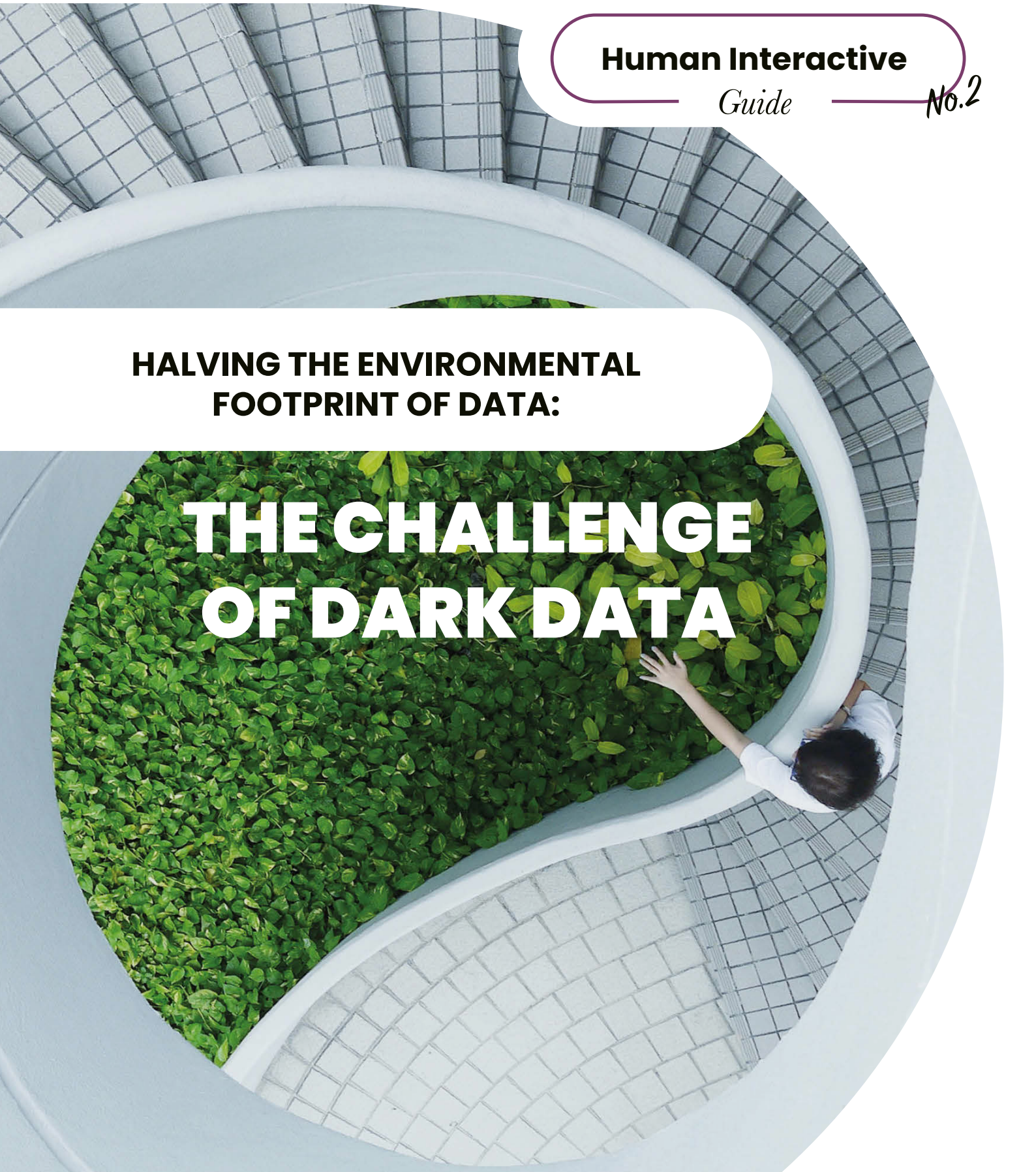
Human Interactive

Guide No.2



HALVING THE ENVIRONMENTAL FOOTPRINT OF DATA:

THE CHALLENGE OF DARK DATA

Innovation&trust

CONTENTS

HALVING THE ENVIRONMENTAL FOOTPRINT OF DATA: IT'S OPEN SEASON ON DARK DATA!

What is Dark Data? It refers to all "cold" data that is not used on a daily basis, that is rarely, if ever, searched and used, that is "dormant" in servers without anyone knowing why.

What are we talking about? Bank documents that regulations require to be kept just in case, records of property and land sales, digital medical files, insurance policies, etc.

In other words, "dormant" data?
Not really. Stored on servers, it takes up space and uses energy.

However... on average 52% of hosted data worldwide is Dark Data, 56% in France, 66% in Germany, 54% in the USA, etc.!

No less than **1.3 billion gigabytes**[1] of Dark Data are generated by businesses every day, the equivalent of 1.3 billion high-definition DVDs, just imagine!

Focus

1.3

billion gigabytes
of Dark Data
generated
by businesses
every day

As we all know, servers that store "hot" or "cold" data, whether they are housed in the basements of businesses or the data centres of Cloud providers, use energy. This use (which does not take into account the manufacture and life cycle of the servers themselves) **accounts for up to 14% of the digital sector's carbon footprint in France, according to the study published by Le Green IT[2].**

This represents almost 2.3 million tonnes of CO_2 emissions, which is huge! According to the same study, if nothing is done to reduce this impact, it could increase by 60% by 2040 to more than 5 million tonnes, and around 0.2% of total CO_2 emissions. It should be reiterated that France's National Low Carbon Strategy, pursued by the Ministry of Ecological Transition, anticipates carbon neutrality by 2050, with a target of -50% by 2030.

It is clearly apparent that, as Dark Data or "cold data" accounts for half of the total amount of data hosted, storing it differently using "inert", non-energy intensive solutions would halve CO_2 emissions relating to data hosting in France: 1.15 million tonnes of CO_2 saved, no less.

This is what this second Human Interactive guide invites you to explore: it's open season on cold data!

1 Digital Decarbonisation
2 Environmental footprint of the global digital economy

FORGOTTEN DATA, THE BLIND SPOT IN CORPORATE ENVIRONMENTAL STRATEGIES

Focus

Key figures Dark Data

175

zettabytes of data **will be generated by the IoT by 2025**

Dark Data figures are likely to increase massively and inexorably in the years to come, as 175 zettabytes[1] of data will be generated by the Internet of Things (IoT) by 2025.

According to various estimates, Dark Data accounts for, on average, 52% of the world's stored data, and generated 6.4 million tonnes of CO_2[1] in 2020. This is equivalent to the annual carbon dioxide emissions of 80 countries, or of a car travelling 575,000 times around the Earth. In France, it accounts for 56% of hosted data, i.e 2.3 million tonnes of CO_2 per year[2], or 352,876 days of gas heating[3].

The environmental impact of Dark Data raises the issue of the nature of this data, and its accumulation processes

Dark Data refers to all unused, unknown and untapped data in a company, generated by the daily interactions of users - employees, customers, external contractors - with countless machines and systems. These files and information collected, processed and stored by businesses, but no longer used, add to their digital environmental footprint[4].

This includes **technical data (server log files) as well as geolocation data, emails and attached files, data from former employees, presentations, financial results, customer call recordings,** etc. This data, considered too old to be valuable, incomplete or redundant, may also be in a playback format that is incompatible with the tools available.

1 The benefits of reducing Dark Data
2 Calculating the carbon footprint of Dark Data:
(CO_2 size of the digital sector x % of carbon footprint of data centres in use)/ % Dark Data
3 Hellocarbo: tonne of CO_2 equivalent
4 Gartner – Dark Data definition
Splunk – Dark Data definition

Taking into account the environmental impact of digital services also requires focusing on the issue of data

Today, the environmental footprint of digital technology is mainly addressed through the prism of IT equipment and terminals: extending their lifespan, reconditioning, using eco-materials, etc.

Yet, data is the fastest growing digital asset. While the health crisis and widespread use of teleworking have led to a tenfold increase in the number of digital terminals deployed (laptops, monitors, etc.), data has increased in an equally spectacular manner: +24% data generated, stored and **consumed between 2020 and 2021 and, according to IDC, the global volume of data will increase 45-fold by 2035 to more than 2,100 zettabytes**.

Data governance and hosting are currently paradoxically overlooked as a way of reducing environmental impact. Optimising the management of energy consumption in data centres and the recyclability of servers - which are likely to multiply to be able to host all this data - will not be enough.

A Marie Kondo inspired data sorting is required!

REGULATORY CONSTRAINTS, LACK OF SKILLS, LACK OF TIME: THE ORIGINS OF DARK DATA

According to an international study by Splunk[1], the top three reasons for the proliferation of Dark Data in organisations are the increase in its volume (39%), the lack of skills to process it effectively (34%) and the lack of resources (32%).

Although businesses' inability to process their data is one of the main reasons suggested, regulatory constraints are also a major challenge.

These constraints - particularly in terms of archiving - also generate Dark Data within businesses.

French law, for example, requires organisations to archive data relating to their activities for a minimum legal period. This period varies depending on the nature of the documents and legal obligations (2 years from the termination of an insurance policy or 30 years for property and land acquisition or sale agreements).

It is worth noting that organisations equipped with a Content Services Platform (CSP)[2] or similar content management system have a better overview of the data subject to these legal obligations, and therefore a potential advantage in managing Dark Data.

Could regulatory changes reduce or even stem the flow of Dark Data storage?

No, because technological developments and the human reflex to keep and preserve will generate more stored data.

When technological development and the precautionary principle contribute to the proliferation of Dark Data

Dark Data is constantly on the rise as a direct result of the inflation in the amount of data generated each year. The increasing use of connected objects (IOT) inevitably leads to a massive production of data. The IOT Analytics company[3] estimates that the number of connected objects worldwide could reach 21 billion by 2025, compared with 12.3 billion in 2021.

Another factor, human this time, conducive to Dark Data is fear. It leads to excessive data retention. For Jérémy Cousin, President of CIV France: *"Everyone wants to keep their data dating back to 2000, just in case. Businesses also apply the same principle by storing data in our data centres without accessing it for years".*

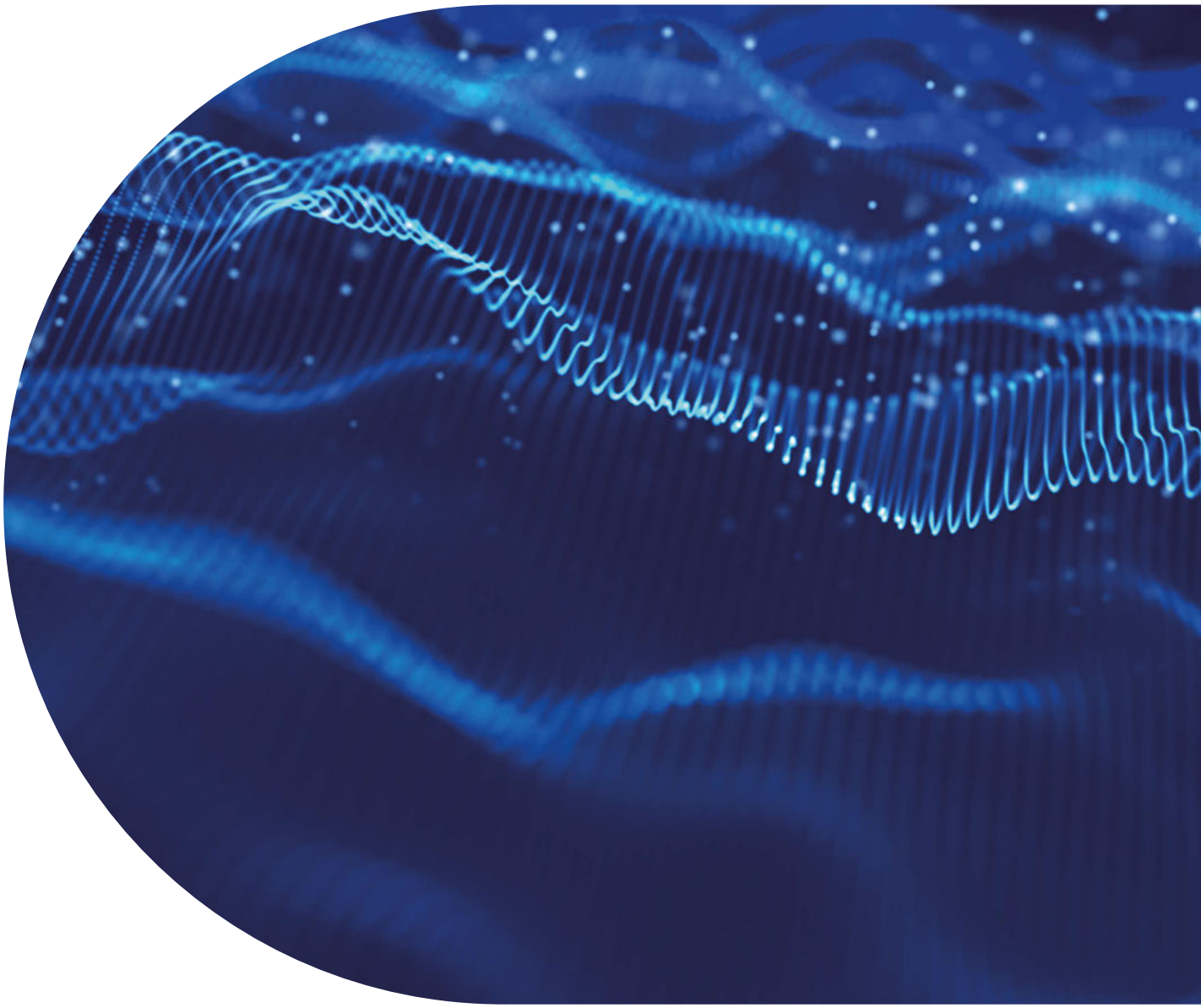
The challenge for companies is not so much to delete this data at all costs - as some of it needs to be kept for archiving reasons for example - as to offer a suitable hosting solution.

1 The state of Dark data
2 Content Services Platform: Gartner's term that replaces Enterprise Content Management (ECM)
3 IOT Analytics research report 2018

So how can we distinguish between data that should be kept within clickable range, data that might be useful from time to time, and data that may be useful one day, if ever?

Today, companies have to face the colossal task of sorting and classifying data on their own. The lack of in-house skills and a limited supply of external resources to guide them through this process make it impossible to contain the proliferation of Dark Data.

"Technological developments and the human reflex to keep and preserve will generate more stored data"



CYBER ATTACKS, STORAGE BUDGET, REGULATORY COMPLIANCE: DARK DATA IS COSTLY TO BUSINESSES

Dark Data generates substantial costs for organisations. Its estimated cost to businesses worldwide is a staggering **€2 billion** each month[1], due to a significant and unnecessary increase in the storage capacity required to host said data.

On top of this is the environmental impact of Dark Data, which in 2020 generated **6.4 million tonnes of CO_2** worldwide. This environmental footprint is borne not only by the businesses that generate this Dark Data but also - and above all - by those that store and process it.

Cloud providers, data analysis and management companies and software vendors are all suffering the consequences of Dark Data.

The Tessi Group, for example, by providing customers with storage solutions, finds itself having to take on the hosting of this data. What proportion of the 2.3 petabytes of customer data held by Tessi should be classified as Dark Data?

Beyond these financial and environmental impacts for businesses, other factors can have significant and sensitive consequences for businesses.

Dark Data, risk factors for businesses

According to a study by HelpNetSecurity[2], 42% of **Dark Data** is confidential information, 1% **sensitive personal information**, **and 9% personally identifiable information.** Failing to process this data therefore exposes businesses to a **cybersecurity** risk.

Dark data does not solely consist of useless data. Some of the data hosted by businesses is of strategic importance. The hacking of these unclassified files stored among many others could have major consequences for businesses and their employees (ransomware, data sold on the dark web, etc.).

The existence of this risk calls for the implementation of a data governance strategy. This approach is validated by Jean-Philippe Balivet, Director of CSP activities at Tessi France: *"Businesses need to be made aware of the governance of data and the life of the document. To reduce the cybersecurity risk, the importance of the data must be analysed and assessed before it is stored."*

This need to regulate data and its use highlights another risk faced by businesses: **compliance in terms of personal data**. Under the GDPR, some personal data must be kept by businesses for a limited period of time (usually 3 years). Beyond this limit, data must be deleted by businesses. However, this data is rarely classified by companies, getting lost among Dark Data, and is generally not deleted after the 3-year period. Businesses may be held liable for this negligence, which puts into perspective the need for data classification, managed by content service platforms (CSPs).

Focus

Dark Data costs businesses an estimated

€2 billion

each month

1 IDC Market Analysis Perspective: Worldwide Managed Support Services 2020

2 HelpNetSecurity study

CARBON ACCOUNTING OF DATA: WHAT IS THE BEST APPROACH TO COMBATING DARK DATA? GOING DIGITAL, WITHOUT FORGETTING TO TRANSFORM

Addressing the ecological transition challenge posed by Dark Data requires finding solutions to assess and sort this type of data while, more importantly, supporting businesses in this sorting process.

Innovation&trust's goal of reducing the volume of Dark Data in businesses by 15% to 20% within five years involves the development of a dedicated system for measuring and managing this "cold data": **"EKODATA" by Innovation&trust.**

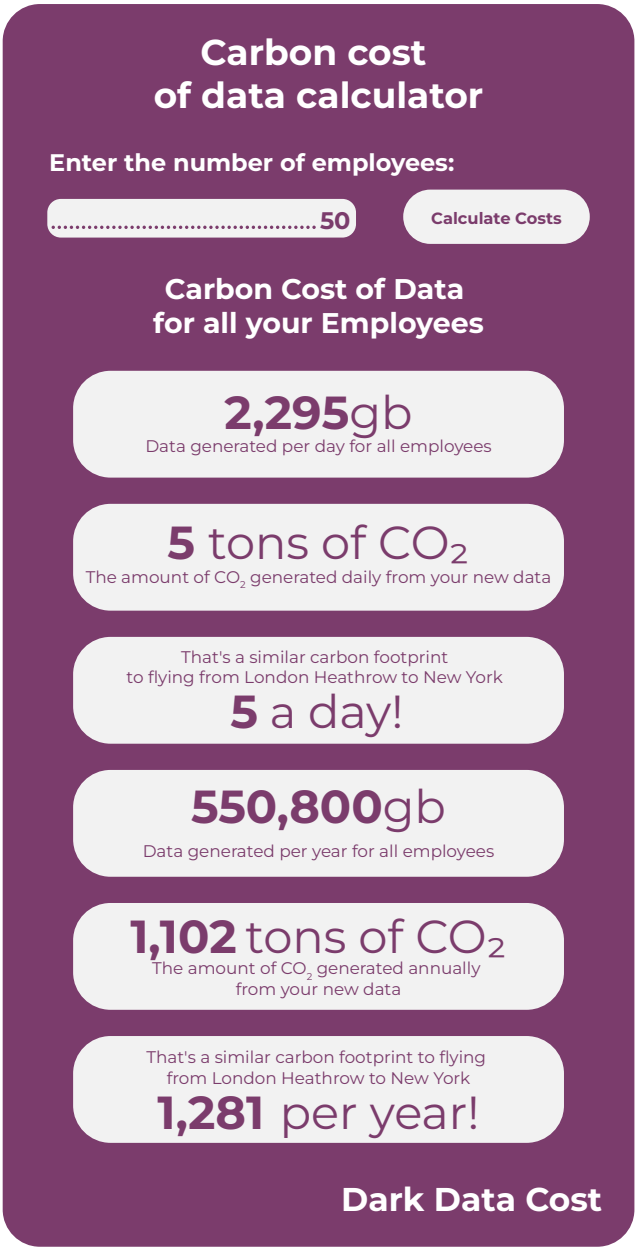
"EKODATA" by Innovation&trust is a platform for assessing, sorting and classifying Dark Data. It aims to keep storage activities in an environmentally responsible framework while helping businesses that handle large volumes of data.

The Innovation&trust method: carbon accounting to keep Dark Data off the servers

"EKODATA" by Innovation&trust is based on several innovative features.

A Dark Data calculator - Track my Data - will allow an organisation to measure and track its estimated carbon and energy footprint and potential savings in terms of hosting costs, based on the volume of data stored by the organisation.

The calculator could be similar to the one provided by the Digital Decarb:



Using Artificial Intelligence (AI) and Machine Learning, this Dark Data management platform scans, identifies, labels and classifies information, ensuring that sensitive or at-risk data is protected. It can also archive data deemed important but less useful in a "glacier" storage system, and lists GDPR-related files with the implementation of alerts or notifications triggered as the expiry of the authorisations received approaches.

This platform is designed for businesses processing large volumes of data, regardless of their size or sector of activity (retail, healthcare, finance, insurance, etc.).

It is the first solution on the market that specifically addresses Dark Data and provides businesses with a resource to manage and sort their data.

Innovation&trust expects it to be commissioned in the first quarter of 2023.

A responsible and committed approach to serving customers

At Innovation&trust, we are committed to combining technological development with economic growth, taking into account the environmental impact of our activities. Committed to a responsible digital programme, we are developing solutions to make the digital and the environmental transitions converge, and wish to guide our customers through this process.

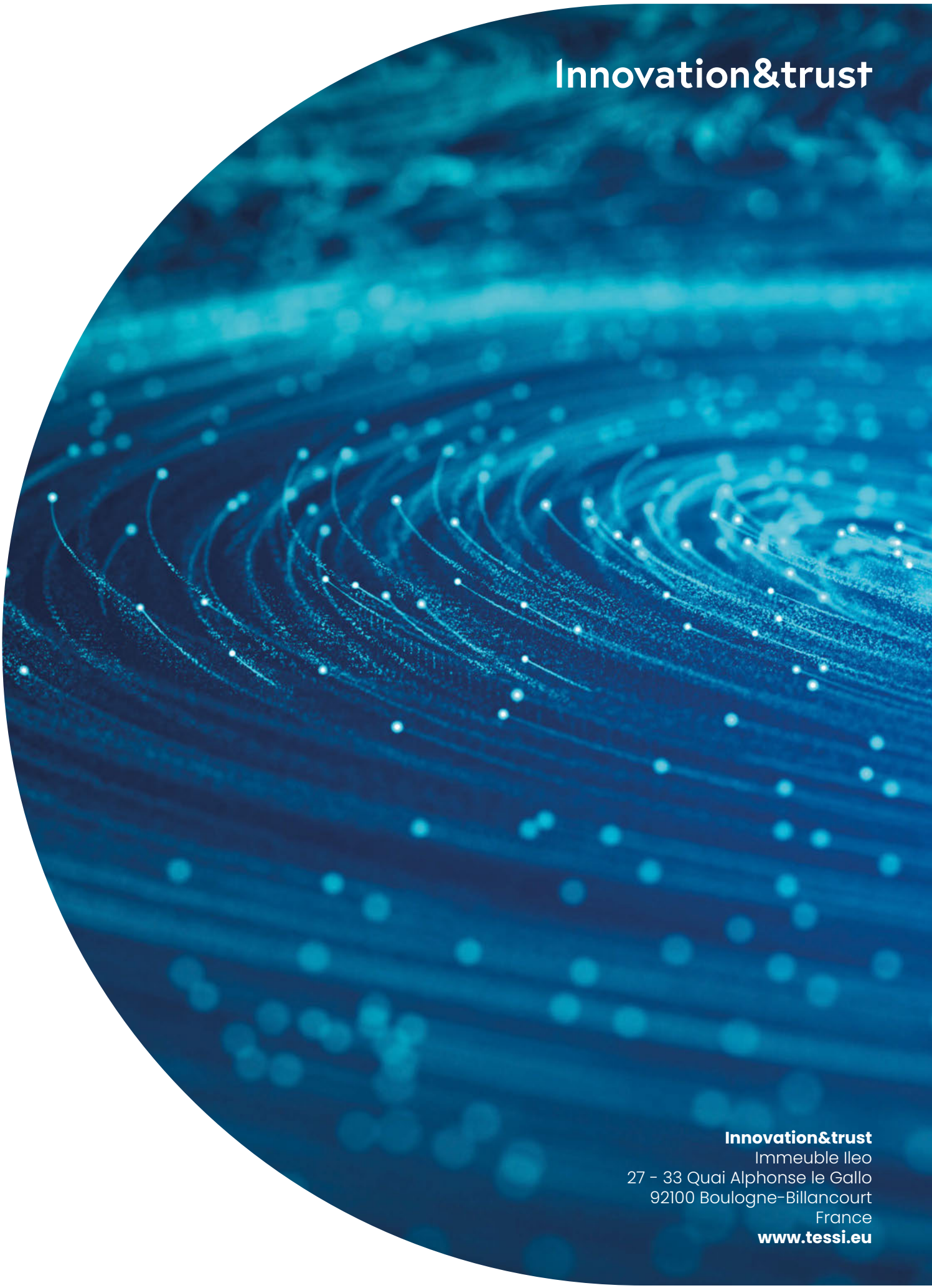
All our solutions are already developed with a responsible approach in mind: a well-coded solution from the first line of code is an eco-designed solution!

Innovation&trust is a subsidiary of the Tessi Group:

The Tessi Group's commitments to a Responsible Digital Technology

- Tessi joined the Planet Tech'Care with NUMEUM (formerly Syntec Numérique) in 2020 https://planet-techcare.green
- Tessi signed the NUMEUM manifesto for the inclusion and the reconversion of women in the digital sector, alongside more than 150 other committed companies
- Tessi has been a member of the *Institut du Numérique Responsable* and a signatory to the Responsible Digital Technology Charter since February 2022

Innovation&trust



Innovation&trust
Immeuble Ileo
27 - 33 Quai Alphonse le Gallo
92100 Boulogne-Billancourt
France
www.tessi.eu