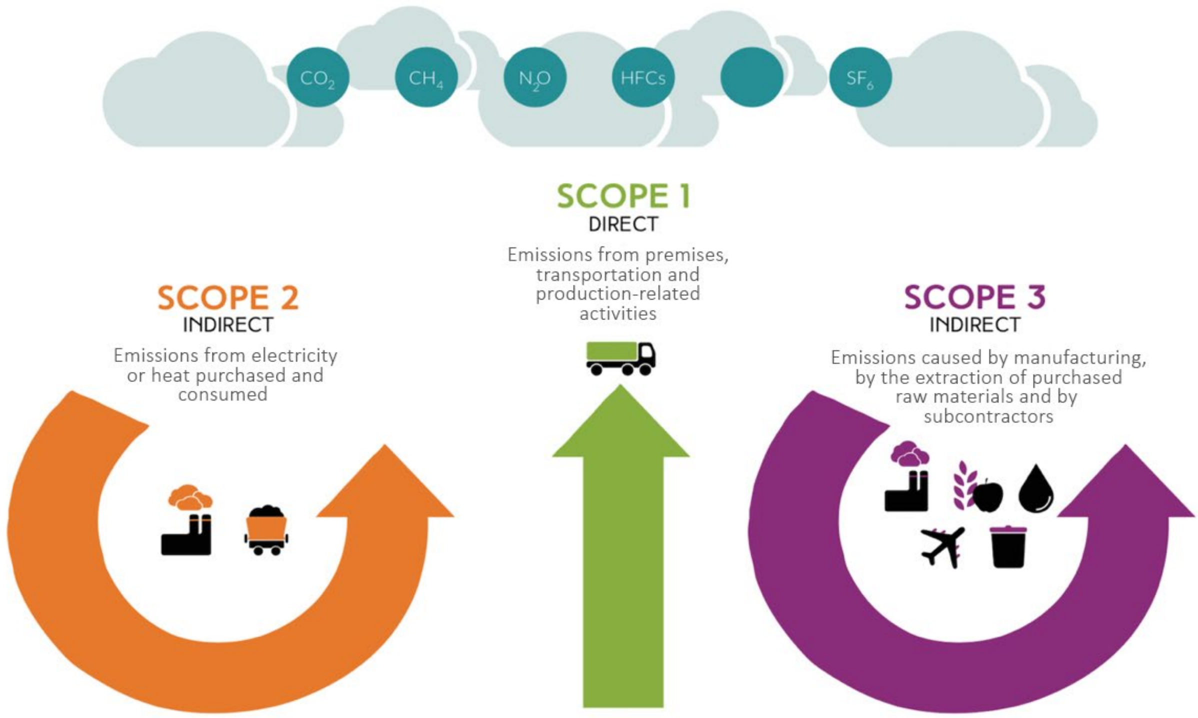

Figure 2: The Greenhouse Gas Protocol

3. Results of the Delpher case study

The total Delpher carbon footprint in 2021 amounted to 53 tonnes of CO_2 equivalents. As the relocation of the servers was started in that year, this is based on the servers as set up in the new situation, i.e. at the ODC B/CIE. This 53 tonnes of CO_2 is equivalent to:

- the CO_2 uptake of 2,650 trees in 1 year;
- 371 flights from Amsterdam to Paris;
- the emissions of more than 5 households (including all indirect emissions of, for example, purchased items).

And looks like:

- 53 hot air balloons of 200 m^2 (the size of a football field) filled with CO_2.

Figure 3: Size 1 tonne CO_2

3.1. Carbon footprint Overview

The carbon footprint of ICT resources is mainly generated by two factors: the manufacturing of the devices themselves and the electricity they use. Which of the two is the most influential factor depends on the hardware in the product. The electricity consumption of Delpher servers has the greatest (indirect) impact and the KB has the most influence on this. Therefore, this will be the main focus of this report.

Figure 4 contains an overview of the Delpher carbon footprint. The total of 53 tonnes of CO_2 equivalents is based on the emissions of green electricity, which means that the footprint of electricity consumption is relatively low. However, even when purchasing green electricity, saving energy is still the best way to reduce indirect CO_2 emissions. The overview therefore also shows the impact that would have been made when purchasing grey electricity. Why this is will be explained below.

3.1.1. Energy consumption and CO_2 impact

Most of the greenhouse gases produced by humans are caused by energy consumption, both directly by burning fossil fuels (for example in a car) and in the form of electricity. The goal of the energy transition is to generate energy sustainably and thus limit the emission of greenhouse gases.

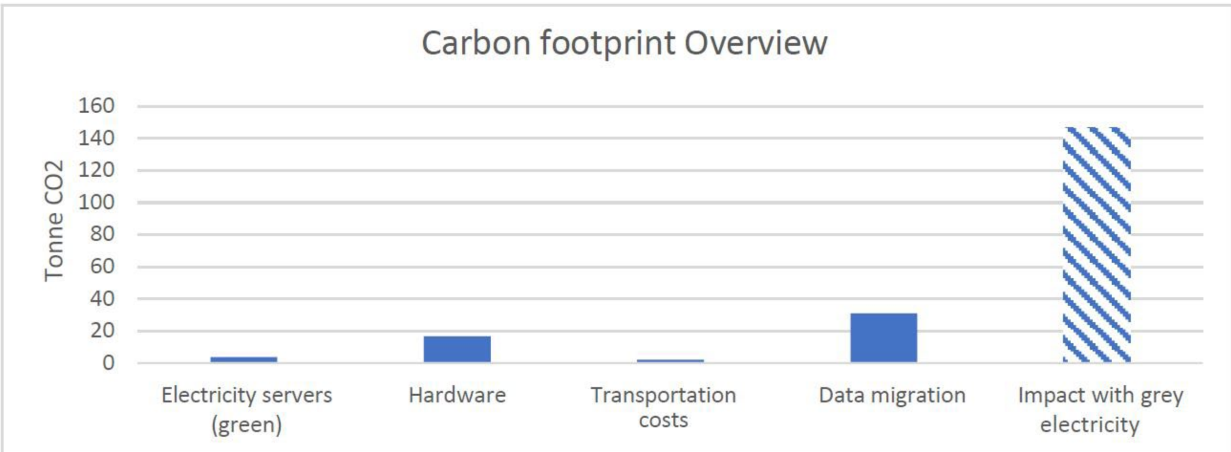

Figure 4: Overview of the Delpher annual carbon footprint. The bar with grey electricity does not contribute to the total Delpher footprint. This is to indicate what the footprint would be if grey electricity had been purchased and what the potential is.

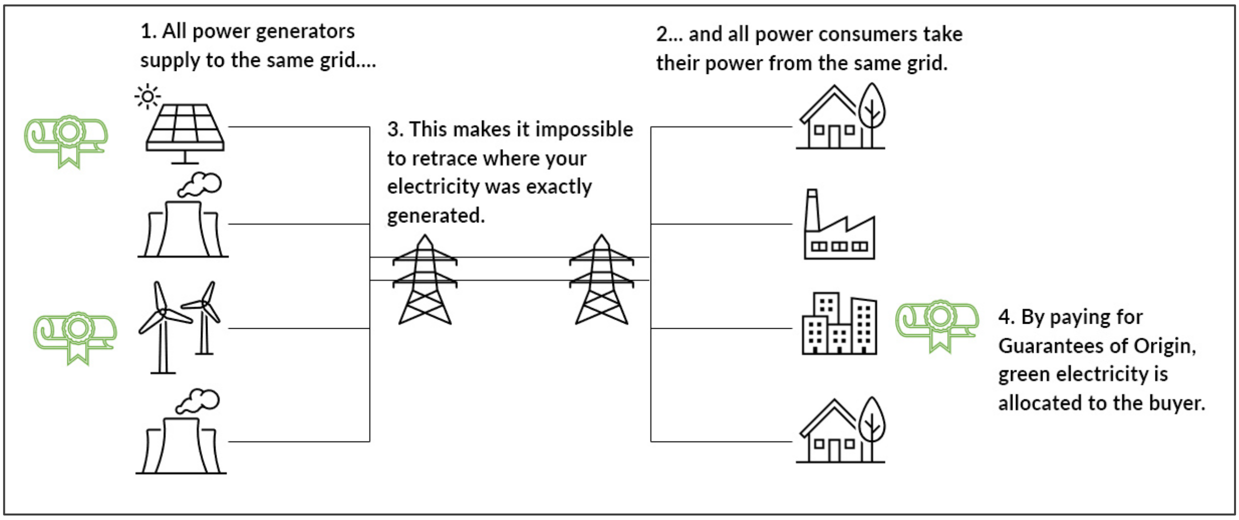

Figure 5: Purchasing green electricity

Electricity is particularly important for this case study because the Delpher servers use a lot of electricity. Indirectly, the use of electricity causes CO_2 emissions. The distinction is made between grey and green electricity.
Grey electricity represents electricity generated from fossil sources, such as coal or gas. One kWh of grey electricity causes on average 0.476 kg of CO_2 emissions at the power plant and another 0.080 kg of indirect CO_2 emissions for the extraction and transport of fossil fuels. Green electricity is generated from a sustainable source, as a result of which CO_2 emissions are much lower. Green electricity causes only a small proportion of indirect CO_2 emissions. These emissions are generated during the construction of wind turbines (0.014 kg CO_2 per kWh) and solar panels (0.093 kg CO_2 per kWh) and are allocated to the electricity generated.

For the servers on which Delpher runs, green electricity is purchased with Guarantees of Origin. These link the generation of green electricity by solar panels and wind turbines to an end user (see Figure 5). Therefore, the total CO_2 impact of Delpher is calculated and reported using the value for green electricity. However, in reality this is a little more complicated, because the electricity market is complex.

As all electricity in the Netherlands is transported over the same power grid, there is no difference between green or grey electricity at power socket level. All the electricity generated is lumped together, so to speak, and customers take their power from this.

Until 100% of the Dutch electricity is generated sustainably, all additional electricity consumption will come from grey electricity. This also means that if the electricity consumption of Delpher's servers decreases, someone else will be able to use more green electricity. Fewer Guarantees of Origin need to be purchased for the use of Delpher and these are therefore available to other customers. In turn, they need less grey electricity. In this way, Delpher (green) energy savings will help reduce the total amount of grey electricity used in the Netherlands. For that reason, energy saving scenarios often calculate the CO_2 emissions of grey electricity instead of green electricity. This has also been done in this report.

Does this mean that green electricity procurement is pointless? No, because by purchasing green electricity, you stimulate the construction of new renewable energy sources, which will enable society to run entirely on green electricity sooner.

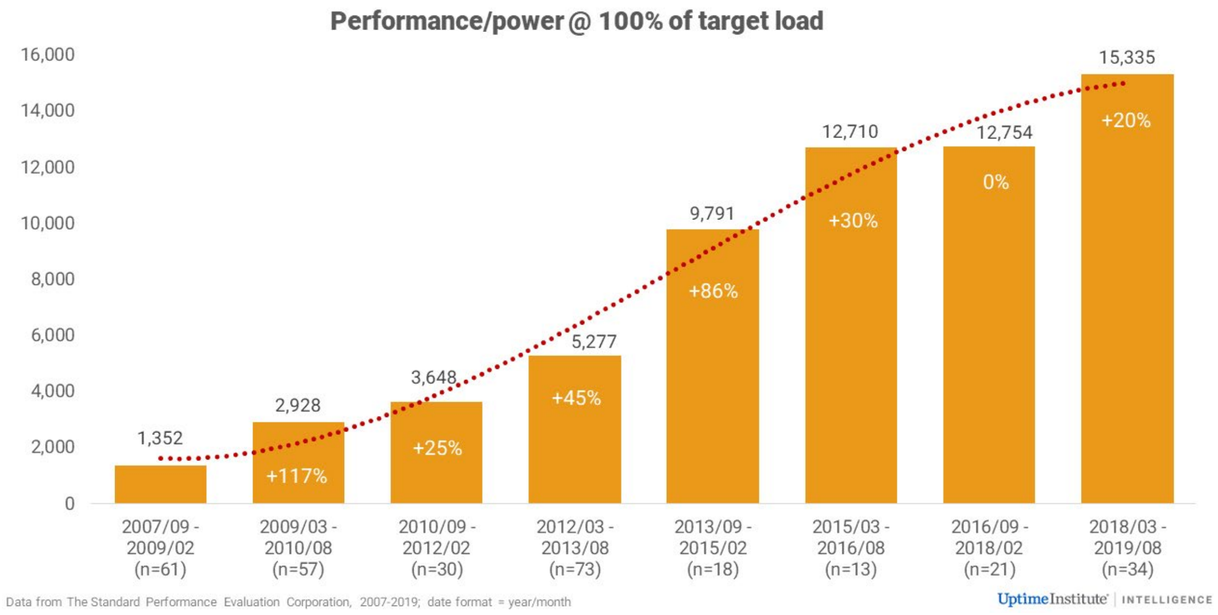

Figure 6: Increase in server efficiency. Source: Uptime Institute

3.2. Energy consumption by servers

Since energy consumption can be a major expense for each ICT department or data centre, reducing electricity consumption is often a high priority. This financial incentive often leads to efficient innovations. As a result, hardware has become increasingly energy-efficient over the years. Figure 6 (based on data from the Uptime Institute[vi]) shows how server performance (=performance) has increased per amount of electricity consumption (=power). Between 2007 and 2019, that performance grew by more than a factor of 10.

The energy consumption of the Delpher servers will be discussed below. But first, the different sources of electricity consumption by servers are discussed.

3.2.1. Computing power and storage

Servers are used for two functions. The first is storing data for retrieval at a later time. This can be, for example, a small text file, a high-resolution video or a digital photo. Storage takes place on tape or disks.

The software used for retrieval is driven by the second function of servers: computing power. In fact, most computers are very large calculators. You give them input to calculate a certain sum and then you will see the result on the screen. Computers have become increasingly advanced in the things they can compute and the (amount of) input they can use for this. As a result, servers have a lot of computing power. To best harness this computing power (also called processing power), *virtual machines* are used. With special software, multiple computers are created, as it were, within one single server, so that capacity can be better utilised. The *virtual machines* use the computing power of the servers. At Delpher, all actions that users can perform, such as a search or the creation of an account, are carried out by these *virtual machines*.

Just as a calculator uses a battery, servers use electricity to carry out calculations and keep data available. Figure 7 shows the electricity consumption of Delpher's computing power, storage and network equipment on the KB servers. This includes the network infrastructure required to use the servers, disks and tapes. The total amounts to 242 MWh (or: 242,000 kWh) per year, which is equivalent to the annual electricity consumption of 98 average Dutch households[vii].

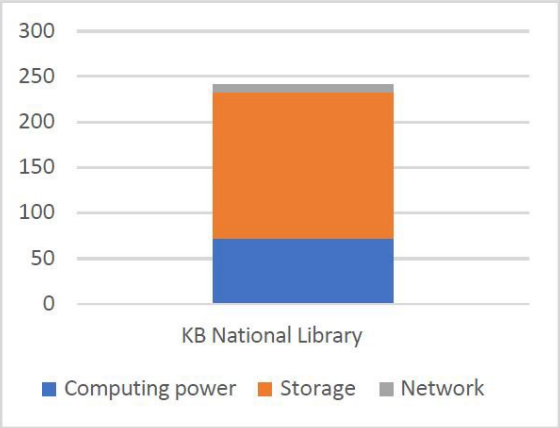

Figure 7: Distribution of power consumption of servers

The amount of electricity used for and by servers depends on many factors. The most important ones can be shown by looking at the relocation of the Delpher servers.

This migration has changed three factors that affect the power consumption of the servers: A different type of storage was chosen, more efficient hardware was deployed, and the environment in which this hardware is located was changed.

3.2.2. Type of storage and hardware efficiency

There are different ways to store data. Tape storage is the most commonly used method for long-term storage that does not need to be used frequently (*cold storage*). The information is stored on magnetic tape, which can be played and read[viii]. Several of these tapes can be combined in a tape robot to store more data, as is the case with Delpher. When the tape is not read, the system does not consume any energy. This is different in the case of a *Hard Disk Drive* (HDD).

HDD or hard disk is a form of storage in which data is stored on a rotating disk[ix] via magnetic polarisation. The big advantage of HDD over tape is that data is available much more quickly. The amount of electricity an HDD consumes depends very much on the type of HDD and the amount of data that can be stored on it.

The more data a HDD can store, the more efficient it is in terms of energy consumption.

Solid State Disk (SSD) is the latest form of data processing that is important to Delpher. The advantage is that the data is stored on chips, which means that it has no moving parts, unlike HDD and tape[x]. This reduces energy consumption and improves the speed of data processing and storage. Energy consumption therefore varies greatly between systems.

In moving Delpher, the primary goal was to make the entire collection readily available. In the old situation at the KB, only a part of the collection (400 Terabytes) was immediately available. At both the KB and the NIBG, the entire collection was on tape and therefore not directly available. In the new situation at the ODC B/CIE, the entire collection is duplicated on HDD.

Electricity consumption would probably have increased dramatically if the same kind of hardware had been used as in the KB. However, this is not the case. In fact, total electricity consumption has decreased, partly by opting for larger HDDs (10 Terabytes per disk instead of 1 Terabyte) and a more efficient driving system.

The difference in energy consumption between the various types of storage can be clearly seen when comparing the situations at the KB, the NIBC and the ODC B/CIE. Delpher's data has always been stored in three places. In the old situation, Delpher used one copy on HDD and two on tape, one of which was located at the NIBG. For the move to the ODC B/CIE, it was decided to store the entire digital collection on HDD twice, with another backup on tape at the NIBG. In the new situation, the temporary storage of the data (*caching*) takes place on SSD, which saves a lot of electricity consumption compared to the situation at the KB.

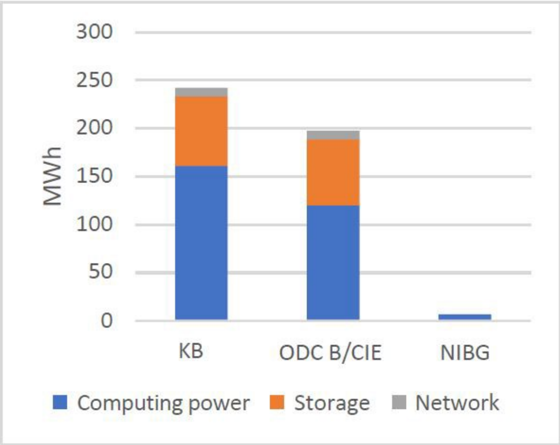

Figure 8: Annual electricity consumption per location. The KB has the entire selection on tape and 400 terabytes on HDD, the ODC B/CIE has the entire selection on HDD twice, and the NIBG has the entire collection on tape only.

To illustrate electricity consumption, the various locations are shown in Figure 8. It is easy to see that the tape backup at NIBG consumes hardly any electricity, compared to the two locations where the data are actually used. This is due to the inactivity of this storage. Furthermore, it can be concluded that in the new situation at the ODC B/CIE, much more efficient equipment was chosen compared to the KB.

Less electricity is consumed for storage than at the KB, despite no longer using tape for storage at that location. This is mainly related to the larger storage per HDD (10 terabytes) compared to the old HDDs (1 terabyte). As a result, less equipment is required. See Chapter 3.3.

Every year, this will save Delpher servers 45 MWh in electricity, the equivalent of electricity use of over 18 average households. Based on grey electricity, this prevents the emission of 25 tonnes of CO_2[xi].

3.2.3. Location of storage

In addition to the servers themselves, the environment in which these servers are located is critical to the total electricity consumption. The building itself also uses electricity to provide a safe environment for the servers. In particular, the server cooling process consumes a lot of energy, as the servers produce a lot of heat due to their activity.

The *Power Usage Effectiveness* (PUE) unit is used to measure the efficiency of energy consumption outside of the servers themselves. The PUE gives an indication of the energy consumption of the facility systems in relation to the IT equipment. If the PUE is 2, the facility systems (such as cooling systems and emergency power supply) consume as much energy as the IT equipment. The closer to 1, the smaller the total electricity consumption. When organisations have servers on their own premises, the PUE is often between 2 and 2.5[xii]. The KB data centre in The Hague has a PUE of 1.9.

Together with the electricity consumption of the infrastructure, the servers on which Delpher ran at the KB consumed 460 MWh in one year.

A colocation data centre, such as that of the ODC B/CIE, houses the servers of several organisations. This up-scaling makes server cooling much more efficient, resulting in a lower PUE (1.34 at ODC B/CIE). Figure 9 shows the difference in PUE between the location of the KB and the ODC B/CIE.

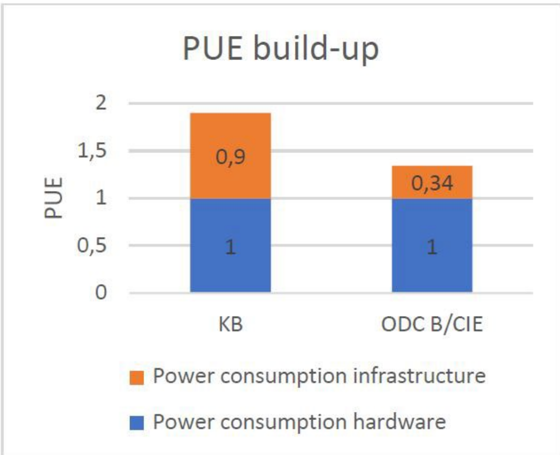

Figure 9: PUE build-up

This efficient infrastructure increases the savings achieved in terms of energy consumption in the new situation. Due to the lower energy consumption of the servers and the lower PUE, the energy consumption of the facilitating infrastructure for Delpher is 151 MWh lower annually than in the old situation at the KB.

Together with the savings from the hardware itself, this amounts to 196 MWh per year, equivalent to the consumption of 79 average Dutch households. This represents a saving of 109 tonnes of CO_2 based on grey electricity. However, green electricity is purchased both at the KB and at the ODC B/CIE. So is the CO_2 savings really that high?

3.2.4. Green electricity

Purchasing green electricity is one of the fastest ways for an organisation to reduce its carbon footprint. Hence the government buys 100% green electricity for its data centres. Figure 10 illustrates the difference in annual CO_2 emissions with green or grey electricity in Delpher's new situation (including infrastructure).

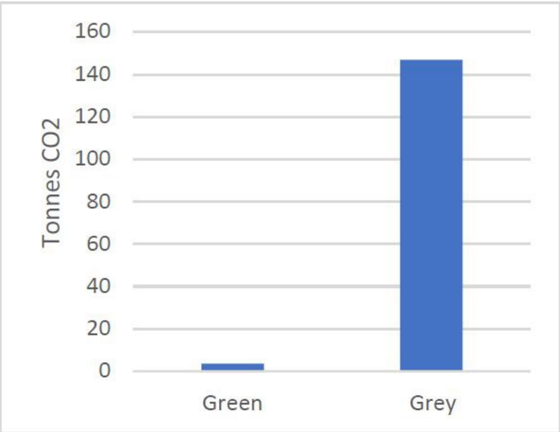

Figure 10: Carbon footprint green versus grey electricity consumption of the ODC B/CIE servers

Because no fossil fuels are used in the generation of green electricity from Dutch wind energy, the CO_2 emissions are many times lower. All that remains are indirect emissions for the construction of wind turbines.

Because green electricity is purchased, the annual carbon footprint of the Delpher servers is very low: less than 4 tonnes of CO_2 per year at ODC B/CIE. The quick conclusion is that the energy savings of the move have had only a relatively limited effect on the carbon footprint: a saving of 2.7 tonnes of CO_2 per year. But if you look at the bigger picture, this energy saving has a much greater effect.

This has been explained in detail in Chapter 3.1.1.

3.2.5. Cloud storage

By storing Delpher data in the ODC B/CIE colocation data centre, energy savings are achieved, in particular due to the improved PUE and more efficient hardware. In addition, the intelligent use of *virtual machines* has made good use of the capacity of the servers. Many other organisations use cloud storage to achieve the same goals. Companies such as Microsoft and Amazon offer this service from huge data centres. These centres have been in the news a lot recently because of their high electricity consumption. What is often not mentioned, is that these data centres are actually very efficient in what they do.

The most important difference with a colocation data centre is that with storage in the cloud, an organisation does not have its own servers. The calculations and data storage of different organisations are carried out on the same servers. This has the advantage that the servers are used much more effectively than in most organisations, since an organisation’s own servers are usually not used for a significant part of the time, but still consume electricity.

Moreover, the PUE at data centres of cloud providers is even lower than most colocation data centres. This is partly due to the way they house their servers. This is done in the smallest possible space so that cooling and maintenance are as effective as possible. This is called *containerisation*.

For heritage institutions that are unable to use their servers in the most optimum manner themselves, cloud storage can be a quick and effective way to reduce the carbon footprint.

Moving to 'the cloud' does have one major disadvantage: you no longer know where your data is physically stored. Since cloud services are offered by large international companies, data may also reside on foreign servers.

Since Delpher contains Dutch cultural heritage, the choice was made to store the data on Dutch territory with a trusted partner.

3.3. Hardware

In addition to electricity consumption, the hardware of the servers also has a hidden CO_2 impact. This is because the production process of servers (and all the other products an organisation purchases) also causes CO_2 emissions. These arise during the various steps of the production process, such as the extraction of raw materials, their processing and transport. Collectively, this produces a quantity of CO_2 that has already been emitted before the server is put into use. The fewer products an organisation purchases, or the more sustainable the products, the lower the CO_2 emissions. Therefore, it is important not only to look at the power consumption of the servers, but also at the hardware of the servers themselves.

A server consists of many different parts and each one may be made of different materials or manufactured by a different brand. All of these factors affect the server's carbon footprint. The largest polluters are the chips that are present in servers; during their production process, a lot of CO_2 is released. Because not every company has mapped the footprint of its products in detail, standards are often used for the carbon footprint of products. For this project, reference is made to the hardware that is present at the ODC B/CIE and how it is distributed over the various components of the servers. Not every specific component and brand is included because of the sensitivity of this information.

In order to compare the footprint of the hidden impact of the servers with other sources that have an impact, the total CO_2 emissions are divided over the expected lifespan (four years). The ODC B/CIE total emission amounts to 16,522 kg CO_2 per year.

This impact is divided into three major groups of hardware: the servers that provide the computing power for Delpher, the storage modules that hold the data for this computing power, and the hardware where the digital collection is stored. The distribution of these is shown in Figure 11.

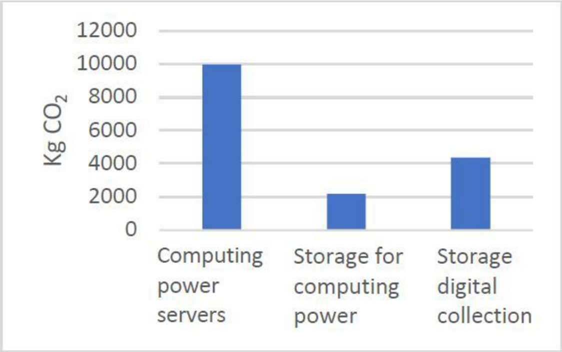

Figure 11: Indirect CO_2 impact of the servers themselves

In the new situation, when storing the digital collection, the hardware generates a relatively smaller footprint. This is because the data in the new situation is on HDDs of 10 Terabytes instead of 1 Terabyte. Even though in the old situation only 400 terabytes of data were stored on HDDs and in the new situation over 2.000, the new situation still requires fewer HDDs. At the same time, the size (and thus the amount of material and the carbon footprint) of the HDDs remains roughly the same. It also reduces the need for peripheral equipment around the HDDs. Together, this results in a lower footprint that arises from the production of the hardware. This is illustrated in Figure 12.

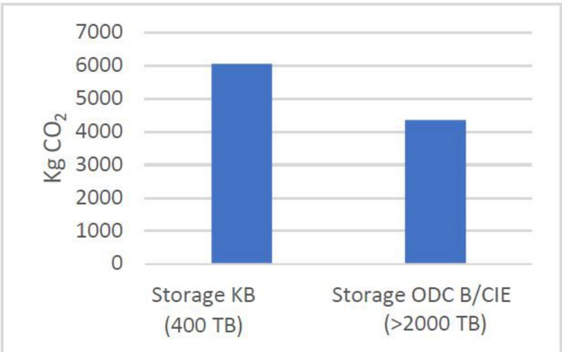

Figure 12: Difference in CO_2 impact of servers at the KB and the ODC B/CIE due to HDDs with more storage

3.4. Data transmission

Not only do the servers and the buildings they are in use electricity, but sending data from one place to another also consumes power. At Delpher, the data from the servers pass through various cables and intermediate stations (meaning: the Internet) and ultimately reach the user's laptop via the Wi-Fi router. The amount of electricity required depends on several factors, such as:

- amount of data;
- distance from the server;
- number of intermediate stations;
- type of transmission, e.g. via cable or Wi-Fi;
- efficiency of all devices.

It is impossible to pinpoint all these factors exactly. Not only are there many different parties involved, but there are also differences per user and data provider. For example, end users can have a different Internet Service Provider, or live much further away, which means that sending the same amount of data has a different electricity consumption. It is therefore not possible to accurately determine the electricity consumption per GB of data sent for one particular service. So this cannot be calculated precisely for Delpher either.

An additional complicating factor is that devices such as a Wi-Fi router also consume electricity when they are not transmitting data. This electricity consumption increases due to the processing of data, but how much this is depends on the device and the amount of data sent[xiii].

Because it is difficult to put a precise figure on data use, estimates of this vary widely. For example, some media have reported that one hour of video streaming generates the same CO_2 emissions as driving 16 kilometres (3.2 kg CO_2)[xiv]. Further investigation has shown that this was estimated far higher than is actually the case, due to various reasons[xv]. Based on various sources, an electricity consumption of 0.056 kWh per GB of data[xvi] has been calculated for this project.

For streaming videos in the Netherlands, this means a consumption of approximately 50 grams of CO_2 per hour.

These are a lot less emissions than some sources would have us believe, which is good news for a platform like Delpher. But in the meantime, our data use is increasing exponentially, mainly because of higher quality photos and videos. The higher the resolution of files, the more data needs to be sent. As a result, it is likely that the impact of data transmission will only increase in the coming years.

3.4.1. Data migration

One situation in which a great deal of data is transmitted all at once is in the migration of a digital file. In the case of Delpher, for example, the migration from the server in the KB to the servers at the ODC B/CIE.

As this involves one Petabyte of data, that is one thousand Terabytes or one million Gigabytes, the data migration generates a significant carbon footprint.

Using the above estimates of electricity consumption for data transmission, this means:

- over 55,000 kWh, which is equivalent to the annual consumption of about 23 average Dutch households;
- with Dutch grey electricity, this results in a carbon footprint of 31 tonnes of CO_2 equivalents.

For data migration, there is also the significant difference whether this is done from tape storage or from HDD. This is because in HDD, the rotating parts consume more electricity than tape. As a result, HDD storage requires about 20 times as much electricity as tape storage during the migration of 1 Petabyte of data[xvii].

Heritage institutions that want to store their data elsewhere because of better hardware and infrastructure should therefore take this into account.

However, if the migration yields structural efficiency, this additional consumption will automatically be reduced later. Because the ODC B/CIE consumes about 195,000 kWh less electricity per year than the KB, more energy was saved after only four months.

In addition to the migration of data by the heritage institutes, data is naturally also sent when the platform is accessed by a user. We will now zoom in on how the use of Delpher contributes to the CO_2 footprint of digital heritage.

3.5. Use of Delpher

In addition to the use of the servers, this project also looked at the use of Delpher itself. Indirectly, the use contributes to the carbon footprint because the ICT resources need electricity to carry out the actions. These include:

- actions in which the Delpher servers deploy computing power;
- loading web pages;
- downloading files.

To investigate this, the user data of 2021 was analysed. Because data was not available from every single day, a selection was made of 132 days, evenly distributed over the weekdays. Data was mapped per action, including loading a web page, downloading files or watching a video.

Figure 13 shows activity on Delpher on different weekdays based on data consumption.

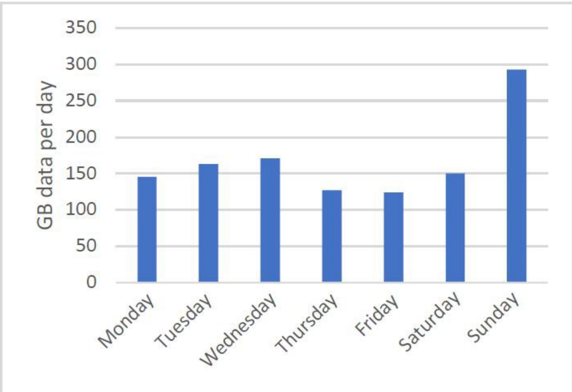

Figure 13: Data usage per weekday (2021)

It can be clearly seen that there is almost twice as much activity on the Delpher platform on Sundays as on other days. This is beneficial for the electricity consumption of the servers because they are also used for other KB software, which runs mainly during weekdays. As a result, the consumption is distributed and there is a low peak consumption. The main advantage of this is being able to make do with a limited number of servers.

Monthly usage for the period January-November 2021 was also considered. Figure 14 shows strong differences between the months.

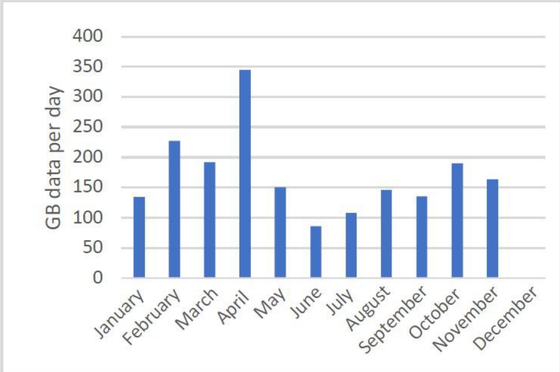

Figure 14: Data usage per month (2021)

Interestingly, the seasons seem to indicate a trend: more use of Delpher in the winter and less in the summer[xviii]. Another possible explanation is the corona measures: usage peaked in the months of the curfew and the lockdown in the third wave.

3.5.1. Carrying out (search) actions

On Delpher's platform, users can carry out many actions such as registering, subscribing to the newsletter or downloading something. One of the most important actions is searching for data in the Delpher database. Based on the user's input, the servers go to work to find a match. These and other actions are carried out by so-called *virtual machines*, as discussed earlier in chapter 3.2.1.

The electricity consumption (of computing power) of servers can only be measured or calculated in its entirety.

As Delpher has 200-250 *virtual machines* running, many of them at the same time, it is not possible to calculate an increase in electricity consumption per generated action. However, it is likely that the search actions require the most computing power, as this involves searching all of the Delpher data.

By limiting searches, the electricity consumption of the servers could be reduced. At present, this is already made possible by offering users the option of searching through only one type of medium. Ways to promote this could include:

- adding additional selection options, such as volumes or types of newspapers or magazines;
- not offering 'search all' as the default option, but allowing users to specify which media should be searched instead;
- adding a request on the web page to limit the scope of the search query. This will make users more likely to use this feature.

3.5.2. Opening and downloading files

After a search has been carried out, the results must be loaded on the user's computer. This involves sending data to be displayed, such as small depictions of the results found. When the user clicks on one of the search results, the object page is loaded. This shows the medium found, often the page from a newspaper, book or magazine.

If users wants to save this file themselves, they can download it to their own computer. There are also many other actions that can be carried out on Delpher, such as watching videos. These four types of action (searching, viewing the object page, downloading and other actions) have been examined for data usage.

Figure 15 shows that in particular viewing newspapers and other actions account for much of the data usage by Delpher users.

In the case of newspapers, this is partly due to the fact that two to six times as many newspapers are downloaded as magazines and books. Moreover, the object pages of newspapers are generally larger in data use than those of magazine and book pages.

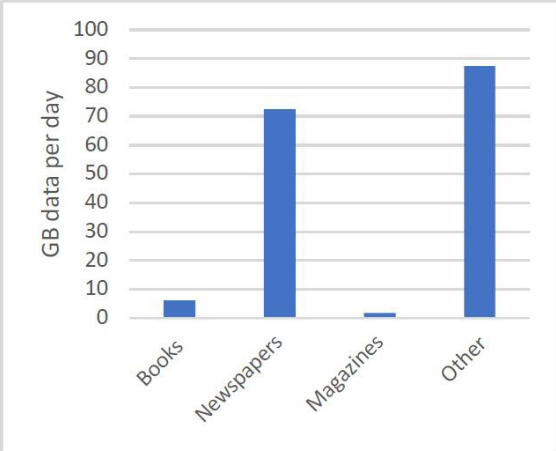

Figure 15: Data usage per medium

It is also notable that in all three media, object pages account for most of the data usage. Only a small part of this is due to the fact that these pages are opened more often (1.4 times for books and newspapers and 2.7 times for magazines) than search pages. The file on the object page is of high quality so that the user can view the file properly. As a result, the amount of data sent to open these pages is a 100 times greater than that of a search page.

The small share that downloaded material has in the data usage is remarkable. Less than 1% of the object pages opened result in a downloaded document. Most likely, users only want to view the object page. Another explanation is that users view multiple object pages until they find (or do not find) the desired information.

To reduce the data usage of object pages, Delpher can offer object pages in lower quality. If a user then wants to study or save the page in more detail, they can download the file in higher quality.

An alternative would be to make the page available via a button on the object page.

By far the most data traffic (98%) of the other actions comes from watching (instructional) videos. These videos (which are not part of the digital collection but explain how to use Delpher) are on YouTube, but can be viewed via the Delpher platform. This means that the videos are stored on YouTube servers and sent to the user from there. As a result, these videos do contribute to the amount of data sent, but this will not be reflected in the electricity consumption of the Delpher servers. Since videos are moving images, watching them costs much more data than an image. It is striking that this category is so large, as Delpher is primarily focused on searching and viewing documents.

Possible measures to reduce data consumption could include:

- offering videos on the platform in a lower quality;
- providing information in a form that uses less data.

4. Recommendations

Storing and making available digital material generates CO_2 emissions. It is therefore important that heritage institutions, which make more and more material available digitally, do so as climate-consciously as possible. This report provides insight into how this can be done, based on a case study of the Delpher platform. The following are a number of recommendations for other heritage institutions that want to reduce their carbon footprint.

4.1. Servers

Servers provide the computing power and storage needed to host and make available a digital collection. These servers are the core of CO_2 emissions. This is due both to the electricity consumption and to the indirect CO_2 emissions from the production of the servers.

- In general, storage on tape or SSD requires less energy than storage on HDD. However, this depends heavily on the type and brand of the equipment. In addition, the fast or slow availability of the medium is also a factor to take into account.
- The more storage per disk (for example, 10 terabytes instead of 1), the more energy-efficient the storage is. Not only the storage itself, but the processors also require less electricity.
- Creating digital compartments in the servers ensures that the capacity of the servers can be used more efficiently. This can be done using *virtual machines*. This improves capacity utilisation, reducing the total number of servers required.

4.2. Server locations

The location of the servers has a major influence on total electricity consumption. If data is stored locally, it is likely that facilitating the servers (including cooling them) will consume as much or even more than the servers themselves.

Measures that can be taken are:

- Sharing servers with multiple parties to use them as effectively as possible. Delpher is a good example of this.
- Housing the servers in an (external) colocation data centre can significantly reduce electricity costs. Because there are many servers here, facility systems such as cooling can do their work much more effectively.
- By making the digital collection available via a cloud environment, control of the servers is transferred to large data centres such as Microsoft or Amazon. The advantage of storage in a cloud environment is that this type of provider is at the forefront of the development of facility systems and the efficient use of the capacity of their servers.

4.3. Collection and use

Retrieving files from a digital collection generates CO_2 emissions. Several measures can reduce this impact, such as limiting searches so that fewer files are searched;

- adding additional selection options, such as volumes or types of newspapers or magazines;
- not offering 'search all' as the default option, but allowing users to specify which media should be searched instead;
- adding a request on the web page to limit the scope of the search query. This will make users more likely to use this feature.

Furthermore, we can look at how to reduce the impact of downloaded objects. This could include:

- Make files available in lower resolution in the object page.
- Giving an option for high or lower resolution, so users can choose for themselves. In addition, a text could be placed on the website that choosing a lower resolution saves CO_2.

This case study focused on the storage and use of digital selections on the Delpher platform. One factor that we have not taken into account in this case study is the impact of a stricter selection and preservation of digital heritage on the total CO_2 emissions of a heritage institution. The exact impact and the measures that would be effective could be investigated further. Naturally, a heritage organisation wants to be as complete as possible in the material it offers its users. However, potentially reducing the digital collection can logically prevent CO_2 emissions in several ways. For example, there is probably less storage capacity needed, which means fewer servers have to be purchased. In addition, electricity consumption will decrease. Furthermore, searches will involve a smaller amount of data, which also results in reduced electricity consumption. Examples of measures to investigate include:

- Preventing and removing duplicates: parts of a collection may be stored in duplicate. Removing this duplication prevents the digital collection from becoming unnecessarily large.
- Making agreements between heritage institutions to prevent duplication. If something is already available digitally elsewhere, is it necessary to digitise and store it again?

With the findings of the report and the recommendations mentioned above, heritage institutions can start to examine the carbon footprint of their digital collection and make choices for a climate-proof future.

5. Sources and endnotes

i https://www.milieucentraal.nl/energie-besparen/inzicht-in-je-energierekening/gemiddeld-energieverbruik/

ii https://www.trouw.nl/duurzaamheid-natuur/wij-consumerende-nederlanders-zijn-de-echte-vervuilers~bbdb1b6c/

iii https://www.environdec.com/home

iv http://idematapp.com/

v For a comprehensive account of the use of emission factors and databases, contact PHI Factory.

vi https://www.datacenterdynamics.com/en/opinions/optimizing-server-refresh-cycles-with-an-aging-moores-law/

vii https://www.milieucentraal.nl/energie-besparen/inzicht-in-je-energierekening/gemiddeld-energieverbruik/

viii https://nl.wikipedia.org/wiki/Tapedrive

ix https://nl.wikipedia.org/wiki/Harde_schijf

x https://nl.wikipedia.org/wiki/Solid_state_drive

xi This includes both direct and indirect CO_2 emissions from the generation of grey electricity. https://www.co2emissiefactoren.nl/lijst-emissiefactoren/

xii https://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf . Page 24

xiii It depends on the amount of data traffic how much the electricity consumption of devices increases compared to no data traffic. For example, the increase in electricity consumption of a router (compared to no data traffic) is about 1.7% when streaming Netflix and 11% (6.5 times more) for heavy use such as gaming, while consuming 9 times more data.
Jens Malmodin, Ericsson Research. https://www.youtube.com/watch?v=Xo0PB5i_b4Y&t=2504s

xiv https://nypost.com/2019/10/28/why-climate-change-activists-are-coming-for-your-binge-watch/

xv https://www.carbonbrief.org/factcheck-what-is-the-carbon-footprint-of-streaming-video-on-netflix

xvi
- Joshua Aslan et al, 2017. Electricity Intensity of Internet Data Transmission: Untangling the Estimates. https://onlinelibrary.wiley.com/doi/full/10.1111/jiec.12630
- Jens Malmodin, Ericsson Research. https://www.youtube.com/watch?v=Xo0PB5i_b4Y&t=2504s
- Malmodin, J. and D. Lundén. 2016. The energy and carbon footprint of the ICT and E&M sector in Sweden 1990–2015 and beyond. Paper presented at (and in the proceedings of) ICT4S, 29–31 August, Amsterdam.
- Pihkola et al, 2018. Evaluating the Energy Consumption of Mobile Data Transfer—From Technology Development to Consumer Behaviour and Life Cycle Thinking. https://www.mdpi.com/2071-1050/10/7/2494
- https://www.carbonbrief.org/factcheck-what-is-the-carbon-footprint-of-streaming-video-on-netflix

xvii https://www.cmd-ltd.com/advice-centre/usb-chargers-and-power-modules/usb-and-power-module-product-help/data-transfer-rate/

http://www.tape-storage.net/en/storage_comparison/article_01/

xviii A caveat here is that only five and nine days of data were available from January and April respectively. For the other months, this was eleven or more, making the extrapolation of those data more reliable. In April, two of the nine days were Sundays which, as we have seen, show a higher consumption. As a result, this month's average may have been higher. On the contrary, there was no data available from any Sunday in January.