CURRENT DATA MANAGEMENT STRATEGIES THAT REDUCE IT ENERGY CONSUMPTION, MITIGATE DATA GROWTH, AND DECREASE CARBON DIOXIDE EMISSIONS

SYSTEMATIC MAPPING REVIEW PROTOCOL

VRIJE UNIVERSITEIT AMSTERDAM

Engelbert Fellinger, Patricia Lago

V1.0, 13 May, 2025

ABSTRACT

This document outlines the review protocol for a systematic mapping study on current data management strategies that contribute to an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions.

KEYWORDS

systematic mapping study, energy-efficient data management strategies, data growth, carbon accounting

CONTENTS

1. BACKGROUND

Data management strategies entail planned approaches that aim to ensure the secure, efficient, and cost-effective generation, storage, and utilisation of data. Several of these strategies, such as data classification and data placement, can potentially help reduce IT energy consumption, manage data growth, and decrease carbon dioxide emissions, either explicitly or implicitly.

Data management involves the people, processes, and technology needed to manage an organisation's data from creation to deletion [27]. The Data Management Association (DAMA), defines data management in the Data Management Book of Knowledge 2nd edition [22] as follows: "Data Management is the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.". Furthermore, data management assists individuals, organisations, and connected devices in optimising their use of data while adhering to policy and regulatory guidelines so that they can make decisions and take actions that maximise the benefit to the organisation [12].

IT organisations in the EU are under increasing pressure to meet environmental targets concerning the use of ICT equipment and services [24], [25]. The growth of edge computing and the shift of data storage from on-premises to cloud solutions are recent developments resulting in more data being generated at an accelerated rate [20], [21]. As the energy efficiency of data centres and data transport improve, the gains are outpaced by the current and projected uncontrolled growth of data, resulting in rising greenhouse gas (GHG) emissions and environmental damage [17]. There is concern that the energy required for digital data may eventually exceed our capacity to produce it [33]. According to IDC [20], the amount of new data created each year is projected to grow annually by approximately 26% between 2015 and 2025. By 2025, it is expected that 175.8 zettabytes (ZB) of new data will be generated, compared to 18.2 ZB in 2015. In terms of enterprise data storage, it is estimated that by 2025, there will be 9 ZB of stored data, up from just 0.8 ZB in 2015. Depending on the context, about 70% of the data stored can be considered digital clutter where the value is unleveraged, lost, or not understood [20].

The goal of this systematic mapping study is to map current data management strategies that contribute either explicitly or implicitly to a diminishing effect on IT energy consumption, data growth, and carbon dioxide emissions. This classification of strategies, or taxonomy will serve as the basis for a field study to develop an integrated decision model intended for decision-makers such as chief information officers, network designers, data architects, and software architects. The classification and resulting decision model will enable IT organisations to formulate the most energy-aware policies by comparing and combining data management strategies and considering economic, technical, social, and environmental implications. Additionally, the classification and decision model are expected to aid relevant stakeholders in gathering requirements and translating them into sustainable, conceptual, logical, and physical data designs.

1.1 EXISTING LITERATURE

The energy consumption through the production and operation of hardware has been relatively well studied, and investigations on the effects of software and data on the overall energy consumption of ICT are beginning to appear as stated by Bieser et al. [16], and Santarius et al.[17]. Research that investigates AI applications increasing costs in terms of energy consumption, and emissions is gaining momentum. Case and point include the 2023 publications about the growing energy demands of AI by Luccioni, Jernite, & Strubell [18], and De Vries [27]. From the data management perspective, surveys on energy-aware data management strategies in cloud-related environments are starting to emerge, as You et al. pointed out in 2020 [19]. The amount of new data generated each year, as well as the projected growth, is well-researched by IDC across multiple years [20]. Nevertheless, Jackson and Hodgkinson [32] note that the topic of data growth is largely overlooked in academic literature regarding digital decarbonisation. Understanding, quantifying, and reporting an organisation's carbon dioxide emissions through carbon accounting (CA) is an established corporate practice [13], [18]. However, the direct and indirect effects of data growth receive minimal attention within this framework. To the best of our knowledge, studies about data management strategies that enable organisations to control IT energy use, data growth, and GHG emissions in an integrated manner, and detail their interdependencies are absent.

1.2 JUSTIFICATION OF THE NEED FOR THE REVIEW

This chapter explains the rationale of a systematic mapping study as a preferred research method. To address the problems caused by uncontrolled data growth, this research aims to bridge the perceived gap between data management strategies, IT energy efficiency, controlling data growth, and CA. The emphasis is on mapping studies that either explicitly or implicitly contribute to the energy-efficient generation, use, and storage of data while managing growth and carbon dioxide emissions. This encompasses architectural strategies, software practices, energy-saving design patterns, related hardware components, and managerial decisions aimed at reducing the data-related energy consumption, carbon dioxide emissions, and data growth from the point of view of IT organisations. However, the scope does not facilitate factors such as the specific types of solar panels used in data centres, renewable energy sources, or compensation schemes such as green roofs.

Currently, there is fragmented information on the types of studies conducted on data management in relation to IT energy-efficiency, data growth, and CA, and how they are defined within their specific application areas. This research aims to fill this gap by analysing both academic and non-academic publications on data management strategies with regards to IT energy use, controlling data growth, and greenhouse gas emissions. This will be done by identifying and categorising existing data management strategies and organising them into a descriptive classification to identify energy-efficiency, data growth control and CA potential. The expected output comprises a visual synthesis and classification of the available studies related to the research question. Intended to provide a broad overview, the classification will aid researchers and practitioners in understanding the characteristics, diversity, and commonalities [4].

Justification of the need and scope of the systematic mapping

- Provides a structured overview of the research area to assess the existing evidence [1], [6].
- Allow us to discover research gaps and trends [1], [6].
- Identify areas which got more emphasis and publicity than others [3].
- The expected output of the systematic mapping study fits the purpose and question of RQ1 [6].
- While a systematic literature review shares commonalities with respect to searching and study selection, synthesising evidence [1] is not a primary objective for RQ1.

- Other methods such as an unstructured literature review may not lead to the desired classification as output.
- The systematic mapping is instrumental for the follow-up research activities as stated in the research proposal.

2 RESEARCH PROCESS

The main objective of a systematic mapping is to provide an overview of a research area by building a classification scheme and structuring evidence on a research field [6]. Throughout the study, we follow the established procedures for conducting systematic mapping studies as defined by Kitchenham and Charters [35], Petersen et al. [1], [2], and Wohlin et al. [10]. The process comprises three main steps: *planning*, *conducting* and *reporting*.

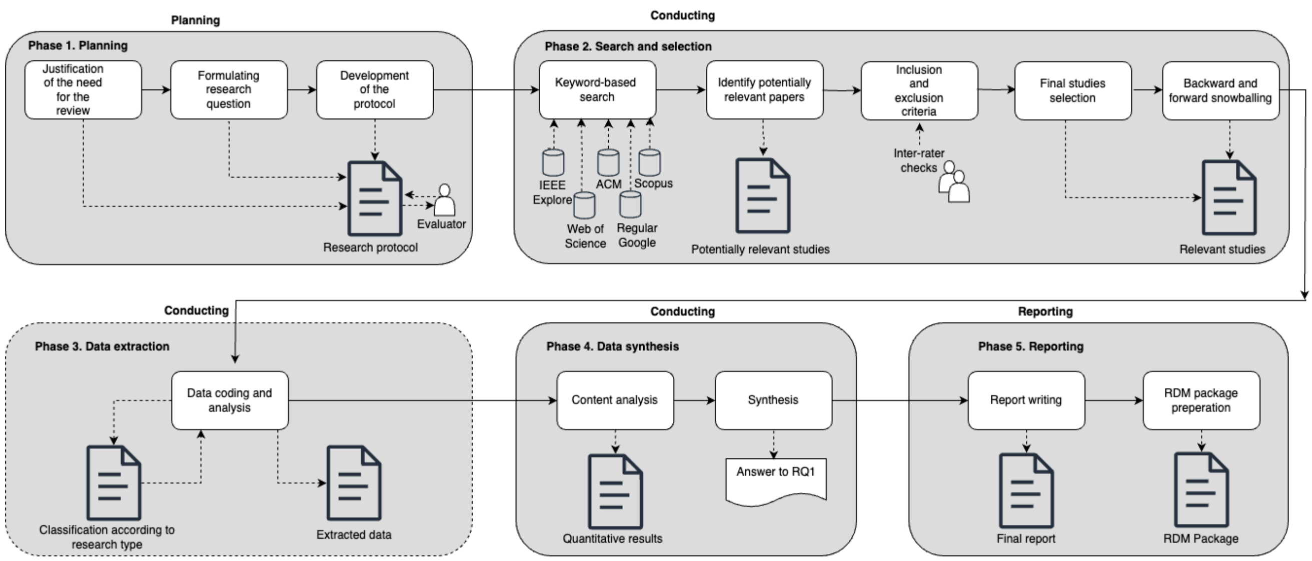

Figure 1 Overview of the study design, based on Kitchenham and Charters [35], Petersen et al. [1],[2], diagram design based on Aberts et al.[34].

In this document, we will describe the planning (2.2), conducting (3), and reporting (4) phases of the systematic mapping study, focusing on the process, activities, and artifacts created for the report in the final phase. The inputs of each stage are the outputs produced by the prior one as illustrated in figure 1.

2.1 PLANNING

The planning phase comprises a detailed research protocol which is described in this document. The first goal of the planning phase is to establish the need to review the current data management strategies that affect energy-efficiency, data growth and CA (section 1.2). The second objective of the planning phase is identifying the research question (section 2.1.2) The third goal is to develop the protocol to be followed by the researcher (section 2.1.3).

2.1.2 RESEARCH QUESTION (RQ1)

The Goal Question Metric (GQM) strategy [23] is used to formulate the objectives of the review: **to analyse** data management strategies, **for the purpose of** gaining a comprehensive understanding of their implications for, among others, energy-efficiency, data growth and carbon accounting **from the point of view** of researchers and practitioners **in the context of** the generation, storage, and utilisation of data.

This systematic mapping study will address the following research question:

What are the state-of-the-art data management strategies affecting IT energy efficiency, data growth, and/or carbon accounting?

This will involve creating a classification scheme to provide an overview of the research area while identifying opportunities for further research.

2.1.3 REVIEW PROTOCOL

This chapter gives an overview of the conducting, and reporting phases of the systematic mapping study.

- *Search and find relevant papers*: We will conduct an automated search using ACM Digital library, IEEE Xplore, Scopus and Web of Science to find relevant peer-reviewed research papers. Additionally, we will perform an automated search for grey literature through regular Google searches. This approach helps us gather perspectives beyond the typical scholarly publication cycle and minimises bias in our results [5]. Grey literature includes theses, reports from reputable organisations, technical documentation, white papers, and policy documents. To identify relevant literature after our searches, we will review the reference lists of the selected peer-reviewed papers and examine both citations that cite them and those they cite. This method is known as snowballing [9].
- *Identify potentially relevant papers*: The identified relevant studies provide practises, solution proposals, concepts, opinions, experiences, or evaluate existing literature that contribute to data management strategies that have an explicit or implicit effect on energy consumption, data growth, and/or carbon dioxide emissions.
- *Screening, quality assessment of the selected papers:* The potentially relevant publications will be filtered according to inclusion, and exclusion criteria (section 3.5 - 3.8) to compile the final list of studies for future review activities.
- *Data coding and data classification*: From the final list of publications, we will code data using a data selection and extraction sheet with set criteria and decision rules for research types in requirements engineering as proposed by Wieringa et al. [28]. This approach allows us to evaluate, and classify the current academic and grey literature according to established decision rules while ensuring consistency and facilitating comparisons with current and future studies [1].
- *Data extraction*: To construct the inventory of current strategies, we employ data from the data selection and extraction sheet and use the aforementioned data classification model. In addition, we will make distinctions between the goals of the strategies, and their means (methods, instruments, and approaches) and organise these into categories [29].
- *Data synthesis:* This phase aims to gather important findings and their implications, which are then used to substantiate the answer to the research question of this study. This involves tabulating and visualising the characteristics of the selected studies using descriptive statistics, along with a synthesis of their goals and means that details the study characteristics and relevant outcome types based on the inventory of current strategies.

- *Reporting (outputs)*: The findings from the systematic mapping study will be utilised, to create the report, and facilitate dissemination. The reporting includes a Research Data Management (RDM) package according to the FAIR (Findable, Accessible, Interoperable and Reusable) principles.
- *Validation:* To mitigate the potential for flaws or researcher bias from the principal researcher, the second researcher conducts checks at critical stages of the research process. Throughout the planning, execution, and reporting phases, the second researcher reviews the procedures and the artifacts generated. To validate the screening of the search results, the principal researcher provides a random sample of 25 studies to both the second researcher and the research advisor. They analyze these studies independently by completing the data selection and extraction sheet according to the established research protocol. Afterward, the researchers evaluate their levels of agreement. We clearly state the achieved rates of agreement in the first round, and subsequent rounds. To prevent generalization of the mapping results, we clearly state the population (N), the methodology used, and the context of the study along with its limitations.

2.1.4 TEAM

Principal researcher: PhD candidate focusing on energy-efficient data management strategies and related fields of research. They will be part of all the activities, including planning, conducting, reporting, and dissemination of the study.
Research methodologist: x
Advisor: x

3. CONDUCTING

In this phase we will carry-out the systematic mapping study by following the predefined protocol.

3.1 SEARCH AND FIND RELEVANT PAPERS

The systematic mapping study will provide an overview of the scope of the area (state-of-the-art data management strategies that have an explicit or implicit effect on energy consumption, data growth, and/or carbon dioxide emissions) and allow us to discover research gaps and trends [1]. Since RQ1 focuses on examining the state-of-the-art of a broad topic area [6], the PICO (Population, Intervention, Comparison, Outcomes) criteria [7] have not been applied. Instead, to maintain the focus on current strategies, a nominal sampling frame (what research has been conducted in the past 5 years?) is used in automated searches and will characterise the literature located within this sampling frame [6]. To ensure the inclusion of essential research beyond the sampling frame of 5 years, we use the reference list of relevant papers to identify relevant peer-reviewed sources, also known as backward snowballing [9].

To identify as much relevant literature to RQ1 as possible, we follow Zhang & Babar [3], based on the framework by Petersen et al. [1], [2]. First, we have conducted multiple manual searches to identify relevant literature to RQ1, validate the keywords, venues, and evaluate the quality of the results. Based on these results, an automated search is developed, refined and executed (Appendix A). To determine relevant literature after the automated searches, we employ both citations that cite selected papers and those they cite, to identify additional peer-reviewed sources, also known as forward and backward snowballing [9]. To validate the performance of the search strategies in relation to the subjects studied, we manually examine the titles and the abstracts of the retrieved material to assess and incrementally improve the coverage and capture of the search results. Coverage or sensitivity for a specific topic is defined here as the proportion of retrieved studies that relate the source's topic to RQ1 [3]. Capture or precision refers to the proportion of retrieved articles that are indeed relevant studies [3]. The aim is to find an optimum that permits finding the most relevant studies with few unwanted results. Given the exploratory nature of this systematic mapping study, we accept a relatively low capture rate to avoid missing potentially relevant studies.

3.2. SOURCES OR VENUES TO SEARCH

Following Petersen et al. [1], [2], the keywords are grouped into sets of full-text automated searches which are performed on ACM Digital library, IEEE Xplore, Scopus and Web of Science to reach an extensive coverage of academic publications [11], and regular Google to include coverage for non-academic publications pertaining to the grey literature search. Where possible, the results of the automated scholarly searches are captured with Harzing's Publish or Perish [14] or downloaded directly and converted to Excel.
The regular Google search is captured with the Google Search Results Scraper extension for use with Google Chrome [15]. The number of results displayed on Google's web page after a standard search is an estimate and can vary. The actual number of results becomes apparent when navigating through the pages, each displaying 10 results at a time. These results, including any labeled as "highly similar" on the final page, are saved to an Excel file.

3.3 SUBJECT, EVIDENCE TYPES TO BE SEARCHED

Subjects of studies under review, scope RQ1
The emphasis is on the energy-efficient generation, usage, and storage of data and managing its growth. This comprises, but is not limited to, architectural strategies, software practices, energy-saving design patterns, related hardware components, and managerial decisions aimed at reducing the data-related energy consumption, and carbon dioxide emissions from the point of view of researchers and practitioners.

Evidence types searched ACM Digital library, IEEE Xplore, Scopus, and Web of Science

- Peer-reviewed academic, and scientific journals, articles, conferences, papers, workshops

Evidence types searched regular Google, grey literature

- Scholarly monographs, edited books and book chapters
- Theses
- Magazines
- Blogs
- Policy documents
- Reports
- White papers
- Technical documentation
- Manuals
- Best practises
- Case studies

3.4 INPUTS TO RELEVANT SEARCH ENGINES

In the searches, we strive for a balance between comprehensiveness and maintaining relevance in the hits. The keywords and their order below are selected to retrieve as many studies related to RQ1 as possible while staying close to the subject of studies.

- Data management strategies
- Energy
- Carbon
- Data growth

To mitigate the fact that this scoped search is only looking for strategies with an explicit focus on energy, we will carry out snowballing. The full text and metadata of the potentially relevant studies must contain the keywords "data management strategies" and “energy” or "carbon" or “data growth” to ensure that the results are relevant to RQ1.

ACM Digital library search-string
AllField:("data management strategies") AND AllField:(energy OR carbon) OR AllField:("data growth")

https://dl-acm-org.vu-nl.idm.oclc.org/action/doSearch?fillQuickSearch=false&target=advanced&expand=dl&EpubDate=%5B20200130+TO+20250130%5D&AllField=AllField%3A%28%22data+management+strategies%22%29+AND+AllField%3A%28energy+OR+carbon%29+OR+AllField%3A%28%22data+growth%22%29

Filter 2020 - 2025 applied
219 results returned
Time of search: 31, January 2025

IEEE-Xplore search-string
("Full Text & Metadata":"data management strategies") AND ("Full Text & Metadata":energy OR carbon) OR ("Full Text & Metadata""data growth")

https://ieeexplore-ieee-org.vu-nl.idm.oclc.org/search/searchresult.jsp?action=search&matchBoolean=true&queryText=((%22Full%20Text%20%26%20Metadata%22:%22data%20management%20strategies%22)%20AND%20(%22Full%20Text%20%26%20Metadata%22:energy%20OR%20carbon)%20OR%20(%22Full%20Text%20%26%20Metadata%22%22data%20growth%22)%20)&highlight=true&returnFacets=ALL&returnType=SEARCH&matchPubs=true&ranges=2020_2025_Year

Filter 2020 - 2025 applied
140 results returned
Time of search: 31, January 2025

Scopus search-string
(ALL ("data management strategies") AND ALL (energy OR carbon) OR ALL ("data growth")) AND PUBYEAR > 2019 AND PUBYEAR < 2026

https://www-scopus-com.vu-nl.idm.oclc.org/results/results.uri?sort=plf-f&src=s&sid=6d03be026dfc67a907f2e6500cc52b79&sot=a&sdt=a&sl=138&s=%28+ALL+%28+%22data+manag

ement+strategies%22+%29+AND+ALL+%28+energy+OR+carbon+%29+OR+ALL+%28+%22data+growth%22+%29+%29+AND+PUBYEAR+%26gt%3B+2019+AND+PUBYEAR+%26lt%3B+2026&origin=searchadvanced&editSaveSearch=&txGid=77edc682c681bbdbed58bb543969b99a&sessionSearchId=6d03be026dfc67a907f2e6500cc52b79&limit=10

Filter 2020 - 2025 applied
141 results returned
Time of search: 31, January 2025

Web of Science
(ALL=("data management strategies")) AND (ALL=(energy OR carbon)) OR (ALL=("data growth"))
https://www.webofscience.com/wos/woscc/summary/434abf5c-8b69-4cc9-bc3f-15651fdbc984-01467d5451/relevance/1

Filter 2020 - 2025 applied
147 results returned
Time of search: 31, January 2025

Regular Google Search-string
"data management strategies" AND "energy" OR "carbon" AND "data growth"

The full text of the potentially relevant studies must contain the keywords "data management strategies" and "energy" or "carbon" and "data growth" to ensure that the results are relevant to RQ1. For feasibility purposes we use a more strict search (AND) to include data growth.

https://www.google.com/search?q=%22data+management+strategies%22+AND+%22energy%22+OR+%22carbon%22+AND+%22data+growth%22&sca_esv=b9c0953971e46747&biw=1439&bih=711&source=lnt&tbs=cdr%3A1%2Ccd_min%3A2020%2Ccd_max%3A2025&tbm=

Estimated 477 displayed/160 actual results returned
Filter 2020-2025 applied
Time of search: 7 February, 2025

3.5 INCLUSION CRITERIA SCHOLARLY SEARCH

- I1 Studies written in English.
- I2 The publications are peer-reviewed research papers, meaning they should have been published in journals, conference proceedings, or workshop proceedings.
- I3 The studies contribute to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions.
- *Examples and their scores on the inclusion criteria are provided in Appendix B.*

3.6 EXCLUSION CRITERIA SCHOLARLY SEARCH

- E1 - Duplicate studies are removed to avoid invalid conclusions or results.
- E2 - Studies not available as full-text since they cannot be analysed.

- E3 - Studies that **superficially mention** data management strategies in the context of diminishing IT energy consumption, data growth, and/or carbon emissions, but lack ends, defined means, detail, methods, implementations, proof, or actionable insights.
- E 4 - Studies that focus on enhancing data precision, expanding data capacity, performance optimisation, security, or efficiency increase but **do not mention** strategies or means for managing data in order to reduce IT energy consumption, mitigate data growth, and/or lower carbon dioxide emissions.
- E6 - Off-topic, retracted
- *Examples of excluded publications and their scores on the exclusion criteria are provided in Appendix B.*

3.7 INCLUSION CRITERIA REGULAR GOOGLE

- I1 Text written in English.
- I3 Text contributes to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions.
- *Examples of included grey literature and their scores on the inclusion criteria are provided in Appendix B.*

3.8 EXCLUSION CRITERIA REGULAR GOOGLE

- E1 - Duplicates are removed to avoid invalid conclusions or results.
- E2 - Not available as full-text since they cannot be analysed.
- E3 - Text that **superficially mentions** data management strategies in the context of diminishing IT energy consumption, data growth, and/or carbon dioxide emissions, but lacks ends, defined means, detail, methods, implementations, proof, or actionable insights.
- E4 - Text that focuses on enhancing data precision, expanding data capacity, performance optimisation, or efficiency increase but **does not mention** strategies or means for managing data in order to reduce IT energy consumption, mitigate data growth, and/or lower carbon dioxide emissions.
- E5 - Multimedia
- E6 - Off-topic, retracted
- *Examples of excluded publications and their scores on the exclusion criteria are provided in Appendix B.*

3.9 IDENTIFICATION AND SCREENING

After gathering an initial collection of papers using the search-strings described earlier, a manual screening of the scholarly search results, and of the grey literature from the regular Google search is conducted to examine, and filter the relevant studies in sequential steps.

While an increased reading effort can yield more valid results when evaluating a potentially relevant study, it is unnecessary to read the full text of studies that are unambiguously excluded [2]. Therefore, we implement the adaptive reading technique towards the chosen level of detail as proposed by Petersen et al., [2]. Adaptive reading involves first examining the title, abstract, and conclusion of a potentially relevant study. If the study does not align with the scope of the research, it is excluded. If the researcher remains uncertain, they will assess the full text of the study in a next round before making a final decision about its inclusion. The data of the searches are tabulated, matched against the criteria, and analysed in

Excel sheets. A potentially relevant study is added to the set of selected studies if it satisfies all applicable inclusion criteria and none of the exclusion criteria. At the end of the procedure, the remaining articles will be merged into a single Excel file and transferred to the stage of data coding.

Source	**Initial search**	**Round 1. Language, full-text unavailable ,or off-topic**	**Round 2. Check for peer review research paper**	**Round 3. After screening of the abstract, and conclusion**	**Round 4. Duplicate removal**	**Round 5. Screening of the abstract, introduction, methodology, conclusion**	**Round 6. Full-text analysis**	**Remaining primary studies relevant to RQ1**	**Snowbal ling**	**Total of studies used**
ACM										
IEEE-Xplore										
Scopus										
Web of Science										
Grey literature Google search			-							

Table 1. Summary of the search and selection process

1. In the first round, we will examine the metadata of the studies to exclude publications in languages other than English (I1) and remove papers that are not available in full text (E2). The title, abstract, and conclusion of a potentially relevant study are read to eliminate off-topic studies that obviously do not contribute to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (E6). A paper with any doubt about its relevance to the study according to the metadata, and conclusion will be included for the next round of selection.

2. In the second round, only the results from the scholarly searches are checked to determine whether the publications qualify as peer-reviewed research papers, i.e. journals, conference papers, or workshop proceedings (I2). Publications that do not qualify as peer reviewed will be excluded.

3. The third round involves reading the abstract, introduction, and conclusion of the remaining peer-reviewed studies to determine whether they align with the I3 inclusion criterion. The study should contain a contribution to data management strategies with an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions. If not, it is regarded as off-topic (E6). If there is any uncertainty regarding the applicability of the E6 exclusion, the paper will be marked as I3 and will be retained for the next round of selection. Since the text structure of grey literature may differ from academic sources, their text will be reviewed by skimming their text structure first, and next scanning their full texts seeking words, phrases and concepts [31] to ensure they align with the I3 inclusion criterion. If there is certainty that in this phase either E3 or E4 applies to the grey literature, they can be marked as such.

4. The fourth round involves removing duplicate papers originating from the automated searches (E1). We will also identify and remove any duplicate works that have been published as separate papers but present the same results, retaining only the most comprehensive publication. NB: in the case of extension papers (e.g., a conference publication followed by a journal extension that adds an evaluation of the proposed novelty), the most mature publication is being considered.

5. As the fifth round, we will read the abstract, introduction, methodology section, and conclusion of the remaining studies to verify whether the E3, or E4 exclusion criterion applies. Any peer-reviewed paper that meets either the E3, or E4 criterion will be excluded from consideration. The text structure of the remaining grey literature will be reviewed by skimming their text structure first, and next scanning their full texts seeking words, phrases and concepts [31], while verifying the E3, or E4 criterion. If there is any uncertainty regarding the applicability of the E3, or E4 exclusions, the paper will be marked as I3 and will be retained for the next round of selection.

6. During the sixth round, the remaining studies will be fully read, filtering out leftover studies that may meet any of the exclusion criteria. In addition, the remaining papers will be marked as primary or secondary study. This process will ensure that only the primary studies relevant to RQ1 will be used for snowballing, and ultimately inform the mapping study.

To validate the identification and screening process, the principal researcher will provide a random sample of x studies to the research methodologist, and the advisor. They will analyse these studies independently by completing the data selection and extraction sheet while following the steps mentioned in chapter 3.9 of the protocol. Afterwards, the researchers will assess their levels of agreement. If any discrepancies arise, they will discuss and resolve them collectively, if necessary.

3.10 SNOWBALLING

In this phase the final results of the automatic searches are complemented with backward and forward snowballing [9] to increase sample size. The references of each considered peer-reviewed primary study are checked during the backward snowballing stage. Studies found online that cite the selected publications, are considered during the forward snowballing procedure. To select relevant primary studies, we will follow the procedure outlined in the previous "Identification and Screening" section. During the snowballing phase, we will only consider peer-reviewed research papers. We also limit our sample to papers that are 10 years old or newer to ensure that our selections are relevant to RQ1 regarding the currentness (state of the art) of the data management strategies involved.

3.11 DATA CLASSIFICATION

To reach the objectives of this systematic mapping study stated in 2.1.2, we employ a fixed, standardised classification scheme for research types in requirements engineering. The six research types as introduced by Wieringa et al. [28], and adapted by Petersen et al.[1], [2], and Wohlin et al. [10] are defined as follows:

- Evaluation Research: This involves the implementation and evaluation of techniques, methods, tools, or other solutions in practice, with an investigation of the outcomes.

- Validation Research: A novel solution is developed and thoroughly assessed in a laboratory setting to determine its effectiveness. Methods include simulation, prototyping, mathematical analysis and proof.
- Solution Proposal: In this type of research, a solution for a specific research problem is proposed, and its benefits are discussed without an extensive evaluation. This may include a proof of concept or an example.
- Conceptual Proposal (or Philosophical Paper): This type structures an area of study into a taxonomy or conceptual framework, providing a new perspective on existing concepts.
- Experience Report: This includes the author's firsthand experiences regarding what occurred in practice and how it unfolded.
- Opinion Paper: This type discusses the author's personal opinions on a specific topic, without relying on related work or established research methodologies.

Intended to provide an overview, the classification according to research types will aid researchers and practitioners in understanding the level of usage in practice, maturity, characteristics, diversity, commonalities, or differences of the various data management strategies, and assess their potential utilisation within a given context relating to managing energy consumption, data growth and CA. The chosen standardised classification scheme allows us to evaluate and classify the current academic and grey literature according to transparent established decision rules, maintaining consistency and facilitating comparisons with current and future studies [1].

Research type **= True**	**Evaluation** Used in practice, implementa -tion	**Validation** Thorough empirical evaluation, lab conditions, not used in practice.	**Solution** Proposes a novel solution, proof of concept, limited or no evaluation.	**Conceptual proposal** Structures an area in the form of a taxonomy or conceptual framework.	**Experience** Personal experiences, lessons learnt.	**Opinion** Expresses an opinion about something without relying on related work and research methodologies.
Publication 1						
Publication 2						
Publication 3						
Publication …						

Table 2. Classification scheme for research types based on Wieringa, et al., 2005 [28], Petersen et al, [1],[2], and Wohlin et al., [10].

Each selected study will be coded according to the matching conditions of the research type. When going through a study in detail at the data coding stage, researchers can agree that a study may lack a meaningful contribution based on the evaluation criteria for the research types set by the classification scheme (Appendix C).

3.12 DATA CODING

In this step, the relevant information pertaining to RQ1 will be coded, and the findings will be tabulated. Data coding refers in this context to "the process of transforming collected information or observations to a set of meaningful, cohesive categories" [30]. First, all relevant results from the scholarly searches, regular Google, and the snowballing method are merged into a single Excel datasheet. To conduct a thorough data coding process, data are systematically and transparently extracted from each eligible study using a predefined data extraction format while reading the full text. For each selected study, we will record the extracted information in a spreadsheet. Each row corresponds to the data from a specific study, the columns will represent these parameters:

- Title
- Author
- Year
- Publication venue
- URL
- Goals, what is supposed to be achieved, specific aims? (i.e. energy efficiency, mitigating data growth, carbon/climate)
- Means, how to achieve the specific aims? (e.g. classification into hot and cold data storage areas with adaptive replication policies.)
- Research type, matching within the classification scheme (evaluation research, solution proposal, validation research, philosophical papers, opinion papers, experience papers),
- Confidence assessment of the paper, all selected studies will undergo a quality assessment using a series of questions regarding the evidence level and the quality of the extracted data, according to the evaluation criteria within the chosen research type category (Appendix C.).

To validate the data coding process, the principal researcher will provide a random sample of studies to the second researcher. They will analyse these studies independently by completing the data coding form while reading the full text. Afterwards, the researchers will assess their levels of agreement. If any discrepancies arise, they will discuss and resolve them collectively, if necessary.

3.13 DATA EXTRACTION

After the data coding stage, we employ the aforementioned classification scheme to group the selected publications according to research type. In addition, we will compare, and sort the identified goals and means of the strategies involved [29] into categories. The means are identified by reading the full text and summarising the identified means toward energy efficiency, mitigation of data growth, or reducing carbon dioxide emissions in two sentences in the selection and extraction sheet. Next, we cluster the means according to the identified patterns and make labels. Following You, et al. [14] we thematically label the identified means in the selection and extraction sheet according to type of strategy (e.g. data classification, data lay-out, data replication) in order to construct an inventory of actionable goals and strategies.

After the data coding stage, we employed the aforementioned classification scheme to group the selected publications according to research type. In addition, we compared, and sorted the identified goals and means of the strategies involved [29] into three categories that reflect RQ: data, energy, carbon. To organise the means, we used a thematic analysis method for qualitative data by Braun and Clark [36]. First, the means are identified by reading the full text and summarising the identified means toward energy efficiency, mitigation of data growth, or reducing carbon dioxide emissions on the selection and extraction sheet. Next, we capture key concepts, patterns, or features with initial codes and label the papers accordingly. This is followed by collating the related means by code to form categories of identified data management strategies that are further explored. After reviewing the initial themes in relation to the coded extracts, we refined the themes to generate clear definitions and names. Subsequently, we labelled the selected articles in the selection and extraction sheet according to the identified type of data management strategy (e.g. data classification or downsampling) to construct an inventory that connects the analysis to RQ and the existing literature.

3.1.4 DATA SYNTHESIS

This phase aims to gather findings and their implications, which are used to substantiate the answer to the research question of this mapping study. This involves tabulating and visualising the characteristics of the selected studies using descriptive statistics, along with a textual synthesis of research types, goals, and means. This synthesis will detail the study characteristics and relevant outcome types based on the inventory of current strategies. The output of this stage will be the classification framework, representing the goals, methods, and aspects of specific data management strategies concerning energy efficiency, data growth, and CA.

4. REPORTING

The reporting stage involves tabulating and visualising the characteristics of the selected studies using descriptive statistics, accompanied by explanatory text to provide context.

The data gathered from the systematic mapping study of academic sources and grey literature will offer a compendium of energy-aware data management strategies, methods related to managing data growth, and associated CA potential, and allow us to uncover research gaps, and highlight emerging trends.

4.1 DISSEMINATION

The results of this research, along with a comprehensive description of the study, will be published in a scientific journal and presented at an international scientific conference.

To provide research transparency and replicability, the raw data and details of the study will be made available to researchers interested in confirming or replicating the study through a Research Data Management (RDM) package according to the FAIR (Findable, Accessible, Interoperable and Reusable) principles.

REFERENCES

[1] K. Petersen, S. Vakkalanka, K. Ludwik, K. Kuzniarz, “Guidelines for conducting systematic mapping studies in software engineering: An update”, Information and Software Technology, Volume 64, 2015, pp. 1-18, ISSN 0950-5849, https://doi.org/10.1016/j.infsof.2015.03.007.

[2] K. Petersen, R. Feldt and S. Mujtaba et al. Systematic Mapping Studies in Software Engineering. 2008. DOI: 10.14236/ewic/EASE2008.8

[3] H. Zhang, M.A. Babar, “On Searching Relevant Studies in Software Engineering.” 2010. DOI: 10.14236/ewic/EASE2010.14

[4] M. Hosseini, A. Shahri, K. Phalp, K. Taylor, R. Ali, “Crowdsourcing: A taxonomy and systematic mapping study”, Computer Science Review, Volume 17, 2015, pp. 43-69, ISSN 1574-0137, https://doi.org/10.1016/j.cosrev.2015.05.001.

[5] University of Cambridge. *Systematic Reviews: Grey Literature* [Online]. Available: https://libguides.cam.ac.uk/systematic-reviews/grey-literature,

[6] A. Booth, "EVIDENT guidance for reviewing the evidence: a compendium of methodological literature and websites". 2016, doi: 10.13140/RG.2.1.1562.9842
[7] M. Petticrew, M., Roberts, H. "Systematic Reviews in The Social Sciences: A Practical Guide", 2008, John Wiley & Sons
[8] GHG Protocol *Standards* [Online]. Available: https://ghgprotocol.org/standards
[9] S. Jalali, C. Wohlin, Systematic literature studies: database searches vs. backward snowballing, in: Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2012, pp. 29–38, https://www.wohlin.eu/ease14.pdf
[10] C. Wohlin, P. Runeson, P. Anselmo da Mota Silveira Neto, E. Engström, I. do Carmo Machado, E. Santana de Almeida, On the reliability of mapping studies in software engineering, Journal of Systems and Software, Vol.86,10, 2013, Pp. 2594-2610,ISSN 0164-1212, https://doi.org/10.1016/j.jss.2013.04.076.
[11] A. Valente, M. Holanda, A. M. Mariano, R. Furuta and D. Da Silva, "Analysis of Academic Databases for Literature Review in the Computer Science Education Field," *2022 IEEE Frontiers in Education Conference (FIE)*, Uppsala, Sweden, 2022, pp. 1-7, doi: 10.1109/FIE56618.2022.9962393.
[12] Oracle. *What is data management* [Online]. Available: https://www.oracle.com/database/what-is-data-management/
[13] Corporate Finance Institute. *Carbon Accounting* [Online]. Available: https://corporatefinanceinstitute.com/resources/esg/carbon-accounting/
[14] Harzing's Publish or Perish [Online]. Available: https://harzing.com/resources/publish-or-perish
[15] Google Search Results Scraper [Online] Available: https://chromewebstore.google.com/detail/Google%20Search%20Results%20Scraper/cmdkhofenchpdlkbpkheddfgfdmmobcc?hl=en
[16] J. Bieser, R.Hintemann, S. Beucker, S. Schramm, L. Hilty, "Climate protection through digital technologies – opportunities and risks, short study," (In German) Bundesverband Informationswirtschaft, Telekommunikation und neue Medien, Borderstep Institut, Germany. Universität Zürich, Switzerland, 2020, https://www.bitkom.org/sites/main/files/2020-05/2020-05_bitkom_klimastudie_digitalisierung.pdf
[17] T. Santarius, J.C.T. Bieser, V. Frick, et al. Digital sufficiency: conceptual considerations for ICTs on a finite planet. Ann. Telecommun. 78, 277–295 (2023). https://doi.org/10.1007/s12243-022-00914-x
[18] A.S. Luccioni, Y. Jernite, E. Strubell. Power hungry processing: Watts driving the cost of ai deployment? In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), pages 85–99. ACM, 2024. doi: 10.1145/3630106.3658542
[19] X. You, X. Lv, Z. Zhao, J. Han and X. Ren, "A Survey and Taxonomy on Energy-Aware Data Management Strategies in Cloud Environment," in IEEE Access, vol. 8, pp. 94279-94293, 2020, doi: 10.1109/ACCESS.2020.2992748.
[20] Seagate & IDC "Rethink Data, Put More of Your Business Data to Work—From Edge to Cloud, A Seagate technology report." Seagate Technology LLC, 2020. https://www.seagate.com/content/dam/seagate/migrated-assets/www-content/our-story/rethink-data/files/Rethink_Data_Report_2020.pdf
[21] M. O'Reilly. (2022). *The Unseen Data Conundrum* [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2022/02/03/the-unseen-data-conundrum/
[22] D. International, DAMA-DMBOK2: Data Management Body of Knowledge. 2017.
[23] V.R. Basili, G. Caldiera, H.D. Rombach. "Goal Question Metric Approach," Encyclopedia of Software Engineering. John Wiley & Sons, Inc. 1994.
[24] EU. "Energy Efficiency Directive." https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-directive_en
[25] EU. "Corporate Sustainability Directive" https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en#legislation
[26] A. de Vries, The growing energy footprint of artificial intelligence, Joule, Vol. 7, 2023, pp. 2191-2194, ISSN 2542-4351, https://doi.org/10.1016/j.joule.2023.09.004.
[27] M. Fleckenstein, L. Fellows "Modern Data Strategy", 2018, Springer International Publishing AG, doi.org/10.1007/978-3-319-68993-7
[28] R. Wieringa, N. Maiden, N. Mead, *et al.* Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements Eng* **11**, 102–107 (2006). https://doi-org.vu-nl.idm.oclc.org/10.1007/s00766-005-0021-6
[29] Howlett, M., & Howlett, M. (2019). Designing Public Policies: Principles and Instruments (2nd ed.). Routledge. https://doi.org/10.4324/9781315232003
[30] Ye, Sun, "Coding of Data." In The SAGE Encyclopedia of Communication Research Methods, edited by Sun, Ye, 149-51. Thousand Oaks, CA: SAGE Publications, Inc, 2017. https://doi.org/10.4135/9781483381411.
[31] D. McGovern, M. Matthews, S.E. Mackay (1994) *Reading*. Harlow: Pearson Education Limited.
[32] Jackson, T.W. and Hodgkinson, I.R. (2023), "Keeping a lower profile: how firms can reduce their digital carbon footprints", Journal of Business Strategy, Vol. 44 No. 6, pp. 363-370. https://doi.org/10.1108/JBS-03-2022-0048
[33] V. Castro, M. Georgiou, T. Jackson, I. R. Hodgkinson, L. Jackson, and S. Lockwood, "Digital data demand and renewable energy limits: Forecasting the impacts on global electricity supply and sustainability," *Energy Policy*, vol. 195, p. 114404, Nov. 2024, doi: 10.1016/j.enpol.2024.114404.
[34] E. Alberts, I. Gerostathopoulos, I. Malavolta, C. H. Corbato, and P. Lago, "Software architecture-based self-adaptation in robotics," *Journal of Systems and Software*, vol. 219, p. 112258, Oct. 2024, doi: 10.1016/j.jss.2024.112258.
[35] B. Kitchenham, and S. Charters, S. (2007) Guidelines for Performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE 2007-001, Keele University and Durham University Joint Report. https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf
[36] V. Braun, and V. Clarke (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

APPENDIX A. SEARCH STRINGS

The numbers after the strings represent the results returned.
Time of search: 31, January 2025.
Filter 2020-2025 applied.

ACM
AllField:("data management strategies") AND AllField:(energy OR carbon OR "data growth") 42
AllField:("data management strategies") AND AllField:(energy OR carbon) OR AllField:("data growth") 219
AllField:("data management strategies") AND AllField:(energy) OR AllField:(carbon) OR AllField:("data growth") 5494

IEEE Xplore
("Full Text & Metadata":"data management strategies") AND ("Full Text & Metadata":energy OR carbon OR "data growth") 141
("Full Text & Metadata":"data management strategies") AND ("Full Text & Metadata":energy OR carbon) OR ("Full Text & Metadata""data growth") 140
((("Full Text & Metadata":"data management strategies") AND ("Full Text & Metadata":energy OR carbon) OR "Full Text & Metadata":"data growth")) 2582
("Full Text & Metadata":"data management strategies") AND ("Full Text & Metadata":energy) OR ("Full Text & Metadata":carbon) OR ("Full Text & Metadata""data growth") 241.794

Scopus
(ALL ("data management strategies") AND ALL (energy OR carbon) OR ALL ("data growth")) AND PUBYEAR > 2019 AND PUBYEAR < 2026 141
(ALL ("data management strategies") AND ALL (energy) OR ALL (carbon) OR ALL ("data growth")) AND PUBYEAR > 2019 AND PUBYEAR < 2026 141
(ALL ("data management strategies") AND ALL (energy OR carbon OR "data growth")) AND PUBYEAR > 2019 AND PUBYEAR < 2026 141

Web of Science
ALL=("data management strategies") AND (ALL=(energy OR carbon OR "data growth")) 10
(ALL=("data management strategies")) AND (ALL=(energy OR carbon)) OR (ALL=("data growth")) 147
ALL=("data management strategies") AND ALL=("energy") OR ALL=("carbon") OR ALL=("data growth") 617.889

Google
"data management strategies" AND ("energy" OR "carbon") AND "data growth" 477
"data management strategies" AND ("energy" OR "carbon") OR "data growth" 74400
"data management strategies" AND "energy" OR "carbon" OR "data growth" 74300
"data management strategies" AND "energy" AND "carbon" AND "data growth" 228
"data management strategies" OR "energy" AND "carbon" AND "data growth" 5560

APPENDIX B. EXAMPLES OF APPLIED INCLUSION AND EXCLUSION CRITERIA

INCLUSION CRITERIA

- I1 Studies written in English.
- I2 The publications are peer-reviewed research papers, meaning they should have been published in journals, conference proceedings, or workshop proceedings.
- I3 The studies contribute to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions.

EXCLUSION CRITERIA

- E1 - Duplicate studies are removed to avoid invalid conclusions or results.
- E2 - Studies not available as full-text since they cannot be analysed.
- E3 - Studies that **superficially mention** data management strategies in the context of diminishing IT energy consumption, data growth, and/or carbon dioxide emissions, but lack ends, defined means, detail, methods, implementations, proof, or actionable insights.
- E 4 - Studies that focus on enhancing data precision, expanding data capacity, performance optimisation, security, or efficiency increase but **do not mention** strategies or means for managing data in order to reduce IT energy consumption, mitigate data growth, and/or lower carbon dioxide emissions.
- E5 - Multimedia
- E6 - Off-topic, retracted, dated before 2020.

#37 Example matching I1, I2, I3

Bastian Havers, Marina Papatriantafilou, and Vincenzo Gulisano. 2024. Research Summary: Enhancing Localization, Selection, and Processing of Data in Vehicular Cyber-Physical Systems. In Proceedings of the 2024 Workshop on Advanced Tools, Programming Languages, and PLatforms for Implementing and Evaluating algorithms for Distributed systems (ApPLIED'24). Association for Computing Machinery, New York, NY, USA, 1–5. https://doi-org.vu-nl.idm.oclc.org/10.1145/3663338.3663680

The study is written in English (I1), and has been published in an ACM peer-reviewed journal (I2). The study focuses on “Enhancing Localization, Selection, and Processing of Data in Vehicular Cyber-Physical Systems”. The conclusion states “We presented novel approaches for processing data in VCPSs more efficiently by focussing on reducing data volumes and involving the vehicles (the system’s edge) more actively than traditional workflows.”. Since none of the exclusions apply, the study therefore contributes to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (I3).

#220 Example matching I1, I2, I3

X. You, X. Lv, Z. Zhao, J. Han, and X. Ren, “A survey and taxonomy on Energy-Aware Data Management Strategies in Cloud environment,” *IEEE Access*, vol. 8, pp. 94279–94293, Jan. 2020, doi: 10.1109/access.2020.2992748

The study is written in English (I1), and has been published in an IEEE peer-reviewed journal (I2). The study focuses on “energy saving-aware data management strategies in cloud-related environments”. Since none of the exclusions apply, the study therefore contributes to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (I3).

#7 Example matching I1, I2, I3, (and possibly E3)
Yusuke Tanimura, Hidemoto Nakada, and Yunzhi Dong. 2023. Towards Building Edge-side Common Data Processing Services on The Computing Continuum. In Proceedings of the 24th International Middleware Conference: Demos, Posters and Doctoral Symposium (Middleware '23). Association for Computing Machinery, New York, NY, USA, 27–28. https://doi-org.vu-nl.idm.oclc.org/10.1145/3626564.3629089

The study is written in English (I1), and has been published as ACM conference proceedings (I2). Although it is a short paper/poster, the study responds to “data growth from edge to cloud” and therefore contributes to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions.
Given the concise character of the paper/poster there might be doubt whether it qualifies for exclusion regarding criterion E3. In these cases the protocol according to chapter 3.9 “Identification and screening”, step 4 states: “A paper with any doubt regarding the exclusions will be marked I3, and kept for the last round of selection”.

#652 Example matching I1, I3
Jennifer Cooke, Phil Goodwin, Ashish Nadkarni, Eric Sheppard, IDC, 2021
https://asset.fujifilm.com/www/mm/files/2023-03/c380054a69154d3902d9e092400b9b87/205532_US48252321.pdf

The IDC white paper is written in English (I1), was part of the regular Google search and has been published in a grey literature venue (≠I2). The study focuses on “the impact of tape storage on energy consumption and carbon dioxide emissions.”. Since none of the exclusions apply, the study therefore contributes to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (I3), and will be included.

#648 Example matching I1, E3
Tim Maiden, Green Small Business 2024.
https://greensmallbusiness.com/the-digital-dilemma-balancing-data-growth-with-the-carbon-costs/

The website is written in English (I1), was part of the regular Google search and can be considered a grey literature venue (≠I2). The text focuses on “raising awareness of the environmental impacts of data storage, and the particular issue of dark data.” which might qualify as I3. However, the text does not go beyond stating “Organisations also need robust data governance frameworks. We provide our clients with a template data retention policy to help foster a culture which prioritises data hygiene and lifecycle management to mitigate the environmental impact of excessive data storage.” Therefore the text qualifies as E3: Studies that superficially mention data management strategies, diminishing IT energy consumption, data growth, and/or carbon dioxide emissions, but lack ends, defined means, detail, methods, implementations, proof, or actionable insights.

#119 Example matching I1, I2, E4
Yu Liu and Junfang Zeng. 2021. Blockchain based Big Data Platform of City Brain. In Proceedings of the 2021 3rd International Conference on Blockchain Technology (ICBCT '21). Association for Computing Machinery, New York, NY, USA, 82–89. https://doi-org.vu-nl.idm.oclc.org/10.1145/3460537.3460561

The study is written in English (I1), and has been published as ACM conference proceedings (I2). While this conceptual article mentions “data on cloud, index on blockchain" that may have efficiency potential within blockchain, there is no further detail toward data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (E4). Its contribution to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions remains implausible.

#41 Example matching I1, I2, E4
Jafar Ali Ibrahim Syed Masood, JAI KUMAR VINAYAGAM, Ranjit Kumar Onteru, UMAMAHESWARARAO KOPPARAPU, and Bulah Pushpa Rani Parabathini. 2025. Harnessing Medical Big Data: Integrating Computational Insights for Enhanced Patient Outcomes. In Proceedings of the 2024 7th International Conference on Digital Medicine and Image Processing (DMIP '24). Association for Computing Machinery, New York, NY, USA, 94–100. https://doi-org.vu-nl.idm.oclc.org/10.1145/3705927.3705944

The study is written in English (I1), and has been published as ACM conference proceedings (I2). While this paper focuses on “integration of big data analytics with clinical practices to deliver personalized medicine, optimize treatment protocols, and improve patient care" there is no reasonable substantiation of data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (E4). Its contribution to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions remains implausible.

#52 Example matching I1, I2, E4
Weihong Yang. 2024. Design of Computer Data Analysis Management System in the Context of Big Data. In Proceedings of the 2023 International Conference on Information Education and Artificial Intelligence (ICIEAI '23). Association for Computing Machinery, New York, NY, USA, 268–272. https://doi-org.vu-nl.idm.oclc.org/10.1145/3660043.3660091

The study is written in English (I1), and has been published as ACM conference proceedings (I2). The paper proposes “a design scheme for a distributed data analysis management system tailored to big data scenarios”. There is no reasonable substantiation of data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (E4) present. Its contribution to data management strategies that (may) have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions remains implausible.

#57 Example matching I1, I2, E4
Inas Sawad, Shaoqing Hu, Hamed Al-Raweshidy, and Rajagopal Nilavalan. 2023. Bandwidth Consideration for Cellular System Upgrade in Developing Countries. In Proceedings of the

2023 15th International Conference on Computer Modeling and Simulation (ICCMS '23). Association for Computing Machinery, New York, NY, USA, 207–211. https://doi-org.vu-nl.idm.oclc.org/10.1145/3608251.3608260

The study is written in English (I1), and has been published as ACM conference proceedings (I2). It focuses on bandwidth efficiency considerations of mobile networks in developing countries. Although network efficiency in itself could conserve energy, the article lacks a reasonable contribution to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (E4). Its contribution to data management strategies that (may) have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions remains implausible.

#13 Example matching I1, I2, E4
Tyler Akidau, Fabian Hueske, Konstantinos Kloudas, Leon Papke, Niklas Semmler, and Jan Sommerfeld. 2024. Continuous Data Ingestion and Transformation in Snowflake. In Proceedings of the 18th ACM International Conference on Distributed and Event-based Systems (DEBS '24). Association for Computing Machinery, New York, NY, USA, 195–198. https://doi-org.vu-nl.idm.oclc.org/10.1145/3629104.3672430

The study is written in English (I1), and has been published as ACM conference proceedings (I2). It focuses on data pipelines that cost-efficiently process large amounts of incoming data. Although the claimed efficiency in itself could conserve energy, the subject of studies may facilitate the opposite, and lacks a reasonable contribution to data management strategies that have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions (E4). Its contribution to data management strategies that (may) have an explicit or implicit diminishing effect on IT energy consumption, data growth, and/or carbon dioxide emissions remains implausible.

APPENDIX C. EVALUATION CRITERIA

Cited from the classification scheme for research types in requirements engineering, as proposed by Wieringa et al. [26] and adapted by Petersen et al.[1], [2] and Wohlin et al. [10]. The classification is used for both peer-reviewed and grey literature.

All selected studies will undergo a quality assessment using a series of questions regarding the evidence level and the quality of the extracted data. The criteria below are guidelines for classifying a paper in a category, and determine the potential contribution to the mapping. Not all checkboxes have to apply. When going through a study in detail, researchers can agree that a study may lack a meaningful contribution to the mapping based on the evaluation criteria below. This is indicated through checking the "Rejected" checkbox and briefly stating the reason.

"Evaluation research (used in practice, implementation)

- Controlled experiment with practitioners
- Industrial case study
- Practitioner targeted survey
- Action research

Is the problem clearly stated?
Are the causal or logical properties of the problem clearly stated?
Is the research method sound?
Is the knowledge claim validated? In other words, is the conclusion supported by the paper?
Is this a significant increase of knowledge of these situations? In other words, are the lessons learned interesting?
Is there sufficient discussion of related work?"

Score .. out of 6
Rejected, reason:

"Validation research (empirical evaluation, lab conditions)

- Simulation as an empirical method
- Laboratory experiments (machine or human)
- Prototyping
- Mathematical analysis and proof of properties
- Ethnography
- Academic case study (e.g. with students)

Is the technique to be validated clearly described?
Are the causal or logical properties of the technique clearly stated?
Is the research method sound?
Is the knowledge claim validated (i.e., is the conclusion supported by the paper)?
Is it clear under which circumstances the technique has the stated properties?
Is this a significant increase in knowledge about this technique?
Is there sufficient discussion of related work?"

Score .. out of 7
Rejected, reason:

"Solution proposal (proposes a novel solution, proof of concept)

Is the problem to be solved by the technique clearly explained?
Is the technique novel, or is the application of the techniques to this kind of problem novel?
Is the technique sufficiently well described so that the author or others can validate it in later research?
Is the technique sound?
Is the broader relevance of this novel technique argued?
Is there sufficient discussion of related work? In other words, are competing techniques discussed and compared with this one?"

Score .. out of 6

Rejected, reason:

"Conceptual proposal

Is the conceptual framework original?
Is it sound?
Is the framework insightful?"

Score .. out of 3

Rejected, reason:

"Opinion papers (opinion about something)

- Is the stated position sound?
- Is the opinion surprising?
- Is it likely to provoke discussion?"

Score .. out of 3

- Rejected, reason:

"Experience papers (personal experiences, lessons learnt)

Is the experience original?
Is the report about it sound?
Is the report revealing?
Is the report relevant for practitioners?"

Score .. out of 4

Rejected, reason: