From the *Cloud* to the Ground
A Data-Driven Research to Build Informative Heritage about Data Centres' Energy Footprint

Fabiola Papini
Politecnico di Milano and Tongji University
Orcid id 0000-0001-6007-4990

Michele Mauri
Politecnico di Milano
Orcid id 0000-0003-1189-9624

Francesca Valsecchi
Tongji University
Orcid id 0000-0002-6453-4605

Keywords
Data Centres, Energy Footprint, Data-Driven Methodology, Communication Design, Information Legacy.

Abstract
By adopting a critical stance towards our technocentric times, the article reflects on the impact of contemporary digital transformation and infrastructures in the broader environmental and climate crisis debate, analysing the Internet's energy footprint. The Internet is often perceived as an intangible and weightless service, as represented by the popular *Cloud* metaphor. This common belief has generated a knowledge gap in users' minds: no correlation exists between digital activities and their footprint on our planet. However, the Internet is a massive network of infrastructures consuming vast resources, contributing significantly to global warming. The article aims to bridge this gap using data visualisation as a research method and design output. It summarises data-driven research with a two-fold output: mapping the known and unknown aspects of the Internet's energy footprint to support transparency and future reflections and identifying visual strategies in the design field for communicating complex information. The research promotes an open-data approach, highlighting the pivotal role of data in understanding the connections between human behaviour, technology, and the environment. Additionally, this type of research has the potential to influence public awareness, engagement, and policy-making, emphasising its broader implications for fostering a more sustainable digital future.

1. Introduction: The Hidden Weight of Digital Transformation

Societies of developed countries are currently experiencing profound digital empowerment, driven by the advent of 5G technology, the pervasive implementation of artificial intelligence (AI) and machine learning, and the extensive deployment of Internet of Things (IoT) devices. This transformation continuously increases data generation and traffic. However, the associated surge in energy consumption is largely overlooked. While networked technology promises to lead to more efficient use of resources, it concurrently masks the operations of an extensive physical infrastructure. This scenario shapes a society where humans become increasingly dependent on data while the connection between digital services and their material aspect is obfuscating. People use the Internet constantly and expect it to be accessible 24/7. However, many people fail to realise that their digital habits are linked to massive digital infrastructure systems that make the Internet possible and functional. In contrast to prevailing narratives and digital metaphors surrounding the Internet, such as the *Cloud*, our contribution explores the physical side of the Internet to materialise its connotation. Reflecting on the material implications of today's digital transformation and inquiring about the climate crisis from a systemic perspective, this article presents a data-driven project that aims to assemble an informative heritage about the Internet's energy footprint.

2. Research Context: Decoding the Internet's Energy Dilemma

In the world of unlimited data, people's lifestyles are increasingly characterised by online operations: businesses, governments and individuals are all persuaded to mass-migrate to an online lifestyle. People may think this is good for the planet, perhaps because they move less, but this does not mean that digital activities are without environmental impact. Whenever we do anything online, such as storing data, watching a movie, participating in a video call, or sending an email, we use resources, consume energy, emit CO2, and leave behind environmental footprints. According to the issue *The Real Climate And Transformative Impact of ICT: A Critique of Estimates, Trends, and Regulations* (Freitag et al., 2021), global greenhouse gas (GHG) emissions from the Information Communication Technology (ICT) sector are between 2.1–3.9%. This wide range considers the divided forecasts published by the most reliable studies on the future trend, as shown in Fig. 1, that reveal the uncertainty around the phenomenon.

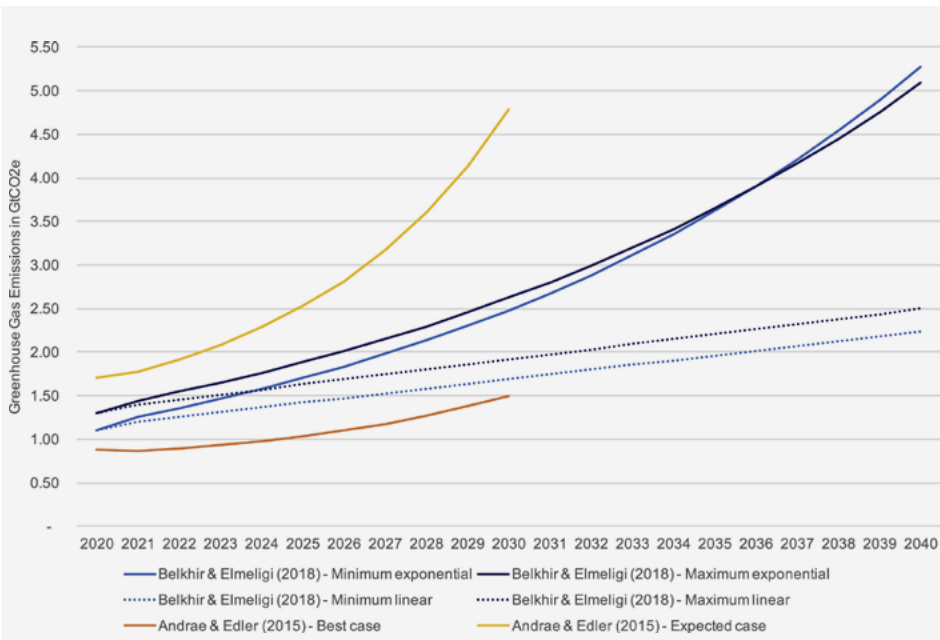

Figure 1. A comparison of different projections of ICT GHG emissions from 2020 according to current literature (Freitag et al., 2021).

The data visualisation project *Dirty Cloud* (Simonetta et al., 2021) is an effective information artefact that makes the current GHG estimates more familiar to the public. By visualising data from the *Global Carbon Atlas* (*Global Carbon Atlas*, 2022), it compares the GHG emissions of the Internet to countries, pointing out that if the Internet were a country, it would fall in the top five after China, the United States, India and Russia. How are the tonnes of CO2 produced when people use digital products and services? The connections between actions and consequences when using digital products are well communicated by Joana Moll's projects titled *CO2GLE* (Moll, 2014) and *DEFOOOOOOOOOOOOOOOOOOOOOREST* (Moll, 2016). In a minimal interface, these digital products show in real-time the amount of CO2 generated by Google requests and the number of trees needed for their compensation. They explore how to trigger thoughts and actions, highlighting the invisible connections between the uses of digital communication technologies and their consequences. In this article the authors exploit the capacity of data to provide an outlook into complexity by adopting a data-driven perspective on an uncertain and unexplored phenomenon. As said by Jussi Parikka in his book *A Geology of Media* (Parikka, 2015), to understand contemporary media culture, we must look for those material realities that precede media themselves, such as the energy on which media depends. An online purchase, a WhatsApp message, a streaming movie, or a chat with chatGPT are examples of online services that respond to our requests via the Internet. When users make a request, a network of infrastructures is activated, allowing it to travel across links and nodes to the source that contains the answer.

The data capable of satisfying the request is stored in special computers, called servers, that send back a response based on the personal account, data and content. The server involved in the request is located in buildings called data centres and thousands of others. In addition to the servers, these contain various components designed to operate and monitor them: power systems, uninterruptible power supplies, ventilation and cooling systems, fire suppression systems, and connections to external networks. Due to anonymous architecture and missing standards, their polluting qualities are far less visible than the billowing smokestacks of coal-fired power stations. However, according to the International Energy Agency (IEA), the data centres industry consumes around 1-1.5% of global electricity use (IEA, 2023). It is growing steadily, driven by the high demand for computing power, especially for artificial intelligence, but there is no complete picture of its dimensions. Data centres are the new bedrock of our digital economy and society, and they are playing a silent and crucial role in the Internet's energy consumption: they are one of the most energy-intensive building types, consuming 10 to 50 times the energy per floor space of a typical commercial office building. Data centres as energy-hungry facilities are the framework of our research. The high energy footprint of the Internet calls for a redefinition of the topic based on a shared value system and new forms of communication and awareness. In this perspective, data visualisation can be a lever for action and innovation: the fragmentary and scarcity of data represents a research opportunity to investigate the current level of available data and contribute to increasing knowledge of the Internet's energy footprint.

3. Research Workflow: A Data-Driven Methodology

This research adopts a Research-Through-Design (RTD) (Gaver, 2012) approach to address the underexplored phenomenon of the Internet's energy consumption, an emerging and underexplored phenomenon, engaging stakeholders from various backgrounds. Given the complex networked infrastructure, including the challenging assessment of energy consumption, a proxy observation is necessary. Therefore, this study focuses on data centres, which are highly energy-intensive and core to the Internet's infrastructure. The study aims to identify design strategies to create visual artefacts that raise awareness among young adults, especially those in higher education disciplines heavily reliant on the Internet or directly involved in its development at various levels. Examples include design students whose work is significantly influenced by the digital realm. The research is structured into two primary methodological approaches due to the fragmented and unevenly documented nature of the subject. The first approach is data-driven, assessing the current knowledge and availability of data on data centre energy consumption. This phase seeks to answer the question, *How much data about data centres' energy footprint is available globally?* The findings from this stage are crucial for establishing a baseline of existing data and are detailed in Section 4, *Unveiling Physicality*. The second approach analyses how data centres are represented in digital media. Understanding how data centres are portrayed online required a data design phase in which suitable methods were necessary to capture this wicked phenomenon. *Digital Methods* (Rogers, 2023) are emerging approaches rooted in media studies and sociology, increasingly adopted in the design field, which views web platforms as societal proxies. In other words, when events or innovations impact society, their effect on the public debate leaves traces on the web, and therefore, it is possible to collect this data as a source of knowledge about the topic. This approach acknowledges the biases related to web platforms as ephemeral media and as economic actors interested in the topics they promote. We drew inspiration from studies on image circulation (Colombo & Niederer, 2021) by examining which images news articles use to depict data centres. Due to the need for a visual compendium to promote news on social media, news outlets are compelled to illustrate each article with an image.

But what kind of image is appropriate when discussing data centres? Photos of infrastructure, sci-fi collages, or portraits of people owning such infrastructure? The research question guiding this second exploration is, *How do Google News' top 100 sources visually represent the data centres' energy footprint?*. The results of this phase are described in Section 5, *Framing Dilemma*. The combination of these two data collection phases informs the design of printed visual artefacts aimed at the target audience. These artefacts, discussed in Section 6 and illustrated in Fig. 2 and 3, are intended for dissemination in academic settings or exhibitions to spur scholarly discourse and encourage proactive engagement with the environmental impacts of digital technologies. By integrating data-driven insights and media analysis, the research not only sheds light on current understandings and gaps but also lays a foundation for future investigations into the environmental entanglement of the digital age.

Figure 2. Cover of the printed artefact (Authors, 2022).

Figure 3. A section of the booklet showing the energy consumption of data centres in different countries (Authors, 2022).

4. Unveiling Physicality: How Much Data about the Data Centres' Energy Footprint is Available Globally?

This section introduces a visual strategy designed to map and understand the global data availability and distribution of data centres. It aims to reveal the physicality behind the digital veil that often obscures our understanding of the Internet, starting from the following research question: *How much data about the data centres' energy footprint is available globally?* In a world where digital infrastructure plays an increasingly critical role, the environmental impact of data centres remains a complex and controversial issue. As outlined in the introduction, the focus on data centres provides an entry point for a better understanding of this vast and intricate phenomenon characterised by the need to ground abstract digital phenomena in tangible, measurable terms. Despite digital corporations' lack of obligation to publish data about data centres, this research leverages the *Datacente.rs world map* – a comprehensive infrastructure dashboard with information on 6,242 data centres – to gather valuable insights. Key metrics such as white space (space allocated for information technology equipment) and gross power (total electrical output) are identified as recurrent and closely linked to the energy consumption of buildings. By adding a geographical component to these two metrics, the first strategy maps the available and unavailable data across countries and regions, generating a final dataset that includes region, country, the total number of data centres, available and unavailable white space and gross power data. Two visual outputs are generated: a scatter plot representing global data availability by region and a multi-set stacked bar chart illustrating data availability by country.

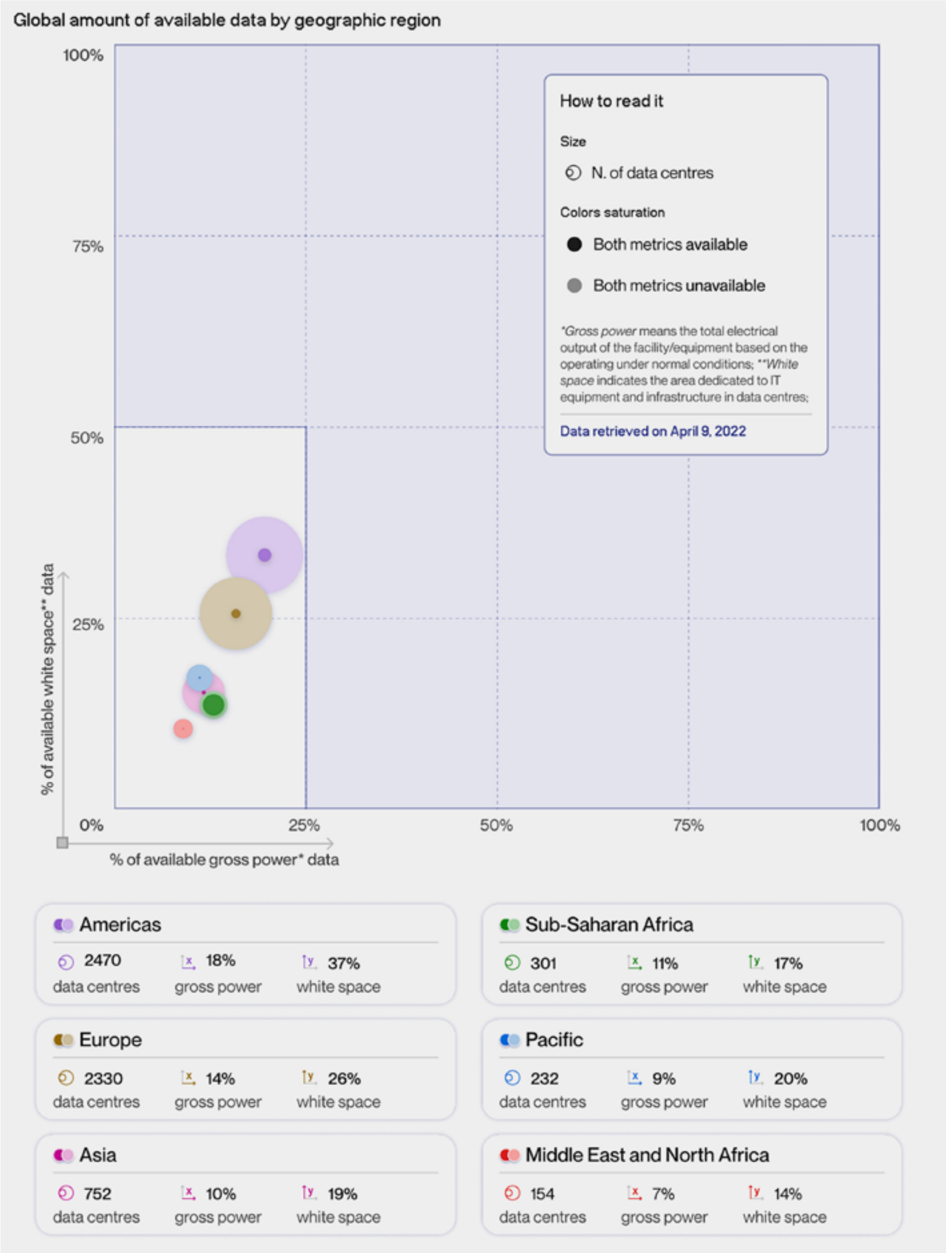

Figure 4. Visualisation correlating the number of data centres and the data availability in terms of white space and gross power per geographic area (Authors, 2022).

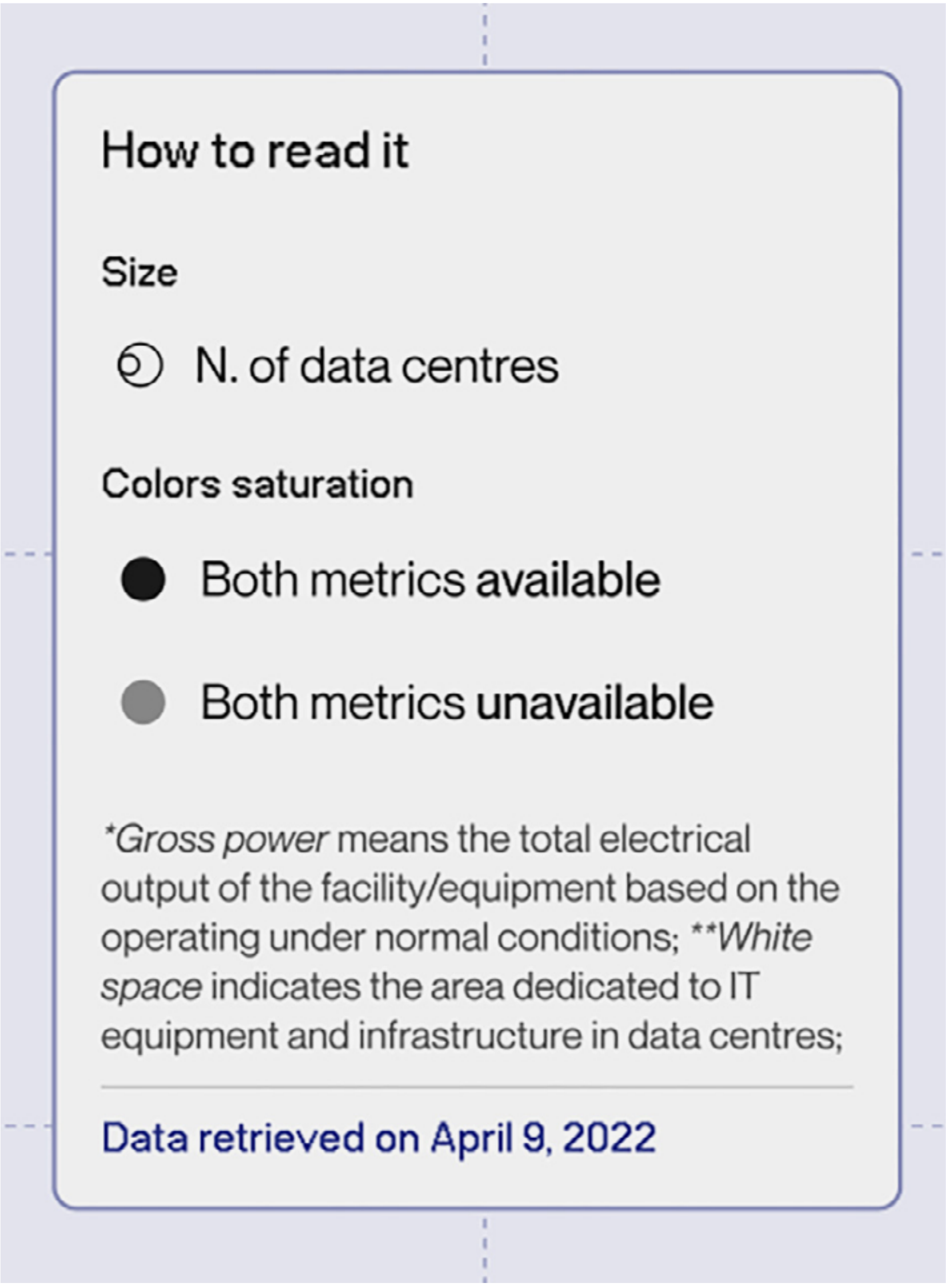

Figure 5. Close-up of the visual model legend (Authors, 2022).

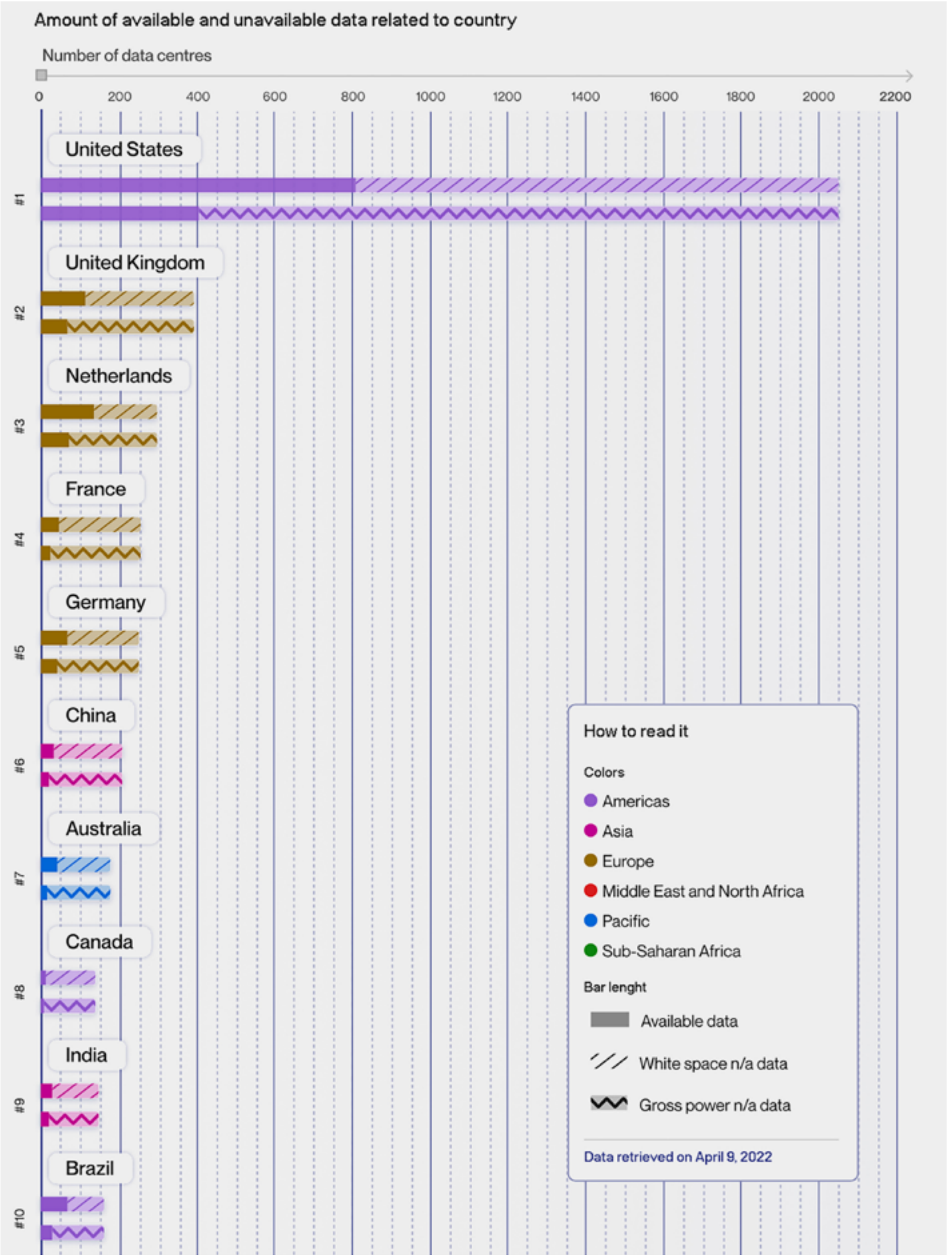

Figure 6. Section of the booklet showing the available and unavailable data in terms of gross power and white space per country (Authors, 2022).

Fig. 4 visualises the global amount of available data by geographical region. The six regions' names follow the Inter-Parliamentary Union (IPU) standards, each associated with a colour. To offer a macro-view of the phenomenon, the chosen visual model is a scatterplot. As visually explained in Fig. 5, each dot represents a region whose colour saturation changes according to the availability or unavailability of white space and gross power data. The dots' size changes according to the number of data centres located, and their position is the result of the correlation between the percentage of available gross power (horizontal axis) and white space (vertical axis) data. The linear grid below the scatterplot is divided into two sections: a white background highlights the area where region shapes are located, while the grey area is empty and does not host any data. The second visualisation in Fig. 6 shows the available and unavailable data related to the country. The chosen visual model is a multi-set stacked bar chart that visualises the number of data centres (horizontal axis) and the country name (vertical axis). Countries are ranked according to the number of data centres in descending order. The bar chart highlights the relationship between available and unavailable data for each metric, as detailed in Fig. 7, through different textures and colours. The information related to white space and gross power is divided into two bars to optimise readability.

This first visual strategy shows over 6,000 data centres across regions and countries. Still, the visualisation can only show a limited view as most data are unavailable. This snapshot is valuable because it highlights the urgency of accurately measuring the sector's impact.

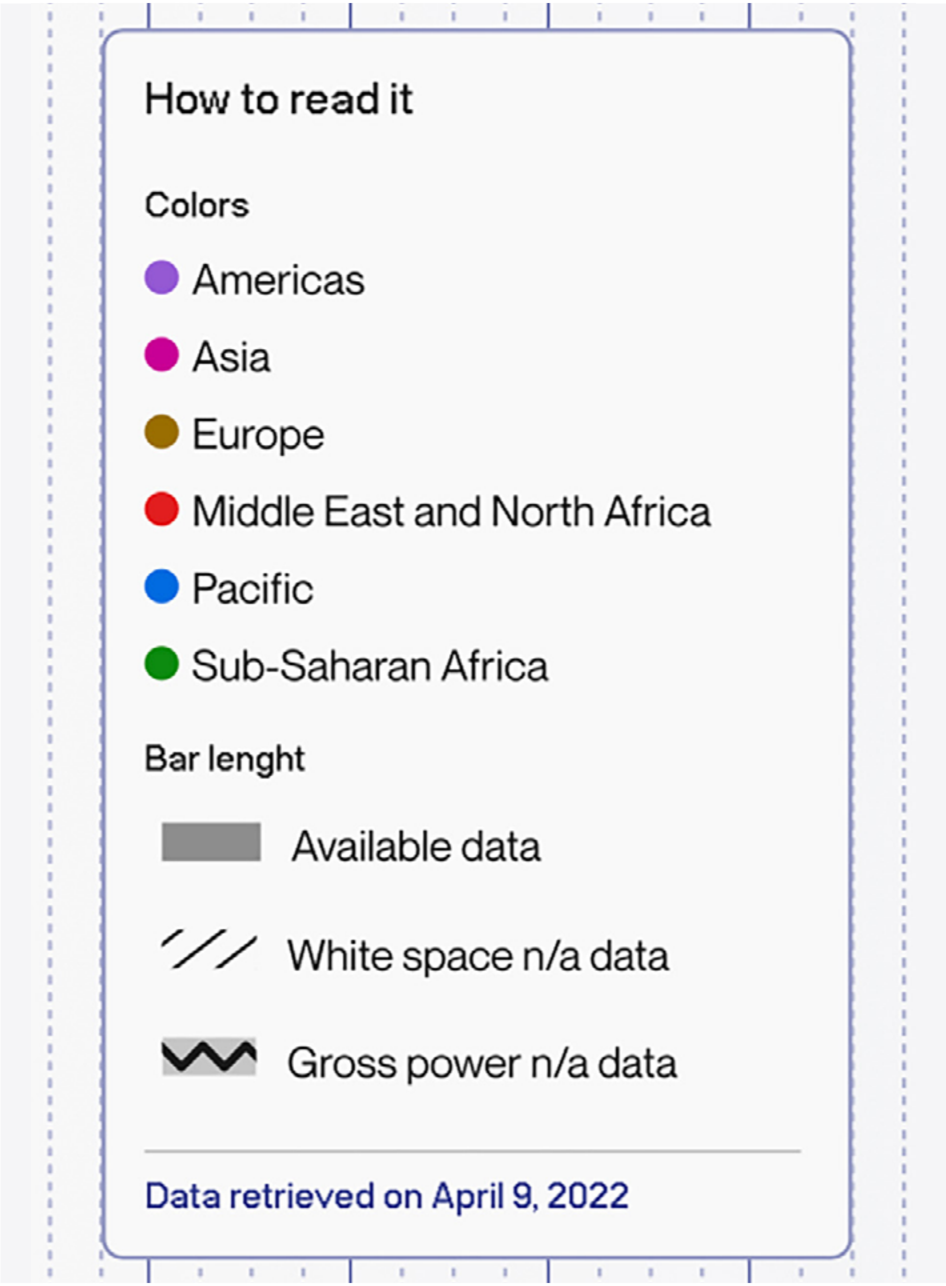

Figure 7. Close-up of the visual model legend (Authors, 2022).

This is consistent with the current concern among researchers and experts who believe it is difficult to measure the energy impact of these infrastructures, which are considered the core of the Internet itself. The output designed by authors serves as tools for unveiling the physicality of data centres, transforming abstract data points into tangible representations of their global distribution and environmental impact. By linking data centre metrics to physical locations, this research establishes a concrete connection between digital performance indicators and their real-world implications. While the visualisations offer valuable insights into data centre energy consumption patterns, they also underscore the challenges and gaps in data availability and the fragmented nature of the current landscape. From the authors' perspective, this visual strategy contributes to a more comprehensive understanding of the Internet's environmental footprint, laying the groundwork for future research and policy initiatives to promote sustainability and transparency in the digital age.

5. Framing Dilemma: How Do Google News' Top 100 Sources Visually Represent the Data Centres' Energy Footprint?

The second visual strategy delves into the challenges and complexity associated with representing the data centres' energy footprint to uncover patterns that shape the visual narratives surrounding the issue. With the escalating significance of data centre's energy consumption, digital media had emerged as influential platforms shaping public perception and discourse on the subject since 2018, when the first article was published on the Nature website under the title *How to stop data centres from gobbling up the world's electricity* (Jones, 2018).

The research involves a meticulous process of collecting and categorising visual content accompanying the articles sourced from Google News. Google News was chosen as the primary source due to its comprehensive coverage and accessibility. By querying “data centres’ energy footprint” on Google News in incognito mode, the authors selected and mapped the top 100 results according to the May 15, 2022, ranking, based on Google News’ algorithms that prioritise factors such as relevance, freshness, and user engagement. The incognito mode was employed to ensure unbiased results, free from personalisation and system preferences. The authors scraped online sources and collected information such as the sources’ links, typology, and depicted subjects. Automated tools, such as Cyren APIs, were utilised for website categorisation. However, the same approach was not feasible for identifying and clustering images’ subjects due to the inherent complexity and nuances in the visual content. This inability to recognise and categorise subjects consistently reveals the intricate nature of the framing dilemma faced in representing the data centres’ energy footprint visually.

In Fig. 8, the final visualisation reveals the relevance of images in this visual strategy. The main visual model is a photographic bar chart, with each bar representing a type of subject depicted in the cover images. In addition, each bar comprises stacked rectangles, the number of which determines their descending order. Each rectangle contains the cover image, converted to greyscale to give them all the same visual relevance, and a bar whose colour identifies the type of source from which the image was taken. This bar can be single or double, depending on the type identified by the API used.

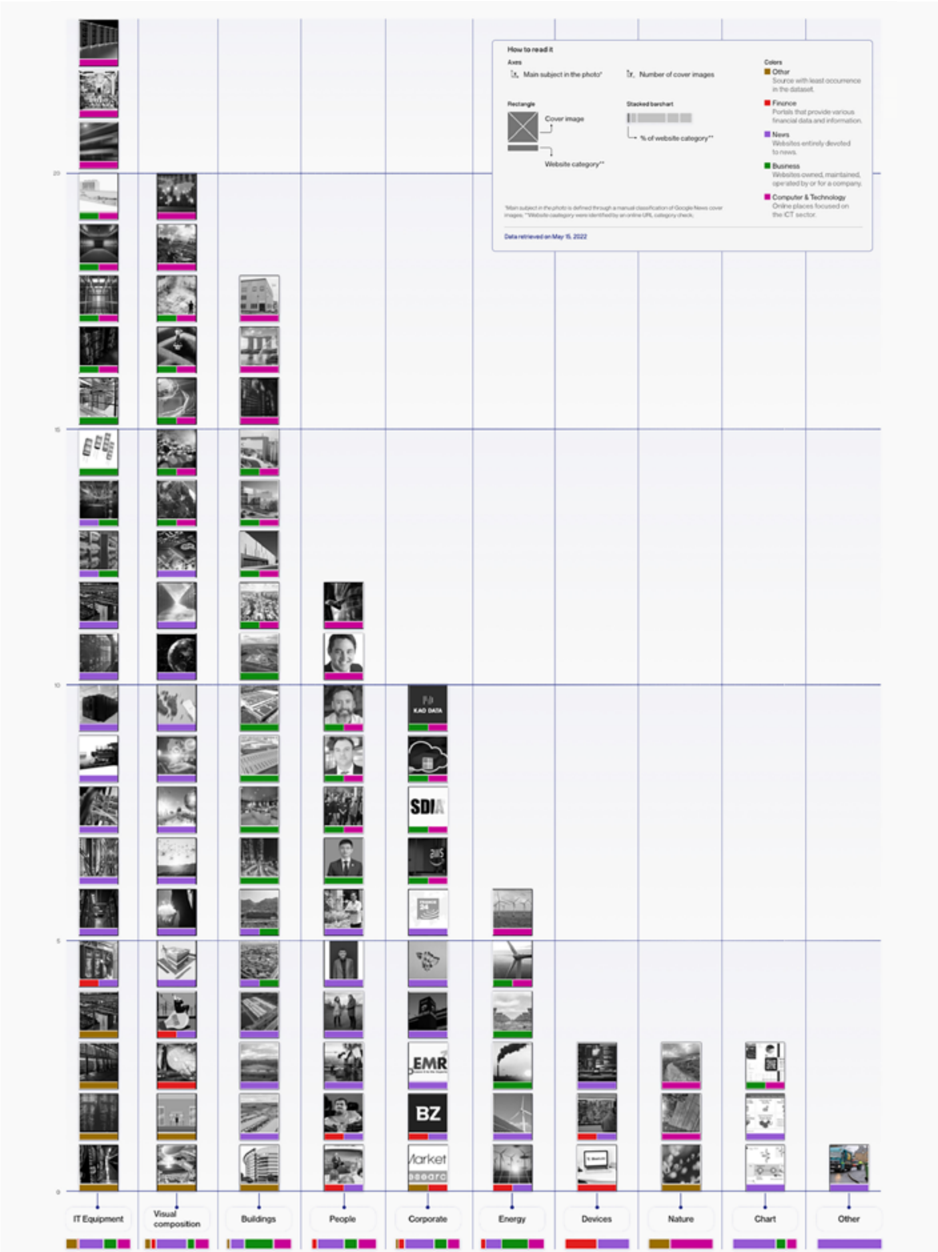

Figure 8. Visualisation of the distribution of Google News source typology and main topics according to cover images (Authors, 2022).

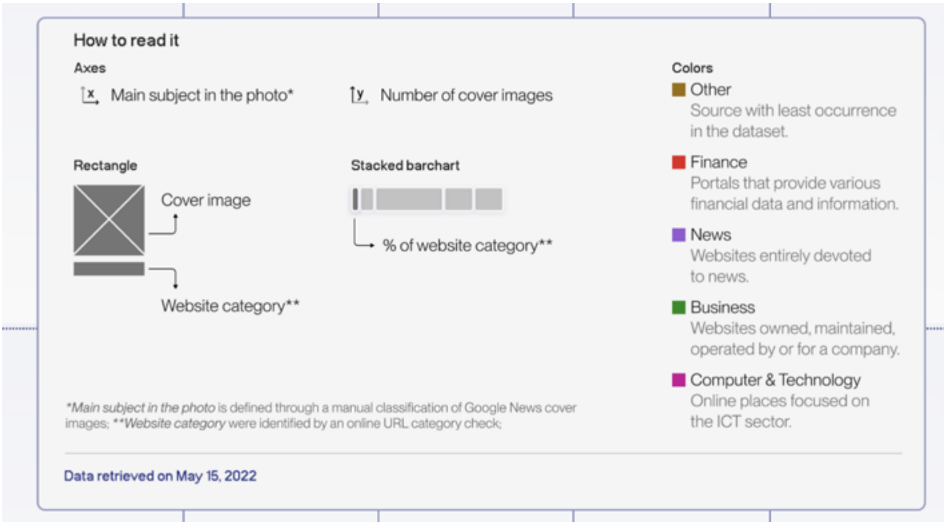

Figure 9. Close-up of the visual model legend (Authors, 2022).

The colour association is shown in Fig. 9, which details the legend. The legend also explains the meaning of the bar chart at the bottom of each bar, which shows the distribution of resource types according to the subject group identified. The second visual strategy provides a comprehensive overview of the recurring themes related to the energy footprint of data centres and the corresponding resources proposing solutions to address this issue. The findings reveal a heterogeneous scenario where visual compositions occupy the second position in the ranking after photographs portraying information technology equipment, such as the interior of server rooms or technical equipment. The first two columns of the analysis indicate a preponderance of resources related to the *News* category, highlighting the significance of the material-abstract contrast in reality. The proposed imagery is diverse and lacks a cohesive view of the phenomenon under consideration.

Unexpected clusters, such as *People* and *Corporate*, assume a prominent role and are very focused on individual subjects, while clusters more inherent to the query, such as *Nature* and *Energy*, collect a small number of images, indicating a potential gap in the visual representation of these aspects.

6. Discussion: A Cultural Shift in Data Centres Energy Awareness

This article discusses an ongoing research project aiming to understand the energy implications of our digital infrastructure. It does so through visualisations, as a design contribution to the understanding of the energy transition, not only as the issue of accessing better energy resources: in a *humanised* energy transition perspective, clarity and awareness of energy consumption is necessary both at individual and system levels. Through visualisation, the research offers insights into the challenges and complexity of visually representing the data centres' energy footprint and the *framing dilemma* faced in striking a balance between physical and abstract connotations of the technology systems we utilise every day. The findings contribute to the existing body of knowledge, highlighting the influence of visual narratives on shaping public perception and discourse on sustainability issues, of which Internet energy consumption is a case.

Future research may extend to visual narratives for public awareness, engagement, and policy-making. Through data collection and analysis, a lack of information about energy consumption related to digital activities emerges. The visualisations advocate for an open-data approach within the data centres and digital services industry as a necessary contribution to the knowledge about the phenomenon of our hyper-connected lifestyles and global energy footprint. By emphasising the interconnected relationship between people, technology and the environment, the project promotes a redefinition of digital dependencies through the lens of environmental responsibility. Such responsibility ought to be built collectively, and visual narratives can complement discursive and textual approaches by introducing accessibility and readability to data.

The research uses data visualisation to inform behavioural change through two distinct strategies to deepen understanding of the phenomenon. The first gateway focuses on unveiling the physicality behind the seemingly intangible world of data. By dispelling the myth of the non-physicality of data, the research reveals the tangible infrastructure that underpins the Internet, highlighting the crucial role data centres play in our digital ecosystem. The project maps data centres' global distribution and environmental impact through data collection and visualisation, clarifying the fragmented landscape of available information and highlighting the need for comprehensive data. The second gateway delves into the challenges and complexities of visually representing the energy footprint of data centres in the digital media landscape.

The research uncovers the prevailing themes and tendencies that shape public perception and discourse on the subject by analysing visual narratives across top Google News sources. This exploration underscores the *framing dilemma* faced in striking a balance between conveying the physical and abstract connotations of the issue and revealing the gaps and inconsistencies in current visual representations. In doing so, it highlights the role of design as an agent of change in promoting awareness and shaping cultural perceptions.

By humanising the energy transition and emphasising the collective sensitivity required for sustainable action, designers can leverage the creative potential of data visualisation to bridge the gap between data, technology, and human behaviour. This approach contrasts technocracy, urging us to move beyond mere resource access to a deeper awareness of consumption patterns and their implications. Whilst the energy transition is an unavoidable challenge and is understood as an environmental, social and economic necessity, more widespread education and information about energy scenarios is needed. By bringing dissonant points of view to the forefront and challenging prevailing narratives, as designers, we can cultivate a culture of sustainability grounded in informed decision-making and collective responsibility.

Acknowledgements

The author's contributions are defined using the CRediT system (https://credit.niso.org/).
Fabiola Papini: Conceptualization, Methodology, Data curation, Visualization, Writing (Original Draft).
Francesca Valsecchi: Supervision, Writing (Original Draft), Writing (Review & Editing).
Michele Mauri: Supervision, Writing (Review & Editing).

References

Colombo, G., & Niederer, S. (2021). Visual Methods for Online Images: Collection, Circulation, and Machine Co-Creation. *Diseña*, (19), 1–7. https://doi.org/10.7764/disena.19.Intro

Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G. S., & Friday, A. (2021). The Real Climate and Transformative Impact of ICT: A Critique of Estimates, Trends, and Regulations. *Patterns, 2*(9), 100340. https://doi.org/10.1016/j.patter.2021.100340

Gaver, W. (2012). What Should We Expect from Research Through Design? In J. A. Konstan, E. H. Chi, & K. Höök (Eds.), *CHI 2012: It's the Experience! The 30th ACM Conference on Human Factors in Computing Systems (CHI). Conference Proceedings* (pp. 937–946). ACM. https://doi.org/10.1145/2207676.2208538

Global Carbon Atlas. (2022). *Global Carbon Atlas*. https://globalcarbonatlas.org/

IEA. (2023). *Tracking Clean Energy Progress 2023*. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

Jones, N. (2018). How to Stop Data Centres from Gobbling Up the World's Electricity. *Nature, 561*(7722), 163–166. https://doi.org/10.1038/d41586-018-06610-y

Moll, J. (2014). *CO2GLE* [Web project]. https://www.janavirgin.com/co2gle.html

Moll, J. (2016). *_Defoooooooooooooooooooorest* [Web project]. https://www.janavirgin.com/deforest.html

Parikka, J. (2015). *A Geology of Media*. University of Minnesota Press.

Rogers, R. (2023). *Doing Digital Methods*. Sage Publications.

Simonetta, B., Alesi, R., Cesaroni, L., & Faraci, L. (2021). Dirty Cloud. *Lab24, Il Sole 24 Ore*. https://lab24.ilsole24ore.com/unirsm-2021/projects/internet/