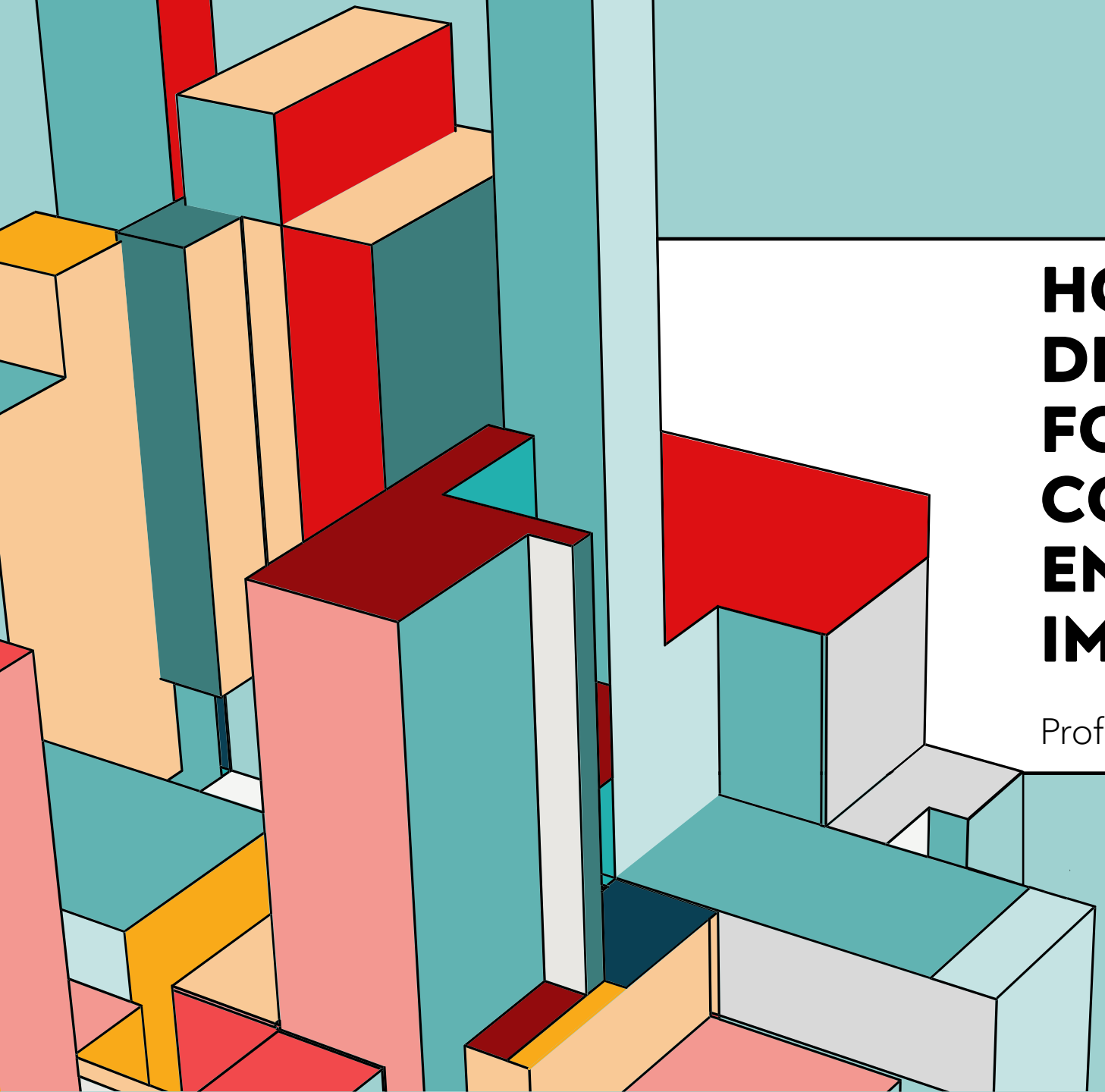
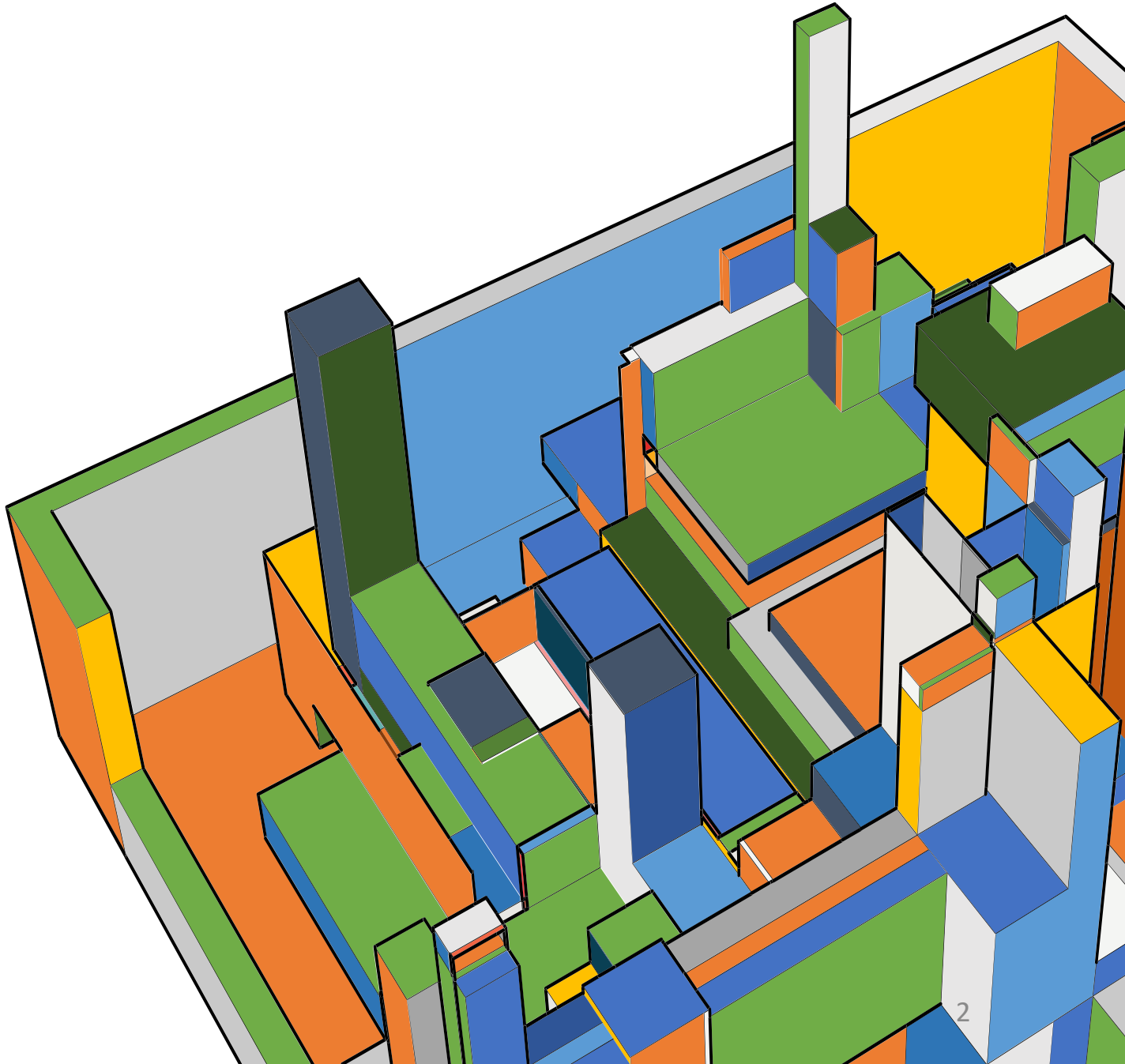


HOW DOES YOUR DIGITAL DATA FOOTPRINT CONTRIBUTE TO ENVIRONMENTAL IMPACT?

Prof. Tom Jackson & Prof. Ian Hodgkinson

ABOUT US

We are thought leaders and pioneers of the digital decarbonization movement (recognized by the World Economic Forum, 2022). Specializing in the measurement of digital data CO2 footprints, our aim is to foster new thinking on how data is discovered, used, and stored. We help organizations and policy makers on their data decarbonization journey, which has been recognized by OECD-OPSI as a critical future focus for accelerating the path to net zero globally.



SETTING THE SCENE

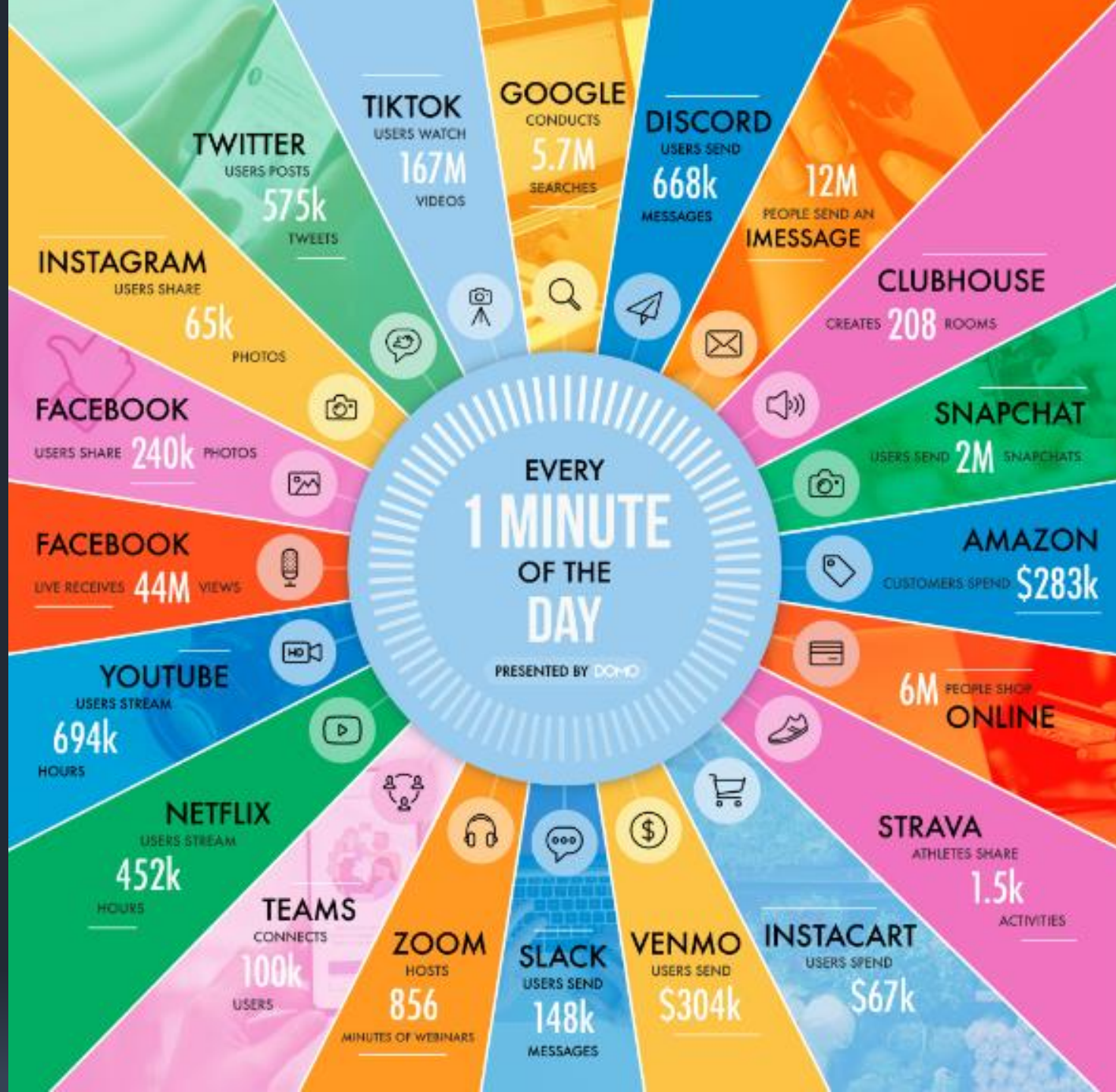
While we may not be able to touch or feel it, the impact of digital on the environment is real. And beyond being real, it is significant.

Digital Data

2.5% - 3.7% of global greenhouse gas emissions are generated from data centres compared with aviation at approx. 2.1% - 2.4% [1]

180 zettabytes Is the expected total global data volume by 2025 [2]

2.5 quintillion bytes of data are produced every day by organizations at our best estimate [3]



The ‘Data’ Challenges

DATA & GHG SCOPES

Organisations are not clear on how data relates to reducing greenhouse gas emissions or how data can be better manged to reduce CO2 reporting.

DATA & C02

The data industry is predicted to account for more carbon emissions than the automotive, aviation and energy sectors combined.

DARK DATA

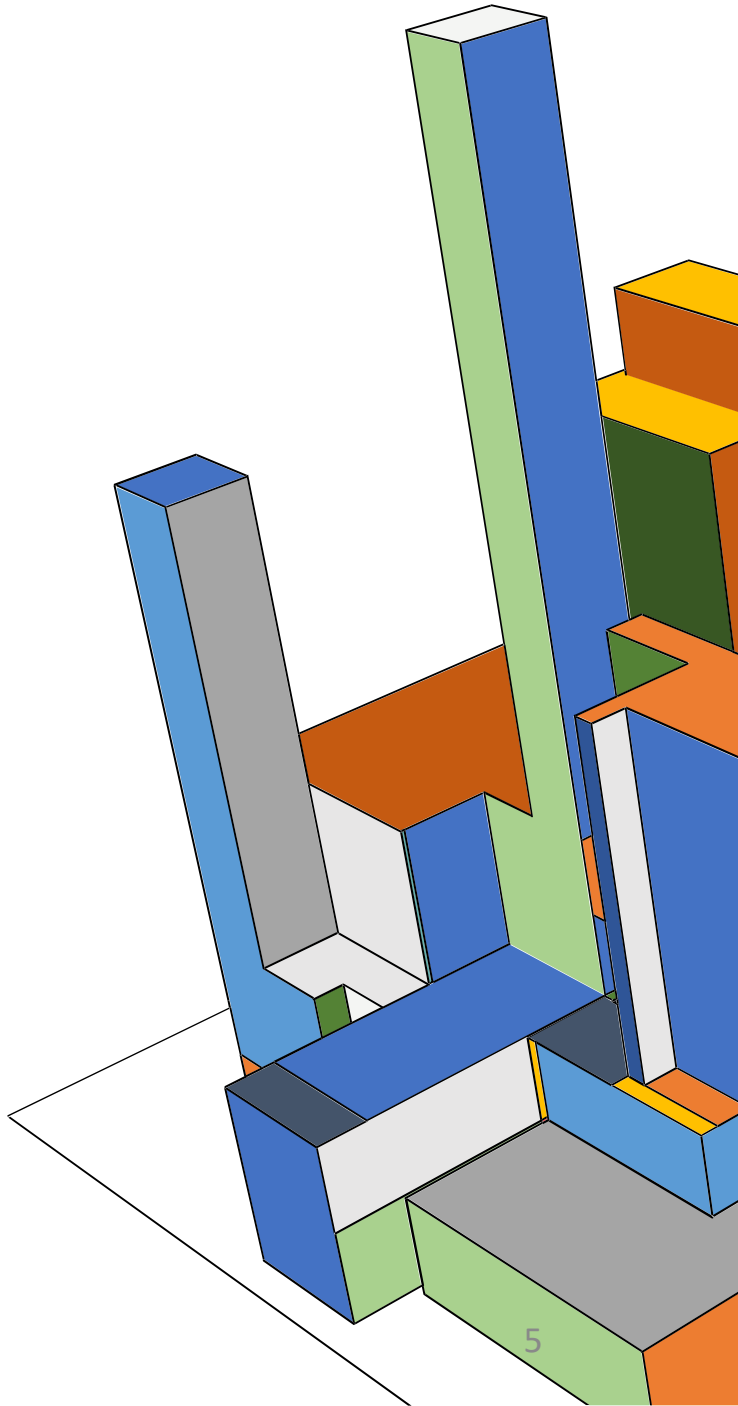
Up to 65% of data generated is never used and up to 15% is out of date, collectively creating a huge energy drain.

POOR QUALITY DATA

IBM estimate that the US lose $3.1 trillion annually due to poor quality data – bad data infiltrating organisations.

UNSTRUCTURED DATA WASTE

Up to 90% of unstructured data sets harvested by organizations for sentiment analysis is irrelevant.



The Problem

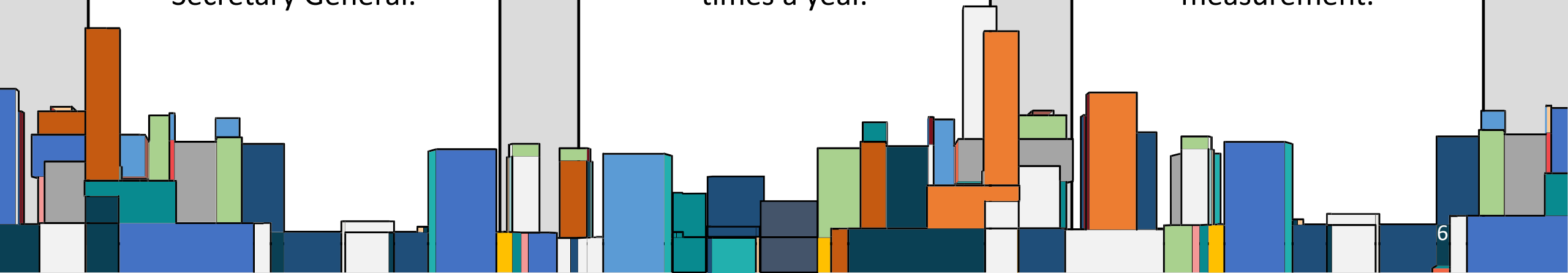
Digital data and digitalization has a CO2 footprint, but why is this significant?

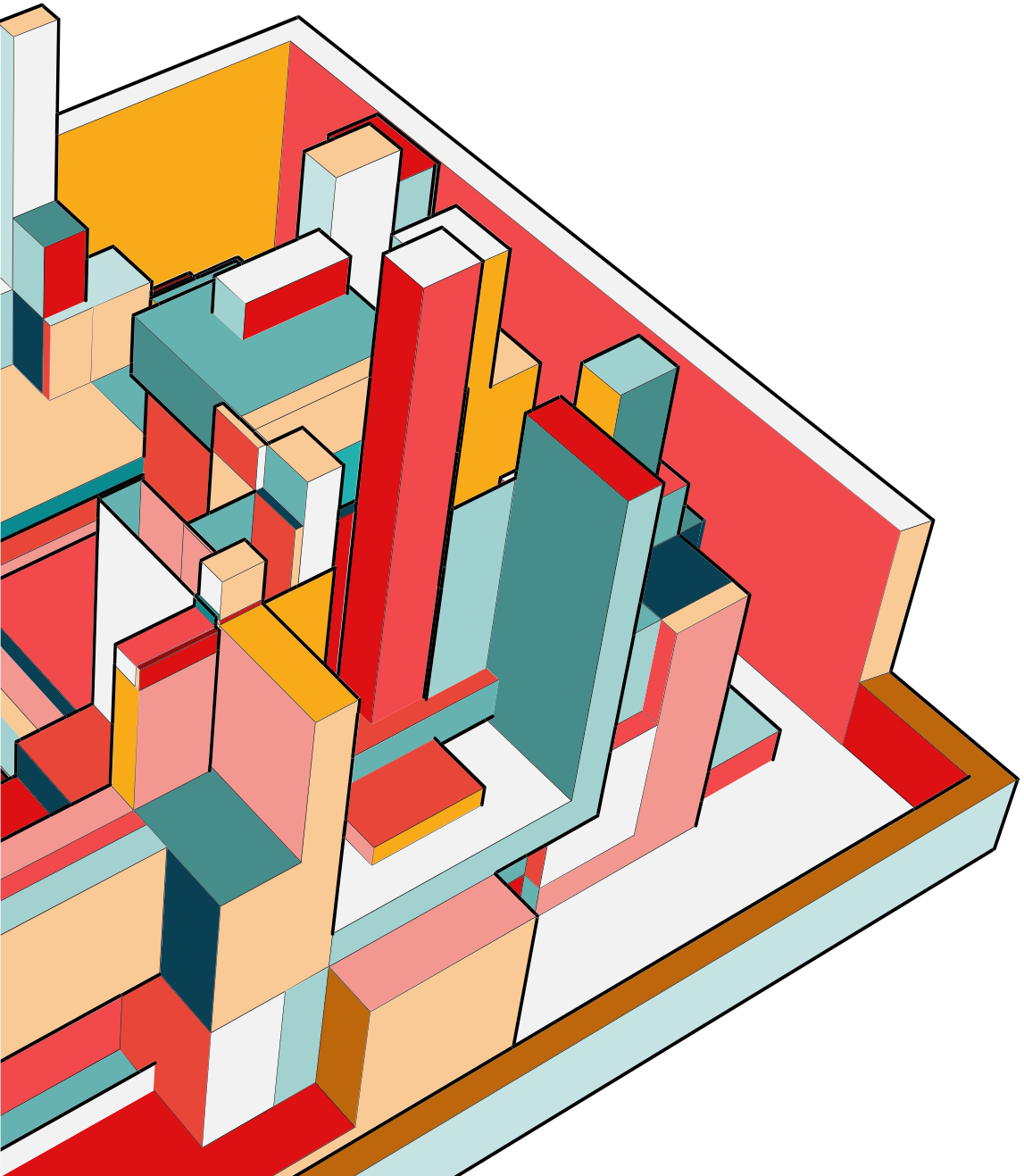
"We are on a highway to climate hell with our foot on the accelerator"

United Nations Chief Secretary General.

The data carbon footprint generated by 100 employees over 12 months is equivalent to flying from London Heathrow to New York 2,562 times a year.

Current policy does not recognize data CO2 as an issue or threat to net-zero. Consequently, there is no agreed current global measurement.





WHAT CAN WE DO?

Most organisations are only just waking up to the fact that their digital data has a CO2 footprint.

This is because, to date, there has been very little information available to spread awareness of the issue.

And there has been no publicly accessible mechanism to begin the net zero data journey …until now.

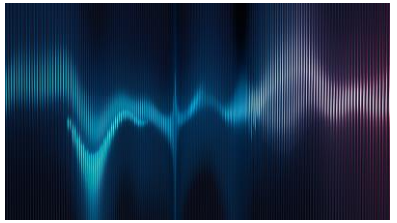
THE DATA JOURNEY

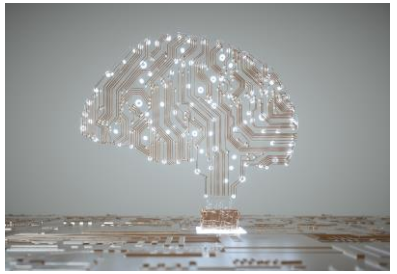
In organizations, data goes on a journey from origin through to its end use. Focusing on this data journey, we track the CO2 footprint that different data-decisions at different stages of the journey can generate.

THE UNDERPINNING THEORY
OF THE DATA CARBON LADDER

POTENTIAL

REALISED





Knowledge Acquisition

The selection of one data set from the data zone (i.e. data that is available to the problem owner)

Knowledge Assimilation

How the data will be (dis)aggregated (e.g. imported, used at host) and the size of the dataset

Knowledge Transformation

The degree of data velocity (e.g. real or near-real time) and the rate of size increase of the data over time.

Knowledge Exploitation

How the data is stored, and the data analytics being used (e.g. descriptive through to cognitive).

TO DISCOVER THE CARBON FOOTPRINT OF YOUR DATA PRACTICES

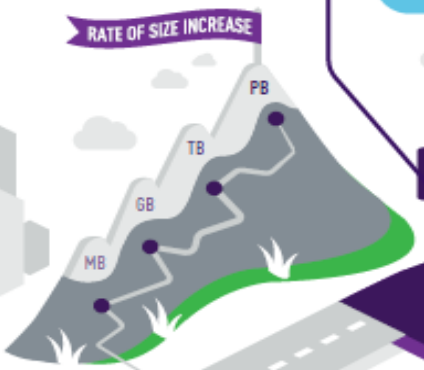
1. DATASET

To diagnose the data carbon footprint along the data journey, the ladder begins with the selection of a dataset. A new ladder is completed for every new dataset, until the minimum viable level of data is achieved to fulfil the task or solve the problem.

3. DATA SIZE

A data carbon score of the proposed dataset can be directly calculated from the size of the dataset, the user can choose to measure dataset size in MB, GB, TB, or PB, depending on appropriateness.

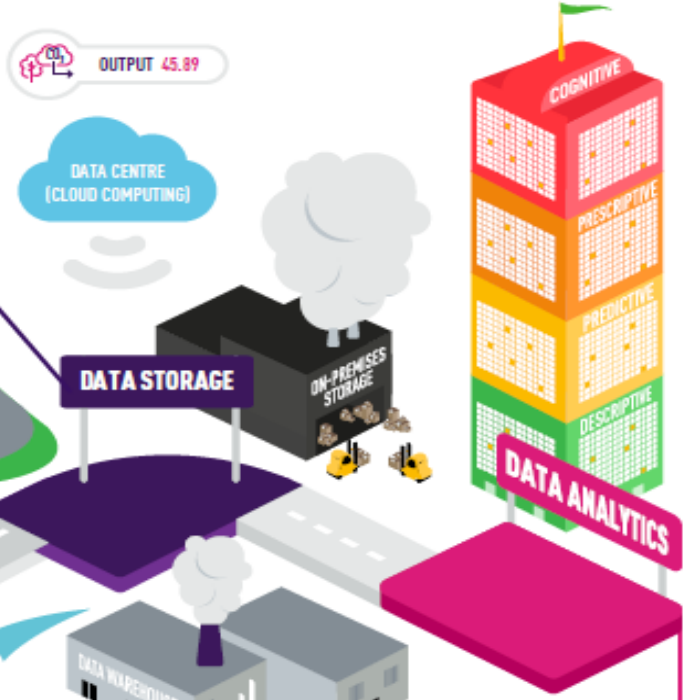
OUTPUT 27.12

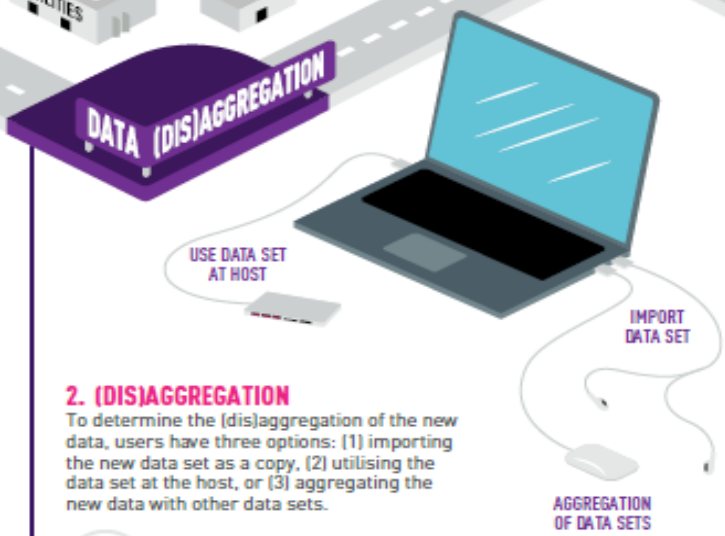


5. STORAGE

When it comes to data storage, users have three options to consider: Store at Host with zero carbon cost, (2) Data Lake, Data Warehouse or Data Centre or (3) On-premises. If the user is unsure where the dataset is being stored the default should be data centre.

OUTPUT 45.89

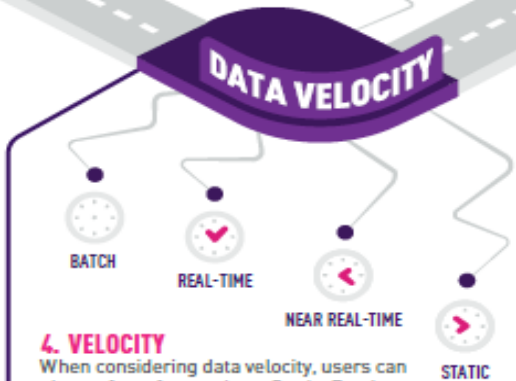




2. (DIS)AGGREGATION

To determine the (dis)aggregation of the new data, users have three options: (1) importing the new data set as a copy, (2) utilising the data set at the host, or (3) aggregating the new data with other data sets.

OUTPUT 12.32



4. VELOCITY

When considering data velocity, users can choose from four options: Static, Batch, Near-real Time, and Real-Time, representing the desired frequency of data processing. The "rate of size increase" can be measured in units MB, GB, TB, or PB.

OUTPUT 33.09

6. ANALYTICS

Users can select the type of data analytics they wish to perform from four options. These options range from descriptive analytics, (e.g. simple spreadsheet), to cognitive analytics powered by AI. Each analytics approach corresponds to an increase in CPU usage.

OUTPUT 53.65

Loughborough University: Tom Jackson and Ian Hodgkinson

The Data Carbon Scorecard

corresponds to the ladder to offer a rapid assessment of the environmental impact of your proposed data project enabling you to:

(1) quickly gauge the CO2 implications of your data project

(2) determine the ‘best’ course of action.

Guided by the data carbon ladder, the data CO2 scorecard provides a quick insight to the green credentials of the proposed dataset.

Through answering just 9 questions, the scorecard provides you with an easy traffic light display of data CO2 hotspots.

The total count for each category is summed to provide a Data CO2 Score for the whole project within minutes.

THE DATA CARBON SCORECARD

Based on the Data Carbon Ladder, the scorecard forecasts a CO2 rating for any future project:

Rating 7+: acceptable CO2.

Rating 4 to 6: moderate CO2, evaluate red & amber choices.

Rating 0 to 3: significant CO2, revisit the problem statement / project brief.

Tick the box that corresponds to your selection...	Green	Amber	Red
Step #1. The first step involves identifying the data necessary to solve the problem...			
Q. Relevant internal data is available (select one option): YES / SOME / NONE	☐	☐	☐
Q. Evolving existing datasets (select one option): YES / SOME / NONE	☐	☐	☐
Step #2. How the data is accessed...			
Q. Data Accessibility (select one option): Utilizing the dataset at host / Combining fresh datasets with pre-existing ones / Bring in dataset from an external source	☐	☐	☐
Step #3. Consideration has to be given to the size of the data set that needs to be accessed...			
Q. Data Size (select one option): Size of dataset is ≤ 2gb OR > 2gb and ≤ 3gb OR > 3gb	☐	☐	☐
Step #4. The next consideration is the frequency of data updates for report generation...			
Q. Velocity (select one option): Data is batch processed / Data remains static / Data receives near real-time updates / Data receives real-time updates	☐	☐	☐
Q. Data cadence (select one option): Annual data increase is ≤ 2gb OR > 2gb and ≤ 3gb OR > 3gb	☐	☐	☐
Step #5. The location for storing data is a significant next variable...			
Q. Storge (select one option): No storage of data / On-premise storage / Data centre, lake, warehouse	☐	☐	☐
Step #6. Decisions regarding the type of AI utilized; what level of analytics is required...			
Q. Analytics used (select one option): Descriptive / Predictive / Prescriptive / Cognitive, Generative AI	☐	☐	☐
Q. Data availability (select one option): Data is accessible to all / Data is accessible to some / Data is not available	☐	☐	☐
COUNT:	_____	n_____	n_____
9 – nR – (nA/2) = Data Project CO2 rating:			

DATA CO2 MAP

	Data Line #1	Data Line #2	Data Line #3	Data Line #4	Data Line #5	Data Line #6	Data Line #...
Data verification							
Data (dis)aggregation							
Data size							
Data velocity							
Data cadence							
Data storage							
Data analytics							
Data reuse							
	Total CO2: _____	Total CO2 _____	Total CO2: _____	Total CO2: _____	Total CO2: _____	Total CO2: _____	Total CO2: _____

Want to Learn More?

https://doi.org/10.1108/JBS-03-2022-0048

Keeping a lower profile: how firms can reduce their digital carbon footprints – **Journal of Business Strategy Article**

https://oecd-opsi.org/blog/digital-decarbonization/

On track for 6.8 billion years of continuous movie streaming: Data, energy & need for digital decarbonization – **OECD-OPSI Article**

https://digitaldecarb.org/

The Digital Decarb Toolkit can be found on our website, a free resource to help organisations on their net zero data journey.

https://doi.org/10.1080/14778238.2023.2192580

Is there a role for knowledge management in saving the planet from too much data? – **Knowledge Management Research & Practice Article**

https://theconversation.com/dark-data-is-killing-the-planet-we-need-digital-decarbonisation-190423

'Dark data' is killing the planet – we need digital decarbonisation – **The Conversation Article**