

How to limit the environmental impact of data

Wednesday November the 20th 2024

Authors:
Pepijn van der Veen (ING)
Ronald Meijer (IBM)
Eloise Zhang (UvA/VU)
Gerrit Ouderkerken (NetApp)

2

Speakers



Ronald Meijer
IBM



Pepijn van der Veen
ING



Gerrit Ouderkerken
NetApp

Table of contents

- 01 Welcome & NCDD
- 02 Quiz / The Dirty Secret of Data
- 03 Introduction
- 04 DMBok and the Lifecycle of data
- 05 The lifecycle phases and sustainability
 - Planning & Design
 - Create & Store
 - Use & Enhance
 - Dispose
- 06 Questions / Discussion



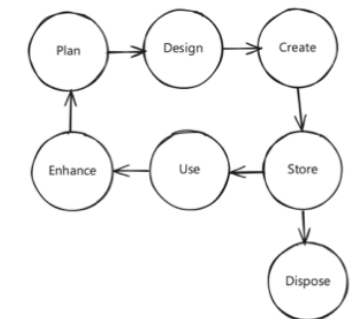
Sustainable data management: Implementing a Sustainable Data Strategy Across the Data Lifecycle

EXECUTIVE SUMMARY

The increasing digitization of our world has led to an exponential growth in data creation, storage, and processing. With this growth comes a significant environmental impact due to the energy consumption and cooling requirements of data centers, as well as the raw materials required for storage solutions. Sustainable data management has therefore emerged as a critical strategy for organizations to address the environmental footprint of their data lifecycle.

Sustainable data management is not just a technical or operational issue; it is a strategic imperative at the executive level as well as a cultural mind shift.

This paper uses the DMBok phases to address the sustainability considerations in all the phases of your data lifecycle, from planning, creation until the data is destroyed. We address the following:



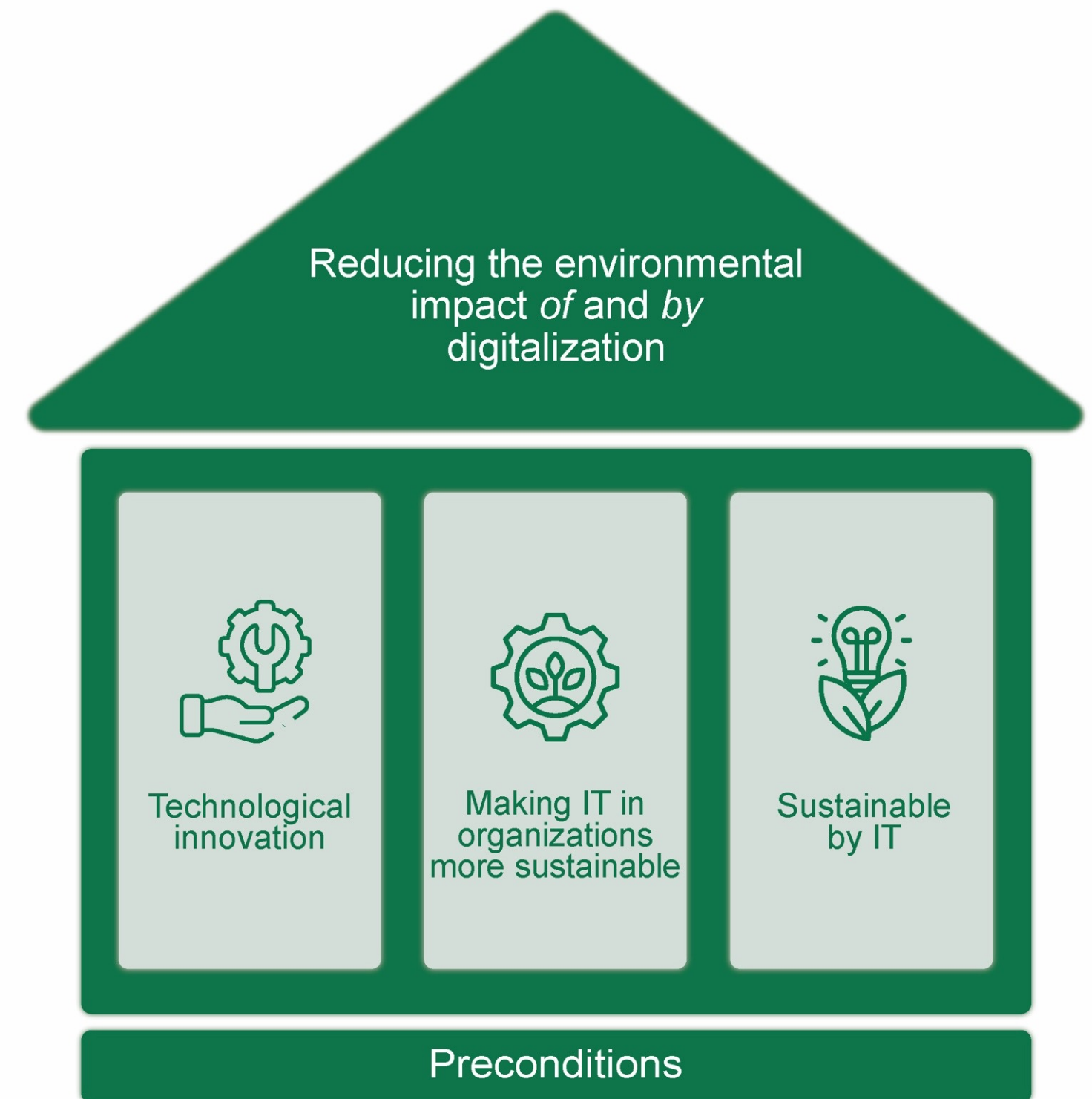
Planning

- Create awareness in your organisation about the sustainability impact of data storage and usage.
- Plan to avoid copies of data in your organisation and use one single version of the truth as a central source of information
- Plan to use the best technology for processing data.

["Sustainable data management: Implementing a Sustainable Data Strategy Across the Data Lifecycle"](#)



HOUSE OF SUSTAINABLE DIGITALIZATION



SCAN THE CODE

The Dirty Secret of Data: How to
limit the environmental impact of
data.



Introduction

The footprint of data is increasing!

Exponential data growth

The increasing digitization of our world has led to an exponential growth in data creation, storage, and processing. With this growth comes a significant environmental impact due to the energy consumption and cooling requirements of data centers, as well as the raw materials required for storage solutions.

The annual electricity report from the International Energy Agency of 2024 says data centers consumed 460TWh in 2022. This figure could rise to 1,050TWh by 2026 in a worst-case scenario due to developments like AI. This increase is equivalent to adding the entire power consumption of a country like Germany.

Need for a strategy

Sustainable data management is a critical strategy for organizations to address the environmental footprint of their data lifecycle. Sustainable data management is not just a technical or operational issue; it is a strategic imperative at the executive level as well as a cultural mind shift. We have developed an architectural framework to support the development sustainable data management.

The lifecycle of data

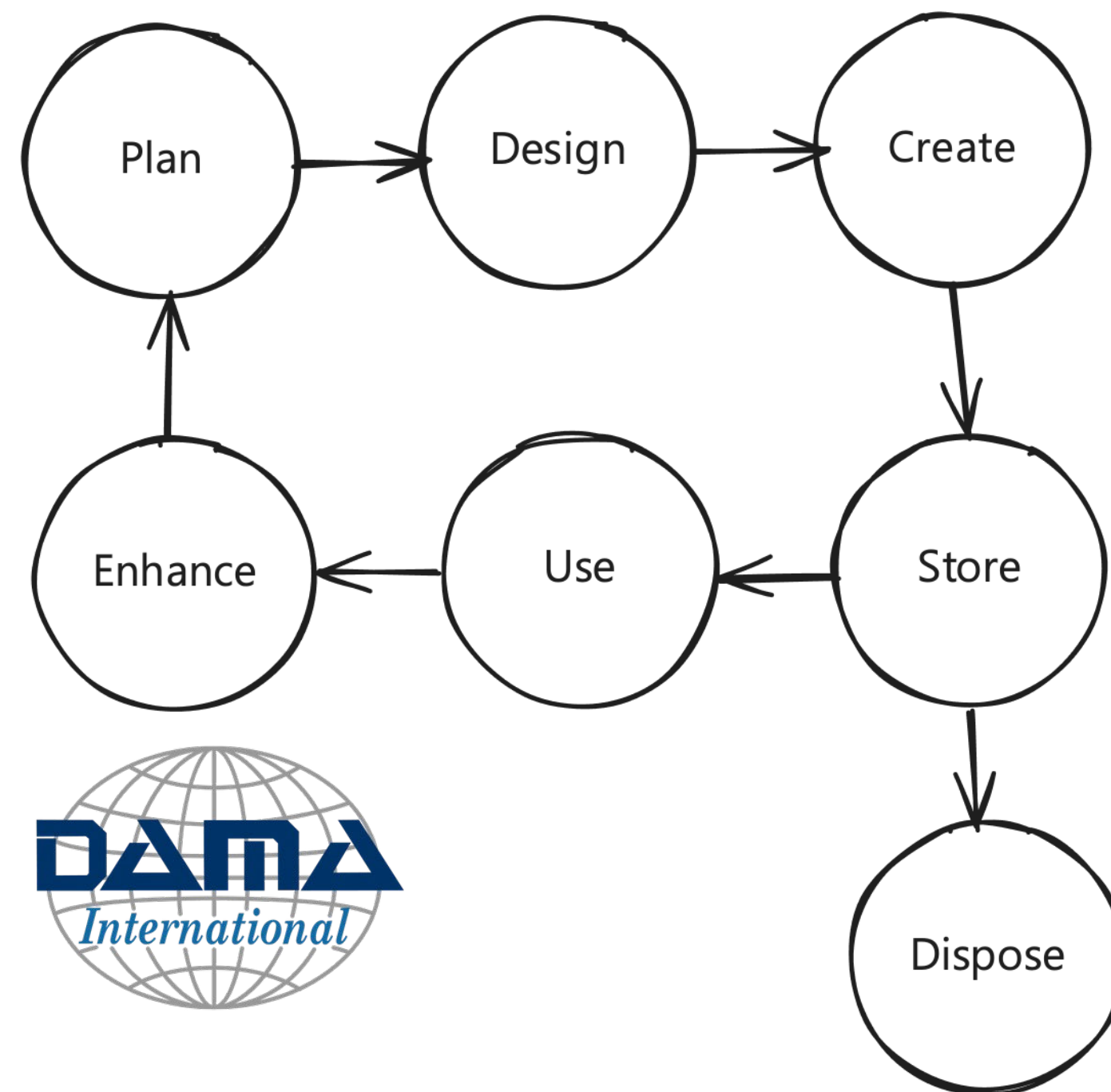
The lifecycle stages from DMBok as the umbrella

Market model developed by DAMA

We adopted the data lifecycle model that is published by DAMA in the DMBok (Data Management Body of Knowledge). This is a widely adapted international standard for data management.

The footprint during the lifecycle

The data lifecycle phases are used to address the sustainability considerations, from planning, creation until the data is destroyed. Every stage has own characteristics and possibilities to lower the impact on our planet.





Planning

Create awareness

Create awareness in your organization about the sustainability impact of data storage and usage.

Plan to avoid copies

Make a plan to avoid copies of data in your organization and use one single version of the truth as a central source of information

Plan on the best technology

Make a plan to use the best technology for processing data.

Een goed
begin is het
halve werk



Design

Use the right patterns

- Compress the data.
- Avoid data duplication in your design.
- Use a tiered storage solution.
- Consider the usage of virtualization techniques.

Classify your data

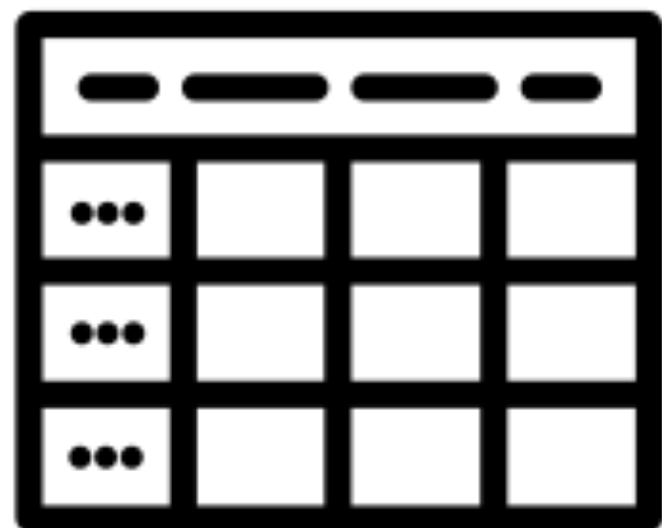
Make sure that you include the right meta data in your design to enable data retention.

Be clean

Include data disposal in your design from the beginning.

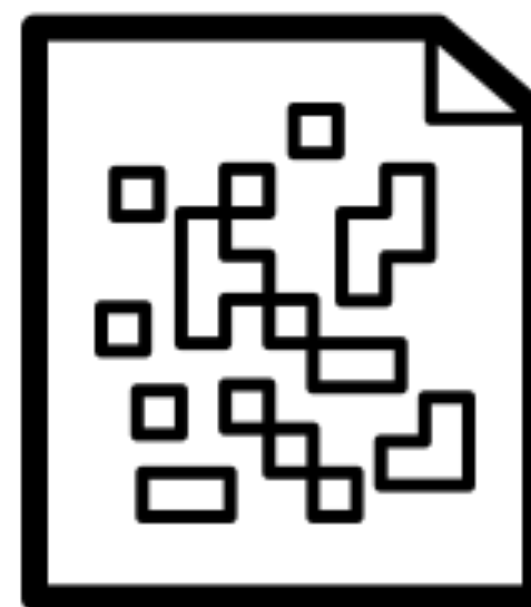
Goed
gereedschap
is het halve
werk

Create



Structured

Over



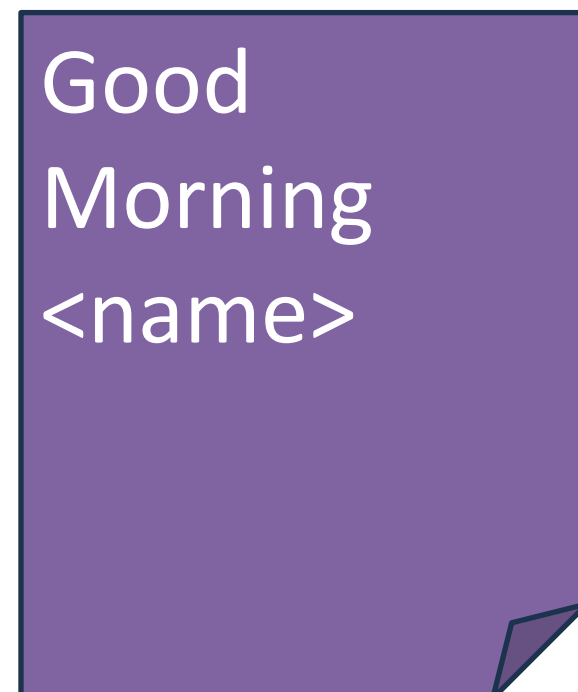
unstructured



Create: Example

From Structured to PDFs

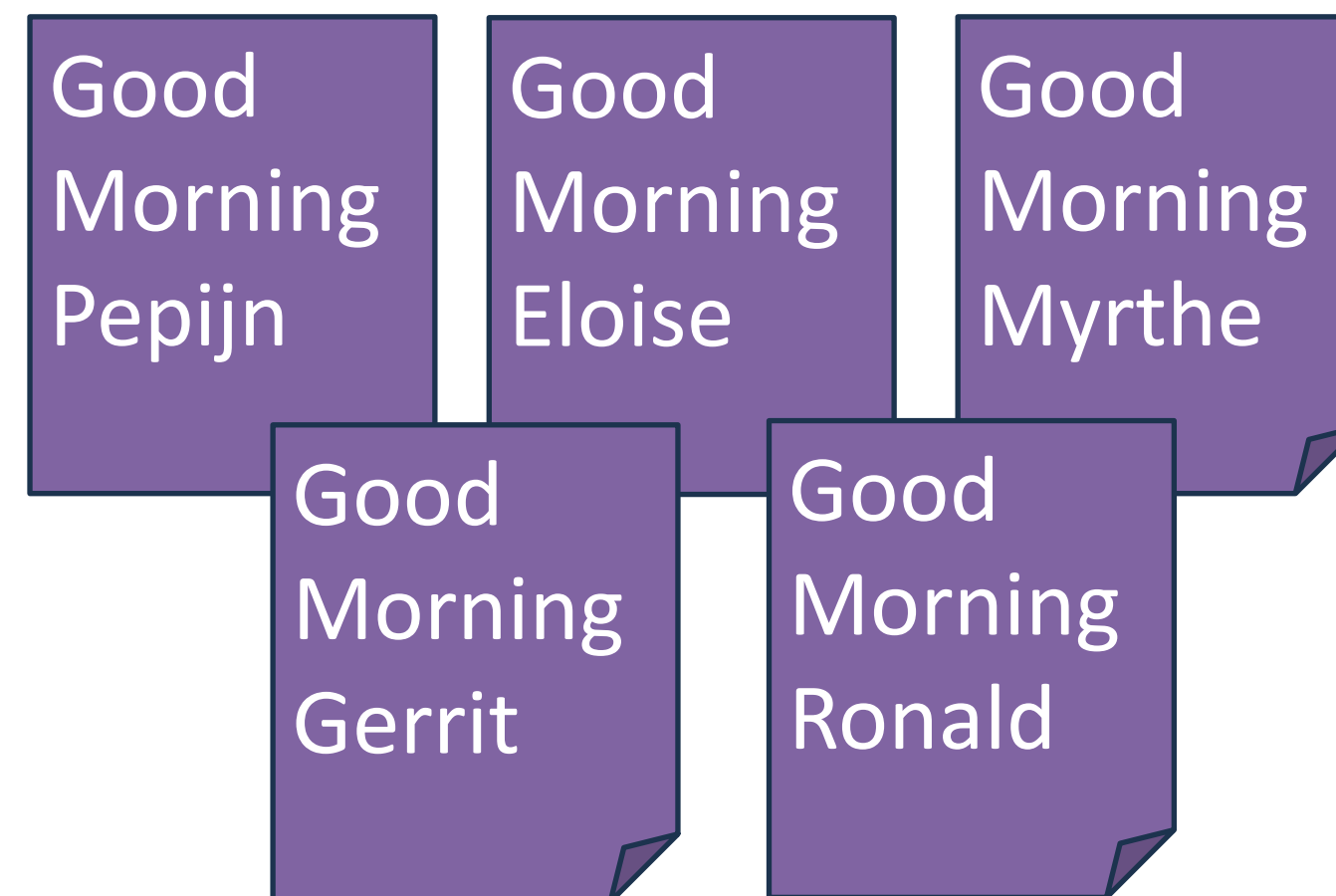
Name
Pepijn
Eloise
Myrthe
Gerrit
Ronald



1 template



to PDFs



72x more space!

Excuse: "It is legally required"



Store

Use the right technology

Implement the best mechanism to store and retrieve data. E.g. RDMS

Implement storage tiering

Select the different storage tiers and implement tiered storage

Compress data

Store the data in a compressed format (data at rest)

Use the right location

Evaluate the place where you store your data and the products that you use from the perspective of sustainability

Everything
Has A place
and There's
A Place For
Every Thing



Use

Compress data in transport

Compress the data while it is moved to be processed and when processing tool place

Only ask for what you need

Be laser focused and only retrieve those data that you need, both in number of documents and rows as well as on the content.

Get most out of your resources

Right level of utilization: Minimize the amount of physical resources

Use the right coding

Use green coding practices to process data

Optimize your search process

Use optimized search algorithms for data processing

*A life
without
purpose, is a
life without
direction*

(Tony Gaskins)



Enhance

Delete the old versions

If you update your data, make sure that you delete the old version

Store delta's

If you must keep history, store the delta's only

Apply retention on all versions

Apply retention policies, both for the new as well as the old version

Aggregate if possible

Aggregate data if you do no longer need the details





Dispose

Minimize historical data



Set the retention policies on
the data



Run the process to clean your
data



Comply and report on your
data retention standards

"The real problem is that we have far more than we need or want"

- Marie Kondo

! Conclusion

There is not so much difference between the lifecycle of your daily household and your data

You buy something because (they told you) that you need it.

You determine your purchase criteria in an efficient way.

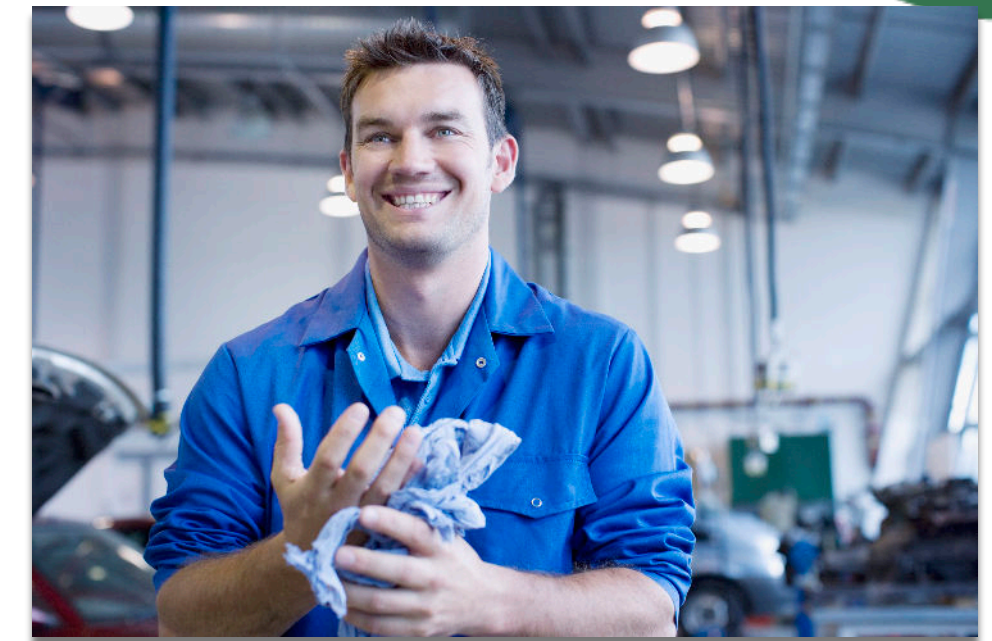
Price, size, quality

You find an optimized way to store it

You maximize its usage

If you need to get out more, you enhance it rather than buying something new

If you do not need it, you get rid of it



Questions?



The publication

"Sustainable data management: Implementing a Sustainable Data Strategy Across the Data Lifecycle"



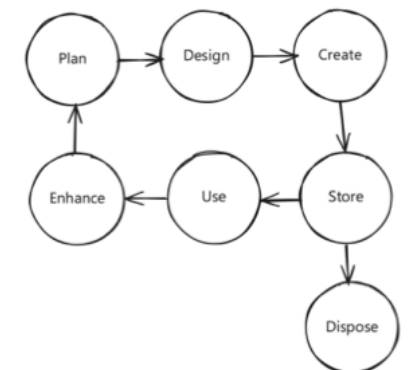
Sustainable data management: Implementing a Sustainable Data Strategy Across the Data Lifecycle

EXECUTIVE SUMMARY

The increasing digitization of our world has led to an exponential growth in data creation, storage, and processing. With this growth comes a significant environmental impact due to the energy consumption and cooling requirements of data centers, as well as the raw materials required for storage solutions. Sustainable data management has therefore emerged as a critical strategy for organizations to address the environmental footprint of their data lifecycle.

Sustainable data management is not just a technical or operational issue; it is a strategic imperative at the executive level as well as a cultural mind shift.

This paper uses the DMBOK phases to address the sustainability considerations in all the phases of your data lifecycle, from planning, creation until the data is destroyed. We address the following:



Planning

- Create awareness in your organisation about the sustainability impact of data storage and usage.
- Plan to avoid copies of data in your organisation and use one single version of the truth as a central source of information
- Plan to use the best technology for processing data.



THANK YOU!

Nationale Coalitie Duurzame Digitalisering