LUISS

Master of Science in
Law, Digital Innovation and Sustainability

Chair of Data Protection Law

Privacy Fatigue: the Challenge of Data Protection Literacy for Digital Citizens

Prof. Filiberto Brozzetti

SUPERVISOR

Prof. Maria Vittoria Catanzariti

CO-SUPERVISOR

Livia Alegi – 631213

CANDIDATE

Academic Year 2023/2024

Table of Contents

Introduction

The digital era hinges itself on data, the so-called "new-oil", making personal data protection a crucial challenge. Every click is turned into data: looking up directions for a friend's house, shopping for home maintenance products, reading a newspaper article, or watching a tv series on a streaming service. These interactions and the personal data they provide, details most would share with only a handful of individuals, become data points in a spreadsheet, ready to be interpreted and used for further action. Everything is quantified and then measured to discover more information and influence behaviour.

How exactly personal data is used is often not clear. Individuals don't know exactly what they're allowing their data to be used for, nor who has access to it. This information is located in privacy policies, explained in verbose technical language, perhaps hidden in page seven of nine. These documents are too difficult to read for large parts of the population and, in any case, few individuals are willing to spend conspicuous time and effort to read a document they will likely accept anyway. After all, services offer a non-choice: to use one, users must accept the privacy policy as there is no way to negotiate or access limited services in exchange for increased privacy. The result is that individuals give up personal data with little resistance.

While spam emails and calls are annoying, they aren't the only issue. Information stored in databases becomes the target of data breaches. In 2023 alone, 40.42 million user accounts were exposed worldwide.[1] Once personal data is obtained, ill-intentioned individuals may use leaked credentials to access applications or services, or even leverage typically reserved information to contact victims and scam them more effectively.

Despite this, individuals repeatedly state that they are concerned about how their personal data are used and care about protecting them. But this is not a simple task: the number of services now available online means individuals are frequently asked to allow cookies, create new accounts, and accept privacy policies. This sequence occurs over and over, especially as time spent daily on the Internet increases and many aspects of life (from banking to school to leisure time) have become digitally focused.

[1] Surfshark, 'Number of User Accounts Exposed Worldwide from 1st Quarter 2020 to 4th Quarter 2023 (in Millions)', Chart (Statista, January 2024), https://www.statista.com/statistics/1307426/number-of-data-breaches-worldwide/.

As these procedures multiply, so does the effort and attention needed to manage personal data. This, coupled with continuous data breaches, leads individuals to feel increasingly overwhelmed and hopeless. For many, protecting personal data simply requires too much effort.

One way to limit fatigue could be to ensure individuals are equipped with the knowledge they need to protect themselves: by boosting data protection literacy, individuals would learn the best practices, measures and tools at their disposal. This empowers them to make the decisions they want to make, rather than defaulting to whatever option online services suggest.

The importance of literacy as a tool for empowerment is not new. It finds a precedent in the rise, for example, of literate (or conscious) consumers: overwhelmed by the choice that supermarkets offer, individuals have gradually become concerned and aware of what they're purchasing. To ensure their shopping carts were filled with products that matched what they were looking for, people learnt to look at labels and understand what the products were telling them. This included learning the difference between an "expiration" or "best before" date, what geographical indication acronyms mean and how to read nutritional labels.

Data protection literacy can similarly become the catalyst for decreasing privacy fatigue. Understanding what dangers one must protect themselves from, knowing what things are and how to do them, means being less susceptible to the variety of issues which arise when navigating on the Internet.

This thesis examines the relationship between data protection literacy and privacy fatigue to see whether an increase in literacy results in lower levels of fatigue. A questionnaire was used to examine the relationship between the two phenomena, and observe beliefs and tendencies which could be expanded upon in future research. Representativeness of the sample was not a requirement as the questionnaire was exploratory in nature; this research provides general insight into the interaction between data protection literacy and privacy fatigue, and cues for future research.

This work is organised as follows. Chapter 1 examines existing literature on the main concepts at the core of this work. This is done by explaining how our society has become datafied, and then proceeding to examine relevant elements for this study. This means describing digital citizens, then defining the concept of data protection (and digital literacy as a macro category) and finally explaining why individuals may experience privacy fatigue.

Chapter 2 describes the empirical methodology adopted for this thesis. This is done by describing the questionnaire at the basis of this work in all its elements: how the questionnaire was

built (taking into consideration data protection principles), how the data analysis process was structured and carried out, and finally the composition of the sample.

Chapter 3 then analyses the questionnaire's results section by section. Each question is analysed and any correlations with demographic markers are highlighted. In doing so, this section highlights both surprising and predictable results. To verify the research question, items from the data protection literacy and privacy fatigue section were directly compared with one another. Finally, four potentially relevant relationships between select survey items were also looked into, with the purpose of understanding participants' opinions on location data, privacy policies, security and advertising.

The results are then discussed in Chapter 4, where they are related to a variety of considerations and existing literature. After reflecting on the relationship between literacy and privacy fatigue, the two phenomena are discussed on their own. This ensures that the relationship between the two is appropriately discussed without, however, forgetting that both phenomena are independently influenced by other factors. The last paragraph of this chapter discusses limitations of this research and avenues for future research.

Finally, the conclusion rounds out this thesis by reflecting on the contents of the previous chapters. It restates the research objective and then summarises the findings, in order to highlight the relevance of this research in a broader context.

1. Literature review

1.1 Introduction

In order to understand how privacy fatigue and data protection literacy affect one another, a broader understanding of data's role in contemporary society is required. This chapter will describe how society has become datafied and what mechanisms are at play, then turn to the discussion of three core concepts referenced throughout this work: digital citizens, data protection literacy, and privacy fatigue. In doing so, this chapter sets the stage for understanding and discussing the results and findings of the questionnaire at the basis of this thesis.

1.2 The datafication of society

As technology has become more pervasive, more and more data has been created; forecasts project that 181 zettabytes – 181 trillion gigabytes – will be created in 2025.[2] This is known as datafication, "the transformation of human life into data through processes of quantification",[3] quantifying activities and enabling the discovery of patterns. It has two crucial components: "the creation of a trace that is recorded and circulated in the form of data beyond that particular moment and place, and the further use of such a trace as a meaningful element in other processes."[4]

Data is valuable because it provides insight and enables prediction. Governments and research labs use data to observe the effects of their projects, guide political and financial decisions; this helps them understand how to foster growth, efficiency and welfare. Companies such as Google, Amazon, Meta, Apple, and Microsoft (commonly referred to as GAMAM) use the large amounts of data gathered from their users to discover behaviour patterns.

Beaulieu and Lionelli (2022) highlight several characteristics that make data valuable. It is a non-rival good (to an extent), meaning it can be used by multiple actors. This in turn implies reusability, but also open-endedness: the same data can be used multiple times and for different purposes than the ones it was initially collected for. Unprocessed, "raw", data is valuable in itself, but

[2] IDC, 'Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2020, with Forecasts from 2021 to 2025 (in Zettabytes).', Chart (Statista, 7 June 2021), https://www.statista.com/statistics/871513/worldwide-data-created/.

[3] Ulises A. Mejias and Nick Couldry, 'Datafication', *Internet Policy Review* 8, no. 4 (29 November 2019), https://doi.org/10.14763/2019.4.1428.

[4] Anne Beaulieu and Sabina Leonelli, *Data and Society: A Critical Introduction* (London: Sage Publications Ltd, 2022), 4.

processing and connecting multiple sets of data from distinct services and activities increases the value.[5] This is made possible, for example, by using the same Google account for Google-owned services (such as Gmail, YouTube, Google Nest or Fitbit) and also non-Google services (such as Canva, The New York Times, Hinge and ChatGPT); being able to track and understand user behaviour between services of different nature further increases the value of data.

Data resources haven't always been this important: they were initially by-products, a side-effect of users' interaction with online services, and data collection (or accumulation) has gradually grown into an objective of its own right. This process is exemplified by Google's evolution and occurred in three steps, as described by Shoshana Zuboff (2019). In the first phase, as Google launched new services and users interacted with them for the first time, new sets of data were created. For example, a simple Google search produces data such as "the number and pattern of search terms, how a query is phrased, spelling, punctuation, dwell times, click patterns, and location."[6] This collateral information was initially ignored. In the second phase, it was used to improve the service and experience for its users. By observing commonly misspelt searches and the subsequent (corrected) searches, Google was able to implement its suggestion feature ("Did you mean…"). Google valued user interactions, analysing them to improve Search and gain an edge over its competitors; in this phase "behavioral data were put to work entirely on the user's behalf."[7] The third phase saw Google use this same mechanism for a different purpose: advertising. The data which had been previously used to improve search results would now be (predominantly) used to target ads to individual users.

At first, this data was "simply 'found,' as a by-product of users' search actions"[8] This has given rise to the popular idea that data is the new oil;[9] while this undoubtedly highlights the value that data holds, this view must be acknowledged as reductive. Implying that data is a resource that is naturally available and ready to be "harvested" means ignoring the process data goes through to become valuable, the tools needed to do so and the entities who make this happen. Though outside

[5] Data can be replicated and shared without limits but requires resources that may be limited, such as storage and devices to access it.

[6] Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (London: Profile books, 2019), 67.

[7] Zuboff, 69.

[8] Zuboff, 93–94.

[9] 'The World's Most Valuable Resource Is No Longer Oil, but Data', *The Economist*, 6 May 2017, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

the scope of this work, it is important to consider that the social origins and technical aspects of technologies directly influence how society is affected by them.[10]

Data is now deliberately sought out and forcefully obtained, and thus the goal of its usage has shifted: this by-product is no longer used exclusively to improve a service; what matters is to analyse data, draw conclusions based on it and act accordingly. This has quickly become a self-sustaining, vicious cycle:

> 'Data analytics' is a way of revealing patterns that would otherwise remain hidden and this is best done through the use of quantified data, systematically obtained from collection tools and platforms. The data and the conclusions drawn based on it are in turn affected by the activities that produce the data - a feedback loop [...].[11]

At the same time, users have been sold the idea that there is value in measuring any aspect of their life that can be measured. In some cases, this seems intuitive: fitness and health data can give individuals peace of mind and a better understanding of their health. Using fitness trackers and smartwatches, individuals can have a well-rounded understanding of their wellness by tracking steps, heart rate, blood oxygen levels and overall fitness levels through scientifically validated tools (such as PAI scores[12]). Devices like the Apple Watch have proven their worth through accounts of lives being saved by their heart rate monitors.[13]

If not lifesaving, data can at least be fun. Spotify, for example, has mastered the art of using behavioural data as a marketing tool: in its annual Spotify Wrapped campaign, the company provides entertaining graphics detailing users' listening habits. It does so by repackaging data it has accumulated (including favourite songs, artists, genres by listening time and more) over the course of a year. This has become a selling point for Spotify: it's both a reason to choose the streaming platform over others, and one to keep using it. However, the information that can be inferred from our listening habits and Spotify's use is not trivial: the service does not solely use the information that someone is a Taylor Swift fan to suggest music by similar artists, such as Sabrina Carpenter, Gracie Abrams or Fletcher. This same information can be used to make inferences about listeners' moods, then deploy

[10] Langdon Winner, 'Do Artifacts Have Politics?', *Daedalus* 109, no. 1 (1980): 121–36.

[11] Beaulieu and Leonelli, *Data and Society*, 152.

[12] 'Personalized Activity Intelligence: A Better Way to Track Exercise?', Harvard Health, 27 January 2017, https://www.health.harvard.edu/blog/personalized-activity-intelligence-better-way-track-exercise-2017012711031.

[13] See, for example: Vanessa Orellana, 'My Apple Watch Saved My Life: 5 People Share Their Stories', CNET, 9 September 2020, https://www.cnet.com/tech/mobile/apple-watch-lifesaving-health-features-read-5-peoples-stories/.

the appropriate ads: for example, a user who listens to songs about heartbreak could be targeted ads for chocolate and dating apps.[14]

Data is becoming central even in areas which were previously untouched. Though individuals associate cars with "'thrill' of driving, the 'joy' of the road, the 'passion' of the collector, the nostalgia for retro designs [...]",[15] cars have become computers on wheels and car companies have a side business as software companies.[16] Vehicles now collect information in a number of ways: through the increasing number of microphones, cameras and other sensors in vehicles; when a phone is plugged into the car's USB port; by accessing a phone through the car's app when it has been downloaded on a phone. In this way, car manufacturers have access to all sorts of data, including but not limited to: demographic data, credit card information, biometric information, sexual activity, medical information, trip start and end location.[17] Unlike the previous examples, however, drivers (and their passengers) are not aware that they are being watched, nor what information is being collected.

Behaviour on search engines, streaming sites, fitness devices, cars and more reveals an incredible amount of personal information. These are intimate details, ones that would typically only be revealed to those who people are close to, such as political preferences, religious beliefs and sexual orientation.

The core issue isn't that users are using free services and becoming products, as is often stated.[18] The issue is rather that users provide the raw material (data) which companies use to create predictions about our behaviour to sell to advertisers. Though companies assure their users that data is anonymised and aggregated, they purchase data from various sources, re-identify it and create one digital profile.[19] This allows precise targeting and influences us to the point that "we are exiles from

---

[14] Jack Morse, 'How to Stop Spotify from Sharing Your Data, and Why You Should', Mashable, 5 April 2022, https://mashable.com/article/spotify-user-privacy-settings.

[15] Mimi Sheller, 'Automotive Emotions: Feeling the Car', *Theory, Culture & Society* 21, no. 4–5 (2004): 5.

[16] Jerry Hirsch, 'Elon Musk: Model S Not a Car but a "Sophisticated Computer on Wheels"', Los Angeles Times, 19 March 2015, https://www.latimes.com/business/autos/la-fi-hy-musk-computer-on-wheels-20150319-story.html.

[17] Jen Caltrider, Misha Rykov, and Zoë MacDonald, 'What Data Does My Car Collect About Me and Where Does It Go?', *Privacy Not Included, 6 September 2023, https://foundation.mozilla.org/en/privacynotincluded/articles/what-data-does-my-car-collect-about-me-and-where-does-it-go/.

[18] Scott Goodson, 'If You're Not Paying For It, You Become The Product', Forbes, 3 March 2012, https://www.forbes.com/sites/marketshare/2012/03/05/if-youre-not-paying-for-it-you-become-the-product/.

[19] Veronica Barassi, *I Figli Dell'algoritmo: Sorvegliati, Tracciati, Profilati Dalla Nascita* (Luiss University Press, 2021).

our own behaviour, denied access to or control over knowledge derived from its dispossession by others for others."[20] It is impossible to escape or resist this continuous extraction mechanism.

Data protection mechanisms become essential in this context: being able to choose what data is shared with whom means retaining control over who we are and determining how we relate to others.

1.3 Digital citizenship

The term "digital citizens" seems self-explanatory, but as Hintz et al. (2018) point out that there is no single definition. A broad, intuitive description would portray digital citizens as individuals who perform a variety of actions online; to do this, they must possess competencies and skills which make them aware of the consequences of their actions. Examples of broad definitions of digital citizens include:

> those who use the Internet regularly and effectively—that is, on a daily basis.[21]

> someone who is skilled in using the internet in order to communicate with others, buy and sell things, and take part in politics, and who understands how to do this in a safe and responsible way[22]

> able to use digital tools to create, consume, communicate and engage positively and responsibly with others. They understand and respect human rights, embrace diversity, and become lifelong learners in order to keep step with evolutions in society.[23]

Literature on this topic has steadily increased over the years; many definitions have been given for this term, creating ambiguity which may cause confusion. This has been mapped in a concept analysis by Choi (2016), which identifies four conceptions in literature: ethics, media and information literacy, participation/engagement, critical resistance.[24] Definitions categorised under ethics refer to appropriate online behaviour (known as "netiquette"). Media and information literacy, instead, covers

[20] Zuboff, *The Age of Surveillance Capitalism*, 100.

[21] Karen Mossenberger, Caroline J. Tolbert, and Ramona S. McNeal, *Digital Citizenship: The Internet, Society, and Participation* (Cambridge, Massachusetts: MIT Press, 2008), 1.

[22] 'Digital Citizen', in *Cambridge Advanced Learner's Dictionary & Thesaurus* (Cambridge University Press), accessed 19 June 2024, https://dictionary.cambridge.org/us/dictionary/english/digital-citizen.

[23] 'The Concept - Digital Citizenship Education (DCE)', Council of Europe, accessed 19 June 2024, https://www.coe.int/en/web/digital-citizenship-education/the-concept.

[24] Moonsun Choi, 'A Concept Analysis of Digital Citizenship for Democratic Citizenship Education in the Internet Age', *Theory & Research in Social Education* 44, no. 4 (October 2016): 565–607, https://doi.org/10.1080/00933104.2016.1210549.

individuals having the skills and capacity to interact and process the information they encounter in online environments. Participation/engagement includes both political participation and personalised participation; critical resistance is similar and thus closely related, though it has more "progressive and radical viewpoints",[25] which in turn lead to deeper engagement. The latter two conceptions, which reference civic or political engagement in an online context, are more in line with the traditional concept of citizen.[26] When digital, however, citizenship does not denote traditional ideas such as membership to a nation-state, but rather participating in society through digital acts.

The variety of issues that digital citizenship encapsulates demonstrate that this is a multi-faceted and multi-disciplinary concept; arguably, the factors at play cannot be separated from one another: the citizens cannot be separated from the activities they carry out and the skills required to do so. In the digital era and in datafied society, the pervasiveness of activities that can (and must) be carried out online mean that everyone is assumed to be a digital citizen. Expectations of connectivity (for example through social media) thus imply a variety of online interactions (i.e. following someone) and activities, and specific polite behaviour (netiquette).

Academic and non-academic resources alike focus mostly on students. Several curricula are available online, with the objective of ensuring that younger generations are "prepare[d] [...] to take ownership of their digital lives."[27] Referencing the concepts mapped by Choi (2016), these curricula mainly interpret digital citizenship as ethics and literacy: this means teaching children, for example, to be kind to others online and be critical of what they see on the Internet. The objective of such resources is to ensure students can participate in society and thus succeed later in life.

Within the broad context of digital citizenship, a distinction can be made between "digital natives" and "digital immigrants", terms first used by Marc Prensky (2001). Digital natives indicates those who are "'native speakers' of the digital language of computers, video games and the Internet."[28] These are individuals who grew up in the information age (generally from 1984

[25] Choi.

[26] "[...] a member of a political community who enjoys the rights and assumes the duties of membership." from Dominique Leydet, 'Citizenship', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta and Uri Nodelman, Fall 2023 (Metaphysics Research Lab, Stanford University, 2023), https://plato.stanford.edu/archives/fall2023/entries/citizenship/.

[27] 'DigCit Curriculum', Common Sense Education, accessed 9 September 2024, https://www.commonsense.org/education/digital-citizenship/curriculum.

[28] Marc Prensky, 'Digital Natives, Digital Immigrants Part 1', *On the Horizon* 9, no. 5 (September 2001): 1–6, https://doi.org/10.1108/10748120110424816.

onwards[29]) and thus are confident in their use of technology because it is extremely familiar to them. Digital immigrants, on the other hand, have had to gradually learn to use technology, incorporating it into their existing habits and adapting to new ways of thinking.

1.4 Digital literacy

In the context of an increasingly complex digital world, literacy is an essential component to understanding what the risks of the online world are and gaining the skills and knowledge to make conscious decisions.

In its most literal and narrow sense, the term literacy refers to the ability to read and write. In a broader sense, it relates to the idea of knowing and understanding a specific sector; this goes beyond acquiring basic skills and instead refers to the consequences of these skills. This duality is shown in the Cambridge Dictionary definition of literacy:

> 1) the ability to read and write; 2) knowledge of a particular subject, or a particular type of knowledge.[30]

This second definition will be used for the purposes of this thesis, as it refers to having area-specific knowledge. However, it is important to understand that the term has also acquired a more holistic meaning. As UNESCO states:

> Acquiring literacy is not a one-off act. Beyond its conventional concept as a set of reading, writing and counting skills, literacy is now understood as a means of identification, understanding, interpretation, creation, and communication in an increasingly digital, text-mediated, information-rich and fast-changing world. Literacy is a continuum of learning and proficiency in reading, writing and using numbers throughout life and is part of a larger set of skills, which include digital skills, media literacy, education for sustainable development and global citizenship as well as job-specific skills. Literacy skills themselves are expanding and evolving as people engage more and more with information and learning through digital technology.[31]

[29] Paul A. Kirschner and Pedro De Bruyckere, 'The Myths of the Digital Native and the Multitasker', *Teaching and Teacher Education* 67 (October 2017): 135–42, https://doi.org/10.1016/j.tate.2017.06.001.

[30] 'Literacy', in *Cambridge Advanced Learner's Dictionary & Thesaurus* (Cambridge University Press, n.d.), https://dictionary.cambridge.org/dictionary/english/literacy.

[31] 'Literacy: What You Need to Know', UNESCO, accessed 15 May 2024, https://www.unesco.org/en/literacy/need-know.

In any of its senses, literacy plays an important role in any person's life: equipping individuals with these skills broadens their spectrum of opportunities, and in turn "reduces poverty, increases participation in the labour market and has positive effects on health and sustainable development."[32]

A crucial aspect of being literate is engaging critically with the content and situations one encounters. This enables learning throughout life and is true for more than one sector. The term literacy can also be applied to specific fields because of these common basic skills, though "media literacy", "financial literacy", "advertising literacy" and "digital literacy" refer to specific sets of skills and knowledge.

Advertising literacy, for example, includes the knowledge and skills needed to understand what advertisements are communicating. Implementing advertising literacy programs in schools ensures that children can critically engage with advertisements, understanding the difference between entertainment and commercial content; this may also reduce their purchase intentions as they grow up.[33]

It is clear, then, that an increasingly digital world requires its citizens to be literate in order to make the most of the online world. The term digital literacy was first used in 1997 by Paul Gilster in his book *Digital Literacy*:

> Digital literacy—the ability to access networked computer resources and use them—[...][34]

> Digital literacy is the ability to understand and use information in multiple formats from a wide range of sources when it is presented via computers.[35]

Twenty-seven years after Gilster's definition, computers and the Internet are not the same. They have evolved and changed, partly due to innovations such as smartphones, making the above definitions outdated and limited. Because technology evolves quickly, definitions of what it means to be digitally literate are constantly changing and crafting the perfect definition becomes hard and, arguably, unnecessary. For example, identifying text and images created by artificial intelligence would not have been part of a definition of digital literacy just five years ago, but is needed nowadays.

---

[32] 'Literacy'.

[33] Laurien Desimpelaere, Liselot Hudders, and Dieneke Van De Sompel, 'Knowledge as a Strategy for Privacy Protection: How a Privacy Literacy Training Affects Children's Online Disclosure Behavior', *Computers in Human Behavior* 110 (September 2020): 106382, https://doi.org/10.1016/j.chb.2020.106382.

[34] Paul Gilster, *Digital Literacy*, Wiley Computer Publishing (New York Chichester: Wiley, 1997), 1.

[35] Gilster, 1.

Gilster identifies core competences needed to critically engage with digital environments: critical thinking first and foremost, with others branching off it, including reading hyperlinked text, knowledge assembly and search skills. These will likely always be part of being digitally literate and will always play a role in learning how to use new technologies. As Gilster himself highlights, "the ability to properly use and evaluate digital resources, tools and services, and apply it to lifelong learning processes"[36] is crucial as it ensures prolonged skill rather than mastery of a restricted set of actions.

It may be tricky to isolate and clearly distinguish digital literacy from other similar terms, such as "information literacy", "computer literacy", "Internet literacy", "social media literacy", and "online privacy literacy". They refer to similar (though not identical) sets of knowledge and skills needed to navigate often overlapping environments. However, digital literacy is the preferred term for the purposes of this thesis as it is wide and general, thus encompasses the largest number of concepts and skills. However, there is an argument to be made for "social media literacy" and "online privacy literacy" to be core components of digital literacy, rather than standalone literacies.

Online privacy literacy (simply "privacy literacy" from now on) indicates knowledge on data and information handling in an online context, and the skills needed to protect one's own information:

> Online privacy literacy may be defined as a combination of factual or declarative ("knowing that") and procedural ("knowing how") knowledge about online privacy. In terms of declarative knowledge, online privacy literacy refers to the users' knowledge about technical aspects of online data protection and about laws and directives as well as institutional practices. In terms of procedural knowledge, online privacy literacy refers to the users' ability to apply strategies for individual privacy regulation and data protection.[37]

The use of "data protection literacy" is rare, perhaps because many of its elements are already present in the idea of privacy literacy. Data protection is arguably more specific, focusing mostly on personal data usage in the context of the online world, while privacy refers to a larger set of behaviours

[36] Two separate sources (Falloon, *From digital literacy to digital competence: the teacher digital competency (TDC) framework*, 2450; Maharana and Mishra, *A Survey of Digital Information Literacy of Faculty at Sambalpur University*, 1) quote this phrase from the 1997 edition of Gilster's *Privacy Literacy*. However, the author of this thesis was unable to find this exact phrase or similarly worded ones. Falloon stated that this quote was from p. 220, which seems unlikely since the section discusses "the merger of the media", not definitions. Mahrana and Mishra attribute it to p. 290, but this is not possible as the book only has 276 numbered pages.

[37] Sabine Trepte et al., 'Do People Know About Privacy and Data Protection Strategies? Towards the "Online Privacy Literacy Scale" (OPLIS)', in *Reforming European Data Protection Law*, vol. 20, Law, Governance and Technology Series (Dordrecht: Springer Netherlands, 2015), 339, https://link.springer.com/10.1007/978-94-017-9385-8.

and actions. However, many aspects of the two overlap, the two terms will be used interchangeably throughout this thesis.

Data protection literacy should not be mistaken for data literacy. The latter includes "the ability to process, sort, and filter vast quantities of information",[38] along with the "awareness of the meaning of data and of how conclusions are drawn from it".[39] Data literacy thus focuses on handling, understanding, using and presenting data rather than protecting one's own personal information.

Being literate does not equal being competent. Literacy typically involves more than just knowing how to use a tool: it is essential to do so with the awareness of both opportunities and risks, and thus consciously engaging in all pertinent activities. For this reason, using technology is not the same as using it competently. Being able to use smartphones, tablets, computers and so on does not mean doing so confidently and being aware of the opportunities and risks that can be found. One example of this is driving: a licence is the minimum requirement to demonstrate that an individual has an understanding of how a vehicle works, can recognise road signs, properly complete several manoeuvres, and drive in a variety of conditions confidently. A driving licence is obtained only once an examiner certifies that an individual can do all of the above; this does not, however, automatically translate into being a good driver. This is just the beginning and drivers' competence continuously increases with further time and practice.

1.5 Privacy fatigue

Rationally, it is clear that data protection is important. However, maintaining one's privacy is not straightforward: there are three phenomena involved in understanding data protection. Privacy fatigue is a consequence of the privacy paradox that individuals encounter after performing a privacy calculus.

When asked to share personal data, for example when signing up to a new service, users evaluate whether to do so. They are performing a privacy calculus, a cost-benefit analysis weighing the usefulness of the service against the risk of sharing personal data. There are two competing (and conflicting) factors: on one hand, the desire to participate in online activities and benefit from the resources the Internet offers; on the other, the knowledge that one's privacy is at risk when performing any action on the Internet.

[38] Clay A. Johnson, *The Information Diet: A Case for Conscious Consumption* (O'Reilly, 2012), 80.
[39] Beaulieu and Leonelli, *Data and Society*, 216.

Individuals make the conscious choice to engage with services that do not respect their privacy because the benefits they might receive (such as "personalization, convenience, economic benefits and social advantages"[40]) outweigh the consequent risk of disclosing personal information. Though logical, a rational perspective does not consider factors such as "time constraints, time inconsistency, immediate gratification and optimistic bias"[41] are also involved and may subconsciously push users towards lower levels of privacy protection. For example, a non-trivial role is played by the complexity of privacy policies: the effort (and time) required to read and fully understand them is often considered disproportionate.[42] A survey found that US citizens believe privacy policies are "just something they have to get past in order to use a product or service"[43] and for this reason can be considered "ineffective for communicating how companies use people's data".[44] This occurs despite users stating that they value their privacy: EU citizens report being concerned about how their personal data is used online. This includes "the use of personal data and information by companies or public administrations"[45] but also "cyber-attacks and cybercrime such as theft or abuse of personal data, ransomware (malicious software) or phishing".[46]

The inconsistencies in this cost-benefit analysis may result in what is known as the privacy paradox. This term refers to the discrepancy between what users state their privacy concerns are and the actual behaviour they engage in. Though people generally state, as described above, that they value privacy and are interested in taking steps to protect their personal information, this does not always occur. Research indicates that this happens because of the difference between disclosure intention and behaviour, and in the end it seems that "risk consequences [...] are not strong enough to influence the actual disclosure behavior."[47] After all, online privacy risks and benefits are abstract

[40] Susanne Barth and Menno D.T. De Jong, 'The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review', *Telematics and Informatics* 34, no. 7 (November 2017): 1045, https://doi.org/10.1016/j.tele.2017.04.013.

[41] Barth and De Jong, 1045.

[42] Barth and De Jong, 'The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review'.

[43] Colleen McClain et al., 'How Americans View Data Privacy' (Pew Research Center, 18 October 2023), https://www.pewresearch.org/internet/2023/10/18/how-americans-protect-their-online-data/.

[44] McClain et al.

[45] European Commission. Directorate General for Communications Networks, Content and Technology. and Kantar., *Digital Rights and Principles: Report* (LU: Publications Office, 2021), 14, https://data.europa.eu/doi/10.2759/30275.

[46] European Commission. Directorate General for Communications Networks, Content and Technology. and Kantar., 14.

[47] Aristea M. Zafeiropoulou et al., 'Unpicking the Privacy Paradox: Can Structuration Theory Help to Explain Location-Based Privacy Decisions?', in *Proceedings of the 5th Annual ACM Web Science Conference* (WebSci '13: Web Science 2013, Paris France: ACM, 2013), 464, https://doi.org/10.1145/2464464.2464503.

concepts with distant effects, while disclosing personal data mostly results in tangible and immediate benefits.[48]

How these two concepts work in real-life situations is seen in relation to social network sites (SNS). Though users have little trust in SNS (mainly Facebook and Instagram, but WhatsApp too), they still choose to engage with them, primarily due to their social benefits. The rational privacy calculus points out shortcomings in terms of user privacy, but biases and heuristics influence users' choice to engage with these platforms is paradoxical. However, the benefit of being able to keep in contact with friends and family, building and maintaining other relationships, and perhaps fear of missing out (known as FOMO) overrides concerns and results in paradoxical behaviour.

Individuals' paradoxical behaviour can be partly explained through the concept of privacy fatigue, defined as "a psychological state of tiredness with the issue of online privacy".[49] Fatigue occurs in situations where individuals face high demands and cannot complete their goals, thus reducing their efforts and disengaging with the task rather than looking for a solution.[50] In the case of privacy fatigue, as the number of digital services increases and so does the effort required to maintain one's information private, users believe there is no valid way of controlling their data. This results in individuals reducing their efforts to protect their data and the attention on privacy issues.

Users who are concerned about privacy engage in a number of behaviours to limit the information that online vendors collect: opting out of data collection mechanisms; avoiding websites and apps with bad data practices; submitting false information. But this means spending large amounts of time and effort, and can be overwhelming. Another way of coping with this concern is to avoid stress by disengaging. Users who experience fatigue minimise their efforts: they tend to accept the default privacy options (instead of personalising them), which results in disclosing more personal information.[51]

Privacy fatigue can be exemplified by several behaviours. Users typically accept privacy policies without reading them: these documents are long and unclear,[52] which means that only highly

---

[48] Barth and De Jong, 'The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review', 1048.

[49] Hanbyul Choi, Jonghwa Park, and Yoonhyuk Jung, 'The Role of Privacy Fatigue in Online Privacy Behavior', *Computers in Human Behavior* 81 (April 2018): 42–51, https://doi.org/10.1016/j.chb.2017.12.001.

[50] Choi, Park, and Jung.

[51] Choi, Park, and Jung.

[52] Kevin Litman-Navarro, 'Opinion | We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.', *The New York Times*, 12 June 2019, sec. Opinion, https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html, https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html.

concerned (and thus motivated) individuals read them. For example, 56% of US citizens always or almost always accept privacy policies without reading them; by comparison, only 18% reported never accepting privacy policies without reading them first.[53] This is because privacy policies are long and full of technical vocabulary; an analysis carried out by the New York Times found that the difficulty level of most privacy policies is college level or higher,[54] which makes them unreadable for the average individual. Reading through these complex documents can feel useless since they must be accepted for services to be accessed by users, so the latter reduce any effort which interferes with reaching their objective.

Passwords are also a valid example. Since they are the most common and basic authentication method, the average number of passwords a person has is 168.[55] To ensure an account is properly protected, a strong password should be used. The European Union Agency for Cybersecurity lists several recommendations:

> 1. Passwords are secrets. Keep them so.
>
> 2. Mix the kind of characters in your passwords.
>
> 3. Use long passwords. Any windows password up to 9 characters can be cracked in seconds using public-dzw3somain tools. The longer the password, the longer it will take for an attacker to crack it. Every added characters [*sic*] increases the cracking time by orders of magnitude. Any password that is not a common word, and is longer than 14 characters cannot be cracked with current computing means.
>
> 4. Use different passwords for different purposes or web sites. That way, even if someone manages to learn or crack one of your passwords, it does not give them immediate access to your other services.
>
> 5. Use a password manager to create and remember random passwords.
>
> 6. If a random password is impractical, use a pass phrase instead.[56]

Following these recommendations create complex passwords that are secure yet hard to remember. Given the important role passwords have, it is no wonder that users report feeling overwhelmed by the number of passwords they have to keep track of.[57] To avoid these issues,

[53] McClain et al., 'How Americans View Data Privacy'.

[54] Litman-Navarro, 'Opinion | We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.'

[55] 'How Many Passwords Does the Average Person Have?', NordPass, accessed 27 August 2024, https://nordpass.com/blog/how-many-passwords-does-average-person-have/.

[56] 'Authentication Methods', Page, ENISA, accessed 27 August 2024, https://www.enisa.europa.eu/topics/incident-response/glossary/authentication-methods.

[57] McClain et al., 'How Americans View Data Privacy'.

individuals often opt for much simpler passwords which can be deciphered in less than a second.[58] This means that passwords are also, potentially, the weakest link in the security chain. To limit these issues, people can use password managers, programs which keep track of passwords. Though these programs help solve a concrete issue, only 32% of Americans report using them.[59] Fatigue is at play all throughout this example: users would like to protect their privacy, but creating secure passwords requires effort that they simply do not believe is worth it. Many could reduce their password fatigue by using a password manager, however this is also viewed as a major effort to avoid.

1.6 Conclusion

This chapter has provided an overview of the core issues of this thesis. Technological advancements have permeated society and, in the process, everyday life has become datafied. Our data is a highly sought-after resource which service providers extract while disregarding privacy. In this scenario, digital citizens must be digitally literate, possessing the understanding and skills to harness the Internet's numerous benefits while mitigating the risks arising from it. They must also be literate in data protection in order to be able to choose how their data is used and shared. Taking into consideration (and putting in practice) the numerous measures and tools which ensure an adequate protection of personal data, however, may be overwhelming. This can result in privacy fatigue, the phenomenon where users give up on protecting their information because they believe it is too difficult and requires too much effort. The following chapter describes the methodological process used to investigate the relationship between data protection literacy and privacy fatigue.

[58] For examples, see 'Top 200 Most Common Passwords', NordPass, accessed 28 August 2024, https://nordpass.com/most-common-passwords-list/.

[59] McClain et al., 'How Americans View Data Privacy'.

2. Methodology

2.1 Introduction

This chapter outlines the methodological framework used to investigate how privacy fatigue impacts digital citizens' data protection literacy. It starts by looking at the research objectives, and then details how the questionnaire was built in order to achieve a holistic understanding of the factors at play in individuals' behaviours and attitudes. Carefully selecting items was a crucial step in validly and reliably measuring literacy, fatigue and concern: for this reason, items were mainly sourced from existing literature to ensure they were tested and reliable. At the same time, a limited number of items were used, as brevity was identified as a way of minimising respondent disengagement, and principles from the GDPR were incorporated, to ensure that respondents' data and rights were respected. The data analysis process is then described: cleaning the data was the first crucial step which allowed a thorough analysis to be carried out. The last section analyses the sample's demographic data, highlighting relevant information while leaving more in-depth observations and cross-analyses to be covered in the following chapter.

2.2 Research objectives and expectations

The aim of the questionnaire is to gain insight into how privacy fatigue and data protection literacy affect digital citizens, thus understanding how this has an impact on their behaviour. Specifically, the study aims to assess participants' thoughts on privacy and its defining elements, and identify whether their data protection literacy and feelings of being overwhelmed by privacy concerns correlate. In doing so, the study looks to find patterns in demographic markers, providing insight into how gender, age, education level, occupation and Internet usage per day may influence and determine perceptions and preoccupations in this field.

Findings are expected to contribute to a deeper understanding of individuals' thoughts regarding privacy, particularly what variables may determine increased disengagement or lower fatigue.

2.3 Building the questionnaire

To ensure the questionnaire effectively captured the nuances of the topics, a systematic approach was taken in its construction, with several factors being kept in mind.

The questionnaire was designed to be brief due to the voluntary nature of participation and lack of compensation for participants. An in-depth questionnaire, while resulting in highly detailed answers, would have been perceived as requiring excessive effort and likely caused participant fatigue and subsequent disengagement. Brevity was identified as the best technique to prevent disengagement. The questionnaire consists of 25 questions, divided into five sections: demographic questions, introductory privacy questions, privacy concern, data protection literacy, and privacy fatigue. Since all questions were mandatory, the questions were arranged to limit respondent fatigue.

Demographic questions are needed to understand who participants are. These questions were placed first as a sort of "tutorial": respondents gained familiarity with the form while answering simple, straightforward questions. This section included both typical demographic questions (age, gender, education level, and occupation status) and two additional Internet usage questions, specific to this field of study. Like traditional demographic items, Internet usage questions are crucial as they enable an understanding of the diversity of respondents and allow the identification of patterns, including within subgroups.

After completing demographic questions, a 3-question introductory section evaluated participants' understanding of privacy and data protection. The first was an open question where participants were asked to explain which aspect of privacy was the most important for them; the objective was to understand participants' thoughts without bias induced from questionnaire items. Open questions require a more complex and involved intricate qualitative analysis because respondents are not limited to a pre-defined set of possible answers; instead, they are allowed to answer freely which ensures their unfiltered thoughts can be studied. This question was placed directly after the demographic items, so that individuals engaged with the answer requiring the most thought and effort at the beginning, when their attention was still at its peak. Furthermore, by making this the first question respondents encountered, influence from the contents of other survey items was avoided.

The following two items evaluated understanding of what personal data are and how to manage them. Q2 drew inspiration from the "What is personal data?" page on the EU Commission's website,[60] with additional options based on CJEU case law.[61] Q3, instead was an item contained in

[60] European Commission, 'What Is Personal Data?', European Commission, accessed 25 August 2024, https://commission.europa.eu/law/law-topic/data-protection/reform/what-personal-data_en.

[61] For some examples, see 'Article 4 GDPR', GDPRhub, accessed 25 August 2024, https://gdprhub.eu/index.php?title=Article_4_GDPR.

Eurostat's 2023 ICT Household Survey.[62] In both cases, participants were presented a list of possible answers, from which they could select as many as they thought applied. These answers were placed after the open question to avoid giving participants any hints or ideas. The remaining questions featured the same type of answers as to limit possible confusion arising from having to repeatedly adjust to new answer formats.

Privacy concern and privacy fatigue items were sourced from Choi et al. (2018), to ensure that this work was built on and referenced well-established literature. Only relevant items were selected to ensure that the survey's brevity did not limit its effectiveness. Answers for these items were on a five-level Likert scale (strongly disagree; disagree; neither agree nor disagree; agree; strongly agree).

The same process could not be done for items on data protection literacy, as no existing questionnaire measured the variable as was needed for this work: some were focused entirely on users being able to carry out specific actions, while others tested broad theoretical knowledge. To overcome this issue, several European Union resources and guidelines were used, particularly the European Commission's Digital Competence Framework for Citizens (also known as DigComp). These resources helped form questions that would properly identify participants' data protection literacy. Following the approach highlighted by literature in Chapter 1, questions were written to ensure that participants' declarative ("I know that") and procedural knowledge ("I know how") regarding data protection were both tested. Items in this section used a five-level Likert scale, as was done for the privacy concern and privacy fatigue items.

Choosing which language to use for the questionnaire was a pivotal issue. Though this work is in English, the survey would be most likely completed by Italians. To ensure clarity and accessibility, Italian was chosen as the survey's language as respondents would likely be more comfortable and able to better articulate their thoughts. Existing questions were translated into Italian, carefully evaluating possible meanings and connotations in order not to distort the meaning or undertones of any words. The questionnaire items and their translations can be found in appendix B. As for original items, these were written with concise and straightforward language, limiting technical terms.

The survey was created and administered online using *tally.so*, a free no-code form builder. The platform was chosen over more well-known alternatives (such as Google Forms, Microsoft

[62] 'ICT Usage in Households and by Individuals (Isoc_i)', Eurostat, accessed 25 August 2024, https://ec.europa.eu/eurostat/cache/metadata/en/isoc_i_esms.htm.

Forms or Typeform) due to its excellent price-feature offer, its simple and pleasant graphical interface, and its availability in Italian. Looking at how form builders addressed data protection was also an important element due to the topic of this thesis. Tally.so has a page on its website explaining how data protection works, and also provides a brief overview on how to create a GDPR compliant form.[63]

2.3.1 Data protection

To ensure that the data collected from the survey was handled properly, a number of steps were taken, following the principles contained in article 4 of the General Data Protection Regulation (GDPR). The GDPR is not strictly applicable to this research: firstly, it is aimed at enterprises who process personal data (recital 156 specifies guidelines for research); secondly, all data collected for the purposes of this research was anonymous from the outset, with no possibility of linking answers to their authors. This approach remains useful in enabling a critical analysis of data collection and management, which is extremely relevant in the writing of this thesis. Furthermore, as the survey creator is also the controller of respondent data, taking these principles into consideration when creating the survey helped verify that data was being handled correctly and no missteps were made.

1. Lawfulness, fairness and transparency

   Before participants could fill out the survey, they were presented with an "informed consent" page to ensure they were aware of what the survey consisted of and how their data would be collected and used. To ensure processing followed the above principles, participants were asked to provide their consent to participation, and were then allowed access to the survey.

2. Purpose limitation

   To implement purpose limitation, the "informed consent" page clearly defined the objectives of data collection, which included research purposes only.

3. Data minimisation

   The survey has been designed to minimise data by only collecting the information required for research. The survey also does not register any personally identifiable information such as name and surname, email address, or IP address. This ensures that data is anonymous and does not permit the identification of the participants, as outlined in Recital 156 of the GDPR.

[63] 'How to Create a GDPR Compliant Form', accessed 22 September 2024, https://tally.so/help/how-to-create-a-gdpr-compliant-form.

4. Accuracy

   No personally identifiable information was collected. In any case, users were provided with the researcher's email should they have wished to amend any of their answers.

5. Storage limitation

   No personally identifiable information was collected through this survey. Since data was anonymous, the storage limitation principle does not apply to this survey's data.

6. Integrity and confidentiality

   There were no particular risks to users for the loss and misuse of their survey data since personally identifiable information is not collected, as explained above.

7. Accountability

   To ensure the researcher was accountable for the survey, an email address was featured on the informed consent page. Participants could direct any and all inquiries to this address, including requests to have their information removed.

2.4 Analysing data

Tally.so provides a graphical interface which summarises survey results, enabling a quick understanding of answers. To perform more complex analysis, however, tools built for this purpose are needed. To do this, survey data was downloaded from tally.so in a .csv format and then imported into Google Sheets.

2.4.1 Data clean up

The data was cleaned up before proceeding with analysis. This was a crucial phase because it ensured that only reliable data was taken into consideration, providing a baseline for the overall quality of the work. The .csv file contained data which was not required for analysis, such as submission ID numbers or time of submission, which were promptly deleted.

Looking through the 232 submissions, one stood out: a participant described themselves as a non-binary individual over the age of 65, who was a sex worker and had begun using the Internet when they were 14. Despite having completed the survey, the participant seemed to have purposefully provided false information and thus their answers could not be taken at face value; their submission was excluded. After the cleaning phase, the total number of submissions was 231.

In some cases, data needed reorganisation. One of such cases was employment status, as 16 participants had selected “other status” and entered custom answers which needed to be sorted into the existing categories. The custom answers entered for the “other employment status” option were the following:

Custom answer	Number of participants
Other status	4
Consultancy and didactic activities	1
Lawyer	1
Teacher	1
University professor	1
Entrepreneur	3
Freelance professional	4
Retired / university professor	1

Table 1. Frequency distribution of custom answers for the employment status demographic question.

Most of these self-reported statuses were in line with existing categories, so the answers were sorted into the appropriate categories. The four participants who did not disclose their employment status were grouped under the “other status” label. The data was reorganised as shown below:

Status	Number of participants
Student	35
Employed	134
Unemployed, looking for a job	3
Homemaker	2
Retired	53
Other status	4

Table 2. Frequency distribution of employment status, after the custom answers were sorted into the existing categories.

The “age of first Internet use” item asked respondents to input their age in a numbers-only field. Eight individuals provided answers in an unexpected format: two used decimal numbers (e.g. 9.1) and six wrote their birth year (e.g. 1986). For the former, the age was rounded to the closest integer (e.g. 9.1 was rounded down to 9). The age of the latter could not be calculated, thus making answers from 6 participants unusable; data from 225 participants was used for this question.

2.4.2 Data analysis

Analysis was carried out in two phases. During the first phase, all submissions were counted and formatted into tables for easy reference and use. Tally.so did provide a summary of results but did not allow export; this was insufficient as data had to be organised in a way that could be analysed and handled later. For most cases, the count.if spreadsheet function was used to count values precisely and limit human mistakes. Pivot tables were used to allow different variables to be measured against one another. More specifically, each item was measured against demographic data to evaluate whether any patterns existed. After this, items from the data protection literacy and privacy fatigue sections were measured against one another to investigate whether (and how) the two phenomena affected one another.

For most items, this was enough as the measurements were quantitative in nature. The open-ended nature of Q1, however, required a qualitative analysis which would code answers in order to identify patterns and trends. This analysis occurred in several steps: first, responses were accurately read to have a general understanding of the contents; respondents' statements were then sorted into temporary categories; finally, statements were sorted a second time to ensure they were in the correct categories and could also be grouped into subcategories. Each statement could only belong to a single category, despite statements often mentioning multiple issues. In these cases, to avoid researchers' preferences or thoughts from influencing results, a simple criterion was used: the issue taken into consideration was the first mentioned, unless the respondent explicitly said another issue was their main concern. Through this process, statements were sorted into one of five categories: types of data, control over use, security, context, and other; within each, subcategories were identified to ensure all nuances could be described.

The second phase of data analysis focused on finding patterns and correlations. Using the data formatted in the first phase, this analysis was carried out using pivot tables as they allow easy and quick comparisons. By selecting variables for rows and columns, data was rapidly and precisely aggregated, enabling comparisons across demographic indicators and items. This flexibility and ease of use facilitated a thorough analysis, which will be discussed in the following chapter.

Due to the sample population being neither representative nor balanced, data was analysed by comparing percentages for each demographic marker rather than absolute values. With 139 respondents being male and only 91 female, relying solely on absolute values would result in skewed results. An example may give a clearer idea: 70 respondents account for 30.30% of the total

population; the same number of respondents also accounts for three quarters (76.92%) of the female population, versus just half (50.36%) of the male population. These percentages paint a clearer, more nuanced picture given the unrepresentative sample.

It is also important to note that two demographic categories had an insufficient size, meaning conclusions could not be drawn based on them. This was the case for the non-binary category and the under 18 category, both of which contain just a single individual.

Finally, to answer the research question and verify whether there was any relationship between data protection literacy and privacy fatigue, a qualitative statistical analysis was carried out. Choi et al. (2018), whom this work drew inspiration from, performed a quantitative analysis by converting Likert scale items into values. While replicating their work would offer valuable insight, this option had to be discarded as the exact values attributed to each item on the scale were unknown. Guessing how these items were valued would have likely resulted in an imperfect comparison with this work and thus invalid conclusions. For this reason and due to the qualitative nature of the questionnaire itself, a qualitative analysis was preferred. Free statistical software "R" was used to analyse the data; despite having a steep learning curve, using a dedicated program meant that statistical analysis was much easier to perform than using spreadsheet software.

The chi squared (χ^2) test was used to determine whether there was an association between questions in the data protection literacy and privacy fatigue sections. This test was used to verify whether a relationship between literacy and fatigue was present or not; in addition to this, it also assessed its statistical significance. After the null hypothesis (H_0) and alternative hypothesis (H_a) were formed, the data was evaluated for a significance level $\alpha = 0.01$ and degrees of freedom = 16, where the χ^2 critical value is 26.30. Only the pairs of questions which were significant ($\alpha < 0.01$) were analysed using a marginal analysis. To do this, a pivot table was used to display the percentage of selections for two responses within cells. These values were compared with the column's marginal distribution in percentage and, if the cell value was greater than the marginal value, an association was found.

2.5 Sample description

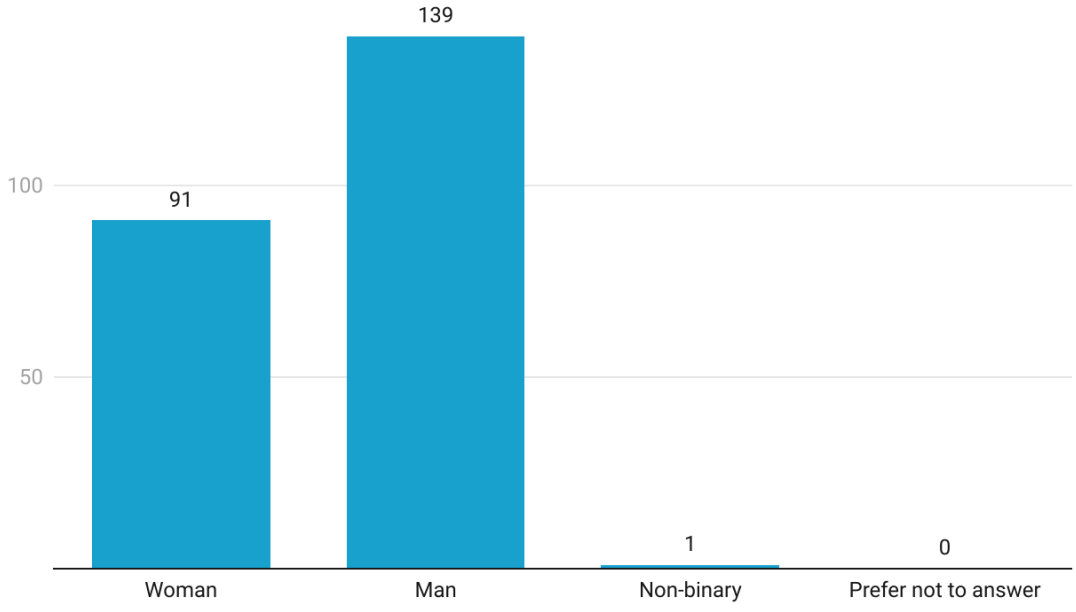
The survey was administered over the course of a three-week period, from the 25th of July 2024 to the 15th of August 2024.

Representativeness within the Italian population or within smaller, more specific populations was not a requirement, since the objective of this questionnaire was to broadly understand the *status quo* of data protection literacy and levels of concern and fatigue.

The questionnaire was spread mainly through messaging apps and social media, with individuals voluntarily taking part. The former was a more direct manner, which helped increase the response rate; social media allowed wider reach but did not ensure high engagement due to short attention spans associated with social media use. A total of 231 responses were analysed.

2.5.1 Gender

No participants refused to disclose their gender. The results were imbalanced: participants were predominantly male (60.17%), with women accounting for just under 40% and just one respondent identifying as non-binary (accounting for 0.43% of participants). This unevenness is likely the result of the distribution method which may result in conclusions drawn on gender being unreliable. The graph below features the distribution of participants' gender.

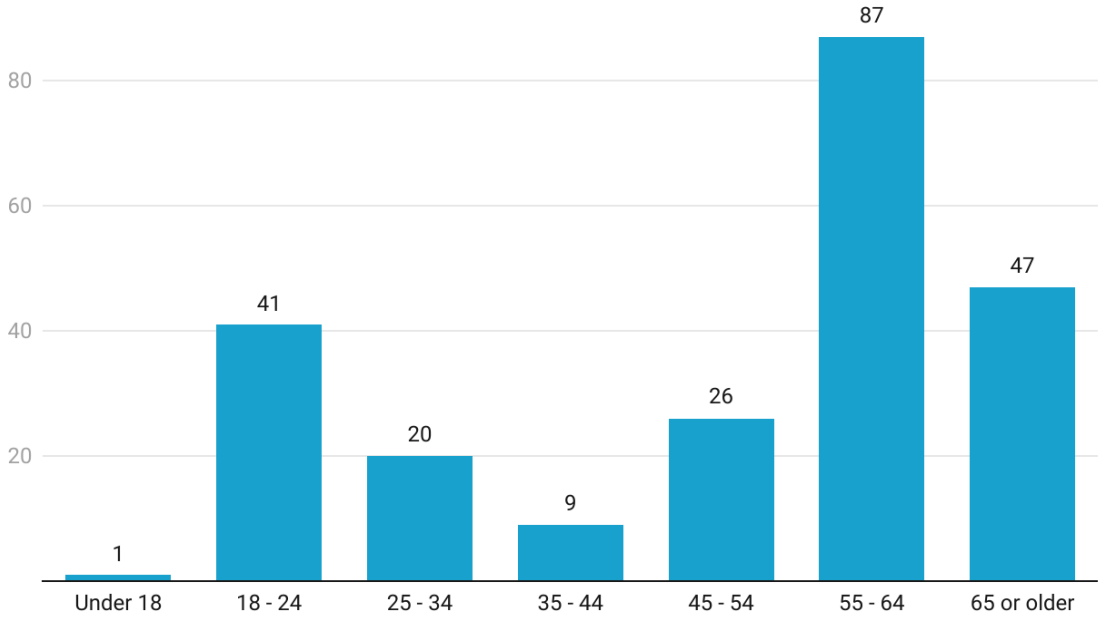


Graph 1. Frequency distribution of participants' gender.

2.5.2 Age

The below graph shows the age of participants. As shown below, most were in the 55 – 64 and 65 or older categories, with these accounting for more than 50% of participants. This uneven distribution may once again be explained by the survey distribution method; however, the large

number of individuals aged 55 or over (compared to younger age groups) is consistent, to an extent, with Italian demography.[64]



Graph 2. Frequency distribution of participants' age.

2.5.3 Education level

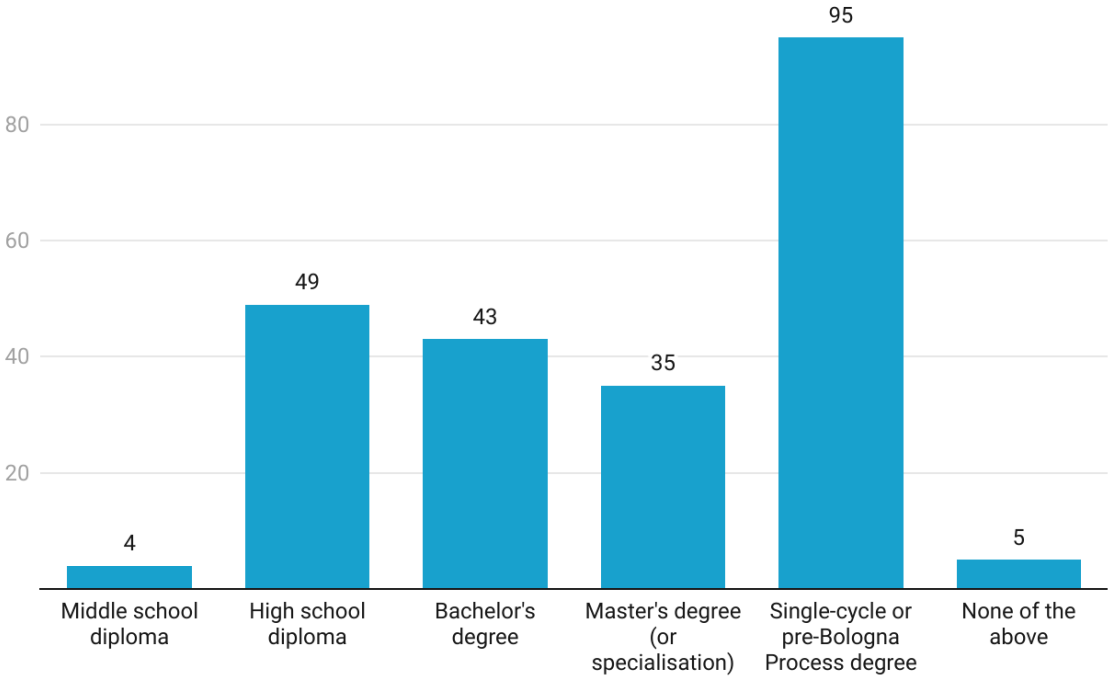
The education level of the sample was high, with almost three in four participants holding a university degree. Of these, 41.13% of participants had a single-cycle or pre-Bologna Process degree,[65] which is coherent with the age distribution of the sample; 33.77% held a Bachelor's degree or higher. This sample has a high education level and this must be taken into consideration when analysing results.

Four participants had solely graduated middle school: currently, this means not having completed compulsory education, however, all respondents with this education level were 55 or older, when compulsory education was completed upon earning a middle school diploma. A further five participants stated their highest education level was not among those listed here; one of these participants listed their occupation as "university professor", which likely means they possess a PhD, and thus their education level was higher than the options presented in the questionnaire.

[64] For an overview of Italian demography, see 'Sette grafici per capire la crisi demografica in Italia', *Pagella Politica* (blog), 12 December 2022, https://pagellapolitica.it/articoli/crisi-demografica-italia. and 'Italy', in *The World Factbook* (Central Intelligence Agency, 13 August 2024), https://www.cia.gov/the-world-factbook/countries/italy/#people-and-society.

[65] Italy's university system was reformed after its signature of the Bologna Accord in 1999.

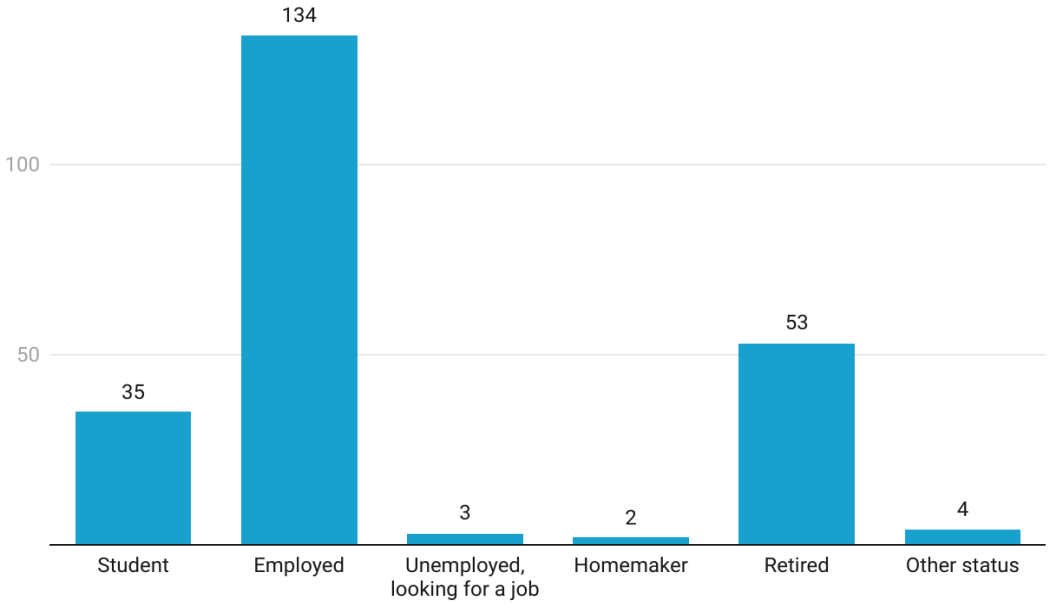
This uneven distribution may skew results: overall, participants had a high education level, which might indicate a higher data protection literacy level.



Graph 3. Frequency distribution of the highest educational qualification achieved by participants.

2.5.4 Employment status

Employment status was consistent with the ages and education levels described above. The largest category was “employed”, with just over half of respondents (58.01%) reporting this status. Retirees were the second most selected category at 22.94% and students accounted for 10.82% of respondents. The graph below shows employment status after the data was reorganised.



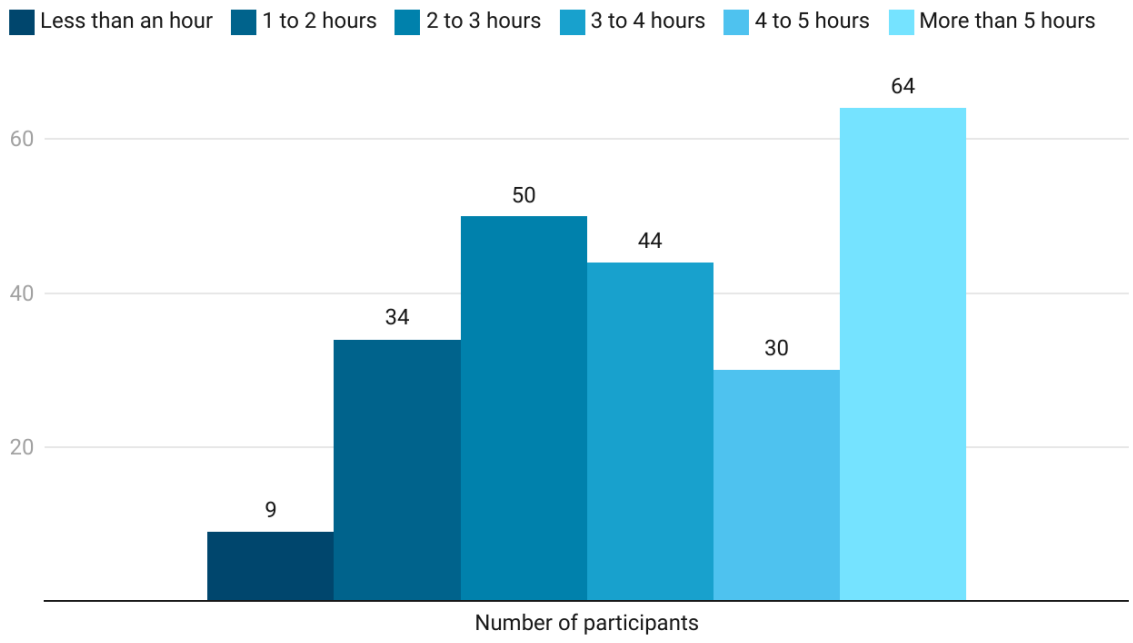
Graph 4. Frequency distribution of participants' employment status.

2.5.5 Internet usage per day

Participants reported varied Internet usage patterns. “5 hours or more” was the most popular option (selected by 27.71% of participants), closely followed by “2 to 3 hours” (21.65%). Given the pervasiveness of the Internet both in the workplace and for leisure activities, the popularity of this lower category is somewhat surprising.

Respondents aged 18 – 24 reported the highest time spent on the Internet per day, with 63,41% (26 participants) spending 4 hours or more. Though more participants (31) in the 55 – 64 category reported spending the same amount of time, this only accounted for 35.63% of individuals in this category. Just 3.90% of overall respondents reported using the Internet less than one hour a day; predictably, these individuals were all 55 or older.

These findings indicate that Internet usage is influenced by age. However, this may also be linked to occupation: 70.73% of individuals aged 18 – 24 indicated they were students or unemployed, which implies having more leisure time compared to those who are employed.



Graph 5. Frequency distribution of time spent by participants on the Internet daily.

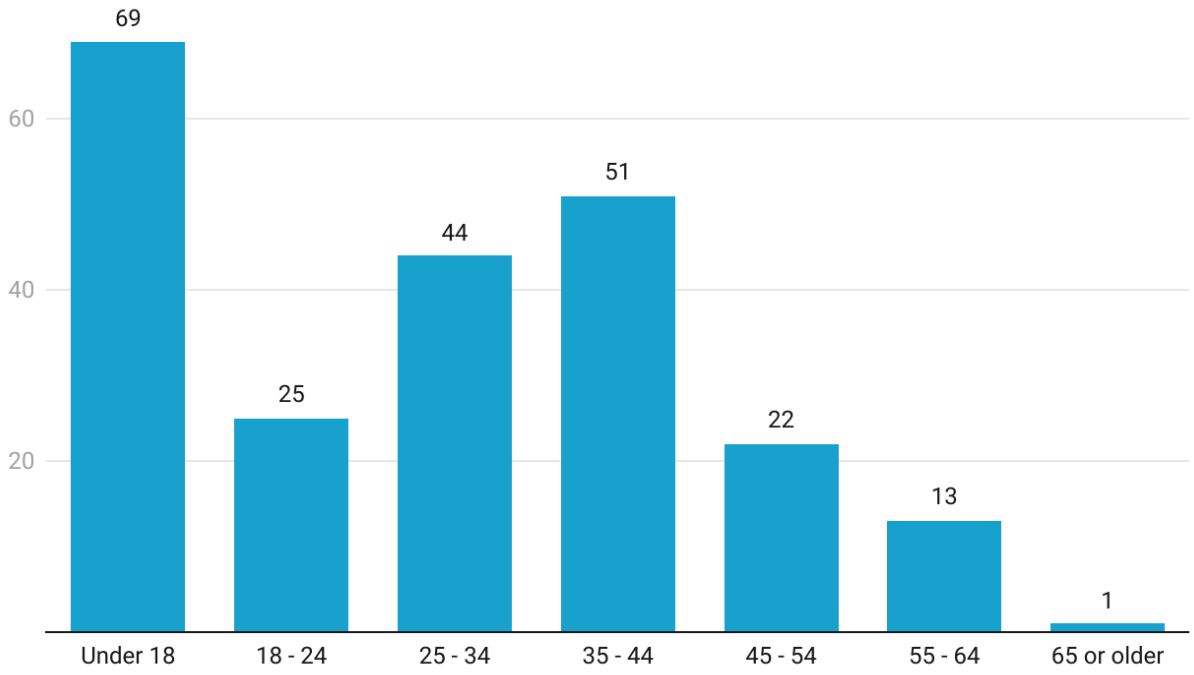
2.5.6 Age of first Internet use

As explained in section 2.4.1, only 225 responses were used for this item as six respondents wrote down the year they first used the Internet in rather than their age at the time. The table below features the ages grouped into the same age categories as those in the demographic questions.

Age category	Number of participants
Under 18	69
18 – 24	25
25 – 34	44
35 – 44	51
45 – 54	22
55 – 64	13
65 or older	1

Table 3. Frequency distribution of participants' age of first Internet use.

The chart below, instead, is a frequency chart of participants' ages.



Graph 6. Frequency distribution of participants' age of first Internet use.

What emerges from this data matches the previous demographic information. Taking into consideration the age of respondents and that the World Wide Web became accessible and popular in the mid-90s, it makes sense that most participants reported first using the Internet in the 25 – 44 age range. There are a few exceptions, perhaps individuals who used the Internet for research purposes, before it was available to the general public. Respondents born in the 1990s and early 2000s (who are now in the 18 – 34 range) had access to the Internet at a much earlier age.

2.6 Conclusion

This chapter has described the methodological process used to investigate the research question. The importance of the design phase cannot be understated: any additional attention in this phase contributes to ensuring results are valid. Accurately selecting and phrasing questions, for example, ensured usefulness and clarity; similarly, translating items into Italian required carefulness to verify that the meaning was not distorted. The same is valid for the extensive focus on the data analysis methodology: clearly defining how data would be analysed ensured the results could be effectively interpreted and, consequently, discussed. Finally, the description of the sample population provides the context needed for the analysis of the data collected through this questionnaire, discussed in Chapter 3.

3. Data analysis

3.1 Introduction

This chapter presents the findings deriving from the data collected through the survey described in the previous chapter. The first paragraph analyses the results from Q1, an open question aimed at understanding respondents' thoughts regarding privacy. Paragraphs 3.3 to 3.5 systematically analyse each section of the questionnaire by describing the data, highlighting variations linked to demographic data and making general observations. Paragraph 3.6 examines key variables related to data protection literacy, privacy concern and privacy fatigue. The last paragraph looks at the statistical significance of the data protection literacy and privacy fatigue sections. This chapter focuses on describing the results, while the following will discuss them in relation to broader discourse, reflecting on issues and challenges for data protection literacy and privacy studies.

3.2 Understanding privacy

3.2.1 What aspect of privacy do you believe is most important to you?

Q1 was an open question aimed at understanding what aspect of privacy respondents believed to be most relevant and important to them, without limiting them to a list of predetermined answers. As explained in Chapter 2, participants' answers were grouped into 5 categories (types of data; control over use; context; security-related issues; other) and several subcategories. Results are listed in the table below.

Category	Subcategory	Number of participants
Types of data		**84**
	Personal data	39
	Sensitive data	17
	Financial data	8
	Various	8
	Navigation data	7
	Photos	5

Category	Subcategory	Number of participants
Control over use		**67**
	Processing	26
	Advertising	14
	Choice	10
	Profiling	8
	Transparency	5
	Limitation	4
Context		**24**
	Private life	12
	Family	4
	Reputation	3
	Combination	5
Security		**20**
	General	7
	Fraud/crime	7
	Transmission	6
Other		**36**
	Privacy	7
	Anonymity	7
	Everything	5
	Rights	4
	Respect	4
	Freedom	3
	Simplification	2
	Don’t know, care, understand	4

Table 4. Frequency distribution of answers to Q1.

Most participants’ answers included specific references to “types of data” (36.36% of answers belonged to this category); references to “personal data” and “sensitive data” were the most common.

This is not surprising as these concepts have become general knowledge; however, these answers were also vague, and in some cases it was unclear what data types individuals believed to fall under these broad umbrella terms, though answers from Q2 may help clear this up. Some individuals mentioned specific data types such as phone number, email address, photos, and localisation data. Interestingly, sensitive data types such as political opinions, religious beliefs, and sexual orientation were only mentioned by five participants.

"Control over use" was the second most popular, with 29% of answers falling into this category. Participants expressed concern over data processing: examples include data being shared without permission, sold to third parties or being tracked. Advertising was also popular, with spam and robocalls being mentioned over any other kind of advertising; this can be easily explained by the pervasiveness of telemarketing calls in Italy in recent times. In this category, the most interesting subcategory was "choice": ten participants stressed that their concern was "being able to choose what to show and what to hide", "[...] whom, eventually, to share my data with".

Roughly 10% of participants mentioned contexts as being the most important aspect of privacy. For most, this meant their private life and family; the "combined" subcategory includes answers which listed multiple contexts, including financial and sanitary. Most contexts were personal ones, which makes sense as those are the ones which require protection from the outside. The presence of the "reputation" category was unexpected, though the need to be favourably viewed by others is a core element of human interaction.

To others, security mattered most. Results here were spread fairly evenly between general reference to security ("I don't want my data to be exposed"), desire not to be involved in frauds or crime, and worry over data being secure when being transmitted or when activities are being carried out.

"Other" was a catch-all category born from the need to collect answers that did not belong to a precise category but belonged to identifiable clusters. This is perhaps the most interesting section, as a variety of conceptions of privacy are present. The two largest sub-categories are "privacy" and "anonymity". The first was slightly redundant, with answers stressing protection and discretion: for example, a participant stated they were most interested in "the actual protection of privacy", and other stating "I don't want anyone to mind my business". The presence of a well-defined "anonymity" subcategory suggests that individuals are apprehensive about being identified while on the Internet,

though respondents made no statements on who they might not want to be identified by or what data they would prefer to hide.

The "everything" sub-category contains answers where individuals state they believe everything to be important; this is quite vague and can be interpreted in two main ways. These participants are either extremely privacy conscious and genuinely care about many issues, or are normal individuals who are unsure about which of the many aspects of privacy they feel apply to them, and thus simply believe that "everything" is important to them.

Only four participants stated they believed rights to be the most important aspect of privacy for them: one individual mentioned a specific regulation (GDPR), while two individuals mentioned the right to be forgotten, and another stated "right to personally access to archived data, right to be forgotten, disclosure of who holds what data, transparency about data transfers, how to defend intellectual property". Though very few respondents chose to mention rights, the last answer exemplifies how a legal approach goes beyond simply technical aspects and tackles a variety of issues.

"Respect" was a theme in four participants' answers. The link between privacy and respect may seem weak but can be found in specific contexts, for example in relation to cyberbullying or, more broadly, on social media. Respect means recognising the existence of boundaries and respecting them, in particular with reference to others' personal information. As an answer in this subcategory points out, differences in opinions are not a valid reason for sharing/exposing personal information.

"Freedom" was the most important aspect for three participants; admittedly, these answers may overlap, to an extent, with the "choice" subcategory within "control over use". This subcategory, however, lacks the reference to data use and rather reflects individuals' desire to do as they please.

The subcategory closest to this thesis' topic, "simplification", was mentioned only by two participants. These answers implied that maintaining privacy is difficult and may be indirectly mentioning fatigue.

The "don't know, care, understand" category collects answers which did not provide a meaningful perspective to this topic. This is true for all but one answer, where a participant stated "I don't care what data companies have as long as they're not public domain".

These categorisations do not correlate strongly with demographic data, however there are certain elements worthy of note. Women stated that "types of data" was the aspect of privacy most important to them; in proportion, this was more than men, who instead were split fairly evenly between "types of data" and "control over use". Similarly, 18 – 24 year olds' answers also highlighted

"types of data" as the most important aspect for them; results in higher age groups, instead, were spread across the five categories. As a consequence, occupation has a similar distribution, though not identical in numbers. Education level did not seem to highlight any particular tendencies. As for time spent on the Internet daily, the answers of those who spent five hours or more were split between "control over use" and "types of data", with a remaining 26% for the other categories (answers were similar for those who spent 1 – 2 hours and 3 – 4 online daily).

3.2.2 Which of these do you consider to be personal data?

Q2 asked participants to select all the data objects they believed to be personal data. Banking data was the most popular option (selected by 93.97% of participants); this can be explained by the perceived (and actual) sensitivity of this kind of information. A similar reasoning can be made for the second and third most popular options, namely home address and health data. E-mail addresses have been given a relatively low importance; a possible explanation for this is that they are required when subscribing to services/apps, which has caused inboxes to become full of spam. If emails are already overrun with undesired messages, perhaps individuals feel that there is no need to further protect them.

Data	Number of selections
Name and surname	151
Home address	200
E-mail address	136
IP address	142
Identification number	130
Photos	162
Banking data	218
Location data	160
Biometric data	167
Health data	190
Other	10

Table 5. Frequency distribution of answers to Q2.

As mentioned in Chapter 2, the predetermined answers for this question were sourced from the EU Commission website. Most of the data types mentioned fall under the personal data umbrella, with the exception of biometric and health data, which are more specifically *sensitive* personal data. The only data which are not regarded as personal at all are photos.

These results indicate that people don't have an accurate idea of what personal data is. Only three participants selected all data except for "photos", 52 participants selected all answers except for "other", six responded by selecting all possible answers; all other answers were combinations of several types of data. When looking at Q1 and Q2 together, what emerges is that individuals care (or perhaps feel that they should) about their personal data, but do not know what this broad term refers to exactly. This may be due to lack of education which leads, in turn, to the term being conflated with others which sound similar, such as "personal information". It is also possible for different individuals' risk attitudes and perceptions to influence which data they believe to be personal or not.

3.2.3 Have you ever done one or more of the following activities to manage access to your personal information on the Internet?

The last question of this section (Q3) asked participants to list which activities they had carried out to manage their personal data on the Internet; the results are mostly unsurprising. The most popular actions were having refused to allow the use of personal data for advertising purposes (90.48%) and having restricted or refused access to their geographical location (89.18%). This is likely the consequence of cookie consent banners and location access requests being common and ubiquitous, often featuring clear language so that users easily understand the stakes. Similarly, the extreme popularity of social media and cloud services permits an intuitive explanation of why 73.16% of users have limited access to profile or content on social networking sites or shared online storage. The above practices are often cited in discussions about privacy and are highlighted as basic precautions one must take, and perhaps this is enough for individuals to become conscious about potential risks and changing their behaviour.

Just over half (58.44%) of participants had read privacy policy statements before providing personal data. While these documents are extremely common because they must be accepted to use online services, research has consistently shown that users typically do not read them. This is because privacy policies are long and verbose documents, requiring time and effort that users often do not have; this is further exacerbated by the fact that users have no choice but to accept privacy policies if

they want to use a given service. Unfortunately, the nature of this answer means that further information on this topic (e.g. how many privacy policies have been read; how often; whether individuals have refused to use a service because of its privacy policy) is not available.

Roughly half (51.08%) of respondents had checked that the website where they provided personal data was secure (e.g. https sites, safety logo or certificate). This is a higher percentage than expected, but still shows that half of respondents lack knowledge on security indicators and may be engaging in risky behaviour because of this.

Predictably, very few participants (12.12%) had asked providers to access the data they hold about them to update or delete it. This is the action which requires the most effort: first of all, individuals must be aware that EU citizens have this right; they must then identify who the data controller is, find their contact information and submit a data access request following the appropriate procedure (different services/websites might have different procedures). If the data controller does not respond within 30 days, individuals must follow up or contact their national data protection authority. Clearly, this procedure is not something most individuals would be aware of or, if they were, choose to engage in because of its complexity; this makes this answer's results overall unsurprising.

Action	Number of selections
Read privacy policy statements before providing personal data	135
Restricted or refused access to your geographical location	206
Limited access to profile or content on social networking sites or shared online storage	169
Refused allowing the use of personal data for advertising purposes	209
Checked that the website where you provided personal data was secure (e.g. https sites, safety logo or certificate)	118
Asked websites or search engines administrator or provider to access the data they hold about you to update or delete it	28

Table 6. Frequency distribution of answers to Q3.

3.3 Privacy concern

Questions Q4 to Q7 were aimed at understanding participants' levels of privacy concern. Items in this section were sourced from Choi et al., 2018.

Levels of concern were comparable for all items, as shown in the table below. The item regarding which participants reported most concern was Q4, with 88.74% selecting "agree" or

"strongly agree" when asked whether they were concerned about information provided to online vendors being misused (Q4). 85.72% of respondents agreed that they were worried about a person finding their private information on the Internet (Q5) and, similarly, 83.12% of respondents reported being worried about giving their data to online vendors, due to what others may use them for (Q6).

Participants reported being least worried about providing personal information to online vendors because it could be used in ways they did not foresee (Q7), though this was practically identical with 82.75% selecting "agree" or "strongly agree".

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Q4	2	5	19	91	114
Q5	2	7	24	96	102
Q6	3	8	28	95	97
Q7	2	7	26	89	107

Table 7. Frequency distribution of participant agreement with questions 4 to 7.

Overall, results for Q4, Q5, Q6 and Q7 were similar. There were no differences in concern based on gender, nor did differences in education level or employment status translate into different levels of concern. The remaining demographic items did not denote correlations either, but there are a few elements worth highlighting. All age categories reported similar levels of concern, though age categories 35 – 44 and 45 – 54 deviated slightly. For Q4, no one in these categories selected "strongly disagree", "disagree", or "neither agree nor disagree"; this was no longer the case for Q5, though in Q6 and Q7 both categories did not disagree with the given statements. Finally, individuals who stated they spent 5 hours or more on the Internet per day also reported the highest levels of concern for all four items in this section.

It is also interesting to note that the only individuals who consistently stated that they were not concerned with the issues mentioned were two individuals over the age of 65 who had retired. The sample does not allow a clear understanding of whether this is correlated with other variables, such as daily Internet usage (one individual reported using the Internet between 1 and 2 hours a day, the other between 4 and 5). Still, this remains an area to explore in future research endeavours.

3.4 Data protection literacy

The objective of questions Q8 to Q13 was to measure the level of participants' data protection literacy. As explained in the previous chapter, the questions were based on the European Commission's Digital Competence Framework for Citizens. Three were aimed at verifying whether participants *knew that* certain things were true, while the other three verified they *knew how* to do certain tasks.

95.68% of participants stated they knew EU citizens have the right to the protection of their personal data (Q8). This is likely due to the high level of exposure that discourse and information on the GDPR has. Just three individuals disagreed with the statement, indicating they did not know of this right.

For Q9, practically all participants (92.67%) stated they knew the purpose of a privacy policy is to explain what personal data is collected and inform users if it is shared with third parties. Compared to the preceding question, more individuals replied they did not know this, with 8 selecting either "strongly disagree" or "disagree".

Q10, where participants were asked if they know that it is good practice to periodically check which applications or services have access to their personal data, sees a shift in confidence. Though "strongly agree" and "agree" are still the most selected options at 81.46%, the former was no longer the most popular answer, with "agree" being selected by 46.55% of participants. The "neither agree nor disagree" answer also grew and was selected by 30 participants. This indicates that individuals were not aware of the importance of this practice, and instead likely adopt a "set it and forget it" approach.

The following three questions tackle "know how", thus looking beyond theoretical knowledge and measuring skills instead. Participants demonstrated less confidence in this section.

Q11 asked users whether they knew how to change browser preferences in order to prevent or limit cookies on any device. Just under 60% of participants reported knowing how to prevent or limit cookies.

When asked whether they knew how to verify that the site that requires them to provide personal information is secure (for example: https sites, logo or security certificate), only 53.87% of participants agreed. As with previous questions, more respondents selected "agree" over "strongly agree", indicating their lower confidence in this area.

The last question of this section asked respondents whether they knew how to modify privacy settings for the websites they used the most (Q13). Once again, just under 60% of respondents reported they knew how to do so, with "agree" being the most popular option.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Q8	2	1	6	106	116
Q9	2	6	8	96	119
Q10	2	10	30	108	81
Q11	12	44	39	85	51
Q12	10	56	40	79	46
Q13	9	42	45	94	41

Table 8. Frequency distribution of participant agreement with questions 8 to 13.

Demographic markers did not suggest any differences in data protection literacy for the first three items of this section, which measured declarative knowledge. Questions measuring procedural knowledge showed different results. For Q11 and Q12, women agreed less (and disagreed more) than men did. Q13 also presented this difference between genders, though it was less pronounced than the two preceding questions. Differences based on age were present in Q11, Q12 and Q13: younger age categories were almost evenly split between agreeing and disagreeing, while older ones had a well-established agreement. Individuals' procedural literacy did not vary according to education level and occupation, though some variations were found and reflected the distribution just described for age groups. Finally, time spent online daily did not seem to impact individuals' procedural data protection literacy.

3.5 Privacy fatigue

The last section of the questionnaire, containing questions 14 to 19, measured privacy fatigue and was sourced from Choi et al., 2018.

Q14 asked participants to state whether dealing with privacy issues made them feel emotionally drained. 59.74% of respondents stated they did; however, it is interesting to notice that

more individuals had a neutral view of this (27.70% selected "neither agree nor disagree") than disagreed. This indicates that, aside from a well-established majority of individuals who feel emotionally drained, individuals feel neutral.

Participants were also asked whether they were tired of issues regarding online privacy (Q15). Results were expected to lean towards tiredness, but participants were almost perfectly split: 39.83% stated they were *not* tired of online privacy, while 41.99% said they *were* tired.

The objective of Q16 was to verify whether participants believed caring about online privacy was tiresome. Here, 56.28% of individuals stated it was. A stronger reaction was expected, but this may be another symptom of fatigue: individuals may dedicate their energies only to the contexts they believe to be the most important and consider this enough.

When asked if they had become less interested in privacy issues (Q17), most participants disagreed. However, these responses made up the minority, with just 48.05% of selections; this could indicate that many still feel confused or overwhelmed by privacy issues.

Q18 measured whether respondents were less enthusiastic in protecting personal information provided to online vendors. For just over half of the population (54.11%), this is not the case, but the remaining individuals were almost evenly split between feeling neutral and agreeing.

The last question of the survey (Q19) measured whether individuals doubt the significance of online privacy issues more often. A majority of individuals, 57.14%, stated they did not have such doubts. Still, this is not a strong percentage and indicates that individuals are, to an extent, disengaging from the issue.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Q14	4	25	64	88	50
Q15	15	77	42	61	36
Q16	9	47	45	91	39
Q17	27	84	58	48	14
Q18	37	88	51	45	10
Q19	48	84	38	41	20

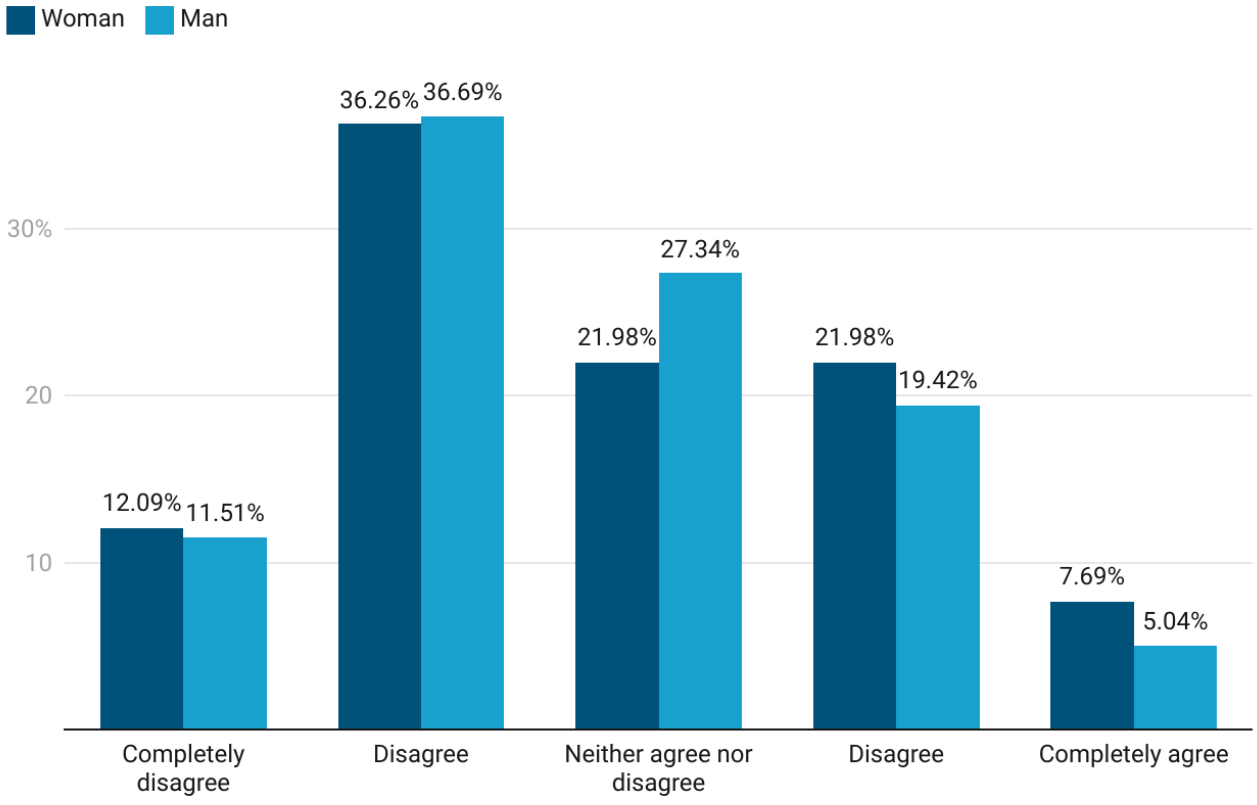
Table 9. Frequency distribution of participant agreement with questions 14 to 19.

This section shows that people do believe that dealing with privacy issues is tiresome. They also report not being less enthusiastic about privacy nor doubting its significance, but these statements are far weaker than those before them.

A few variations correlated with demographic markers are worth highlighting. For Q14, women reported being more emotionally drained than men (64.84% of women agreed against 56.83%). For this question, age seemed to be slightly correlated, with reports of feeling emotionally drained increasing with age: only 51% of 18 – 24 year olds stated they were drained, while over 3 in 4 individuals (78.72%) over the age of 65 stated the same. Similar trends, though less pronounced, were found for education level and occupation, based on age groups.

No patterns or anomalies were found for Q15, while Q16 saw age-based differences: once again, the older the age groups, the more individuals stated that being interested in online privacy is tiring.

Gender once again presented some slight differences in Q17. While the same percentage of men and women disagreed with the statement (48.35% for women and 48.20% for men), a majority of the remaining women agreed that they were less interested in privacy issues, while men responded neutrally. This is shown in the graph below.



Graph 7. Participants' agreement with Q17 in percentage, by gender.

There is little to highlight for Q17 through Q19. There were no variations linked to gender, unlike the previous questions. However, across all three, the 45 – 54 age group consistently disagreed the most, indicating they experienced limited fatigue.

3.6 Data exploration

This section looks at several potentially relevant relationships between select survey items. Instead of following a predetermined structure and cross-referencing all questions with one another, the focus is on connections that seemed noteworthy throughout the main analysis process. This approach allows an exploration into less evident areas of interest and provides a more nuanced look at the collected data. This ensures a more holistic approach to the data analysis phase, which does not stop at immediately visible data.

3.6.1 Location data

Location data seemed a relevant aspect to explore given that it was mentioned in two questions within the survey. Participants were asked whether they considered geolocation data to be personal data (Q2) and whether they had, at any time, limited or refused access to their location (Q3). The data for this relationship is shown below.

	Location data is *not* personal data	Location data is personal data
I have *not* limited or refused access to my location	18	7
I have limited or refused access to my location	53	153

Table 10. Comparison between answers to Q3 (rows) and Q2 (columns).

70% of participants believe that location data is personal data and 89.18% have limited or refused to access their location. What arises from cross referencing this data is that there are four groups of individuals: two are coherent, two are incoherent. The coherent individuals are those who believe that location data is personal data and have limited access to their location, and also those who believe that it is *not* personal data and thus have *not* limited access.

The incoherent groups are quite interesting. A small percentage of those who believe location data is personal data have not taken actions to avoid sharing their location: this could be attributed to not knowing how to do so, or believing that protecting this type of data is too difficult/not worth the

effort. Another possibility is that, after performing a privacy calculus, they might have decided that sharing location data enabled convenient services which would not be available otherwise. The second group of incoherent individuals includes those who do not think location data is personal data but have limited access to it. The reason for this choice is harder to explain, but not because data that is not personal is not worth protecting: individuals may simply not want others to know where they are. A specific study into this topic may lead to a more complete understanding of this phenomenon and reveal interesting information about how individuals conceptualise their data.

3.6.2 Privacy policies

Privacy policies are relevant as they provide crucial information on how personal data is collected, processed and shared by companies. Users who read them can make informed decisions and take the appropriate steps to protect their data. In the questionnaire, participants were asked whether they had read a privacy policy before providing personal data (Q3) and whether they knew what a privacy policy is meant to do (Q9). Table 11, below, shows cross-referenced data for these two questions.

		I have read a privacy policy before providing personal data	I have *not* read a privacy policy before providing personal data
I know that the privacy policy of an application or service should explain what personal data is collected and inform if it is shared with third parties.	Strongly disagree	2	0
	Disagree	3	3
	Neither agree nor disagree	1	7
	Agree	57	39
	Strongly agree	72	47

Table 11. Comparison between answers to Q9 (rows) and Q3 (columns).

Most respondents (93.07%) indicated they knew what a privacy policy should contain; however, this does not seem to convince individuals to read privacy policies as 40% of individuals who agreed with the statement had not done so. Surprisingly, more than half of respondents (58.44%) stated they had read a privacy policy; surveys show that this is typically done by only a minority of individuals.

Some cases are worth highlighting. Seven of the eight individuals who responded neutrally to Q9 indicated that they had not read a privacy policy, which makes sense. Instead, six participants stated that they had read a privacy policy but did not know what a privacy policy should explain. While it is true that reading one single document does not automatically imply general knowledge of the category, this is still perplexing. This could be attributed to the well-documented complexity in reading privacy policies, but also due to gaps in data protection literacy: this information is inaccessible to those who do not have a certain type of basic knowledge.

3.6.3 Security

Adopting adequate security measures online prevents unauthorised access and malicious actions from being taken. This section focuses on verifying whether individuals knew a security measure and, if so, whether they applied it.

Security was mentioned twice in the questionnaire: participants were asked to state whether they had checked that the website where they provided personal data was secure (Q3) and whether they knew how to check (Q12).

		I have checked that the website where I provided personal data was secure.	I have *not* checked that the website where I provided personal data was secure.
I know how to verify that the website where I provided personal data was secure (e.g. https sites, safety logo or certificate)	Strongly disagree	0	10
	Disagree	20	36
	Neither agree nor disagree	12	28
	Agree	49	30
	Strongly agree	37	9

Table 12. Comparison between answers Q12 (rows) and Q3 (columns).

The results from this cross-reference highlight that individuals were not taking enough steps to protect their security online. Slightly over half (54.11%) of participants knew how to check whether the website was secure; this low percentage was predictable but still worth noting. Similarly, just 37.23% of participants had checked whether the website was secure. Websites which are known to

be secure (such as Amazon, eBay, subito.it, etc) make up the majority of online e-commerce visits.[66] This might lead some to underestimate the importance of verifying whether websites are secure but, given how common online scams are, this should not be overlooked.

3.6.4 Advertising

To test whether participant behaviour followed concern, the 14 answers sorted into the "advertising" subcategory from Q1 were taken into consideration. When cross referenced with one of the possible answers for Q3, individuals' behaviour was consistent with their concern, as thirteen individuals stated that they had refused permission to use personal data for advertising purposes. This seemingly indicates that individuals' concerns affect their behaviour, though the small sample does not give certainty.

3.7 Data protection literacy and privacy fatigue

To answer the research question, a direct comparison of the results from the data protection literacy (Q8 – Q13) and privacy fatigue (Q14 – Q19) sections must be made. To verify the statistical significance of the interaction between data protection literacy and privacy fatigue items, the chi squared (χ^2) test was carried out. The following hypothesis were formulated:

H_0 – There is no relationship between data protection literacy and privacy fatigue.

H_a – Data protection literacy has an effect on privacy fatigue.

With this data taken into consideration, for a test of significance at $\alpha = .01$ and degrees of freedom = 16, the χ^2 critical value is 26.30. Pairs of questions with a p-value > .01 were considered statistically non significant and thus were not analysed. The following table lists the p-values for the pairs of questions.

[66] SimilarWeb, 'Most Popular E-Commerce and Shopping Websites in Italy in December 2023, Based on Share of Visits.', Chart (Statista, 1 January 2024), https://www.statista.com/statistics/1256072/italy-visit-share-leading-ecommerce-websites/.

	Q14	Q15	Q16	Q17	Q18	Q19
Q8	0.115	0.426	0.137	0.046	0.044	0.189
Q9	0.002	0.016	0.008	0.023	0.004	0.018
Q10	0.000	0.027	0.006	0.000	0.010	0.105
Q11	0.010	0.060	0.025	0.000	0.018	0.157
Q12	0.001	0.017	0.023	0.000	0.002	0.040
Q13	0.001	0.000	0.015	0.000	0.000	0.000

Table 13. P-value of pairs of questions from the data protection literacy (Q8 - Q13) and privacy fatigue (Q14 - Q19) sections.

The 19 pairs of questions highlighted in blue in Table 13 had a p-value greater than 0.01. This sample came from a population where these pairs of questions were independent, meaning that the null hypothesis could not be rejected. The remaining 17, instead, supported the alternative hypothesis to varying degrees and thus indicated that data protection literacy does have an effect on privacy fatigue. The questions were analysed to understand how these dependencies work, and more specifically which aspects of literacy and fatigue interact.

The tables that follow show marginal distributions for the pairs of questions which were deemed statistically significant. The columns represent questions from the privacy literacy section (Q14 – Q19), while the rows feature questions from the data protection literacy section (Q8 – Q13). Each cell represents the percentage of respondents who selected a particular combination of responses. The “total” row at the bottom, instead, shows the total percentage distribution of responses for the corresponding column. The cells in blue highlight values that are greater than the column’s total distribution, indicating where associations are present.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree			50.00%		50.00%
Disagree	16.67%		16.67%	50.00%	16.67%
Neither agree nor disagree			12.50%	37.50%	50.00%
Agree	1.04%	7.29%	32.29%	50.00%	9.38%
Strongly agree	1.68%	15.13%	25.21%	28.57%	29.41%
Total	1.73%	10.82%	27.71%	38.10%	21.65%

Table 14. Marginal distribution of Q9 (rows) and Q14 (columns).

The table above is for the Q9 – Q14 pair, which relates whether individuals knew what a privacy policy should explain and whether they felt emotionally drained from dealing with privacy issues. For Q9, "strongly disagree" was associated with "strongly agree" and "neither agree nor disagree" responses in Q14. Similarly, "disagree" is associated with agreement with Q14, with a percentage being the exception and associating with "strongly disagree". Agreement with Q9, instead, is associated with both "disagree" and "strongly agree" in Q14. Overall, lack of knowledge regarding EU citizens' right to protection of personal data was associated with feelings of emotional exhaustion because of privacy issues; on the other hand, knowledge of this right was not exclusively linked to a lack of emotional exhaustion.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree				50.00%	50.00%
Disagree	16.67%		16.67%	50.00%	16.67%
Neither agree nor disagree	12.50%	12.50%		25.00%	50.00%
Agree		23.96%	25.00%	44.79%	6.25%
Strongly agree	5.88%	19.33%	16.81%	35.29%	22.69%
Total	3.90%	20.35%	19.48%	39.39%	16.88%

Table 15. Marginal distribution of Q9 (rows) and Q16 (columns).

Responses for Q9 and Q16 are shown above. The former concerns knowledge of the EU right to the protection of citizens' personal data, while Q16 measures whether individuals feel it is tiresome for them to care about online privacy.

"Strongly disagree" for Q9 was associated with agreement with Q16. "Disagree", instead, was associated with both "agree" and "strongly disagree". The neutral option was associated with the two extreme responses ("strongly agree" and "strongly disagree"). Agreeing with Q9 was associated with disagreement, neutrality and agreement with Q14. Lastly, "strongly agree" for Q9 was associated with both "strongly disagree" and "strongly agree". A clear association between disagreement with Q9 and agreement with Q16 emerged, but the remaining responses were fairly spread out. Individuals' knowledge of what a privacy policy should explain did not seem to indicate whether an individual felt it was tiresome for them to care about online privacy.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree				100.00%	
Disagree	16.67%	16.67%	16.67%	50.00%	
Neither agree nor disagree	25.00%		25.00%	50.00%	
Agree	7.29%	40.63%	25.00%	23.96%	3.13%
Strongly agree	22.69%	40.34%	20.17%	10.92%	5.88%
Total	16.02%	38.10%	22.08%	19.48%	4.33%

Table 16. Marginal distribution of Q9 (rows) and Q18 (columns).

The table for Q9 and Q18 shows associations for knowledge of EU citizens' rights to personal data protection and whether individuals doubt the significance of privacy issues more often.

Strongly disagreeing with Q9 was strongly associated with agreement with Q18; this is the only clear association for this pair of questions, as all other responses for Q9 are associated with a variety of responses for Q18. "Strongly agree", for example, is associated with "strongly disagree", "disagree" and "strongly agree".

Overall, there does not seem to be a precise indication of whether high values for one question determine the same (or the opposite) for the other question; this points to the fact that knowing what a privacy policy should explain does not make individuals any less (nor more!) enthusiastic about protecting personal information provided to online vendors.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree			50.00%	50.00%	
Disagree			10.00%	70.00%	20.00%
Neither agree nor disagree	3.33%	10.00%	23.33%	43.33%	20.00%
Agree	1.85%	9.26%	34.26%	46.30%	8.33%
Strongly agree	1.23%	14.81%	22.22%	20.99%	40.74%
Total	1.73%	10.82%	27.71%	38.10%	21.65%

Table 17. Marginal distribution of Q10 (rows) and Q14 (columns).

Table 17 shows percentages for the Q10 and Q14 pair. These related knowing that it is good to periodically check which services have access to personal data and feelings of emotional exhaustion.

Strongly disagreeing with the first statement was associated with either neutrality or agreement Q14; the "disagree" response was strongly associated with agreeing with Q14. Q10's neutral response was associated with "strongly disagree" and "agree". Agreeing with the literacy question was then associated with "neither agree nor disagree" and "agree" for the question from the fatigue section. Finally, "strongly agree" was associated with "disagree", and strongly associated with "strongly agree".

This indicates that individuals do feel emotionally drained from dealing with privacy issues, and those who reported not knowing that it is good to periodically check which services have access to personal data were especially drained.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree				50.00%	50.00%
Disagree			20.00%	60.00%	20.00%
Neither agree nor disagree	3.33%	6.67%	23.33%	46.67%	20.00%
Agree	1.85%	24.07%	23.15%	44.44%	6.48%
Strongly agree	7.41%	23.46%	13.58%	27.16%	28.40%
Total	3.90%	20.35%	19.48%	39.39%	16.88%

Table 18. Marginal distribution of Q10 (rows) and Q16 (columns).

Comparing Q10 and Q16 meant comparing individuals' knowledge that it is a good practice to periodically check which services have access to personal data and whether they feel that it is tiresome to care about privacy. This pair of questions showed a much stronger and identifiable pattern than the previous ones.

Disagreeing with Q10 was strongly associated with agreeing and strongly agreeing with Q16. Answering "disagree", similarly, was associated with "agree", "strongly agree", and the neutral option for Q16. Neutral answers for Q10 replicated this. "Agree" was then associated, albeit not particularly strongly, with "disagree", "neither agree nor disagree", and "agree". Finally, "strongly agree" was associated with disagreement in Q16, though the association with "strongly agree" cannot be ignored.

Generally speaking, there is an identifiable negative relationship between the two questions, which shows that knowing that it is good to periodically check which services have access to personal data limits the feeling that it is tiresome to care about online privacy.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree		50.00%		50.00%	
Disagree		20.00%	10.00%	40.00%	30.00%
Neither agree nor disagree	16.67%	23.33%	30.00%	23.33%	6.67%
Agree	2.78%	43.52%	23.15%	27.78%	2.78%
Strongly agree	23.46%	33.33%	28.40%	7.41%	7.41%
Total	11.69%	36.36%	25.11%	20.78%	6.06%

Table 19. Marginal distribution of Q10 (rows) and Q17 (columns).

Q10 was then compared to Q17, measuring whether individuals have become less interested in online privacy issues. The results here were spread out, with a weak pattern emerging. The association for Q10's "strongly disagree" option was split between "disagree" and "agree" for Q17. Individuals who disagreed with Q10 then agreed with Q17. Neutrality was instead associated with "strongly disagree" and "neither agree nor disagree". Agreeing with Q10 was associated with both "disagree" and "agree" for Q17, while strongly agreeing showed an association with "strongly disagree" and "neither agree nor disagree".

Disagreement with Q10, overall, does seem to be associated with agreement with Q17. With data being this spread out, it is difficult to draw accurate conclusions on whether knowing that it is good to periodically check which services have access to personal data implies being less interested in online privacy issues.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	50.00%			50.00%	
Disagree		30.00%	20.00%	40.00%	10.00%
Neither agree nor disagree	16.67%	33.33%	33.33%	16.67%	
Agree	10.19%	37.96%	24.07%	25.93%	1.85%
Strongly agree	24.69%	41.98%	16.05%	8.64%	8.64%
Total	16.02%	38.10%	22.08%	19.48%	4.33%

Table 20. Marginal distribution of Q10 (rows) and Q18 (columns).

When relating Q10 to Q18, which measures whether individuals have become less enthusiastic in protecting personal information, a negative relationship (with notable exceptions) appeared.

Overall, individuals who disagreed with Q10 then agreed with Q18 (and vice versa). However, in both cases "strongly disagree" and "strongly agree" were associated with the same answer in the opposite question, meaning that this negative relationship cannot be fully supported. Here, there is no clear indication of whether individuals who know it is good to periodically check which services have access to their personal data also do not lose enthusiasm in protecting personal information.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree			16.67%	58.33%	25.00%
Disagree		6.82%	20.45%	63.64%	9.09%
Neither agree nor disagree	2.56%	10.26%	28.21%	33.33%	25.64%
Agree	2.35%	8.24%	32.94%	36.47%	20.00%
Strongly agree	1.96%	21.57%	27.45%	17.65%	31.37%
Total	1.73%	10.82%	27.71%	38.10%	21.65%

Table 21. Marginal distribution of Q11 (rows) and Q14 (columns).

Q11 and Q14 measured the relationship between knowing how to change browser settings to restrict cookies and feeling emotionally drained due to privacy issues. As shown in the table above, the marginal distribution analysis revealed that disagreement with Q11 was consistently associated with agreement with Q14. A notable exception was Q11's "strongly agree" response, which was associated with "strongly agree" for Q14 as well. Similarly, the former's neutral response was associated with the latter's "strongly disagree".

Overall, associations for this pair of questions indicates that knowing how to change browser settings results in individuals feeling less emotionally drained from dealing with privacy issues.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	8.33%	16.67%	33.33%	16.67%	25.00%
Disagree	4.55%	29.55%	15.91%	40.91%	9.09%
Neither agree nor disagree	2.56%	35.90%	46.15%	12.82%	2.56%
Agree	14.12%	44.71%	16.47%	21.18%	3.53%
Strongly agree	21.57%	33.33%	29.41%	9.80%	5.88%
Total	11.69%	36.36%	25.11%	20.78%	6.06%

Table 22. Marginal distribution of Q11 (rows) and Q17 (columns).

Q11 and Q17 also interacted similarly. Once again, lower levels of literacy resulted in higher levels of fatigue: "strongly disagree" was associated with "neither agree nor disagree" and "strongly agree"; disagreeing was associated with "agree" and "strongly agree". "Agree" and "strongly agree" followed a similar distribution, with the exception of the former also associating with "agree" for Q17. The notable exception for this pair of questions, however, was undecided responses for Q17, which were above the marginal value for "strongly disagree", "neither agree nor disagree" and also "strongly agree".

On the whole, however, results still show that individuals who stated they knew how to change their browser settings also stated they had not lost interest in online privacy.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree			20.00%	40.00%	40.00%
Disagree	1.79%	12.50%	17.86%	58.93%	8.93%
Neither agree nor disagree		5.00%	50.00%	22.50%	22.50%
Agree	2.53%	11.39%	25.32%	41.77%	18.99%
Strongly agree	2.17%	15.22%	26.09%	19.57%	36.96%
Total	1.73%	10.82%	27.71%	38.10%	21.65%

Table 23. Marginal distribution of Q12 (rows) and Q14 (columns).

Q12 and Q14 looked at whether knowing how to verify that a website is secure results in feeling emotionally drained by privacy issues. A negative relationship between literacy and fatigue emerged: "strongly disagree" and "disagree" were associated with the agreement responses for Q14; agreeing and strongly agreeing with Q12, conversely, was associated with disagreeing and strongly disagreeing for Q14. Notable exceptions are Q12's "disagree", which was associated with "disagree" for Q14, and "strongly agree", which was associated with Q14's "strongly agree".

Overall, the data supports the idea that knowing how to verify that the site they are providing their personal information is secure means experiencing less emotional exhaustion from dealing with privacy issues.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	10.00%	30.00%		30.00%	30.00%
Disagree	5.36%	35.71%	14.29%	39.29%	5.36%
Neither agree nor disagree	2.50%	30.00%	50.00%	12.50%	5.00%
Agree	12.66%	44.30%	24.05%	17.72%	1.27%
Strongly agree	26.09%	30.43%	23.91%	8.70%	10.87%
Total	11.69%	36.36%	25.11%	20.78%	6.06%

Table 24. Marginal distribution of Q12 (rows) and Q17 (columns).

A negative correlation was also present between Q12 and Q17. Answering "strongly disagree" for the former was associated with "agree" and "strongly agree" for the latter; "disagree" was associated with "agree". Agreeing with Q12, instead, was associated with disagreeing with Q17. The only exception was Q12's "strongly agree", which was associated with both "strongly disagree" and "strongly agree".

This pair of questions revealed that respondents who stated they knew how to verify that a website was secure primarily reported that they had not become less interested in online privacy.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	30.00%	10.00%	10.00%	20.00%	30.00%
Disagree	5.36%	44.64%	23.21%	23.21%	3.57%
Neither agree nor disagree	12.50%	45.00%	25.00%	17.50%	
Agree	17.72%	36.71%	21.52%	22.78%	1.27%
Strongly agree	26.09%	32.61%	21.74%	10.87%	8.70%
Total	16.02%	38.10%	22.08%	19.48%	4.33%

Table 25. Marginal distribution of Q12 (rows) and Q18 (columns).

The table above shows distribution for Q12 and Q18. “Strongly disagree” for Q12 was associated with “agree” and “strongly agree”, but also with “strongly disagree” for Q18. Disagreeing with Q12, similarly, was associated with disagreeing, agreeing and remaining neutral with the latter. Agreeing and strongly agreeing with Q12 was associated with both agreement and disagreement with Q18, as shown above.

For this pair of questions, an association cannot be made between individuals knowing how to verify that a site is secure and becoming less enthusiastic in protecting their information.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree			11.11%	55.56%	33.33%
Disagree	2.38%	9.52%	23.81%	59.52%	4.76%
Neither agree nor disagree	2.22%	4.44%	31.11%	35.56%	26.67%
Agree	1.06%	12.77%	34.04%	37.23%	14.89%
Strongly agree	2.44%	17.07%	17.07%	17.07%	46.34%
Total	1.73%	10.82%	27.71%	38.10%	21.65%

Table 26. Marginal distribution of Q13 (rows) and Q14 (columns).

Q13 and Q14 verified whether knowing how to change privacy settings on the sites individuals use most limited whether they felt emotionally drained. An association between high agreement for Q13 and low agreement for Q14 appears, though there are two exceptions. “Disagree” for Q13 was associated with “strongly disagree” for Q14, and the two questions’ “strongly agree” were also associated with one another.

Overall, there did seem to be a negative correlation between individuals who knew how to change privacy settings on their most used sites and feeling emotionally drained from dealing with privacy issues; the two exceptions, however, cannot be overlooked.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	33.33%		11.11%	22.22%	33.33%
Disagree	2.38%	42.86%	11.90%	40.48%	2.38%
Neither agree nor disagree	4.44%	33.33%	15.56%	28.89%	17.78%
Agree	4.26%	36.17%	21.28%	27.66%	10.64%
Strongly agree	12.20%	24.39%	21.95%	7.32%	34.15%
Total	6.49%	33.33%	18.18%	26.41%	15.58%

Table 27. Marginal distribution of Q13 (rows) and Q15 (columns).

The Q13 and Q15 pair of questions verified whether knowing how to change privacy settings on users' most used sites affected whether they felt tired of online privacy issues. "Strongly disagree" for Q13 was associated with both "strongly disagree" and "strongly agree" for Q15; similarly, "strongly agree" in the former question was associated with strongly disagreeing, remaining neutral and strongly agreeing with the latter. The remaining responses were also associated without particular patterns, resulting in a table without a clear relationship. For this reason, it is difficult to state that knowing how to change privacy settings limits feeling tired of online privacy issues.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	11.11%	33.33%		22.22%	33.33%
Disagree	4.76%	30.95%	23.81%	38.10%	2.38%
Neither agree nor disagree	8.89%	20.00%	40.00%	22.22%	8.89%
Agree	7.45%	53.19%	19.15%	18.09%	2.13%
Strongly agree	31.71%	21.95%	29.27%	7.32%	9.76%
Total	11.69%	36.36%	25.11%	20.78%	6.06%

Table 28. Marginal distribution of Q13 (rows) and Q17 (columns).

For Q13 and Q17, a negative association returned. Disagreement with Q13 was consistently associated with agreement for Q17, and vice versa. The only exception were the "strongly agree" answers being associated with one another, though not with particular strength. Overall, this pair of questions indicated that knowing how to change privacy settings on websites was not associated with losing interest in online privacy issues.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	22.22%	11.11%	11.11%	22.22%	33.33%
Disagree	4.76%	42.86%	21.43%	28.57%	2.38%
Neither agree nor disagree	22.22%	26.67%	22.22%	28.89%	
Agree	11.70%	46.81%	24.47%	14.89%	2.13%
Strongly agree	29.27%	31.71%	19.51%	9.76%	9.76%
Total	16.02%	38.10%	22.08%	19.48%	4.33%

Table 29. Marginal distribution of Q13 (rows) and Q18 (columns).

Q13 and Q18's marginal distribution is shown in the table above. Though there are associations, these do not show a clear pattern. "Strongly disagree" in Q13 is associated with "strongly disagree", "agree" and "strongly agree" in Q18. "Disagree" is then associated with "disagree" and "agree". Q13's neutral response is associated with "strongly disagree", "neither agree nor disagree" and "agree" for Q18.

Overall, this pair does not show that an increase in data protection literacy is associated with lower levels of privacy fatigue.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Strongly disagree	33.33%	11.11%		33.33%	22.22%
Disagree	11.90%	33.33%	23.81%	30.95%	
Neither agree nor disagree	20.00%	28.89%	4.44%	28.89%	17.78%
Agree	20.21%	46.81%	20.21%	10.64%	2.13%
Strongly agree	29.27%	29.27%	17.07%	4.88%	19.51%
Total	20.78%	36.36%	16.45%	17.75%	8.66%

Table 30. Marginal distribution of Q13 (rows) and Q19 (columns).

Q13 and Q19 had a number of associations, though a discernible pattern was not present. Q13's "strongly disagree", for example, was associated with "strongly disagree", "agree" and "strongly agree" in Q19. Similarly, "strongly agree" was then associated with "strongly agree", "neither agree nor disagree" and "strongly agree".

The association between the two "strongly disagree" and "strongly agree" mean that, Overall, we cannot state that when individuals are more literate, they do not doubt the significance of online privacy issues.

3.8 Conclusion

This chapter has described the results of the questionnaire at the heart of this work. Results for the first three questions revealed respondents' general attitudes. Q1, the only open question in this survey, showed that most respondents associated privacy with specific "types of data", with very few instead referencing "rights" or "anonymity". When asked which data they believed to be personal data (Q2), the most popular choice was "banking data". Responses for Q3, instead, showed that the action most individuals had carried out to manage their personal data was refusing to allow its use for advertising purposes.

Levels of concern (Q4 – Q7) were fairly constant, with little to no differences based on gender, education level or employment status. Age and time spent on the Internet daily, instead, seemed to impact concern in some cases.

Data protection literacy (Q8 – Q13) was fairly consistent, albeit with some differences between declarative and procedural knowledge items. No correlations with demographic markers were found for the former, while the latter showed variations for age and gender.

Privacy fatigue (Q14 – Q19) showed that respondents believed that dealing with privacy issues is tiresome, but did not report feeling cynical regarding its significance. Different questions within this section were correlated with different demographic markers, though mainly age and gender.

Finally, paragraph 3.7 compared questions from the data protection literacy and privacy fatigue sections. Not all pairs of questions were deemed statistically significant for the chosen level; those which were generally showed a negative correlation between the two phenomena, albeit not consistently.

4. Discussion

4.1 Introduction

The purpose of this chapter is to critically examine the findings of the questionnaire; the results described in Chapter 3 are analysed through the lens of Chapter 1's conceptual baseline. This discussion explores the significance of the results, identifies potential implications and observes their relevance in the relationship between data protection literacy and privacy fatigue.

The first paragraph addresses the research question, discussing the extent to which data protection literacy does have an impact on privacy fatigue, along with additional information that was found through the data analysis process. Paragraphs 4.3 and 4.4 then explore select topics within literacy and fatigue which arose in the data analysis phase. Focusing on the two variables separately ensures a well-rounded understanding of both literacy and fatigue as standalone phenomena. The last paragraph highlights limitations of this study, both from a methodological point of view, but also in the sample itself. Scenarios on how future research can both avoid these and build upon this work are also briefly outlined.

4.2 Data protection literacy and privacy fatigue

The key relationship explored in this work is the connection between data protection literacy and privacy fatigue, with the intention of understanding how knowledge influences individuals' attitudes towards privacy.

Literacy levels were higher than expected. However, a clear difference could be noticed between declarative and procedural knowledge. Participants' responses demonstrated that they had more than sufficient theoretical knowledge on data protection, but then did not know how to concretely take the steps to ensure their information was protected. This brings light to the existence of a knowledge gap: though individuals are aware of privacy risks and are concerned about how their data is used, this does not translate into having (or applying) the practical skills needed to address such issues and implement the appropriate protection mechanisms. For example, individuals may be aware of best practices regarding passwords, but may not be following them on a regular basis or doing so effectively.

Respondents were highly concerned about privacy issues, but also reported feeling overwhelmed by them. Despite this, stronger feelings of fatigue were expected. These levels of

fatigue were also linked to low cynicism, with individuals indicating that their belief in the importance of privacy had not diminished. Privacy fatigue was clearly present, though not constant. Participants did report that they felt drained from dealing with privacy issues, that they were tired of doing so, and that they felt that it is tiresome to do so; however, they also stated that they were not less interested in online privacy, nor did they doubt the significance of these issues.

To answer the research question, a comparison between the data protection literacy section and privacy fatigue sections was carried out. The results described in section 3.7 of the previous chapter show that a clear, unequivocal relationship did not emerge: not all pairs of questions were statistically significant. Among those which were statistically significant, responses mostly followed a negative relationship: high values for data protection literacy questions were associated with low values in the privacy fatigue section; high levels of privacy fatigue, instead, occurred when literacy levels were low. Several pairs of questions, however, also featured non-trivial exceptions.

Overall, the findings suggest that individuals who have a greater understanding of core issues in privacy and are familiar with data protection measures have the skills and tools needed to manage their personal information, which in turn leads to lower levels of exhaustion. These individuals will also be less cynical and continue to see the value of engaging in privacy-conscious behaviours. Those who have limited knowledge or none at all, instead, will experience frustration and confusion, and consequently higher levels of fatigue. The burden of privacy, in this case, will cause them to disengage and adopt a "why bother?" attitude. The former case is the desired effect of education, as outlined in Chapter 1: individuals who achieve basic literacy are aware of existing issues and possess skills needed to protect themselves, which in turn increases their chances of succeeding in life. Additionally, in the context of a digital society, this ensures they can engage with others and access many services safely and consciously, enabling participation in crucial activities for them to successfully engage with society. This result is encouraging, as it shows that education can play a pivotal role in boosting data protection among digital citizens.

Data analysis also showed, however, that some individuals deviated from this negative relationship and did not neatly fit this dichotomy. This was the case for individuals who indicated they had low levels of data protection literacy and low levels of fatigue, but also for those who were knowledgeable about data protection and still experienced exhaustion. "Ignorance is bliss" is a simplified explanation for those who lack knowledge and fatigue. A possible interpretation for this pattern is that limited understanding of data protection and unfamiliarity with tools or measures

needed to protect one's personal information may cause individuals not to feel any stress about making decisions regarding data protection. Briefly put, low data protection literacy levels may also have the effect of preventing individuals from experiencing the burden of being concerned about privacy and, in this sense, limit privacy fatigue.

The other side of the coin, some of the individuals who were knowledgeable went on to experience high levels of fatigue. At higher literacy levels, individuals are more aware of the existing risks and how challenging it is to effectively protect personal information. This makes them aware of the complexity and effort needed to keep track of data usage, read privacy policies, adjust privacy settings as they wish and, more generally, adopt whatever measure they believe to be necessary. The attentiveness and care for details required to do so can overwhelm individuals, causing them to experience fatigue.

The presence of these two categories that do not fit the previously discussed dichotomy is a potential issue and, at first glance, may cast doubts on validity of results. However, a more nuanced interpretation can understand these categories as natural. After all, the existence of these fringe categories occurs naturally due to the law of large numbers: for most, the presence of fatigue may be related to their data protection literacy levels, but there will likely still be some individuals who do not perfectly fit the model. To confirm this, however, similar research focusing on the existence (and perhaps persistence) of these attitudes would have to be conducted on a larger sample and across countries.

Having noted the existence of the relationship between literacy and fatigue, complementary reflections on factors affecting the two phenomena, or in any case related to either, can be useful. This ensures that potential issues in the way technology is dealt with currently in society are highlighted and discussed.

Literacy undoubtedly increases awareness on existing issues. When individuals know what risks they might be exposed to and how they can limit them, they are also exposed to more analysis and decision-making. This increases their mental load, which means they are affected by privacy fatigue. Generally speaking, does not seem to cancel out the importance that they attribute to safeguarding their personal information. However, simply restating the influential role of literacy is not sufficient: reflecting on how data protection literacy, and more broadly digital literacy, can be achieved in practice is also necessary.

For the purposes of this study, data protection literacy was evaluated with a minimal set of questions. This was done by asking participants about their declarative and procedural knowledge; each aspect of literacy was evaluated with three questions. These criteria were specifically chosen for the purpose of this work, with some items representing "basic" knowledge and others more advanced aspects of data protection. However, many digital skills/literacy assessment methods exist and measure different variables, or the same ones in different manners. While this questionnaire revealed a slight negative relationship, other literacy assessments could provide different insights, or perhaps more nuanced ones.

For example, the International Certification of Digital Literacy (ICDL, formerly known as European Computer Driving Licence) offers a variety of programmes which assess different skill categories. "Digital Citizen", for example, ensures individuals have the basic skills required to use a computer for everyday activities. This includes being able to identify computer types and their main parts, understanding how to use the desktop, documents and the Internet, and also communicating using a variety of methods.[67] This programme is undoubtedly sufficient for day-to-day, simple activities. To properly engage with the workplace, more advanced concepts such as those in the "Computer and Online Essentials", instead, are needed; this includes managing files, network concepts, accessing information online and more.[68] These are both valid assessments and users who pass either of them will be considered literate but, crucially, will have different skill levels.

One of the issues with these assessments is that they measure individuals' literacy levels in a single point in time. Technology changes rapidly and a well-defined yet static set of digital skills may not be sufficient for individuals to keep up, adapt and properly address concerns which will arise from future technology. The "continuum of learning and proficiency"[69] that UNESCO identifies for traditional literacy is extremely relevant for digital skills too. The ability to understand and critically evaluate new developments in technology may reduce the extent to which individuals feel disoriented and overwhelmed, thereby reducing fatigue. It is much easier to learn step-by-step, independently, by building upon pre-existing skills, than having to repeatedly learn completely new sets of measures, tools and best practices.

[67] 'Digital Citizen', ICDL Global, accessed 12 September 2024, https://icdl.org/digital-citizen/digital-citizen/.

[68] 'Computer and Online Essentials', ICDL Global, accessed 12 September 2024, https://icdl.org/workforce/computer-and-online-essentials/.

[69] 'Literacy'.

Another question goes hand-in-hand with the above: how are individuals being taught digital skills and, in particular, about data protection theory and practice? Most individuals learn on their own,[70] simply by interacting with the devices and applications at their disposal or may need to use for a variety of reasons (including school, work or personal administrative/tax purposes). This is especially true for younger generations who have had devices within reach for longer periods of their lives. The issue is that this only makes people familiar with very specific uses; a literate individual, instead, has the ability to navigate and use unfamiliar services without specific training. The ideal context for developing digital skills is in school (preferably before university), and educational systems have offered such classes, though a decline in both quality and quantity has been a topic of ongoing discussion.[71]

This is linked to the issue of how digital skills can be achieved outside of the school context, as opportunities are not as widespread nor accessible. Organisations who administer tests, such as ICDL, frequently offer courses. Their intrinsic link with certifications, however, may result in individuals learning how to pass the specific test rather than ensuring they actually acquire the skills meant to be certified.

Discussing the relevance of literacy in this relationship can be sterile if one does not look into how literacy itself can be achieved. Literacy is not something that happens on its own and suddenly, but takes effort over time. This must not to be underestimated nor a simple afterthought; which is why the following paragraph describes patterns in literacy which emerged during the data analysis process.

4.3 Understanding literacy

Having reflected on the relationship between literacy and fatigue, it is worth focusing on details of the individual phenomena which emerged from data analysis. This paragraph looks at how individual elements within data protection literacy play a role in its levels.

This questionnaire showed that age does not seem to correlate with literacy, contrary to popular belief. The results from Q8 – Q13 showed that, when comparing ages, there were no differences in declarative knowledge and only minor variations in procedural knowledge. While the

[70] Sally Foster, 'Australian Undergraduate Internet Usage: Self-Taught, Self-Directed, and Self-Limiting?', *Education and Information Technologies* 5, no. 3 (2000): 165–75, https://doi.org/10.1023/A:1009602617991.

[71] Tony Koppi et al., 'The Crisis in ICT Education: An Academic Perspective', in *Hello! Where Are You in the Landscape of Educational Technology?*, 2008, http://www.ascilite.org.au/conferences/melbourne08/procs/koppi.pdf.

non-representative nature of the sample must be acknowledged as a possible reason for this result, this data could also be interpreted as an indication that the widespread belief in digital natives' innate ability to use computers and other digital devices has limited validity.

This finding is in agreement with recent literature, such as Kirschner (2017), which has disproved previous beliefs that younger individuals are inherently more adept with technology. Younger generations are certainly more "tech-savvy" than older ones; young adults often help parents order from e-commerce websites or watch films on streaming sites, or perhaps set up their grandparents' phones and teach them how to send texts. But this savviness might simply be born from an increased usage of newer technologies and thus familiarity, rather than from an innate deep knowledge of them.

Younger generations, indeed, are familiar with smartphones and social media apps, for example, which they might have learned to use organically by being introduced to them at an early age and through frequent use. These individuals are proficient in the simpler tasks such as browsing the web, emailing and using basic office suite functionalities[72] and can confidently use intuitive devices, such as smartphones and tablets, rather than computers. However, these skills are insufficient in an ever-digital world and this digital skills gap may impact their ability to be successful digital citizens. This has implications that reach beyond simple generational divides: if ill-equipped with digital skills, serious issues can arise within the workplace. For example, unfamiliarity with devices such as printers and scanners or lack of understanding how hierarchical file systems work may be seriously limiting, if not damaging.[73] For example, placing a file in the wrong directory could mean sharing confidential information with people should not have access to it. Employment is not the only context where these considerations matter, nor is it the end-all-be-all metric; it serves as a mere example to visualise how a lack of data protection skills may have consequences.

Other two demographic markers in this survey saw consistent literacy levels, likely linked to the lack of variation in age. This is the case for education level and occupation, which mostly presented similar results as age. This makes sense seeing as, for example, younger individuals have reached lower education levels and their "employment" is being students.

[72] Kirschner and De Bruyckere, 'The Myths of the Digital Native and the Multitasker'.

[73] Alaina Demopoulos, '"Scanners Are Complicated": Why Gen Z Faces Workplace "Tech Shame"', *The Guardian*, 28 February 2023, sec. Technology, https://www.theguardian.com/technology/2023/feb/27/gen-z-tech-shame-office-technology-printers.

Age may not be an explanation for digital literacy levels, but other factors may be linked to them. Higher income is one: those who are more well-off are able to purchase digital devices, use them in their leisure time and thus become more skilled. The contrary happens for individuals with lower income: if they cannot afford devices or have limited access to them, they will have fewer possibilities to develop these skills, which in turn will impact their literacy levels in the long term. Though income is a classic demographic marker, this study cannot make assertions based on income since its participants were not asked to disclose this information.

Time spent on the Internet daily also did not translate into competence. This is once again surprising as it seemingly goes against common knowledge, where those who spend copious amounts of time using their devices are believed to be more skilled. Part of the reason was that those with the highest daily Internet usage were participants aged 18 – 24, and age's lack of significance for literacy has already been discussed above. As Debbie Irish, HP's head of UK and Ireland human resources, (rather condescendingly) put it: "[...] neither watching TikTok videos nor playing Minecraft fulfills [*sic*] the technology brief."[74] Anecdotal evidence points to the fact that younger generations mainly use software which is designed to be user-friendly and intuitive, rather than having to read instruction manuals to learn proper usage. However, the extent to which is true is unknown as literature on these topics is limited.

Aside from this, what arises from this data is that the quantity of time spent on the Internet does not mean that an individual is digitally literate nor, as it naturally follows, knows how to protect their data. Quality must then be the key rather than quantity; more specifically that individuals need to be properly taught and not expected to learn by themselves by simply interacting with technology.

Literacy levels were also slightly different for men and women, with the latter displaying lower levels of confidence. Exploring the reasons why can provide insight into how digital curricula can be improved. Firstly, the field of technology is mostly viewed as a male field. Though women had an "unusually prominent role in the history of computer programming, especially in its earliest decades",[75] societal perception has changed to now perceive computers as stereotypically masculine. Young boys are encouraged to develop an interest and skills for anything relating to technology, while girls aren't; over time, this results in a 'digital skills gap'.

---

[74] Oliver Pickup, 'Gen Z Workers Are Not Tech-Savvy in the Workplace – and It's a Growing Problem', *WorkLife*, 14 December 2022, https://www.worklife.news/technology/myth-buster-young-workers-are-not-tech-savvy-in-the-workplace-and-its-a-growing-problem/.

[75] Nathan Ensmenger, '"Beards, Sandals, and Other Signs of Rugged Individualism": Masculine Culture within the Computing Professions', *Osiris* 30, no. 1 (January 2015): 38–65, https://doi.org/10.1086/682955.

Another possibility is also that women stated they were less capable because of lower self-esteem rather than actual lower abilities. The tendency women have to underestimate their own abilities is sometimes referred to as the "confidence gap"; women are far more critical of themselves than men are and also have lower self-esteem.[76] In the context of this questionnaire, it is possible that, in reality, literacy levels were equal but self-evaluation skewed them due to differences in how gender influences and determines self-confidence. This would be consistent with Ehrlinger and Dunning (2003), who found that women predicted lower performance than men, despite achieving comparable scores.[77]

4.4 Understanding fatigue

The elements that arose when analysing fatigue are similar to those observed in the analysis of privacy fatigue, which supports the hypothesis of a relationship between the two phenomena. However, some additional reflections, unique to fatigue, can be underlined.

Results from the survey showed that older individuals feel more tired than younger age categories. Surveys show that lack of knowledge is a barrier to use of technology for older generations;[78] a possible explanation for their fatigue, then, is that these individuals are simply not confident with current technologies and thus feel overwhelmed more easily. They may also have a different perception of privacy, as they grew up in times when personal information wasn't nonchalantly shared and easily observable by many.

If the natural environment for young individuals to learn digital skills is in school, the same cannot be said for those who are not of schooling age. While those in the workforce likely learn the skills required by their jobs on site, or through company training programs, elderly people do not have the same opportunity. As more services transition to online versions, the elderly may be cut off from participation in society due to their limited digital skills. This may cause them to view technology as something difficult, and in turn induce them to experience increased levels of fatigue.

[76] Wiebke Bleidorn et al., 'Age and Gender Differences in Self-Esteem—A Cross-Cultural Window.', *Journal of Personality and Social Psychology* 111, no. 3 (September 2016): 396–410, https://doi.org/10.1037/pspp0000078.

[77] Joyce Ehrlinger and David Dunning, 'How Chronic Self-Views Influence (and Potentially Mislead) Estimates of Performance.', *Journal of Personality and Social Psychology* 84, no. 1 (2003): 5–17, https://doi.org/10.1037/0022-3514.84.1.5.

[78] McKinsey & Company, 'Barriers to Use Technology among Older Adults Worldwide in 2023, by Type and Age Group', Chart (Statista, May 2023), https://www.statista.com/statistics/1387165/older-adults-barriers-technology-use/.

Women also reported more fatigue than men. This echoes what was stated for literacy and likely goes hand-in-hand with the above discussion. Women are less interested in technology (whether for cultural or personal reasons) than men, thus using computers and other devices less; the natural consequence is that they develop less skills. Furthermore, they are generally less confident in their own (digital) skills. These elements, together, likely impact women's tolerance and thus threshold for privacy fatigue.

This is not the sole possible explanation. Time availability could also offer a useful perspective: women experience more time poverty than men, as their "paid and unpaid work combines into a longer working day than men's";[79] among the consequences of time poverty is emotional exhaustion.[80] In this context, pre-existing fatigue may influence women's tolerance for dealing with privacy issues, and lead them to experience more fatigue: not only do they feel less capable, but their longer working days (and thus shorter leisure time) means they may also not have the time to care about privacy.

4.5 Limitations and avenues for future research

While this study offers insights into the relationship between data protection literacy and privacy fatigue, and how these phenomena interplay, it also has limitations that should be considered when interpreting the results. In order to obtain more accurate and relevant data, future research should aim to correct or limit this work's shortcomings.

The purpose of this study was to gain an understanding of individuals' perceptions and knowledge of data protection and measure their privacy fatigue. However, the imbalances in respondents gender and age may have an effect on the generalizability of the results in this exploratory analysis. An unintentional bias also occurred in the demographic section: by not including PhD as an educational level, some respondents may have not been able to accurately report their highest academic achievement. For example, a participant reported being a university professor, yet stated that they had not achieved any of the educational levels listed (the highest among which was a Master's degree); this individual may have a PhD but not have been able to report it properly.

[79] Caroline Criado-Perez, *Invisible Women: Exposing Data Bias in a World Designed for Men* (London: Chatto & Windus, 2019).

[80] Katja Teuchmann, Peter Totterdell, and Sharon K. Parker, 'Rushed, Unhappy, and Drained: An Experience Sampling Study of Relations between Time Pressure, Perceived Control, Mood, and Emotional Exhaustion in a Group of Accountants.', *Journal of Occupational Health Psychology* 4, no. 1 (January 1999): 37–54, https://doi.org/10.1037/1076-8998.4.1.37.

Similarly, collecting data on which devices are used to access the Internet may provide another useful variable in understanding factors affecting literacy and fatigue. Further research should aim at increasing sample size and participant variety (e.g. including more students, individuals from a wider range of backgrounds etc.).

Another limitation arises from respondents self-reporting their concerns and habits; this modality introduces a number of issues related to validity. Individuals tend to overestimate their abilities or, on the contrary, minimise their negative habits. For example, in the context of this survey, this may result in users overestimating how digitally literate they are and minimising the steps they do not take to protect their privacy. Reasons for such an occurrence include a lack of self-awareness or a desire to answer in a socially desirable way. Furthermore, individuals might struggle to accurately remember past behaviours or experience, once again resulting in inaccurate answers. As explained in section 4.3, this may be particularly true for women; this should be taken into account in further studies.

The length of the questionnaire may also limit the validity of results. As mentioned in Chapter 2, brevity was required to ensure that individuals were willing to participate and did not become tired while completing the questionnaire. However, this may also have consequences on the validity of the questionnaire: by limiting the number of questions included in the survey, key variables may have been omitted, leading to an incomplete picture of respondents' experiences. This, in turn, may oversimplify and under-explain complex issues at play in this field, missing nuance and limiting the depth of findings.

Aside from addressing the above, a potential avenue for future research is to go beyond simply verifying whether a relationship exists between data protection literacy and privacy fatigue. An evaluation of whether variations in literacy levels determine variations in privacy fatigue could help identify which elements of literacy have the greatest impact, and thus help focus literacy efforts in particular directions. In this sense, using a validated and known scale (or several) could be useful and ensure that these core skills are identified. Finally, employing this method could also help investigate whether scoring higher on literacy tests results in lower levels of self-reported fatigue.

4.6 Conclusion

This chapter has built upon the previous, taking the results from the questionnaire this work is based on and discussing potential explanations and implications.

Given the high threshold chosen for the test of significance, not all questions from the data protection section, when compared with questions in the privacy fatigue section, returned statistically significant results. However, those that were significant showed a negative relationship between the two phenomena, indicating that literacy does affect privacy fatigue. Some individuals reported feeling no fatigue despite having little to no knowledge of data protection, while others were literate but still fatigued; these appear to be an acceptable exception, not necessarily negating the findings of this research.

Privacy fatigue is a pressing issue for digital citizens and may limit their possibility of participating in a variety of activities. Researching what the best paths to literacy are for digital citizens and how to ensure that lifelong learning can occur for most means investing in future generations' success.

Conclusion

In a datafied society, digital citizens interact every day with services and applications which are looking to obtain their data. Every day, individuals allow cookies, agree to share their personal data with advertisers they've never heard of, and accept privacy policies they haven't read. For most, these are the unavoidable conditions that must be accepted to access the online world. This is true to an extent, as users cannot haggle with Internet companies and reach a compromise on which data they will allow access to.

This does not mean that digital citizens are not concerned about how their data is used, who has access to it and what they do with it. Though some are tired of managing data, this questionnaire's respondents have stated that the difficulty is not something that makes them value privacy any less. Perhaps this is because the importance of privacy has become almost ubiquitous, with various actors promoting it. In any case, this attitude needs to be leveraged to ensure that people actually do act upon it.

It is not a matter of avoiding datafication: the convenience and efficiency they bring in a number of contexts often outweighs their downsides. This societal transformation was described in Chapter 1, and only serves to show the relevance of literacy: it is important to possess the digital literacy needed to know how to use devices and access the Internet, but also know how to evolve our own skills and apply them to ever-evolving technology. In particular, data protection literacy becomes crucial in ensuring that individuals are aware of how their personal information is used and can choose to do so only in the ways they agree with. Data use, production and reliance are widespread and often play an active role in society's regular functioning, whether employed by online vendors or public administration. Though individuals repeatedly find themselves performing a privacy calculus, they find they often *must* agree to unfavourable conditions which results in paradoxical behaviour. It is easy to see why they would experience fatigue and disengage from privacy issues.

To look into whether these feelings can be mitigated by boosting data protection literacy, an empirical approach was deemed the most effective to verify participant concerns, behaviours and attitudes in a way that did not cause them to disengage. This thesis aimed to establish whether a relationship between data protection literacy and privacy fatigue exists, and to understand whether the former could limit the latter. A questionnaire was administered to an exploratory sample with the

purpose of researching this relationship in a non-representative manner. Chapter 2 has discussed the methodology for both the design and analysis phases.

The qualitative analysis carried out in Chapter 3 found that respondents were concerned about their privacy; data protection literacy levels were fairly consistent; privacy fatigue was present, while cynicism was low. When cross-referenced, a negative relationship between data protection literacy and privacy fatigue, though weak and inconsistent, did appear. This shows that teaching digital citizens how to use digital devices beyond basic use, and in particular ensuring they "know how" and "know that", does have an impact on how fatigued they feel.

These results prompted a discussion on data protection literacy, reflecting on how (and when) individuals are taught essential digital skills such as data protection and how these skills are assessed; Chapter 4 discussed how these factors may impact literacy levels. Patterns within literacy and fatigue were identified, and potential reasons and factors determining these correlations were examined. Taking these differences into account is crucial in ensuring that data protection education is effective: though everyone must be educated, noticing which demographic factors affect literacy ensures those at a disadvantage can achieve the same level of competence and not suffer the consequences of these differences later on in life.

Future research should aim to go beyond simply rectifying this work's limitations and verifying whether a relationship between data protection literacy and privacy fatigue exists. Though a number of paths can be envisioned, exploring if different literacy levels result in varying levels of fatigue could offer practical insight on how to make effective data protection programmes.

This thesis has shown that privacy fatigue has an impact, to an extent, on how individuals approach the issues of data protection and how literacy in this field has a critical role in countering disengagement. As more complex technologies are introduced, implementing effective data protection literacy frameworks and systems will ensure individuals have the knowledge and abilities to adapt to these technologies and protect their data how they want to.

Bibliography

Barassi, Veronica. *I Figli Dell'algoritmo: Sorvegliati, Tracciati, Profilati Dalla Nascita*. Luiss University Press, 2021.

Barth, Susanne, and Menno D.T. De Jong. 'The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review'. *Telematics and Informatics* 34, no. 7 (November 2017): 1038–58. https://doi.org/10.1016/j.tele.2017.04.013.

Beaulieu, Anne, and Sabina Leonelli. *Data and Society: A Critical Introduction*. London: Sage Publications Ltd, 2022.

Bleidorn, Wiebke, Ruben C. Arslan, Jaap J. A. Denissen, Peter J. Rentfrow, Jochen E. Gebauer, Jeff Potter, and Samuel D. Gosling. 'Age and Gender Differences in Self-Esteem—A Cross-Cultural Window.' *Journal of Personality and Social Psychology* 111, no. 3 (September 2016): 396–410. https://doi.org/10.1037/pspp0000078.

Caltrider, Jen, Misha Rykov, and Zoë MacDonald. 'What Data Does My Car Collect About Me and Where Does It Go?' *Privacy Not Included, 6 September 2023. https://foundation.mozilla.org/en/privacynotincluded/articles/what-data-does-my-car-collect-about-me-and-where-does-it-go/.

Choi, Hanbyul, Jonghwa Park, and Yoonhyuk Jung. 'The Role of Privacy Fatigue in Online Privacy Behavior'. *Computers in Human Behavior* 81 (April 2018): 42–51. https://doi.org/10.1016/j.chb.2017.12.001.

Choi, Moonsun. 'A Concept Analysis of Digital Citizenship for Democratic Citizenship Education in the Internet Age'. *Theory & Research in Social Education* 44, no. 4 (October 2016): 565–607. https://doi.org/10.1080/00933104.2016.1210549.

Common Sense Education. 'DigCit Curriculum'. Accessed 9 September 2024. https://www.commonsense.org/education/digital-citizenship/curriculum.

Council of Europe. 'The Concept - Digital Citizenship Education (DCE)'. Accessed 19 June 2024. https://www.coe.int/en/web/digital-citizenship-education/the-concept.

Criado-Perez, Caroline. *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Chatto & Windus, 2019.

Demopoulos, Alaina. '"Scanners Are Complicated": Why Gen Z Faces Workplace "Tech Shame"'. *The Guardian*, 28 February 2023, sec. Technology. https://www.theguardian.com/technology/2023/feb/27/gen-z-tech-shame-office-technology-printers.

Desimpelaere, Laurien, Liselot Hudders, and Dieneke Van De Sompel. 'Knowledge as a Strategy for Privacy Protection: How a Privacy Literacy Training Affects Children's Online Disclosure Behavior'. *Computers in Human Behavior* 110 (September 2020): 106382. https://doi.org/10.1016/j.chb.2020.106382.

'Digital Citizen'. In *Cambridge Advanced Learner's Dictionary & Thesaurus*. Cambridge University Press. Accessed 19 June 2024. https://dictionary.cambridge.org/us/dictionary/english/digital-citizen.

Ehrlinger, Joyce, and David Dunning. 'How Chronic Self-Views Influence (and Potentially Mislead) Estimates of Performance.' *Journal of Personality and Social Psychology* 84, no. 1 (2003): 5–17. https://doi.org/10.1037/0022-3514.84.1.5.

ENISA. 'Authentication Methods'. Page. Accessed 27 August 2024. https://www.enisa.europa.eu/topics/incident-response/glossary/authentication-methods.

Ensmenger, Nathan. '"Beards, Sandals, and Other Signs of Rugged Individualism": Masculine Culture within the Computing Professions'. *Osiris* 30, no. 1 (January 2015): 38–65. https://doi.org/10.1086/682955.

European Commission. 'What Is Personal Data?' European Commission. Accessed 25 August 2024. https://commission.europa.eu/law/law-topic/data-protection/reform/what-personal-data_en.

European Commission. Directorate General for Communications Networks, Content and Technology. and Kantar. *Digital Rights and Principles: Report*. LU: Publications Office, 2021. https://data.europa.eu/doi/10.2759/30275.

Eurostat. 'ICT Usage in Households and by Individuals (Isoc_i)'. Accessed 25 August 2024. https://ec.europa.eu/eurostat/cache/metadata/en/isoc_i_esms.htm.

Foster, Sally. 'Australian Undergraduate Internet Usage: Self-Taught, Self-Directed, and Self-Limiting?' *Education and Information Technologies* 5, no. 3 (2000): 165–75. https://doi.org/10.1023/A:1009602617991.

GDPRhub. 'Article 4 GDPR'. Accessed 25 August 2024. https://gdprhub.eu/index.php?title=Article_4_GDPR.

Gilster, Paul. *Digital Literacy*. Wiley Computer Publishing. New York Chichester: Wiley, 1997.

Goodson, Scott. 'If You're Not Paying For It, You Become The Product'. Forbes, 3 March 2012. https://www.forbes.com/sites/marketshare/2012/03/05/if-youre-not-paying-for-it-you-become-the-product/.

Harvard Health. 'Personalized Activity Intelligence: A Better Way to Track Exercise?', 27 January 2017. https://www.health.harvard.edu/blog/personalized-activity-intelligence-better-way-track-exercise-2017012711031.

Hirsch, Jerry. 'Elon Musk: Model S Not a Car but a "Sophisticated Computer on Wheels"'. Los Angeles Times, 19 March 2015. https://www.latimes.com/business/autos/la-fi-hy-musk-computer-on-wheels-20150319-story.html.

'How to Create a GDPR Compliant Form'. Accessed 22 September 2024. https://tally.so/help/how-to-create-a-gdpr-compliant-form.

ICDL Global. 'Computer and Online Essentials'. Accessed 12 September 2024. https://icdl.org/workforce/computer-and-online-essentials/.

ICDL Global. 'Digital Citizen'. Accessed 12 September 2024. https://icdl.org/digital-citizen/digital-citizen/.

IDC. 'Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2020, with Forecasts from 2021 to 2025 (in Zettabytes).' Chart. Statista, 7 June 2021. https://www.statista.com/statistics/871513/worldwide-data-created/.

'Italy'. In *The World Factbook*. Central Intelligence Agency, 13 August 2024. https://www.cia.gov/the-world-factbook/countries/italy/#people-and-society.

Johnson, Clay A. *The Information Diet: A Case for Conscious Consumption*. O'Reilly, 2012.

Kirschner, Paul A., and Pedro De Bruyckere. 'The Myths of the Digital Native and the Multitasker'. *Teaching and Teacher Education* 67 (October 2017): 135–42. https://doi.org/10.1016/j.tate.2017.06.001.

Koppi, Tony, Fazel Naghdy, Joe F. Chicharo, Judy Sheard, Sylvia Edwards, and David Wilson. 'The Crisis in ICT Education: An Academic Perspective'. In *Hello! Where Are You in the Landscape of Educational Technology?*, 2008. http://www.ascilite.org.au/conferences/melbourne08/procs/koppi.pdf.

Leydet, Dominique. 'Citizenship'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2023. Metaphysics Research Lab, Stanford University, 2023. https://plato.stanford.edu/archives/fall2023/entries/citizenship/.

'Literacy'. In *Cambridge Advanced Learner's Dictionary & Thesaurus*. Cambridge University Press, n.d. https://dictionary.cambridge.org/dictionary/english/literacy.

Litman-Navarro, Kevin. 'Opinion | We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.' *The New York Times*, 12 June 2019, sec. Opinion. https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html, https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html.

McClain, Colleen, Michelle Faverio, Monica Anderson, and Eugenie Park. 'How Americans View Data Privacy'. Pew Research Center, 18 October 2023. https://www.pewresearch.org/internet/2023/10/18/how-americans-protect-their-online-data/.

McKinsey & Company. 'Barriers to Use Technology among Older Adults Worldwide in 2023, by Type and Age Group'. Chart. Statista, May 2023. https://www.statista.com/statistics/1387165/older-adults-barriers-technology-use/.

Mejias, Ulises A., and Nick Couldry. 'Datafication'. *Internet Policy Review* 8, no. 4 (29 November 2019). https://doi.org/10.14763/2019.4.1428.

Morse, Jack. 'How to Stop Spotify from Sharing Your Data, and Why You Should'. Mashable, 5 April 2022. https://mashable.com/article/spotify-user-privacy-settings.

Mossenberger, Karen, Caroline J. Tolbert, and Ramona S. McNeal. *Digital Citizenship: The Internet, Society, and Participation*. Cambridge, Massachusetts: MIT Press, 2008.

NordPass. 'How Many Passwords Does the Average Person Have?' Accessed 27 August 2024. https://nordpass.com/blog/how-many-passwords-does-average-person-have/.

NordPass. 'Top 200 Most Common Passwords'. Accessed 28 August 2024. https://nordpass.com/most-common-passwords-list/.

Orellana, Vanessa. 'My Apple Watch Saved My Life: 5 People Share Their Stories'. CNET, 9 September 2020. https://www.cnet.com/tech/mobile/apple-watch-lifesaving-health-features-read-5-peoples-stories/.

Pagella Politica. 'Sette grafici per capire la crisi demografica in Italia', 12 December 2022. https://pagellapolitica.it/articoli/crisi-demografica-italia.

Pickup, Oliver. 'Gen Z Workers Are Not Tech-Savvy in the Workplace – and It's a Growing Problem'. *WorkLife*, 14 December 2022. https://www.worklife.news/technology/myth-buster-young-workers-are-not-tech-savvy-in-the-workplace-and-its-a-growing-problem/.

Prensky, Marc. ‘Digital Natives, Digital Immigrants Part 1’. *On the Horizon* 9, no. 5 (September 2001): 1–6. https://doi.org/10.1108/10748120110424816.

Sheller, Mimi. ‘Automotive Emotions: Feeling the Car’. *Theory, Culture & Society* 21, no. 4–5 (2004): 221–42.

SimilarWeb. ‘Most Popular E-Commerce and Shopping Websites in Italy in December 2023, Based on Share of Visits.’ Chart. Statista, 1 January 2024. https://www.statista.com/statistics/1256072/italy-visit-share-leading-ecommerce-websites/.

Surfshark. ‘Number of User Accounts Exposed Worldwide from 1st Quarter 2020 to 4th Quarter 2023 (in Millions)’. Chart. Statista, January 2024. https://www.statista.com/statistics/1307426/number-of-data-breaches-worldwide/.

Teuchmann, Katja, Peter Totterdell, and Sharon K. Parker. ‘Rushed, Unhappy, and Drained: An Experience Sampling Study of Relations between Time Pressure, Perceived Control, Mood, and Emotional Exhaustion in a Group of Accountants.’ *Journal of Occupational Health Psychology* 4, no. 1 (January 1999): 37–54. https://doi.org/10.1037/1076-8998.4.1.37.

The Economist. ‘The World’s Most Valuable Resource Is No Longer Oil, but Data’. 6 May 2017. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

Trepte, Sabine, Doris Teutsch, Philipp K. Masur, Carolin Eicher, Mona Fischer, Alisa Hennhöfer, and Lind. ‘Do People Know About Privacy and Data Protection Strategies? Towards the “Online Privacy Literacy Scale” (OPLIS)’. In *Reforming European Data Protection Law*, Vol. 20. Law, Governance and Technology Series. Dordrecht: Springer Netherlands, 2015. https://link.springer.com/10.1007/978-94-017-9385-8.

UNESCO. ‘Literacy: What You Need to Know’. Accessed 15 May 2024. https://www.unesco.org/en/literacy/need-know.

Winner, Langdon. ‘Do Artifacts Have Politics?’ *Daedalus* 109, no. 1 (1980): 121–36.

Zafeiropoulou, Aristea M., David E. Millard, Craig Webber, and Kieron O’Hara. ‘Unpicking the Privacy Paradox: Can Structuration Theory Help to Explain Location-Based Privacy Decisions?’ In *Proceedings of the 5th Annual ACM Web Science Conference*, 463–72. Paris France: ACM, 2013. https://doi.org/10.1145/2464464.2464503.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile books, 2019.

Appendix A – Demographic data

Measure	Item	Count	%[81]
Gender	Woman	91	39.39
	Man	139	60.17
	Non-binary	1	0.43
	Prefer not to answer	0	0
Age	18 or younger	1	0.43
	18 – 24	41	17.75
	25 – 34	20	8.66
	35 – 44	9	3.90
	45 – 54	26	11.26
	55 – 64	87	37.66
	65 and over	47	20.35
Education level	Middle school diploma	4	1.73
	High school diploma	49	21.21
	Bachelor's degree	43	18.61
	Master's degree	35	15.15
	Single cycle degree or pre-Bologna reform degree	95	41.13
	None of the above	5	2.16

[81] Percentages may not add up to 100% due to rounding.

Employment status	Student	35	15.15
	Employed	134	58.01
	Unemployed, looking for a job	3	1.30
	Homemaker	2	0.87
	Retired	53	22.94
	Other status	4	1.73
Time spent on the Internet daily	Less than 1 hour	9	3.90
	1 – 2 hours	34	14.72
	2 – 3 hours	50	21.65
	3 – 4 hours	44	19.05
	4 – 5 hours	30	12.99
	5 hours or more	64	27.71
Age of first Internet use	Under 18	69	29.87
	18 - 24	25	10.82
	25 - 34	44	19.05
	35 - 44	51	22.08
	45 - 54	22	9.52
	55 - 64	13	5.68
	65 or older	1	0.43
	Invalid answers	6	2.60

Appendix B – Questionnaire items

Variable	Item	Source
General	Q1. What aspect of privacy do you believe is most important to you? Quale aspetto della privacy ritieni essere più importante per te?	Own elaboration
	Q2. Which of these do you consider to be personal data? Quali tra questi ritieni essere dati personali?	EU Commission
	Q3. Have you carried out one or more of the following to manage access to your personal data on the Internet? Hai mai fatto una o più delle seguenti attività per gestire l'accesso ai tuoi dati personali su Internet?	EUROSTAT
Privacy concern	Q4. I am concerned that the information I submit to online vendors could be misused Mi preoccupa il fatto che le informazioni che invio a venditori online possano essere utilizzate in maniera impropria	Choi et al., 2018
	Q5. I am concerned that a person can find private information about me on the Internet Mi preoccupa il fatto che una persona possa trovare informazioni private su di me su Internet	
	Q6. I am concerned about providing personal information to online vendors, because of what others might do with it Mi preoccupa fornire dati personali a venditori online, per via di ciò che altri potrebbero farci	
	Q7. I am concerned about providing personal information to online vendors, because it could be used in a way I did not foresee Mi preoccupa fornire dati personali a venditori online, perché potrebbero essere usati in modi che io non abbia previsto	
Data protection literacy	Q8. I know that EU citizens have a right to the protection of their personal data. So che i cittadini dell'Unione Europea hanno il diritto alla protezione dei propri dati personali.	Vuorikari et al., 2022
	Q9. I know that the privacy policy of an application or service should explain what personal data is collected and inform if it is shared with third parties. So che l'informativa della privacy di un'applicazione o di un servizio dovrebbe spiegare quali dati personali	

Variable	Item	Source
	vengono raccolti e informare se vengono condivisi con terze parti. Q10. I know that it is good to periodically check which applications or services have access to my personal data. So che è bene controllare periodicamente quali applicazioni o servizi hanno accesso ai miei dati personali. Q11. I know how to change my browser settings to prevent or restrict cookies on any device. So come cambiare le impostazioni del mio browser per prevenire o limitare i cookie su un qualsiasi dispositivo. Q12. I know how to verify that the site that requires me to provide my personal information is secure (for example: https sites, logo or security certificate). So come verificare che il sito che mi richiede di fornire i miei dati personali sia sicuro (ad esempio: siti https, logo o certificato di sicurezza). Q13. I know how to change the privacy settings on the sites I use most. So come modificare le impostazioni della privacy sui siti che uso maggiormente.	
Privacy fatigue	Q14. I feel emotionally drained from dealing with privacy issues in an online environment. Mi sento emotivamente esausto dall'affrontare questioni riguardanti la privacy in ambienti online. Q15. I am tired of online privacy issues Sono stufo di questioni relative alla privacy online. Q16. It is tiresome for me to care about online privacy. È stancante interessarmi alla privacy online. Q17. I have become less interested in online privacy issues. Mi interesso meno alle questioni riguardanti la privacy. Q18. I have become less enthusiastic in protecting personal information provided to online vendors Sono meno entusiasta di proteggere le informazioni personali fornite a venditori online. Q19. I doubt the significance of online privacy issues more often Ho dubbi sull'importanza di questioni della privacy online più spesso.	Choi et al., 2018