DaPIS: a Data Protection Icon Set to Improve Information Transparency under the GDPR

Arianna Rossi[1]

Università di Bologna, CIRSFID, via Galliera 3, 40121 Bologna, Italy

Monica Palmirani[2]

Università di Bologna, CIRSFID, via Galliera 3, 40121 Bologna, Italy

January 21, 2019

[1]Electronic address: arianna.rossi15@unibo.it
[2]Electronic address: monica.palmirani@unibo.it

Contents

Abstract

Privacy policies are known to be impenetrable, lengthy, tedious texts that are hardly read and poorly understood. Therefore, the new EU legal framework for data protection, the General Data Protection Regulation (GDPR), introduces provisions to enhance information transparency and suggests icons as visual means to clarify data practices. Notwithstanding the many benefits in terms of e.g. comprehension that legal visualizations demonstrably provide, visual communication can take many shapes and show its effectiveness on different levels. The scientific debate around graphical symbols for legal concepts is still in its infancy, whilst both the creation and consequent evaluation of icons depicting abstract or unfamiliar concepts represent a challenge. Chances of misinterpretation would undermine the final goal of transparency. Moreover, precision of representation can support the individuals' sense-making, but at the expense of simplicity and usability.

In this report, we present the research that led to the design of DaPIS, the Data Protection Icon Set that we created and evaluated through human-centered methods drawn from the emerging discipline of legal design. Firstly, we have organized rounds of participatory design sprints where designers and lawyers collaborated side by side. Then, we ran some user studies to empirically determine strengths and weaknesses of the icon set as communicative means for the legal sphere.

The icon set is modeled on PrOnto, an ontological representation of the GDPR, and is organized around its core modules: personal data, roles and agents, processing operations, processing purposes, legal bases, and data subjects' rights. In combination with the description of a privacy policy in the legal standard XML Akoma Ntoso, such an approach makes the icons machine-readable and semi-automatically retrievable. Icons can thus serve as information markers in lengthy privacy statements and help the reader to navigate through the text. In this way, we aim to map and connect different representations of legal information to enhance its comprehensibility: the lawyer-readable, the machine-readable, and the human-readable levels.

Acknowledgements

The research described in this report has been carried out in cooperation with the following institutions:

Stanford University , Legal Design Lab. We gratefully acknowledge Margaret Hagan;

Accademia di Belle Arti di Bologna . We gratefully acknowledge Prof. Danilo Danisi and (in alphabetical order) Sara Cominelli, Grazia Coppolecchia, Nicoletta Guastalla, Francisca Lessio, Valentina Marafioti, Emy Michelin, Alyssa Scroffa, Vladimir Senyuk, Francesca Velleca, Nikolas Vigorito and Federica Zanasi;

University of Bologna , CIRSFID. We gratefully acknowledge (in alphabetical order) Giorgia Bincoletto, Dino Gerardi, Dina Ferrari, Michele Martoni, Maria Livia Rizzo, Michela Rossi and Salvatore Sapienza;

University of Luxembourg . We gratefully acknowledge Prof. Leon van der Torre.

The research has also been supported by the Marie Sklodowska-Curie Mining and Reasoning with Legal Texts (**MIREL**) project and we gratefully acknowledge Livio Robaldo.

We would also like to thank Chiara Cavazzuti from the Accademia di Belle Arti di Firenze and Claudia Busetto from Commonground Srl for their valuable contributions.

The following individuals have provided precious comments to the research described in this report and are gratefully acknowledged (in alphabetical order): Colette R. Brunschwig, University of Zurich; Rossana Ducato, Université Catholique Louvain; Helena Haapio, Aalto University and Lexpert; Stefania Passera, Aalto University. Finally, all the participants in the "Law & Design for Privacy" workshops and in the user studies, alongside those that we have not mentioned, are gratefully acknowledged.

The following organizations have sponsored the research described in this report and are gratefully acknowledged:

EACEA : the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Commission that financed the Joint International Doctoral Degree (PhD) in Law, Science, and Technology (LAST-JD);

SIIG : Società Italiana Informatica Giuridica (SIIG).

Executive Summary

This report describes DaPIS, the Data Protection Icon Set that was explicitly designed to respond to the General Data Protection Regulation's (GDPR) call for machine-readable, standardised icons to give "in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing" (Article 12.7). Indeed, privacy communication is generally considered lengthy, cumbersome and unable to inform data subjects about the collection and processing of their personal data or about their rights. Icons can contribute to the aim of improving the transparency of privacy statements.
Indeed, graphical symbols are deemed capable to convey meanings universally. However, their ease of interpretation depends on several factors. At the individual level, dimensions such as familiarity, semantic distance, concreteness, and complexity of the icon must be explored, whereas also discriminability and coherence across symbols are criteria that should be considered if an icon set is created. Moreover, user's characteristics such as culture and level of experience also influence the interpretation process. Icons for legal matters present an additional challenge compared to the majority of graphical symbols in use, that depict concrete objects: legal icons mostly convey abstract meanings and represent unfamiliar referents. For this reason, even their evaluation can be demanding, since arbitrariness or lack of familiarity cause low recognition rates at first exposures. There have been some attempts to design icon set for data protection (most notably, the PrimeLife project), on which DaPIS partially build. However, DaPIS is designed to represent those concepts that are proper of the new EU legal framework for data protection in a machine-readable manner. It is therefore modeled on an ontological representation of the GDPR, that can be connected and, thus, provide meaning to a machine-interpretable representation of documents containing information on privacy and data protection. The icon set is therefore composed of modules derived from the ontology and considered relevant for the data subject: types of data, agents and roles, processing operations, data subjects' rights, processing purposes, and legal bases for processing.

Such a computational approach is coupled with the human-centered methods of the emerging discipline of Legal Design in a twofold manner, both in the phase of creation of the icon set and in the subsequent phase of evaluation. DaPIS was generated through a series of participatory design workshops, where interdisciplinary teams (mostly composed of lawyers and designers) confronted themselves with the tough challenge of creating small graphical symbols to convey complex and nuanced legal meanings. Two subsets of icons were thus created and each underwent evaluation in a user study to determine their legibility and fitness for the corresponding referent. Since some of the symbols are inherently arbitrary, thus their meaning could not be immediately evident, it was also researched whether the user could understand the reasons behind certain iconographical choices, i.e. if she could align her mental model with that of the designers.
It came at no surprise that the symbols that received higher scores represent concrete objects, familiar concepts or are based on familiar representations (e.g. the 'i' signifying information). Conversely, the concepts behind the icons that scored worst are vague, general, and abstract (e.g. the purpose of provision of the service). There is, thus, the necessity to determine a threshold of acceptability, by especially considering the intercultural nature of the EU residents and their varying levels of experience with data protection matters. Some iconographical choices are inescapably arbitrary, while some others are uninformative if it is the underlying referent to be unknown. For such reasons, standardization of the symbols and education of data subjects about digital life are essential measures to be adopted at the EU level. This report presents the initial stages of the research on icons for data protection, but several other directions outlined at the end must and will be explored.

Introduction

Privacy policies are aimed to inform individuals about the practices of collection and processing of their personal data by a certain organization. Such information is necessary to understand and, if necessary, challenge the processing operations carried out on their personal data [7]. However, privacy notices are known to be unclear, lengthy, impenetrable, tedious texts that are hardly read and poorly understood (*see* e.g. [9, 28, 47]. As in much legal drafting, the information around the collection and processing of data is written to merely fulfill the legal requirement of mandated disclosure, instead of effectively inform data subjects about the collection and processing of their personal data [23].

The General Data Protection Regulation[1] [16], among its many objectives, aims to rebalance the power and information asymmetry between data subjects and data controllers, namely those organizations that collect personal data. This is why transparency is introduced as an overarching obligation that applies to any communication addressed to data subjects and as a crucial element of the principles of fairness and lawfullness of processing [7]. For the first time, under the GDPR, not only the content, but even the quality, accessibility, and comprehensibility of any communication addressed to data subjects assume an unprecedented role to demonstrate compliance with the principle of transparency, which can be effected not only through verbal means but also through visualisation tools [7].

Under this light, the GDPR suggests to provide information in combination with machine-readable, standardised icons (Article 12.7) to give "in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing"[2]. This approach is aimed at reducing excessive amounts of written information [7]. Although eventually it will be the role of the European Commission to adopt delegated acts to give directions on the creation of these icons, the need of expert advice is emphasized in Recital 166 GDPR and by the Article 29 Working Party, which encourages an "evidence-based approach" and "extensive research" [7, p.26] to inform the development and application and determine the efficacy of icons in this context.

This report intends to be a contribution to such preparatory work and to the scarce discussion around data protection icons. Chapter 1 describes the re-

[1] hereafter: GDPR

[2] Similarly, Article 8 of the proposed ePrivacy Regulation

search background that motivated the directions of this research, provides some introductory remarks about the nature of icons, in particular of legal icons, and presents previous attempts to design and evaluate data protection icons. Chapter 2 presents DaPIS, the Data Protection Icon Set designed for the application of the GDPR's transparency principle, and explains the methodology behind its creation, based on a computational ontology and participatory design methods. Chapter 3 describes two user studies for the evaluation of several dimensions of the icon set and their results, based on which the final DaPIS is presented in Chapter 4. Chapter 5 presents conclusions and limitations of the research, whereas Chapter 6 introduces promising future work directions. Lastly, the appendices gather DaPIS (Appendices C and D), excerpts from the user studies (Appendices F and G), and an exemplifying layout for privacy policies with the icons from DaPIS (Appendix **??**).

Chapter 1

Background and Research Scenario

This chapter briefly introduces the functions of icons as communicative devices, but also their potential limitations, especially when used in the legal sphere. Although icons are commonly regarded as elements that can convey meanings effortlessly across cultures, their effectiveness actually depends on many (intrinsic and extrinsic) factors. Such elements must be taken into account during the evaluation phase to properly reveal strengths and weaknesses of the icon set.

1.1 On the Nature of Icons

Icons are attractive communicative devices because they can be easily recognized, processed and memorized. They can serve as memory devices and help in the classification of content [37]. Icons are deemed to communicate in a non-verbal manner quickly, concisely, and across languistic and cultural differences [35]. However, one must be cautious about these claims (*see* e.g. [14]). For instance, **familiarity** with the icon is a relevant dimension to preview ease of access to memory and time of recognition, and has a twofold dimension: it is affected by previous experience with the graphical symbol, but also with the symbol's underlying concept.

Icons can be also highly diverse in terms of their **semantic distance**, which determines their interpretability and ease of learning [35]. They can be placed on a continuum that ranges from resemblance icons (that depict objects), to exemplar icons (that portray an individual of a class), to symbolic icons (that convey a concept on a higher level of abstraction) to arbitrary icons (that have no relationship to objects or concepts) [37]. Moving towards the end of the spectrum, the semantic transparency of the symbols shrinks [40] and their meaning must be learned rather than deduced [29]. Moreover, the function assigned to a graphical symbol by its designer can be different from the meaning attributed to it in practice, i.e. there can be misalignments between the designers' intentions

and the sense-making activity of the user.

For new exposures, ease of identification also depends on the icon's **concreteness**, which is the extent to which real objects, materials, or people are depicted [35]. This effect diminishes as users gain experience over the icons, though. An additional aspect that must be considered is **complexity**, i.e. the amount of icon's details, that influences search activity, but not identification. Finally, when creating a set of icons, attention should be devoted to the degree of discriminability of one icon from the others of the set and to coherence across set elements [14]: the same graphical symbol should be used consistently to signify the same meaning, whereas it should be enough distinctive to be easily identified.

By considering these dimensions and the visualizability of the underlying concept [53], an icon's cognitive effectiveness (speed, ease, and accuracy of interpretation) [40] can be roughly estimated even before an empirical evaluation. If these dimensions are not carefully considered, there exists the risk that users will process the visual representations more slowly, with more difficulty and with less success compared to written text. By doing so, obscurity in lieu of transparency would be achieved.

1.2 On the Nature of Legal Icons

The Creative Commons[1] and the highway code are classical examples of visual means that univocally convey legal matters: the first about intellectual property rights, whilst the second about permissible or mandated dviers' behaviour. Contrary to what is commonly believed, however, these iconic systems do not constitute an instant language because of their universal communicative power, but rather thanks to campaigns of standardization and the stablishment of international conventions [15]. Moreover, individuals get acquainted with these codes over time. For what concerns street signs, they get explicit education about the meanings of the graphical language, on which they must even take an exam to be considered expert users.

Indeed, icons have limited self-explanatory nature [31]: decoding these pictograms requires context and learned knowledge (e.g. cultural knowledge). Icons that convey abstract meanings, such as data practices, might not be universally understood if they are not accompanied by some textual explanations [51]. Usability tests [31, 46, 32] show that "critical confusions" [52], namely misinterpretations opposite to the intended meaning, are possible due to multiple reasons: misalignment between designers' intentions and users' expectations on the icon meaning and differences in individuals' level of education, age, and cultural background. This matter has great relevance if individuals take legally-binding decisions on the basis of the visualizations, such as entering into a contract with a service provider or giving consent to certain data practices. Indeed, legal meaning encoded in pictures is open to multiple interpretations [11] and there exist serious concerns that pictograms cannot represent legal concepts and

[1] `https://creativecommons.org/`.

norms in terms of details and adequateness as words would do. Thus, businesses do not reasonably want to risk misunderstanding of the pictorial representations or oversimplification of their privacy terms, because this might cause liability issues [26].

As recalled earlier, familiarity is a critical component to determine ease of recognition of an icon. This is why good practice for icon design is to rely on an established visual vocabulary [29]. However, this proves difficult in the legal sphere because there exist only a few, overly preponderant, law-related symbols, such as the scale and the gavel. As for what concerns data protection, only a few symbols around (cyber)security are well-known, such as the shield. Different is the case of technology-related visuals, since the widespread use of graphical user interfaces has favored the creation of mental references between a number of icons and their functions (e.g. a pencil for the edit function).

Also familiarity with the concept underlying the icon plays a fundamental role: if the concept is unknown to the interpreter, as it is generally the case with legal matters, then the icon must possess a low level of arbitrariness to easily shed light on its underlying meaning. An additional difficulty is posed by the fact that legal concepts are usually abstract in nature, so it becomes even more difficult to depict them.

Such characteristics also challenge classical evaluation methods, which are mainly suited to determine the comprehensibility of graphical symbols whose referent is known to the user (see also Chap. 3). Differences of comprehension rates are to be expected, because they depend on the intrinsic icon's characteristics, i.e. familiarity, concreteness, and semantic distance, but also on the characteristics of the person that interprets them, i.e. culture, age, etc. Finally, researchers underline how important is the provision of contextual cues that mirror the actual usage situation of the icons (ecological validity [35]) to support the sense-making process of individuals. Without such precautions, low recognition scores would falsely indicate that more design and test work is necessary [54].

1.3 On the Role of Data Protection Icons

Multiple visualization patterns for the legal matters have been designed and experimented [24], each having its own way of encoding (different kinds of) information and of conveying knowledge. Icons belong to the category of the visual representation patterns, which help to explain, complement, and disambiguate the (legal) text: it has been indeed demonstrated that the presence of visual elements can improve comprehension accuracy and speed, compared to pure text [45]. "Companion icons" are "graphic symbols that represent the meaning or function of the textual element they accompany" [24, p. 26]. They help readers to search and find relevant information quickly, especially in long and undifferentiated texts such as privacy policies. Thus, icons can in principle highlight and quickly communicate the key aspects of the privacy practices of

an organization [23]: it is commonly believed that "a privacy icon[2] is worth a thousand-word policy[3]".

However, it is a common misconception in the legal sphere that icons, or visual elements more in general, should substitute words and text completely [15]. Rather than substituting the legal text, data protection icons can integrate it and act as information markers, namely to help the reader to quickly navigate or skim through long texts [45]. Used in combination with a structured layout, they can help data subjects to quickly find specific information items and, thus, to exercise strategic reading. They can also attract the attention of the reader, fight information fatigue, and help to memorize information. In principle, they can even provide a short summary of the privacy practices at a glance. It is however questionable whether they should also provide a judgment on the fairness of terms (see also next Section and [46, 22]).

In the interpretation of the Article 29 Working Party, the icons are meant to enhance transparency by reducing the extreme amount of information and, upon standardization, to be used across the continent as universal shorthand for that information [7]. However, reducing the complexity and the potentially infinite combinations of linguistic terms into a limited set of icons is impossible and non-functional (for critical remarks, see also [39]).

1.4 Previous Efforts to Design and Evaluate Data Protection Icons

A few privacy-related icon sets already exist and vary deeply in nature: (i) as for what concerns the types of information that they represent; (ii) whether or not they represent a legal assessment about the fairness of the terms they represent, and for (iii) the regulatory framework they refer to (EU or USA) [49]. Most of them have not undergone any evaluation with users about their ease of comprehension, nor other dimensions. For the purposes of European data protection, two attempts need to be mentioned, which also provide the opportunity to discuss the evaluation methods applied to data protection icons.

1.4.1 The PrimeLife project

The PrimeLife project [22] is notably the most structured attempt to design and evaluate icons for data protection in the European context. The first icon set produces during the project comprises symbols representing data processing steps of various kinds, types of data, processing purposes and categories of recipients (the latter only for social network). However, most of the icons produced during this first step of the project were discarded during the testing

[2]The literature generally refers to icons depicting concepts related to data practices as "privacy icons". However, they mostly represent concepts of data protection, thus the term is inexact. In the present report, the expression "data protection icons" will be preferred.

[3]see the "privacy icon pattern" on `https://privacypatterns.org` and `https://privacypatterns.eu`.

phase. Although the user study [19] highlighted how visual vocabulary depends on culture, therefore calling for intercultural user audiences, it is difficult to determine its ecological validity since it did not seem to provide much context to support the sense-making of the test participants. In general, icons with labels were better understood than the same icons without labels. No specific numbers about the results are provided, but it comes at no surprise that the icons that scored best (medical data, payment data, storage, deletion, etc.) refer to more concrete and more familiar referents, whilst the less recognized icons (such as anonymization, user tracking, etc.) depict less familiar and concrete concepts.

Then, another test with a wider audience was conducted [22]. It was asked either to decide between two/three possible alternatives or to rate icons according to their comprehensibility, clearness, and feasibility. Participants could even add comments on their own and elaborate on reasons for critique or approval of the icons [31]. Some principles that emerged were: simplification of the elements is crucial, as well as uniformity of the design styles. The PrimeLife's researchers end with a negative note: given the low results in both user studies, only a few icons were deemed appropriate to be included in the final icon set. Among their tested icons, the researchers suggest only four of them: third party sharing, storage period, third party tracking and behavioral targeted advertising (plus three icons about data disclosure in social network sites). Many other processing steps, data types and recipient groups were deemed too hard to illustrate. Although the goal of the PrimeLife project was the creation of icons for an interface, it seems that no test in context was carried out. This could have sparked higher results.

1.4.2 Icons in the Draft Report on the Regulation Proposal

During the parliamentary discussion on the GDPR, a table with 6 icons [17] was proposed to summarize the main data practices of a data controller. The display of such icons would have constituted a legal obligation for websites, were the amendments approved[4]. Instead of "neutrally" translating privacy notions into visuals, these icons symbolize assessments about the website's compliance with six basic data protection guarantees.

The comprehensibility test carried out on these icons [46] shows some shortcomings as for what concerns its ecological validity, as the research's author acknowledges. The first part of the test asked for the icons' meaning without providing any contextual reference. The second part of the test was a matching task between icon and correspondent textual explanation, where no one-to-one association was asked to determine if multiple matches, meaning confusion, would occur. The test also pointed out important results: for example, the use of the combination of two icons for a double negation was not easily understandable.

[4]Trace of this proposition can be found in the GDPR's call for icons.

1.4.3 Neutral Representation of Concepts or Assessment on Fairness?

As pointed out by the Primelife project [22], we also share the idea that icons should have a headline function, rather than make a statement about the fairness of processing to reach global acceptance. It can be argued that it would be more meaningful for data subjects (but also for supervisory authorities) to be provided with a visual summary of the risky or less lawful practices conducted by an organization on their data (i.e. a rating), in order to support their decision-making, e.g. if to use a certain service or head elsewhere. However, such an approach encounters the problem that a decision about the lawfulness and fairness of certain practices should be taken and it is questionable who should take it and on the basis of which principles. Moreover, such an approach would be probably opposed by many organizations, since the GDPR proposes icons as a possibility and does not impose an obligation upon controllers: icons rating lawfulness would therefore very difficultly be widely adopted. These are the reasons why the present research adopts a more neutral approach, namely depict notions of data protection to act a information markers, whereas it will be the individual to decide for herself whether to engage with a certain service. As the next Chapter illustrates, depicting concepts is also the most suitable integration with an ontological formalization of legal knowledge.

Chapter 2

The Design of DaPIS

The analysis reported in the previos Chapter highlighted some limitations and issues that the present research aims to tackle, by proposing a methodology to design a privacy icon set [42] that represents core concepts of European Data Protection Law. Notwithstanding the dismal results of the usability tests on previous privacy-related icons, that led the researchers to discard the majority of them, it is hereby argued that previous experiences can be improved and icons to convey legal meanings can be more successful, based on results in comparable research [45]. Therefore, the present Chapter presents DaPIS: the Data Protection Icon Set. The project described in the following employs experimental, human-centered design practices and semantic web technologies to satisfy GDPR's legal requirements about transparent information provision.

The approach proposed in the following has the final aim to semi-automatically display the icons in correspondence of the matching privacy terms (see Appendix **??**). The underlying hypothesis, derived from [45, 24], predicts that this would make statements on specific topics in privacy policies easier to find and understand. Clickable icons can also be employed to signify a data subject's explicit consent to certain practices [8]. A methodology based on three steps is here proposed (see also [42]):

1. Formalization of legal knowledge: Sect. 2.1 defines the objects of representation;

2. Participatory design methods: Sect. 2.2 describes the design process of DaPIS over some participatory design workshops;

3. Evaluation: Chap. 3 describes methods and measures for the assessment of DaPIS.

,

2.1 Formalization of Legal Knowledge

Past data protection icon sets are not based on a systematic formalization of knowledge, but rather focus on data types and a handful of processing operations (see Section 1.4). Moreover, the GDPR introduces several new information items that must be presented to data subjects to enforce the principle of transparency (see Artt. 13-14), such as their rights, for which no graphical representation has been proposed yet.

The GDPR provides a European legal framework that defines concepts of data protection, relations among them, and a common vocabulary to describe them. Since the Regulation demands "machine-readable" icons[1] if presented on electronical means (Art. 12.7), DaPIS is modelled on a computational ontology, i.e. a knowledge base of semantic concepts and logical relations extracted from the GDPR. When a privacy policy is marked up with tags linked to the ontological instances, its semantic content can be described in a machine-readable manner. Icons can be associated to the concept they represent and, hence, be semi-automatically summoned by the semantic tags[2]. Furthermore, the explicit representation of legal knowledge results in the disambiguation of uncertain concepts and the clarification of meanings [43]. The description of legal information in a machine-interpretable format also allows automated reasoning on the legal text (e.g. to draw inferences to match expressions in natural language to the corresponding ontological instance). Finally, an ontology is independent from language, which counts as an additional strength: the same icon can be provided for text spans expressed in different languages, that however refer to the same ontological concepts, whilst correspondent labels in different languages can be provided for the same icon.

PrOnto, an ontological representation of the GDPR, is currently under development [41], based on the analysis of the normative text by legal knowledge engineers. This has been then integrated with expert feedback and authoritative sources, and finally framed in foundational ontologies and existing ontology design patterns [21, 20]. PrOnto has been developed around a few interconnected conceptual cores of EU data protection law (see Table **??**): personal data, roles of agents, processing operations, processing purposes, legal bases, and rights and obligations. Processing operations are carried out on personal data of heterogeneous nature for specific purposes and must rely on a specific legal basis to be lawful. Natural or legal persons can assume different roles depending on different contextual situations, such as data controller or data subject. From the roles derive different rights and obligations. The so-defined basic conceptual modules were fundamental to design the icon set, complemented by the requirements set forth in Articles 13-14.

[1] Although the GDPR does not provide a definition of machine-readable, Recital 21 of Directive 2013/37/EU17 defines it as "a file format structured so that software applications can easily identify, recognize and extract specific data, including individual statements of fact, and their internal structure" see also [7]

[2] Provided the development of such a tool, which was not the goal of the present project, though.

Superclass	**Class**
Personal data types	Original personal data Derived personal data Inferred personal data
Agents' roles	Data subject Data controller Data processor Supervisory authority Third party
Processing operations	Copying Pseudonymization Anonymization Direct marketing Automated decision-making Profiling Encryption Transfer of personal data to third countries
Data subject's rights	Right to be informed Right of access Right to rectification Right to erasure Right to withdraw consent Right to data portability Right to restriction of processing Right to object to processing Right to lodge a complaint
Processing purposes	Research purpose Statistical purpose Purpose of information security Purpose of provision of the service Purpose of service enhancement Marketing purpose Profiling purpose
Legal bases for processing	Consent Legal obligation Vital interest Public interest Legitimate interest Contract

Table 2.1: Conceptual cores of the GDPR ontology, on which DaPIS is based

2.2 Participatory Legal Design Workshops for the Design of DaPIS

The risk of misalignments between designers' intentions and the sense-making activity of individuals oriented the research towards participatory design methods [50] for the creation of the icon set. Indeed, previous research has found that it is arduous for experts to think like non-experts and, thus, symbols created by the target audience are more likely to be correctly interpreted by other members of the target audience, since they share similar mental models and cognitive profiles [13]. In this framework, expert knowledge becomes only one among the many resources to be drawn from, but also active participation by laypeople is sought.

Collaborations among experts in different areas and laypeople can, on the one hand, leverage on the multiple skills and knowledge of the different stakeholders involved. On the other hand, this reduces the chances of personal bias because reciprocal understanding is delibrately sought [10]. Diverse mental models and visual vocabularies derived by different backgrounds and experiences are thus considered. In the three participatory, multi-stakeholders' workshops organized to create DaPIS, multi-disciplinarity was a critical element, thus motley working groups were formed. Each background represents an asset [49]: (1) legal experts that help with the correct interpretation of data protection concepts and that can explain and exemplify their meaning; (2) computer scientists or participants with similar backgrounds that have the technical expertise to understand and explain technical notions included in the data protection law; (3) graphic designers and other professionals from visual disciplines that know the techniques and tools to produce functioning visualizations for the intended audience and the intended medium; and lastly (4) laypeople that add non-expert, but also non trivial views and knowledge to the design process, for instance about the visual conventions they are familiar with.

In participatory design cycles, multiple ways of representing the same concept are collaboratively examined. In our experience, the level of detail of the visualization was source of discussion and represented the main tension between precision of representation (favored by lawyers to avoid oversimplification) and the simplicity (endorsed by designers to attain usability). Unlike other disciplines, design thinking does not aim for a prescriptive theory to generate a single "right" image or layout, because it could be unsuitable for individual needs. Preferable is a collaborative and creative process that tends towards a visualization that "works", given information type and goal of the design [10]. This is why, discussions about low fidelity prototypes in small and then bigger groups is encouraged, as well as intercept interviews with individuals that have not participated in the creation phase. By doing so, shortcomings and strengths of the proposed ideas are identified before the actual creation of high fidelity

icons and chances of failure at later stages (e.g. during the evaluation phase) can be minimized.

The DaPIS was developed during three participatory legal design workshops that are briefly described in the following Sections.

2.2.1 The Design of a First Icons' Subset for Data Types, Agents' Roles, and Processing Operations

A first, exploratory workshop [36] was held with Margaret Hagan at the Legal Design Lab [25] at Stanford Law School in July 2017 . The workshop was structured around the design cycle, over 6 hours. After a presentation of the GDPR and its call for icons, previous attempts to create data protection icons were reviewed. Interdisciplinary groups were formed to create icons for the following classes of concepts (see Table 2.1): (a) data types (e.g. processed data, inferred data, etc.), (b) agents' roles (e.g. data subject, data controller, etc.), (c) processing operations (e.g. copying, transfer of data outside of the EU, etc.) and, finally, (d) the right of access and the right to data portability. At this stage, workshop participants worked with pencils and papers to explore different possible visualizations of such concepts and to generate low-fidelity visuals. These prototypes were tested internally, with the entire workshop group for a critical review, and externally, by conducting five to ten intercept interviews on Stanford campus to get early, unstructured feedback. Eventually, a workshop plenary discussion identified those iconographical elements that deserved to be kept, those that deserved further elaborations and those that needed to be abandoned. A graphic artist then collaborated with the researchers to render the draft icons digitally and to harmonize their style, also thanks to additional, recursive small sample testing conducted on campus.

As a result of the workshop, some basic building blocks were identified to compose the visual vocabulary of data protection and to originate more complex icons or pictograms. Such a compositional approach also derives by the underlying ontological modelling of the concepts. For instance, an arrow with gears means "processing", whereas "personal data" is represented by a folder with a user figure outline atop it. When personal data is processed, the basic personal data folder is combined with the arrow and a more graphically elaborated personal data folder to show the result of the processing activity (e.g. anonymized data). The visual narrative guides the reader from the left-hand side, where the basic personal data folder is showed, to the right-hand side, where the result of the processing activity is shown (e.g. Img. 2.1). Thus, a standard way to combine the visual elements to achieve consistency across the icon set was developed. The need for coherence, precision and completeness across the icon set resulted in some complex icons, that could be defined more as pictograms and visual narratives rather than as single icons. This also derives from one of the tendencies that emerged during the workshop: some groups tried to generate metaphors for some complex notions (e.g. for derived data), which were, however, read literally by users and not swiftly nor correctly interpreted. Thus, literal representations were chosen over their metaphorical counterparts.

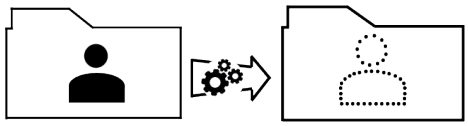


Figure 2.1: Example of icon composition for "anonymization": personal data (the folder on the left-hand side) are processed (the geared arrow) and result in anonymized data (the folder on the right-hand side).

This first icon set is composed of 18 icons (see Appendix C):

1. three icons describing different types of personal information, whose difference is especially relevant in the exercise of data subjects' rights: original, processed, and derived personal data;
2. five icons describing the main agents involved in data processing: data subject, data controller, data processor, third party, and supervisory authority;
3. eight icons describing processes carried out on the data: profiling, direct marketing[3], copying, automated decision-making, encryption, anonymization, pseudonymization, transfer to third countries;
4. two pictograms describing two data subjects' rights: right of access and right to data portability.

The evaluation of this icon subset is described in the next Chapter (see Sect. 3.2).

2.2.2 The Design of the Second Icons' Subset for Data Subjects' Rights, Processing Purposes, and Legal Bases

Two workshops were held at CIRSFID[4], University of Bologna, in March 2018. The first workshop [2] aimed to complete the icon set with the missing classes of concepts identified in the ontology and relevant for the information to be provided to the data subject, whereas the second one [3] aimed to harmonize the design style among the two icon subsets. The vast majority of participants were legal experts or designers from the Academy of Arts in Bologna and the Academy of Arts in Florence.

The two workshops were structured around the design cycle explained above, over 8 hours, and could build on the strengths and weaknesses of the icon subset produced in the first workshop. Thus, the central elements of the visual

[3] Profiling and direct marketing can also be processing purposes.

[4] Interdepartmental Centre for Research in the History, Philosophy, and Sociology of Law and in Computer Science and Law [1]

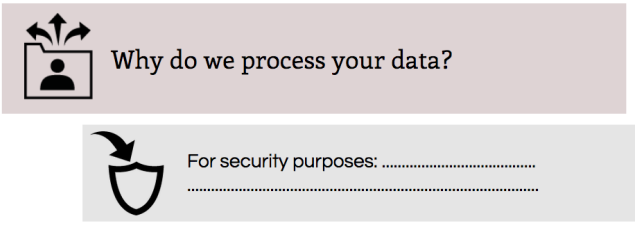


Figure 2.2: Example of icons' systematic use: the arrow exiting the personal folder stands for the general concept of "processing purposes" and heads towards a specific purpose, here e.g. "security purposes".

vocabulary were provided (i.e. the data subject as a user, processing as gears, etc.) to be reused in the creation of the new icons. In order to create a coherent set of icons from the very beginning, three groups composed of a balanced mix of designers and legal experts were formed and to each group was assigned on of the following three classes of concepts: (a) data subjects' rights, (b) processing purposes, and (c) legal bases. Each group received a simplified definition and a practical example for each concept. Furthermore, the intended icons' context of use was specified: an exemplifying privacy policy with a structured layout was distributed (based on the layout shown in Appendix **??**). Providing concepts organized in classes and providing the layout were icons would be inserted served to generate a coherent set. For instance, the very abstract concept of "processing purpose" must be seen in a global view where arrows exit a personal data folder and move towards a specific purpose (see Image 2.2). Finally, instructions to generate ideas and sketch them out on limited space were given to the workshop's participants so that icons would be simple and ready to be displayed in small dimensions, such as on mobile devices.

Attention to balance among simplicity of representation, distinctiveness of some traits, but also coherency of elements across icons was also distributed among the guidelines. For instance, a hand's palm facing up was chosen to indicate any data subject's right. The metaphor underlying the holding hand represents the concept of being in control or having the power over the element located above it. This palm recurs in every data subject's right as common denominator among concepts belonging to the same class, but the meaning of the icon as a whole is specified by the element held by the hand: e.g. a bin for the right to erasure, a pencil for the right to rectification, etc (see Fig. 2.3). The hand element needed to be sufficiently big to be noticed as common denominator across icons, but at the same time sufficiently small to make individuals focus their attention on the distinctive elements (in order to avoid similarity that neutralizes icon's distinctiveness).

The tension among icons' usability and their supposed informative value, which had prompted criticism during the evaluation of the first icon set (see Sec. 3.2), re-emerged persistently during this second workshop. On the one hand,

(a) The icon for the right to erasure

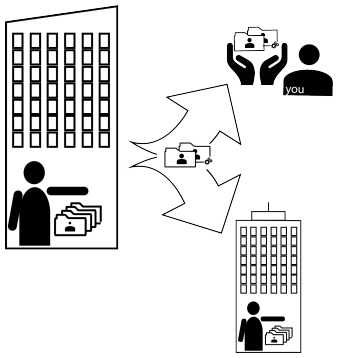
(b) The icon for the right to rectification

Figure 2.3: The icons present a common element, the hand, which stands for the class of subject's rights.

legal experts warned of the risk of misrepresentation or oversimplification when data protection concepts. were visualized. On the other hand, exemplifying individuals representing an entire class and icons containing few details must be preferred to achieve ease of recognition and adaptability to different contexts. For example, the low rates received by the representation of the data controller in the first user study provoked a long and heated discussion. In the previous workshop and after some small sample testing, it was decided to represent the data controller as a person inside a building to convey the idea of a person in charge, inside of an organization. Following the first user study's results and the fact that the controller is a basic element that, combined with others, composes more complex icons (i.e. the legitimate interest of the controller, the contract, etc.), a simpler, even if arguably imprecise, exemplification of the controller was chosen: that of a person dressed as a business man.

Given the previous research on graphical symbols outlined in Chapter 1, it did not come as a surprise that concepts with a higher degree of concreteness (e.g. "contract"), for which an exemplification could be easily provided (e.g. "research purposes"), and that could rely on established visual convention (e.g. the bin to signify erasure in the "right to erasure" icon) were more quickly and effortlessly visualized. Conversely, abstract or general concepts such as "rights", "processing purposes"," service enhancement" and "service provision" were object of thoughtful consideration and, even, intense debate. Decisions that appeared arbitrary to some group members had to be taken. To cope with these difficulties, an assumption that had emerged in the previous workshop was expressly challenged, i.e. the fact that metaphors must be avoided to enhance clarity and reduce the openness of interpretation. Indeed, literal pictograms produced in the previous workshop (e.g. the right to data portability), despite their concreteness and their arguable informational value, can not work in small dimensions and were considered too complex by user during the evaluation phase. On the contrary, a metaphor in which one idea is understood in terms of another is well suited to convey much through minimal elements. For these reasons, for instance, a folder in the shape of a suitcase was proposed to more compactly recall the right to data portability (see Fig. 2.4). Thus, many explanatory details of the first icon were lost but, if icons need to be usable and scalable

elements, some specifications must be left to the written privacy terms that they complement. Conversely, if the visual elements want to fulfill an explicative function, then different visualizations, such as pictograms, illustrations, and even animations, can be proposed.



(a) The icon that emerged from the first workshop and that was designed to depict the right to data portability in a literal manner.

(b) The icon emerged from the second workshop, that metaphorically depicts the portability of personal data as a suitcase with wheels.

Figure 2.4: The two icons realized for the right to data portability.

Furthermore, we realized how metaphors are nevertheless used consistently throughout the entire icon set to convey meanings, i.e. a data folder to indicate "personal data"; an arrow leaving a circle of stars to signify "transfer outside the EU"; binary code to express something that is not readable by humans, thus "encrypted data"; gears to represent a functioning machine and, hence, "data processing", although it is not a mechanical processing, etc. At different extents, they all are metaphorical depictions of a concept.

Iconographical choices, especially if metaphorical, were discussed at length in the individual groups and then in plenary, when harmonization with the visual elements generated by the other groups was also sought. Eventually, some of the icons produced during the first workshop that needed refinement (i.e. supervisory authority, controller, third party, right of access, right to data portability, etc.) were also presented and discussed, to gather feedback and alternative ideas concerning possibly more functioning solutions. Comments, doubts, and ideas for promising visual solutions were recorded and later transcribed in a workshop report that was distributed to the participants some days afterwards in view of the subsequent workshop that aimed at the harmonization of the two icon subsets and at their digitalization.

During this last workshop, a grid composed of squares of 16x16 mm (64x64 px ca.) was provided to transform the draft icons into digital form, by following the privacy policy template provided. Some visual solutions that worked in paper were not depictable in a smaller digital form, thus some icons' details had to be discarded. At the end of the workshop, the previous data protection icon

subset had been enriched by the following icons:

- 11 icons representing data subjects' rights: data subject's rights (as superclass), right of access, right to data portability, right of rectification, right of erasure, right to be informed, right to withdraw consent, right to lodge a complaint to the supervisory authority, right to restrict processing and right to object to processing (for this concept two different alternatives were produced because it was impossible to elect one best alternative internally);
- 7 icons representing legal bases for processing: legal basis (as superclass), consent, contract, legitimate interest of the controller, public interest, vital interest, legal obligation;
- 6 icons representing some common processing purposes[5] purposes (as superclass), statistical purposes, research purposes, security purposes, purposes of service provision, purposes of service enhancement.

These icons are displayed in Appendix D. An English translation of the simplified definitions provided to the workshops' participants is also shown, together with the reasons behind each iconographical choice. Also the first icon subset was re-elaborated following the visual conventions and the elements' dimensions established during the day. However, the lengthy and vivid discussion generated by the concept of third party (defined in comparison with other roles in the GDPR) made it unfortunately impossible to find a suitable visualization for this element, which however deserves a solution since it is fundamental in almost every data transaction and often appears in consent requests. The evaluation of this icon set is illustrated in Sect. 3.3.

[5]These are the basic, recurring processing purposes identified in our analysis of the GDPR and inserted in its ontological formalization (together with the less frequent judicial, humanitarian, health-related, and journalistic purposes). However, service providers usually list many additional and more precise purposes in their privacy policies to justify the processing operations they carry out on data subjects' data. A comprehensive analysis of these could be carried out to discover if they all are individuals that can be attributed to one of these few ontological classes.

Chapter 3

Icon Set Evaluation

A two step-based approach for the evaluation of DaPIS was chosen (see also [42]), given the specific conditions of the present research. Firstly, icons must be evaluated as stand-alone elements, i.e. according to dimensions such as comprehensibility and legibility. In the second place, icons must also be evaluated for their function in context, as information markers that support the navigation through large amounts of information and increase speed and accuracy of comprehension.

In the following, the first stage of evaluation is described, i.e. the evaluation of icons as stand-alone elements. The first section (sect. 3.1) provides an exposition of classical assessment measures, followed by the description of two subsequent user studies that were carried out to gauge the effectiveness of DaPIS (Sect. 3.2 for the first icon set and 3.3 for the second icon set).

3.1 Evaluation Measures for Graphical Symbols

Given the prominence of graphical symbols e.g. on graphical user interfaces (GUIs) or in public spaces, there exists a body of literature concerning methods and relative measures to assess the effectiveness of symbols along different dimensions.

3.1.1 Ease of Understanding

Ease of understanding is "the most important single index of a symbol's effectiveness" [14, p. 292]. A typical measure of evaluation is hit rate, i.e. the number of correct matchings between an icon and its referent. The only international standardized existing methodology for the comprehension of graphical symbols (ISO 9186-1.2014 [**?**]) is unsuitable because it has been designed for symbols meant to be employed in public spaces (e.g. airport) and whose referent, i.e. the entity to which the symbol refer, is known to users (e.g. airplane). ISO 9186-3.2014 is aimed at assessing the ease of association between icon and

referent, after familiarity training for unknown referents has been carried out. However, the process of learning the referents' meanings takes time, which increases with big number of referents, as in the present case. Therefore, such a test will be taken into account for the future, when familiarity will be rehearsed and an appropriate experience to increase motivation to learn will be designed (see also Chapter 6). As for what concerns the ETSI Multiple Index Approach [34], it was meant to evaluate symbols for telecommunication interfaces of the early 1990s (it has not been updated since), and to elect the best alternative among several icons for the same referent.

As recalled earlier (see Sect. 1.2), DaPIS has some distinctive qualities that make such standardized evaluation methods ineligible for this context. As a matter of fact, individuals are usually not familiar with the referents (e.g. the concept of 'pseudonymization'), whereas the icons might entail low concreteness and high semantic distance. Moreover, many icons are only marginally based on a shared visual vocabulary, thus familiarity with some graphical conventions is expected be low or even non-existent. It is hard then to reach high rates of comprehensibility on first exposures and to set a high bar for acceptance. However, the provision of sufficient contextual information about where the icons might be found and about their function in context should lead to better results, as shown in different contexts [53]. Nevertheless, before testing the icons' functionality in context, icons must be evaluated as single elements to provide detailed insight into the mental models and the line of reasoning behind the interpretation process.

3.1.2 Legibility

Legibility, namely the ease of recognition of the elements that compose an icon determines the ease of recognition of the icon as a whole. This dimension is important because if some elements are not easily visible (e.g. for their size) or recognizable (e.g. for the way they are designed), they could hinder the comprehensibility of the icon's meaning.

3.1.3 Subjective rates

Subjective certainty can be also considered to be included in the test because high uncertainty can reveal higher possibilities of incomprehension. Qualitative feedback should also be encouraged to understand the rationale behind certain preferences or rejections and even to increase comprehension scores [53].

3.1.4 Best alternatives

If there are multiple alternatives for the same concept, it is good practice to ask for a preference among them (see e.g. [34, 22]) In many cases, however, the icons of our icon set have been created for newly introduced concepts and there are no alternatives. Furthermore, some alternatives for the same referent have been already discussed and discarded during early stages of the design process.

3.2 Evaluation of the First Icon Subset: Data Types, Agents' Roles, and Processing Operations

3.2.1 Introduction

The first user study (see Appendix F), carried out in August 2017, did not focus exclusively on the first icon subset (see Sect. 2.2.1), but also other channels to convey the same data protection concepts were explored: simplified definitions, real-world scenarios, and classical legal terms. For the scope of this report, however, only the data around the evaluation of the icons will be reported. The study was conducted through in-person observations and interviews with participants. The subjects had to perform different tasks, explained below, and follow a think aloud protocol. They were thus asked, throughout all the tasks, to verbally express their thinking process and the reasons behind their choices. When they went silent, they were encouraged to verbalize their thoughts.

3.2.2 Participants

The research study was arranged for 20 participants, from as diverse as possible demographics, in terms of origin, gender, age, education, and profession. Of the 16 participants that showed up, 7 males and 9 females, all of them described themselves as "having lived most of their life in the US". Indeed, the participants were recruited in the Bay Area (San Francisco), where the study was carried out, and received a 30$ Amazon card gift as an incentive to take part in the study. Their educational background was diverse, but still medium-high: all the participants were at least high school graduates. Equally diverse was their profession, and nobody had a legal background. Their age also varied, with a minimum of 19 and a maximum of 76, with an average age of 41 and a median of 39.

3.2.3 Tasks

For the scope of this report, only three tasks are described:

Task 1: Rephrase 18 simplified data protection definitions. The test participants received a piece of paper with 18 simplified definitions listed on one side which described the concepts of the 18 icons listed on the other side. The order of the definitions and of the icons was randomized for each participant to prevent possible order effects. After a brief contextual introduction, the participants were asked to read and restate in their own words the definitions or mention examples relating to the definitions. This step had the goal of assessing users' understanding of the definitions and to point out possible flaws. The underlying idea is that anyone should be able to understand the definitions, even without any previous knowledge of the topic. To ensure their comprehensibility, from the lexical point

of view, the 4000 most commonly used words according to the Collins dictionary were exclusively employed. Also the syntax and the lexicon according were adjusted to reach the highest possible level of readability, according to measures such as the Flesch Reading Ease Score (FRES) and the Flesch-Kincaid Grade Level (FKGL), so that the definitions would be as understandable as possible (min. FRES = 59.6, max. FRES= 92.9; min. FKGL= 2.2, max. FKGL=8.4). The simplified definitions were double checked by the team's legal experts to ensure the correspondence of their meaning to the definitions in the Regulation.

Task 2: Match between simplified definition and correspondent labelled icon. Then, the participants were asked to find the best icon match for each definition, in a one-to-one correspondence, among the 18 icons listed on the other side of the sheet of paper. However, the participants were also allowed to choose even more than one icon per definition, or viceversa, when they felt that there was no best match. They were encouraged to explain the reasons behind the icon choice. The icons reported in Appendix C were coupled with the correspondent labels taken from the ontology, since it is recommended to join the visual representation with a descriptive keyword to reduce chances of misinterpretation. We deliberately decided to employ the legal terms used to describe the concept, in order to explore how easily these terms can be understood by average users.

Task 3: Post-study self-reported effort rating. Finally, the users were asked to rate the effort for each communication modality on a seven-point Likert scale, and to explain us the reasons for a certain score. Room for further comments, questions, or suggestions was also allowed, especially concerning icons.

3.2.4 Analysis

Qualitative analysis

Firstly, an analysis of the interviews' transcripts was carried out with a twofold purpose. firstly, to gather the qualitative feedback on icons and other communication modalities, e.g. common types of problems or patterns, and, secondly, to determine the subjects' level of understanding of the simplified definitions. In fact, whereas usual icon recognition tests employ familiar referents that participants need to associate with icon, in this case many referents are specific of European data protection law, thus they are very unlikely known to test participants. It would have been, therefore, very difficult to determine whether a wrong association between referent and icon would have depended on the characteristics of the visual element or on the lack of familiarity with the concept itself. Thus, the rephrased sentences for Task1 were extracted from each interview, compared, and rated.

Measures

In this first user study, classical measures for icon evaluation were reproduced.
For Task 1 (rephrasing the simplified definition):

Answer accuracy: measure that considers the correctness of each answer. For each simplified definition, the accuracy was calculated as the sum of the scores accrued from each participant. The measure ranges between 0 (no participant could understand the simplified definition or the scenario) and 16 (all participants could understand the simplified definition or the scenario) for each item. Each answer could receive a score of 0 (in case of wrong, irrelevant or unknown answer), 0.5 (in case of vague, incomplete or partially wrong answer), or 1 (in case of precise, correct answer or relevant example). The maximum score was assigned only when the subject restated the sentence clearly, precisely, specifically, correctly in their own words, or provided a correct example, or mentioned the concept to which the original sentence refers. For example, for the term 'data subject', with the original sentence: "this is the person to whom personal data refer", a maximum score was assigned to the participant (P9) that restated: "That would be me, because it's my data so it's referring to me". A low score indicates that the definition was difficult to understand.

For Task 2 (match between definition and correspondent icon):

Hit rate: (correct matches between referent and icon + partial correct matches *0.5) / total number of given matches. Partial correct matches are those cases where there was a wrong match alongside a right match. A low score reflects the fact that the association was difficult to make.

Error rate: number of wrong matches between referent and icon / number of total given matches. The closer to zero, the lower the number of errors, thus the more understandable the simplified definition and its connection to the icon.

Missing values: number of lacking matches between referent and icon / number of total actual matches. This measure reflects the level of incertitude.

For Task 3 (post-study self-reported effort rating):

User experience measure: post-study self-reported rating with values ranging between 1 and 7. The user experience of the different communication modalities (icons, simplified definitions, scenarios, legal terms) was evaluated in terms of self-reported effort rate.

3.2.5 Results

Task 1: Understandability of simplified definitions

Overall the understanding average of the definitions was 52% (min= 25% and max=78%). Exactly half of the definitions are above 50%, while the other half

below. In Image 3.1 are reported the single rates of understanding for each simplified definition.

'Direct marketing' was the better understood definition by the participants. Throughout all the tasks, indeed, most of the subjects made continuous reference to online and offline direct marketing practices, one reason probably being that they have firsthand experience of it. definition Right to data portability' ranked second: this result is somehow unexpected since it is a newly introduced right by the GDPR. However, portability as transfer of information from one entity to a new one can be a familiar concept, as the examples provided by the subjects reveal: the transfer of data to a new device or the transfer of medical records to a new practitioner. Then followed 'data subject', 'supervisory authority', and 'profiling'.

Conversely, the definitions of 'processor', 'third party', 'processed personal data', and 'pseudonymization' ranked last. This can either be due to the lack of clarity of the definition itself or to the understandability of the concept described. Since the participants were only given a minimal context at this stage of the test, the results are not dissatisfactory on the whole. Indeed, some subjects commented that the definitions were rather vague or that they would need some more context to understand what the definitions referred to (see 3.2.6).

Task 2: Match between simplified definitions and icons

The average success rate of the matching between icons and simplified definitions was 69%, on the whole, whereas the success rate for each icon is displayed on the graph in Image 3.2. 'Copying' was recognized 100% times (possible reasons being that the definition had the lowest readability score among all the others and it was worded similarly to the icon's label), closely followed by transfer to third countries, and direct marketing. The matches that scored worst are 'processor' and 'controller' (but see conceptual confusion between the two in the Discussion of results, Sect. 3.2.6), 'derived personal' data and 'pseudonymization'. The latter also received the highest number of missing values, therefore indicating high incertitude, perhaps not surprisingly (see Discussion of results).

As can be noticed from the graph in Image 3.3, in almost the totality of cases the subjects performed better in the matching task than in the rephrasing task. The reasons can be several. The simplest one is that, by having tried to rephrase and thus to reflect upon the definitions beforehand, the participants had already built a mental model of what the definitions referred to, so it was easier to associate the definition to the corresponding labelled icon. However, we also observed that the participants used some visual elements of the icons to determine the meaning of the labels or to perform the match with the definition. For instance, in the case of transfer of personal data to third country, whereas the results from the rephrasing task highlighted that the definition must be improved, the presence of an arrow exiting a circle of stars made it clear that there was 'movement of personal data outside of the European Union'.

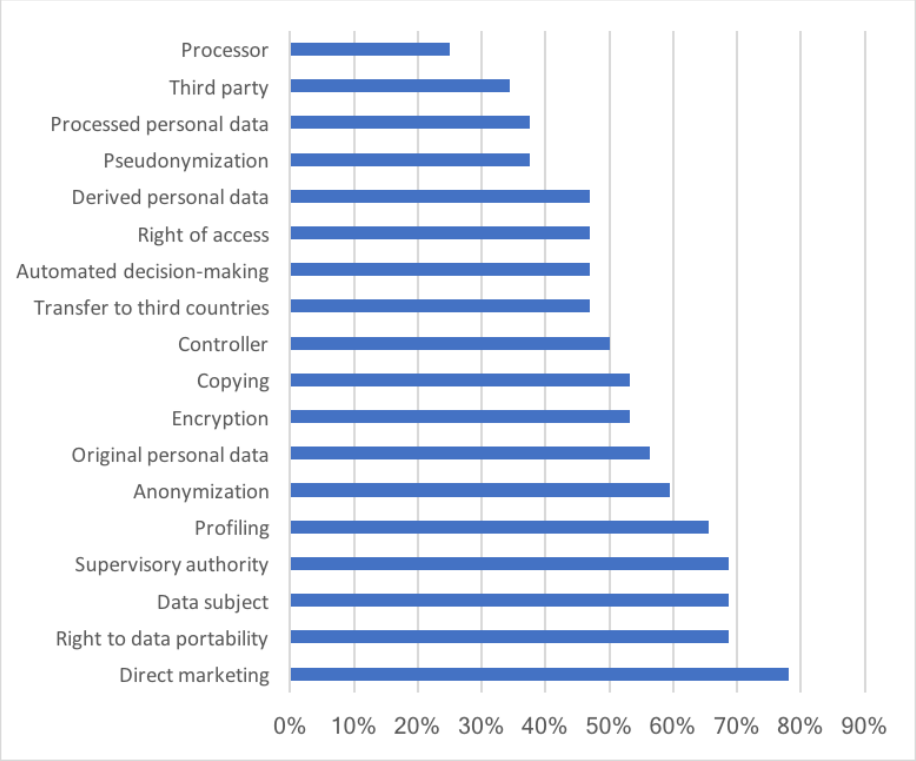


Figure 3.1: Chart that represents the understandibility relative to each simplified definition (Task 1), where the closer to 0 the less correct rephrasing

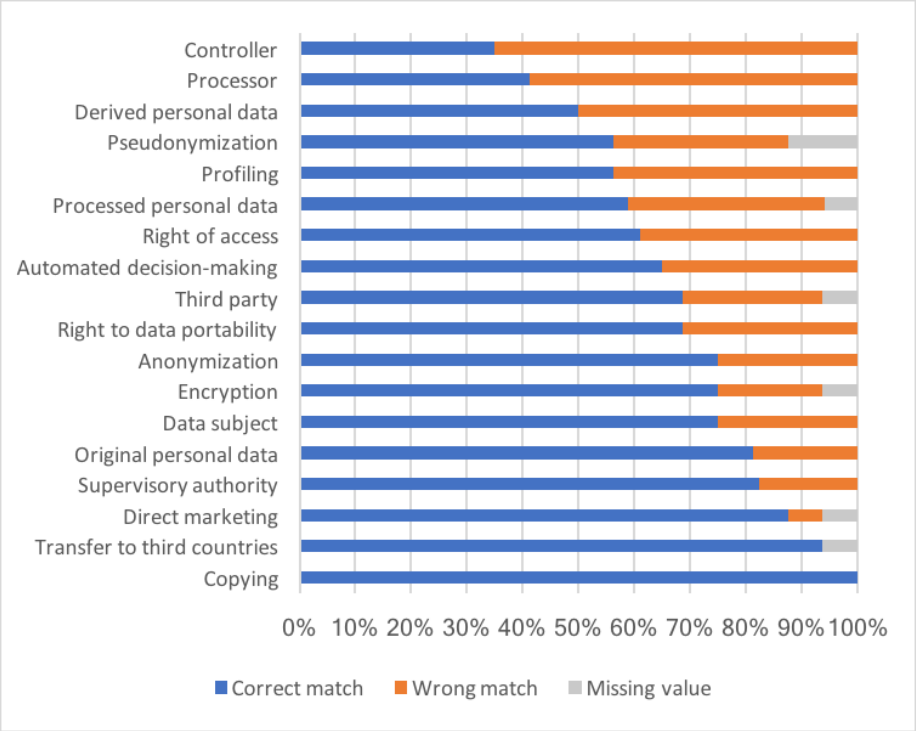


Figure 3.2: Chart that represents the correct matches, the wrong matches, and missing values for the association between icons and simplified definitions (Task 2)

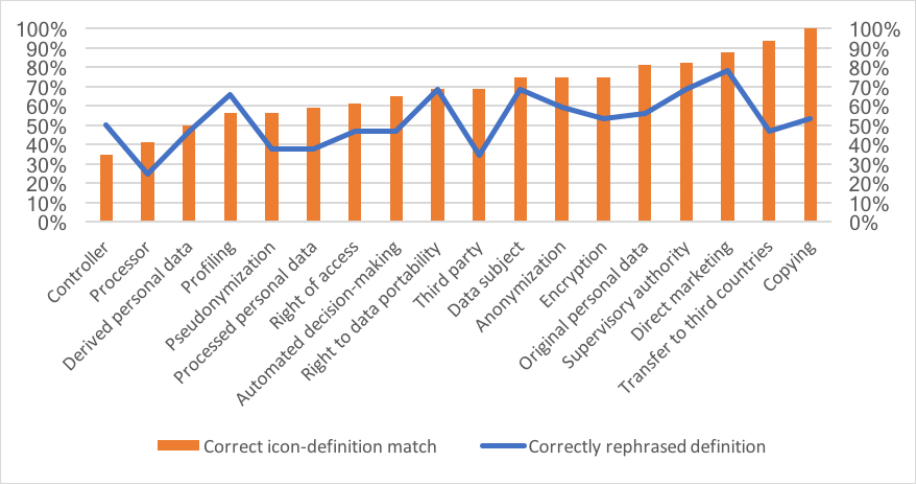


Figure 3.3: Chart that compares the correct rephrasing of the simplified definitions (Task 1) and the correct matches between icons and definitions (Task 2)

Task 3: Subjective evaluation of effort per communication modality

A Friedman test on the overall effort ratings returned a significant result (2= 12.867, p=0.004933), meaning that there is a detectable difference in the effort ratings that the subjects gave for the four different communication forms (i.e. icons, simplified definitions, scenarios, and legal terms). A Wilcoxon signed-rank test (adjusted with Holm's sequential Bonferroni correction) revealed a statistically significant pairwise difference (p=0.0135827) between the ratings of icons and scenarios. In general, as also the boxplot (Img. 3.4) shows, our icons were perceived to be more difficult to understand than the other communication modalities (see Sect. 3.2.6).

3.2.6 Discussion of Results

The self-reported evaluation draws attention to the fact that there were contrasting opinions about the icon set that we developed. From a thorough analysis of the interviews' transcripts, common patterns and recurring comments on the icon set were gathered.

Low Familiarity with the Icons' Referents: The analysis revealed that one of the reasons of recognition or non-recognition of the icon was the familiarity with its corresponding referent. For instance, the subjects named many examples of marketing practices throughout the whole test, referring to their own personal experience. Indeed, 'direct marketing' was among the more easily identified icons. This tendency is even more ev-

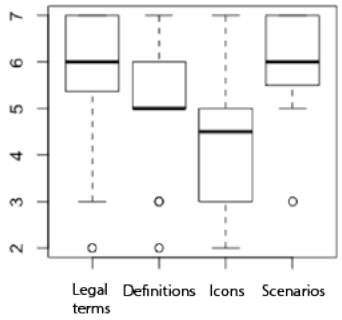


Figure 3.4: Boxplot graphs that represent the effort rate on the different communication modalities. On the vertical axis is displayed the effort rate on a scale from 1 to 7 (Task 3)

ident among the least recognized icons. 'Pseudonymization' constitutes perhaps the most emblematic case in this sense. Many subjects struggled to understand the concept underlying it. Unlike anonymization, the notion of pseudonymization is reasonably unknown to the general public. Finally, many participants encountered problems even only when reading the label associated with the icon and some asked for explanations of its meaning. From this follows that, even if the icon is well designed, people notably rely on the associated label to interpret the visual element. Thus, even the label must be carefully conceived with attention to user-friendliness to foster understanding, instead of resorting to the legal terms.

Shared Visual Vocabulary: It was also observed that the participants relied on some icons' features to determine their meaning, especially when familiarity with the concept was high. For instance, in the case of 'transfer of personal data to third country', the presence of an arrow exiting a circle of stars was correctly interpreted as movement of personal data outside of the European Union. The use of the file folder for 'personal data' and of the user outline for 'data subject' were positively evaluated, since they rely on a graphical language that is already part of the shared knowledge of computers' and social networks' users. The word "you" on the data subject figure was also positively regarded, since it helped participants to understand where the data subject is situated in complex dynamics, such as in the representations of her rights. These results confirmed that the more the icons rely on an established visual language and common mental models, the more recognizable they are. We identified some elements of the shared visual language on data protection, but more must be integrated into a more easily recognizable or learnable set of icons. However, if it is the concept to be unknown, it is difficult to find a sufficiently good representation that can straightforwardly communicate it.

Distinguishing Features: 'Processor' was frequently confused with 'controller', and viceversa. This is due not only to the fact that these two roles were not readily distinguished at a conceptual level, but also to their visual representations. Both processor and controller are depicted as user figures inside a building, the former overlooking file folders, whilst the latter overlooking processing gears. The participants' comments highlighted that the difference between the two icons was perceived as too subtle, thus went in some cases unnoticed. From this follows that consistency of the elements across the icons is important, but greater relevance must be given to the distinguishing features of the icons. Similarly, the distinction between the different personal data depicted in the rights was ignored. Thus, a more visible, straightforward way to show the distinction among these similar concepts must be envisioned, perhaps even by recurring to words alongside or instead of visuals.

Chances of Misinterpretation: From the user study, one risk concerning misinterpretations emerged: 'profiling' was repeatedly mistaken for 'direct marketing', and vice versa. On the one hand, this shows that individuals are aware that the two processes are closely interlinked. However, this poses a serious problem when consent is asked, since individuals can choose to give consent to none, either one, or both operations. Being able to distinguish between them not only on the conceptual level, but also on the visual level is therefore crucial, if consent was to be asked through clickable icons [8]. Also 'automated decision-making' exposed a similar problem: although the computer depicted in the icon helped some subjects to correctly match it to the concept, it was sometimes confused with profiling and was linked to artificial intelligence. Although such exchange might indicate a non-trivial understanding of this technical notion, it can also cause some issues: from a legal perspective, automated decision-making refers solely to those decisions that have significant effects on data subjects, such as the eligibility for money loans. The icon was therefore not able to be as precise as it should have been to clearly represent this distinction. This may entail considerable consequences, because it is connected to the possibility of exercising the right to object to an automated decision.

Combination of Icon and Label: In general, however, not only the users relied on the combination of icon+label for the interpretation, but in many cases they principally relied more on the textual cue than on the visual element. Half of the participants even mentioned the fact that without labels or some extra information (e.g. a clickable, pop-up definition), the images would be hard to understand or ambiguous. Some participants (P1, P6) even pointed out that, if users are not familiar with the icon's referent, the visual is not going to help. Thus, a user-friendly label must always be associated to the visual representation of a data protection concept, especially for first exposures to less semantically transparent icons. Thus, the provision of both textual and visual cues in a solution should

be preferred, so that individuals can leverage either one or both of them to understand the communication.

Non-Native Speakers: Nevertheless, there are cases where this is not true: participant 8, whose native language was not English, relied strongly on the visual elements to to match it with the definition, especially when she did not know the meaning of the label associated to the icon (e.g. 'encryption', 'anonymization', 'pseudonymization', etc.). And in those cases, her assumptions were in fact correct. If it was possible to give an easily intelligible visual representation of complex data protection concepts, it could e.g. benefit individuals with lower literacy levels.

Simplicity Versus Precision: Probably the most important insight is that some subjects deemed some of the icons as "too complicated" and "too crowded", such as the pictogram representing the right to data portability. However, the level of detailedness helped some of them to follow the embedded narrative and, thereby, understand how this right unfolds and distinguish it from the right of access, which presents similar elements. A trade-off between accuracy and coherence of representation, necessary for legal purposes, and simplicity, as users require, clearly emerged as future direction of research.

3.2.7 Limitations of the study and future work

The study described in the previous pages present some limitations. For instance, a focus was placed on the users' verbal skills in the rephrasing tasks, whereas in the future it would be interesting to discover the effects of other learning styles (i.e. the preferred way for individuals to process and remember information) on the understanding of different communication modalities.

It must also be acknowledged that the match between icons and definitions was finite. Therefore, as some participants even pointed out, they carried out the task through an elimination process, instead of abstaining from those matches they were most uncertain about. However, in the real world the set of referents for a certain icon or a certain definition is not closed, it is rather infinite.

Moreover, the subjects sample was not very diverse in terms of educational background and nationality. It would be therefore necessary to carry out a user testing with people with lower education levels and different nationalities (especially Europeans). The logistics of the study, physically based in the US, caused the sample to be exclusively American. Nevertheless, there is value in US participants' feedback to DaPIS, since it aims to become a standard set of icons that can be understandable across nationalities, language, and cultural backgrounds. This might mean encountering different privacy attitudes, or different levels of savviness on data protection, or knowledge of technology. The goal is to reconcile different testers' feedback to the icons to create a set that can by themselves communicate effectively to multiple audiences. Alternatively, icons that adjust depending on the audience might be recommended.

Finally, some more limitations are discussed in the next section, since they guided the design of the evaluation phase for the second subset of DaPIS.

3.3 Evaluation of the Second Icon Subset: Data Subjects' Rights, Legal Bases, and Processing Purposes

3.3.1 Introductory Considerations

The evaluation of the second subset (see Sect. 2.2.2) was carried out across different dimensions compared to the first user study, in order to avoid its limitations. Firstly, the focus was solely on icons to gather as much feedback as possible, whilst other communicative devices were ignored. Relevance was given to three aspects: icons' legibility, correspondence between icon and it underlying concept, and alignment between users' and designers' mental models.

Legibility of the icons was examined to ensure that all the elements could be easily visible and recognizable even at small dimensions, since low legibility levels can influence recognition and interpretation.

A second dimension that was researched was ease of understanding of the icons. As recalled at the beginning of this Chapter, specific characteristics of DaPIS render it hard to replicate existent standard evaluation methodologies that employ quantitative measures such as hit rate. Such a measure was used to evaluate the first icon subset, but showed limitations since it was sometimes hard to determine whether wrong associations depended on the icon or rather on the individuals' limited knowledge about the underlying concept. The rephrasing task was adopted as solution, but the concrete test revealed that in certain cases it was arduous to establish whether low comprehension rates were due to flaws in the definition itself or to difficulties in grasping the underlying concept (see Sect. 3.2.5). Although the definitions were translated into simple language, strictly following readability measures in some cases transformed the descriptions into texts that were too simplistic to be easily understood.

For these reasons, a different approach was adopted in the second user study described in the following pages: the same simple definitions provided to the designers during the icons' creation were displayed next to the icons and, instead of measuring efficiency of association between concept and icon, the primary focus was placed on the process of interpretation of the visuals. The adoption of such a method had three main goals: firstly, to clarify the difference between poor rates caused by icons' representation or design, and poor rates derived by lack of understanding of the underlying concept. Secondly, no interpretation strategy based on the exclusion of previous associations was deployed since this strategy could hide the level of subjective certainty about an association: in the previous user study, the icons selected and associated first were in most cases those that the data subjects could more easily recognize (e.g. copying). Finally, this strategy allowed to explore whether the rationals behind the iconographical

choices made during the design phase could be grasped, i.e. if alignment between designers' and users' mental models was possible, especially on less semantically transparent icons.

Moreover, unlike the first test, the icons were not associated with the corresponding label because it was found that subjects would base their interpretation most prominently on the latter than on the visual cues. Nonetheless, a label was placed above the definitions to which the icon corresponded, since it represents a contextual aid that can support the interpretation.

An additional consideration that drove the design of this second user test sparked from the fact that, especially for abstract concepts, any visual representation could be potentially suitable[1] . After repeated exposure, individuals will eventually learn and memorize the association between an icon and its referent. It is for this reason that it was deemed necessary to assess the capacity of an icon to convey its meaning in isolation (to replicate the conditions of those that will briefly glance at the icons, instead of reading the privacy policy's terms), while research on the effectiveness of contextual elements (e.g. labels, definitions, text, etc.) at complementing, reinforcing or even guiding the interpretation process is left for the future (see Future directions in Chapter 6).

3.3.2 Participants

16 participants, 11 females and 5 males, were recruited in Bologna, where the study took place, through paper ads and online ads, and received a 20 euros Amazon gift's card as reimbursement for their time. The participants' age span ranged from 20 to 29 years old and the great majority were university students, more than two thirds already with a Bachelor's degree, indicating an high educational background. A self-assessment of the participants' digital and legal skills was also asked. Three quarters (n=12) of the subjects described themselves as having intermediate digital skills, while two fell on both side. Half of the participants (n=8) also claimed to have basic legal competences, while two declared them to be non-existent and 6 intermediate. All of them had native or comparable levels of Italian: this was a necessary prerequisite to carry out a task where also comprehension of legal written definitions was requested.

3.3.3 Tasks

The study was conducted as in-person observations and interviews by three researchers at the University of Bologna, Italy, in March 2018. The test was constituted by a predefined set of questions (see Appendix G) along the dimensions explained above. The subjects were asked to answer such questions and were actively encouraged to follow a think aloud protocol. Researchers intervened when the users asked for explanations or examples around a certain concept, since the understanding of the concepts' meaning was critical for the experiment (see earlier at Sect. 3.3.1). To provide contextual elements to help

[1] see e.g. the data protection icons available at `http://www.lawinfographic.com/`

the participants to create a mental model similar to the actual icons' context of use, a brief explanation about the research was given at the beginning of the test and a mock-up of a visualized privacy policy was shown. The participants were asked to record their answer in written form next to each question (see Appendix G). The researchers took notes of participants' behaviours or of comments that were not recorded by the participants themselves.

Four tasks were designed:

Task 1: Icons' legibility. legibility was estimated by asking the test participants to name the elements composing each icon. Also some icons from the first subset were evaluated in this sense, since they were harmonized during the third workshop and had not received any legibility evaluation in the previous test. To replicate non optimal conditions (i.e. the worst case scenario), the icons' legibility at reduced sizes was explored. Hence, they were printed out at 16x16 mm to reproduce small settings and low resolution, as they could appear on paper-based privacy policies or devices' boxes. Furthermore, research shows that human beings make sense of information differently on screen than on paper (e.g. [38]), even if contrasting results exist (e.g. [48]), although these studies are rather focused on textual information than on visual information.

Task 2: Subjective rating on the icons' correspondence with its underlying meaning: for the reasons outlined above (Sect. 3.3.1), a subjective rating on a Likert scale that ranged from 1 to 5 (where 1=strongly agree, 3=don't know, 5=strongly disagree) about the ability of a certain icon to represent the correspondent concept, expressed through a label and a simplified definition, was chosen. Explicit explanations for the mark were asked and recorded in written form.

Task 3: Alignment between users' mental models and designers' mental models: the participants were asked to attempt to provide an explanation for the visual choices made by the designers. This task was meant to find out whether users could understand the reasons behind the choices and thus align their line of reasoning with the designers', despite their opinion on the appropriateness of a certain icon for a certain concept, evaluated in the previous task. In order to avoid influence by different wordings, the same identical simplified definitions that were distributed to the designers to spark the design process, were also provided to the test participants. Neither examples nor further explanations were given with the hand-outs, but rather provided to the subjects orally by the researchers if needed, similarly to the design phase.

Task 4: Best Alternative: one single alternative choice between two icons representing the concept of 'right to object to processing' (see Table D.1) was asked.

3.3.4 Analysis

The data collected in written form by the participants was gathered and integrated with the notes taken by the researchers during the study. This data was then analyzed by one of the interviewer in search for common patterns. Given the nature of the tasks and the study goals, qualitative analysis was the main source of data.

For legibility (task 1):

Hit rate : For each icon, there was a lower bound equal to a minimal set of elements that had to be identified in the icon for the answer to be considered correct. Every wrong identification of elements and common mistakes were recorded.

For the correspondence between icon and concept (task 2):

Average : mathematical mean across the users' self-reported marks was computed for each icon;

Frequency : since means can hide details, frequencies for each mark were computed instead.

Furthermore, the explanations provided by the participants to motivate their choice were analyzed to find common lines of reasoning, but also to find out explicitly about the words that were employed to describe why some associations were easier/harder than others.

For the alignment between users' and designers' mental models (task 3):

Hit rate : number of correct explanations provided by the participants with respect to the actual designers' reason behind a certain iconographical choice for each icon's element (e.g. 'right to erasure': one mark for the hand and a different mark for the bin symbol). Since some elements appear in more than one icon (e.g. the hand in all data subject's rights), the score for each element is computed as: number of correct matches/(number of responses given * number of icons where the element appears). Higher scores correspond to better user's understanding of the designers' reasons for a certain iconographical choice.

Error rate : number of wrong explanations provided by the participants with respect to the actual designers' reason behind a certain iconographical choice for each icon's element. The overall score for each element is computed as number of wrong matches/(number of responses given * number of icons where the element appears). Higher scores correspond to worse user's understanding of the designers' reasons for a certain iconographical choice.

For these last two measures, marks were assigned as following: 1 if the user's explanation coincided with the designer's intention; -1 in case of wrong explanations or if, even if the icon's referent has been understood, the reason why that

visual element is chosen is unclear; 0 when no explanation is provided, or the explanation is incomprehensible or vague (i.e. "because it's clear, well-known, intuitive" etc.), or the participant admits that she does not know the reasons behind a certain visual choice.

For the best alternative between the two icons for the "right to object to processing" (task 4):

Higher number of preferences : the best alternative was chosen simply by counting which of the two icons got the majority of votes. Reasons for the choice were also recorded.

3.3.5 Results

Task 1: Legibility

Positive results were achieved on almost all icons, meaning that the elements were simple enough to be easily recognized, even in small dimensions. This confirmed that the simplification of icon design brought favorable results. The only icons that show lower rates of legibility are:

1. right to data portability: only one fourth of the respondents identified the bag-shaped file folder, whereas the great majority did not recognize it; three interpreted the drawing as a padlock;
2. right to lodge a complaint to a supervisory authority: almost all test participants could not correctly identify the file folder below the supervisory authority and almost half of them interpreted the gears as a key;
3. controller: only two subjects expressly noticed that the silhouette has a white shirt and is slightly different that a usual user silhouette;
4. legal basis: more than half of the participants could not determine that the element under the hammer is a column;
5. vital interest: some minor doubts (three people) on the graph within the hands;
6. encryption: almost everybody could detect the written characters in white, but only half could safely assume that it was binary code;

Although it was not a goal encompassed by this first part of the test, some participants attempted to provide a free interpretation of the icon even in this phase and the investigators let them free to do so (see also 3.3.5).

Task 2: Rating on Fitness of Correspondence Between Icon and Concept

The results of the assessment on the icons' capacity of representing the underlying concept are reported in Img. 3.5. The results reported in the following paragraphs are organized in three groups, according to the average value obtained by each icon:

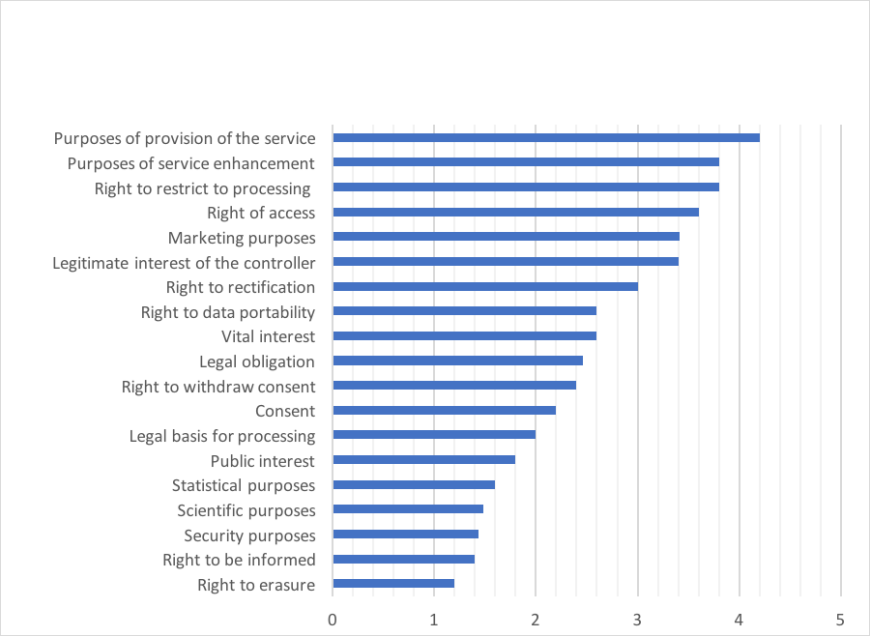


Figure 3.5: Means of the self-reported values for ease of understanding, where the closest to 1, the better the results

1. best rated icons: average value ranging between 1 (=completely agree with the fitness of correspondence between icon and concept) and 2 (=agree), see Fig. **??**;

2. medium rated icons: average value ranging between 2 (=agree) and 3 (=uncertain), see Fig. **??**;

3. worst rated icons: average value ranging between 3 (=uncertain) and 4 (=disagree), see Fig. **??**.

Best rated icons: Among the icons that scored best, the symbol of a bin to signify "cancellation, erasure" **right to erasure** and of an "i" to signify "information" in **right to be informed** were described as "universal, immediate, instantly recognizable, clear, intuitive, unmistakable" because "grounded in our culture, codified and common on application software".

The **security purposes** and **research purposes** icons were also rated positively, since the shield is "stereotypical" for security, defense, and (antivirus) protection, whereas the microscope is "emblematic" of science and research. The bar graph "intuitively" recalls statistics (**statistical purposes)**, whilst the presence of three user silhouettes seem to be easily associated to the idea of "public", group of people or community (**public interest**) (for the meaning of the hands that sparked some doubts and

(a) Right to erasure

(b) Right to be informed

(c) Security purposes

(d) Scientific purposes

(e) Statistical purposes

(f) Public interest

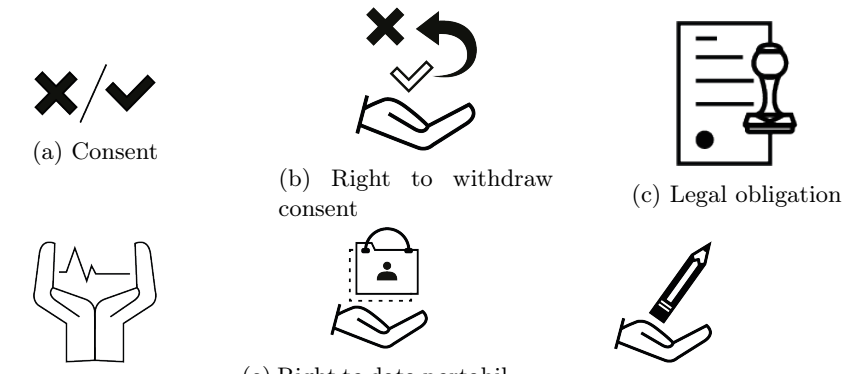
(g) Legal basis for processing

Figure 3.6: The best rated icons in Task 2

lower scores see below). In the **legal basis for processing**, only one respondent was able to link the column to the basis (but this might be caused by the low recognizability of this element), whereas the hammer was unequivocally associated to the legal sphere, which might be a rather vague association, or even more often to justice, which is incorrect. Nevertheless, the legal and juridical sphere overlap in the common sense and, hence, in their visual representations.

Medium rated icons : **Consent** was expressly symbolized as a choice between accept and refuse (as opposed to more common passive acceptance). Although half of the respondents noticed that emphasis was put on the possibility, i.e. right of choice, between accept/refuse or agree/disagree, which corresponded to the designers' intention, the lower rates are due to the fact that the tick and the cross can be ambiguously associated also to the dichotomy of right/wrong, yes/no, true/false. Although it is reasonable to assume that the provision of more contextual cues would shrink the number of possible interpretations, this must be proved. Similar critics received the icon for **the right to withdraw consent**, based on the same symbols, combined with an arrow to signify the transition from given consent to withdrawn consent, which was understood by almost all the respondents.

Mixed opinions have been gathered on the icon for **legal obligation**: whereas a few endorsed the stamp has symbol for official, thus per extension legal, some others expressed their doubts since the stamp can be also linked to the administrative domain, such as a certificate, and a few also



(a) Consent

(b) Right to withdraw consent

(c) Legal obligation

(d) Vital interest

(e) Right to data portability

(f) Right to rectification

Figure 3.7: The medium rated icons in Task 2

suggested to use the hammer to be more legally specific. The variety of opinions gathered and the good arguments provided suggest that in order to be more easily and unambiguously understood, more appropriate alternatives for this icon should be created and tested. The graphical symbol for **vital interest** also received comments of mixed nature: although the EFC conveys the idea of vital importance, thus concerning life and death, proposed enhancements would possibly specify the icon in context. Indeed, "it can look like an audio file" (P16) and "it can be confused with the device' life" (P9). In addition, it could be too strictly linked to the health domain because it's the health metaphor for anyone" (P3).

The folder in shape of a suitcase, a metaphor symbolizing the **right to data portability**, was positively embraced by three quarters of the subjects. However, a few respondents specified that the transfer of data from an entity to another should have been better specified, e.g. with arrows. These suggestions can be integrated into the icon, since the movement represented by the dotted line was not readily legible.

Although the marks given to the **right to rectification** are not high, which contradict the researchers' expectations, the explanations from all the participants mention the pencil as a clear symbol for modification (borrowed by the many applications that use the same symbol for the modify function). However, one fourth of the respondents expressed the need to specify the object of modification, e.g. by adding a data folder (P18), whereas at least in two cases the low grade was given because it was the concept of "rectification" that was not grasped, but once explained the symbol was considered appropriate.

Worst rated icons : At the lower end of the spectrum, the icons that were rated more poorly, for instance the **legitimate interest of the controller**. Although one fourth of the respondents noticed the controller's

specific clothing referring to the "higher social status of the person" (P7), "an authority" (P14), "elegantly dressed, so maybe in a position of power" (P18), more than one third could not distinguish the controller from the user - a legibility problem already highlighted in the first part of the test. We also suspect that these low rates are due to the two hands to signify the interest, but see discussion below. The icon for **marketing** was also poorly rated, mainly because it was deemed too similar to any other kind of communication, thus too general to be exclusively attributed to the advertisement sector. In a couple of cases it was wrongly interpreted as dissemination of personal data.

(a) Legitimate interest of the controller

(b) Marketing purposes

(c) Right of access

(d) Right to restrict processing

(e) Purposes of service enhancement

(f) Purposes of service provision

Figure 3.8: The worst rated icons in Task 2

The **right of access** also got mixed marks. On the one hand, half of the participants correctly interpreted the magnifying glass as metaphor for looking into the folder, thus accessing the files. Criticism was raised, however, on the fact that it is unclear that it is the data subject's data (and not someone else's) that someone else owns (and not the data subject). See the discussion below on this point.

The **right to restrict processing** ranked among the worst icons for a series of reasons: firstly, half of the participants underlined how the gears' symbol is usually employed to signify the device settings, so it does not unequivocally recall the processing. Secondly, the idea of restriction/limitation symbolized by the difference in gears' color was not easily grasped, either because the difference was too subtle or because it could not be traced back to its meaning. Nevertheless, half of the participants could understand the designers' intentions, even those that rated the icon poorly.

The icons for **purposes of service enhancement** and **purposes of service provision** that ranked second to last and last, respectively, can be discussed together since they are almost identical and therefore present similar problems. Although the idea of an exchange represented by the

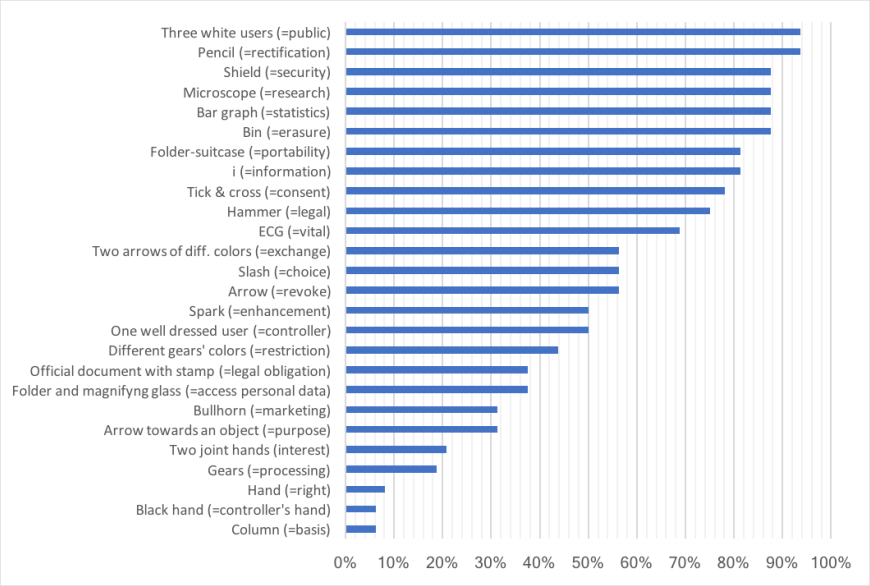


Figure 3.9: Percentage values of correct explanations for designers' iconographical choices

two arrows in opposite direction was approved by more than half of the subjects, the icon was described as "incomprehensible", "not intuitive", "unclear", "vague" and a few respondents clarified that they would need an explicit explanation. However, the star/plus element was positively rated and easily interpreted by three quarters of the subjects as symbol for enhancement.

Task 3: Alignments with Designers' Intentions

Figure 3.9 displays the results for each visual element appearing in DaPIS that has self-contained meaning. The best results (e.g. a group of users meaning public, a pencil that stands for rectification, a shield signifying security, etc.) seem to reproduce the icons that scored better in the previous task, whereas the elements towards the end of the graph (e.g. the column, the controller's black hand with white sleeves, the rights' hand, the gears for processing) were assigned more frequently a wrong association with the designer's intended meaning, thus misinterpretations.

Task 4: Best Alternative for the Right to Object to Processing

Among the two icons produced for the right to object to processing, three quarters of the respondents preferred the icon with sharply separated gears (see final icon set in next chapter) because it could more easily suggest a complete break

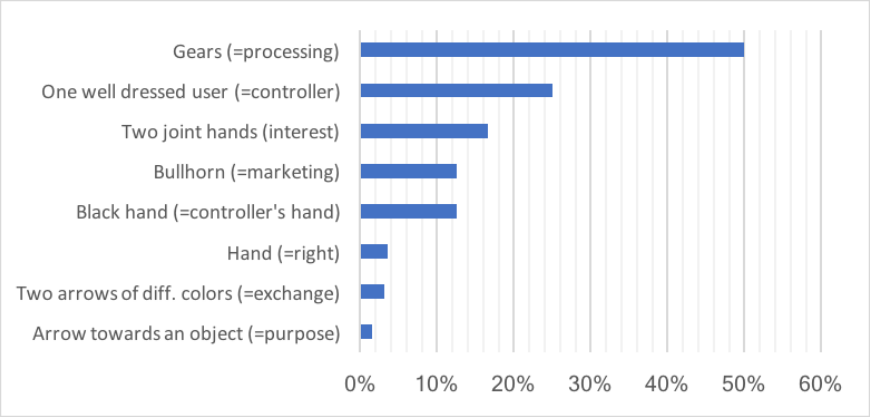


Figure 3.10: Percentage values of wrong tentative explanations provided for designers' iconographical choices. For the elements not shown on the graph, no wrong explanation was provided

that can not be fixed.

3.3.6 Discussion of Results

Legibility Concerning legibility, the few icons reported below need to be enhanced, in order of severity: 1. right to lodge a complaint to a supervisory authority: there are too many elements, some of which are too small to be effortlessly recognized; 2. legal basis: either make the column more recognizable (it is hard to tell whether the metaphor for basis was understood due to its low legibility) or take it away; 3. right to data portability: the folder needs to resemble more closely to a bag or a suitcase in order for the metaphor to be grasped; 4. controller: as some comments in the recognition task also confirm, its difference from the normal user's silhouette needs to be more prominent, e.g. by widening its white parts or adding additional distinctive marks; 5. vital interest: make the ECG more distinctive or provide some contest around it.

Ease of Understanding and Alignments between Mental Models Not surprisingly, the icons that scored best represent concrete objects, familiar concepts or are based on familiar representations (e.g. information, erasure), which is also reflected by the users' explanations displayed above. Conversely, the concepts behind the icons that scored worst, e.g. for the provision or enhancement of the service, are vague, general, and abstract. During the design phase, close scrutiny and long discussions originated around possible ways of representing these notions. Although for such concepts a semantically transparent solution might never be reached, alternatives can be explored and compared to DaPIS to determine whether

they are more readily graspable. The analysis of the frequencies also indicates that, whereas for better ranked icons judgments were more compact, marks are distributed along the five possible marks as the icons become less familiar or less concrete. Individual characteristics might be the cause.

A general tendency that can be noted is that, in those cases where the association between icon and referent was deemed more appropriate, the users' explanation more frequently coincided with the reason behind the design choice. For instance, participant (P2) explains her high mark for the right to be informed with the fact that the "i" is an unquestionable symbol for information that everybody knows and this is why the designers chose it. For what concerns symbols that were rated poorly, in some cases the designer's intention was nevertheless understood and explained. For instance, another participant (P3) expresses her doubts on both icons for purposes of service provision and of service enhancement, by saying that thy are not intuitive and that she would not grasp the underlying meaning from the image alone. Nevertheless, she is able to provide correct and accurate explanations for the designers' choices: the arrows signify an exchange, whereas the star stands for enhancement. Also the explanation provided for the right to data portability is exemplary in this respect: "I would not click on this icon to receive or transfer my data. Intuitively, I would have accepted a symbol of entrance/exit" provides as reason for her mark expressing incertitude. However, when asked about the supposed line of reasoning behind the folder, she effortlessly identifies the folder with the handle as metaphor for transportability. An additional emblematic case is presented by the icon for the right to restriction of processing: notwithstanding the poor evaluation given to the difference in gears' color, meaning limitation, almost half of the participants could explicitly and correctly associate it with its intended meaning. This seems to indicate that, even if some visual choices are not readily grasped, some consideration can guide the interpretation process and align the mental models. This is also shown explicitly by some users' comments, e.g. P19 when she notices the recurrence of the joined hands to signify the interest: "it is not easy to link the hands to the concept of interest. [...] once established that the interest is represented by the hands in this position, then it is easy to identify the controller in the picture". In other cases, however, the participants were confused about the reasons behind certain iconographical choices and could not follow the line of reasoning: for instance, the reason why the icon with gears has been selected to signify processing, whereas it is the usual icon for settings, or the reason why the joint hands signify interest.

The use and interpretation of the hands symbol must be commented. The hand with the palms facing up (the "holding hand") has the metaphorical extension of "being in control" or "have the power over" to indicate the possibility granted by a right to its holder. Even in the legibility phase multiple interpretations were offered for the symbol: a offering hand (P2),

a welcoming hand (P3), a helping hand (P14), a requesting hand (P17), a hand offering a possibility (P9 e P14), a protecting hand (P17). During the second task, less interpretations were provided since the participants focused more on the distinctive element placed above the hand than on the element in common, i.e. the hand. Only four individuals explicitly referred to the hand's intended meaning and all of them after a few exposures, thus after having noticed that it is the recurring element to signify a right. Others seemed to interpret the hand by drawing inferences from the other elements of the icon or from the definition provided, e.g. for the right of access: P2 provided the explanation "me, data that I own", P10 referred to "grab to open", for P15 it signifies "data in our hands", whereas P17 said that it represents "someone external that holds that about you". Although the symbol is rather arbitrary, thus variations of interpretation are inevitable, the metaphor seems to not have been smoothly decoded. Alternatives could therefore be researched and compared to this to find out if they perform better. The joined hands to signify interest were also oppositely interpreted: either they were recognized as metaphors for protection, taking care on the elements in-between or as synonym of power exerted over the elements. A symbol resembling those in the Figures 3.11, 3.12, 3.13 might be more unambiguously interpreted as interest, protection.

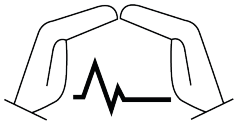


Figure 3.11: Possible alternative icon for the vital interest

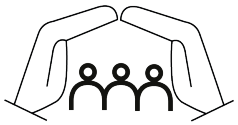


Figure 3.12: Possible alternative icon for the public interest

Figure 3.13: Possible alternative icon for the legitimate interest of the controller

Some minor modifications suggested during the study will be taken into account to enhance the comprehensibility of DaPIS. For instance the icon for marketing was poorly rated, but it might be improved by simply adding a euro symbol or a shopping cart in the speech balloon, as suggested during the study, to better specify its meaning. For some other icon, however, the solution might not come from minimal adjustments: for instance, more straightforward visual solutions for the "provision of the service" could be found in order to avoid easy misinterpretations (e.g. Fig. 3.14).

The comments on the icon for the right of access, i.e. that it is unclear to whom the data belongs and who owns them, seem to echo the reasons that brought to the literal representation of this concept in the first icon subset (see Discussion about the first icon subset in Sect. 3.2.6). The same conclusion, however, holds: some details need to be sacrificed and

Figure 3.14: Possible alternative icon for provision of service

left to textual provisions for the sake of icons' usability.

Many comments, even from the legibility testing phase, highlighted that the gears in isolation are immediately linked to settings rather than processing. This metaphor has been consistently used across the icon set, although it must be reckoned that the original version saw a composition of arrow and gears to signify a transformation of personal data into something else: some kind of processing, indeed (see also Img. 2.1). It must be therefore researched if some element must be added to specify the gears' meaning or, rather, if a completely different metaphor must be found.

Finally, there is a specific case that can shed light about how previous knowledge of concepts can positively influence icon understanding. During the legibility task, one of the study participants freely provided correct interpretations of a number of icons: not only on the more familiar ones (i.e. right to be informed, right to erasure of data, data transfer to third countries), but also on less immediate icons, such as right to data portability, data controller, encrypted data, pseudonymized data. Although the participant described herself as having intermediate legal knowledge, her answers clearly indicate an accurate knowledge of the topics.

3.3.7 Limitations of the Study

The participants in the two user studies were really diverse, as well as the types of tasks that they carried out. For these reasons, the results are not comparable. In the second study, although a greater variance was expected, almost the entire totality of the subjects were individuals in their twenties with a high educational background. However, involve very wide, differentiated, international audiences that are representative of European population is admittedly out of our reach (see also the Conclusions in Chapter 5).

From this also follows that, as in other lab researches, the tasks of this user study presume a serious consideration over icons that might not closely mirror the (presumably quick) sense-making process carried out in real-world conditions. However, such a limitation can be overcome only if organizations start to employ the icons in different contexts (online versus offline, paper versus

digital, in combination with text versus as stand-alone elements, etc.) and"in the wild", i.e. on online platforms, social networks, etc.

Moreover, since the interpretation keys were provided, it was impossible to explore the informational value of the icons and it is plausible to assume that some icons' received a better score that they would have in a matching task. Although on the one hand this strategy partially simulated the research activity of a user that wants to find a specific piece of information that she already knows (from concept to icon), on the other hand it is the opposite of facing unknown symbols for the first time (from icon to concept). We believe that both directions must be researched.

3.4 Evaluation of the Third Icon Design

3.4.1 Introductory Considerations

After the last redesign of DaPIS (*see* Section **??**), it was deemed necessary to run a further evaluation of these icons, on the model of the last one. Moreover, some of the icons of the first icon set, for example those related to the processing operations, had not been previously evaluated.

It was decided to keep a similar pool of users, i.e. highly educated young people (20-35). For this reason, motivated individuals owning at least a master degree were recruited across some universities, mostly at the University of Luxembourg, which also guaranteed a more international audience than the two previous tests. Another constraint was a high level of English proficiency, to ensure easy comprehension of the questions and of the legal definitions, and to ensure sufficient linguistic means to provide detailed and elaborated answers. The study was carried out in an online environment to be more easily distributed to the participants and also to experiment if, given the lessons learned from the previous studies, it is possible to conduct such a test online and at distance, within the view of future large-scale distribution for a final evaluation (*see* next Chapter). The organization of in-person studies, in fact, and the collection of results are usually extremely time-consuming and unfeasible on large scales.

3.4.2 Participants

10 participants took part in this online study, all having at least a master degree and advanced English level. All the participants described themselves as having intermediate or advanced digital competencies, whereas their legal competences are placed on two opposites, i.e. either advanced or basic. Their origins are Italian, Armenian, Iranian, Canadian and Greek. Non-EU residents have lived in Europe at least since one year. Their age ranges from 28 till 33, with an average of 31 years old.

3.4.3 Tasks

For the reasons anticipated earlier, this last user study was carried out online, on the website that documents the research[2], in July 2018. This study replicated the same kind of tasks of the second user study (*see* Section 3.3): subjective rating on the correspondence between icon and concept, and alignment with designers' intentions for those icons that had not received previous evaluation or that were completely redesigned based on the results of the last user study. Moreover, it was asked to choose among two or more alternative icons for the same concept in 5 cases, i.e. the more problematic cases in terms of simplification versus completeness of representation (as in the icons for contract and legal obligation) or in terms of comprehensibility (as in the icons for provision of a service, sharing with third parties, and the hands symbol for interest as in public interest).

Since the study was not undertaken with the presence of a researcher, it was fundamental to provide very clear explanations for the tasks, especially about how to provide meaningful answers. Illustrative answers were thus provided at the beginning of the test and detailed explanations about the desired answers were given, e.g. "Please try to provide a precise answer that will help us understand what you think and why you think it. Avoid general answers like 'because it's clear/not clear' ". More importantly, as previous experience shows, the icons are meaningful only if considered in context: firstly, the icons and their elements have to be understood as part of a set because only in this manner some elements become understandable (e.g. the recurring hand symbol to signify the data subjects' rights); secondly, the icons' function as information-markers in privacy policies must be made clear. This is why the icons were displayed in groups according to their conceptual category and the relevant section of privacy policy (see was displayed right above the icons pertaining to that section to provide enough contexts to the respondents. The choice to display the mock-up of a visualized privacy policy was also motivated by the need of reproducing similar conditions to the second user test (*see* Appendix A. Early feedback about the questionnaire design and the questions' wording from two colleagues helped to set up a smooth experience for respondents that had no familiarity with the icons and the typology of questions.

Three tasks were designed:

Task 1: Subjective rating on the icons' correspondence with its underlying meaning: a subjective rating on a Likert scale that ranged from 1 to 5 about the ability of a certain icon to represent the correspondent concept, expressed through a label and a simplified definition, was chosen. Since the researcher was not present, an explanation for the marks was provided: 1) Strongly disagree (you find it impossible to associate the icon with its meaning); 2) Agree (with some changes or efforts of interpretation, the icon could work); 3) Neither agree nor disagree (you do not have sufficient elements to express your opinion); 4) Agree (the icon can work,

[2] http://gdprbydesign.cirsfid.unibo.it/

but it needs minor improvements); 5) Strongly agree (you could not think of a better icon for the concept). Explicit explanations for the mark were asked and recorded in written form in a dedicated space below.

Task 2: Alignment between users' mental models and designers' mental models: the participants were asked to attempt to provide an explanation for the visual choices made by the designers. This task was meant to find out whether users could understand the reasons behind the choices and thus align their line of reasoning with the designers', despite their opinion on the appropriateness of a certain icon for a certain concept, evaluated in the previous task. In order to avoid influence by different wordings, the same identical simplified definitions that were distributed to the designers to spark the design process, were also provided to the test participants.

Task 3: Best Alternative: five alternative choices among two or more icons were asked.

3.4.4 Analysis

The data collected in written form by the participants was gathered and analyzed by one of the interviewer in search for common patterns. Given the nature of the tasks and the study goals, qualitative analysis was the main source of data. 14 icons were evaluated: whereas some underwent assessment for the first time, for some others it was the best alternative among multiple options the focus of the assessment.

For the fitness of correspondence between icon and concept for the results of data processing operations (i.e. "anonymized data", "pseudonymized data", "encrypted data", "profiling", "automated decision-making", "storage of data inside of the EU") and for the rights of the data subject (i.e. "rights of the data subject", "right to object to processing", "right to lodge a complaint to a supervisory authority") (task 1):

Average : mathematical mean across the users' self-reported marks was computed for each icon;

Furthermore, the explanations provided by the participants to motivate their choice were analyzed to find common lines of reasoning, but also to find out explicitly about the words that were employed to describe why some associations were easier/harder than others.

For the alignment between users' and designers' mental models of all the icons, i.e. the ones listed above in task 1 and below in task 3 (task 2), hit rate and error rate were computed as in the second user study (see Section 3.3).

For the best alternative between the alternative icons for the concepts of: "contract", "legal obligation", "public interest", "purpose of provision of the service", "data sharing with third parties" (task 3) (*see* Figures in Section 3.4.5):

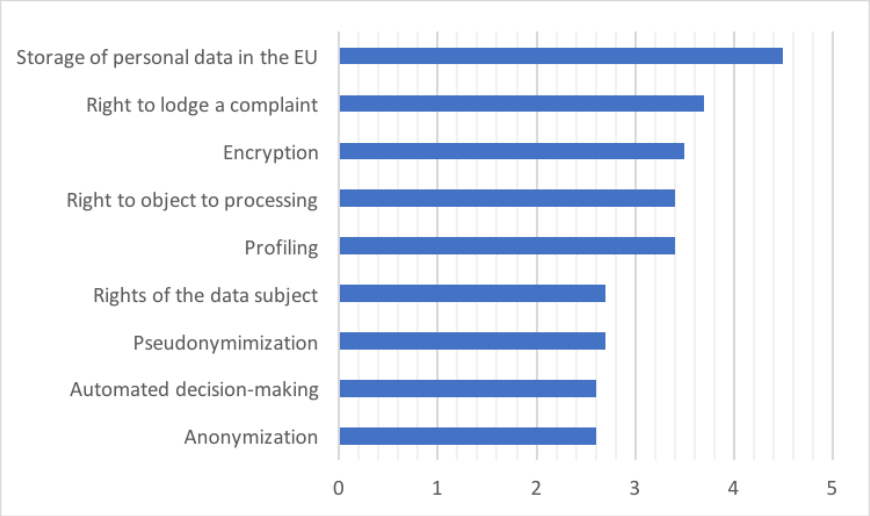


Figure 3.15: Means of the self-reported values for ease of understanding, where the closest to 5, the better the results

Higher number of preferences : the best alternative was chosen simply by counting which icons got the majority of votes. Reasons for the choice were also recorded.

3.4.5 Results

Task 1: Rating on Fitness of Correspondence

The results of the assessment on the icons' capacity of representing the underlying concept are reported in Fig. 3.15. The results reported in the following paragraphs are organized in three groups, according to the average value obtained by each icon:

1. best rated icons: average value ranging between 5 (=completely agree with the fitness of correspondence between icon and concept) and 4 (=agree);
2. medium rated icons: average value ranging between 4 (=agree) and 3 (=uncertain);
3. worst rated icons: average value ranging between 3 (=uncertain) and 2 (=disagree).

Best rated icon: the only icon that received very positive ratings is the symbol representing the **storage of data in the European Union** (*see* Fig. 3.16), which is the counterpart of the transfer of data outside the EU. Every participant was able to recognize the stars in circle as the "EU flag" or "the symbol of the EU" or similar, while the personal folder placed in

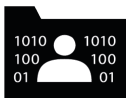
Figure 3.16: The best rated icon in Task 1 is the storage of data inside the EU

the middle was understood as data "physically stored inside the European Union" (P4).

Medium rated icons: there are four icons that received marks between "I agree" and "Neither agree nor disagree" (Fig. 3.17). The icon for **right to lodge a complaint to a supervisory authority** was quite positively evaluated because a person sitting on an armchair and behind a table was seen as "person with power or authority" (P2) or "someone who is responsible and can make authoritative decisions" (P4). **Encrypted data** ranked next, because the combination of zeros and ones was interpreted as "content that is not readable by everybody" (P4), "binary language that is symbolic [...] and expresses a code" (P5). Those participants that gave lower scores mainly appointed it to the fact that they would have expected a padlock (*see* Section 3.4.6). As for what concerns the **right to object to processing**, the gears were positively welcomed as expressing processing (although a few participants pointed out that they are more readily associated to settings), while the fact that they are broken was also easily understood as interruption, although a few people also suggested alternative symbols (*see* discussion). Similarly scores received the icon for profiling, with participants appreciating the idea of the puzzle pieces relating to "different aspect of personal information" (P1) through which "it is possible to reconstruct the individual identity and preferences" (P2) or "reconstruct the behaviour" (P10) in a process of "profile-creation" (P3). However, four participants pointed out that the icon could be interpreted as "decomposition" (P9) of the data folder, as if the pieces were "separated" (P5), instead of composed together.

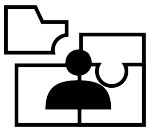
Worst rated icons: four icons received grades between "Neither agree nor disagree" or "Disagree" (*see* Fig. 3.18). Although the diamond in the **rights of the data subject** was correctly interpreted as "my rights are important" (P10), "something precious" (P8), "invaluable right" (P4), "something valuable and important" (P2) or "relevant" referring to the rights, not everybody understood the metaphor, while the hand was interpreted with difficulties. The icon for **pseudonymized data** and **anonymized**

(a) Right to lodge a complaint to a supervisory authority



(b) Encrypted data

(c) Right to object to processing

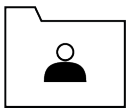


(d) Profiling

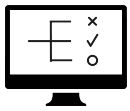
Figure 3.17: The medium rated icons in Task 1

data also received similar low marks, mostly because the symbolic differences of colors and the icon representing the personal data folder was grasped, but deemed difficult to readily associate with the intended meaning and to distinguish among similar icons. Lastly, **automated decision-making** was deemed very difficult or impossible to associate to its underlying concept, with explanations like "it takes a good effort" (P7), "I cannot *see* neither the 'decision-making', nor the 'automated' concept" (P9), "I do not *see* the connection" (P10).

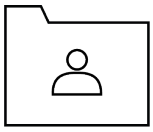
(a) Rights of data subjects



(b) Pseudonymized data



(c) Automated decision-making



(d) Anonymized data

Figure 3.18: The worst rated icons in Task 1

Task 2: Alignments with Designers' Intentions

Figure 3.19 displays the results for each visual element with self-contained meaning appearing in the batch of icons analyzed in this last phase. The EU flag stars, the changing colors of the user's silhouette and the puzzle pieces were more

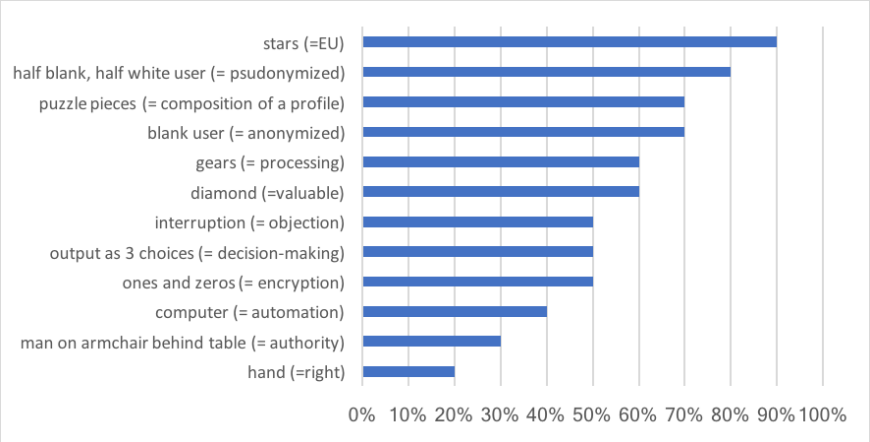


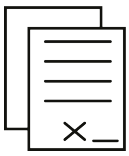
Figure 3.19: Percentage values of correct explanations for designers' iconographical choices

correctly associated to their intended meaning, whilst the designer's intention behind the hand, the authority, and the computer was not readily grasped. The results also show some errors of interpretation that will be discussed in Section 3.4.6.

Task 3: Best Alternatives

Following the dismal results from the second user study, two alternative icons were designed for the concept of **contract** (*see* Fig. 3.20): one that more precisely represents the legal relation between the user and the controller (Fig. 3.20a) and another that, for the sake of usability, only represents the written agreement (Fig. 3.20b) . Seven out of ten respondents preferred the representation showing the legal bound between two entities because the document alone "could be anything" (P7) and "not necessarily a contract" (P4), but "could be a simple letter" (P5).

(a) Precise version

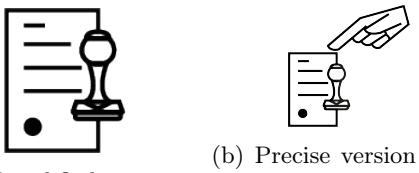


(b) Simplified version

Figure 3.20: The two alternative icons representing a contract

Since the second user evaluation showed that the icon for the **legal obligation** (Fig. 3.21a) was deemed too similar to a certification and not enough specialized in the legal sense, two alternative icons were tested together with the

original one (*see* Fig. 3.21). The original idea for the icon, that had been simplified for usability reasons, was recovered: a pointing hand was thus added to the icon in two different versions, a simpler icon without the stamp (Fig. 3.21c) and a more elaborated one with it (Fig. 3.21b). The preferences of the respondents are distributed almost evenly among the three alternatives: three respondents preferred Fig. 3.21a because "a sealed document shows an obligation" (P6) and "the hand in the other icons was not meaningful" (P9); four people elected Fig. 3.21c "because of the hand (authority) and also a hand pointing at the rules making it look stricter" (P8) and "the rubber stamp gives me the idea of a 'legal' process" (P10); finally, three participants chose Fig. 3.21b, citing similar reasons for the hand (e.g. "an external intervention (law in this case) that obliges me to do something" (P2) and "the hand makes me think of something mandatory" (P5)), but the stamp recalls a registered contract (P2), is unnecessary (P4) or is unrecognizable (P5). The combination of the answers suggest that the pointing hand indicating the obligation can be meaningful, while the details of the sealed document can be simplified, as long as the symbol is not mistaken with another official document (e.g. a contract, a certification, etc.).



(a) Simplified version

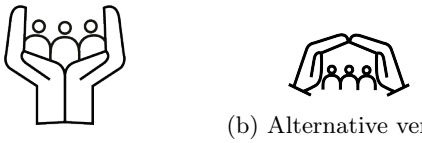
(b) Precise version - Option A

(c) Precise version - Option B

Figure 3.21: The three alternative icons representing a legal obligation

The third symbol that was investigated in this phase is the hand related to **interest**, which is the basic element that, combined with others, generates the icons representing "public interest", "interest of the controller", and "vital interest". In order to focus participant's attention on the hand element, the evaluation was carried out on the example of "public interest", which had shown less interpretative issues compared to the other two in the preceding user test. The original icon (Fig. 3.22a) was thus compared to an alternative version with the hands placed in a different manner (Fig. 3.22b). The original version gathered seven preferences out of ten, with the symbol of hands interpreted as "supporting a group of people" (P2), while the second icon was interpreted univocally as "protection" by all participants, a concept not corresponding to the "promotion" of the public interest (P5). This clearly means that, although the meaning of the "supporting hands" might not be totally transparent, the "protecting hands" is certainly not a good alternative.

As the discussion of results in the last section highlighted, one of the most controversial graphical symbols was the one attempting to represent the **purposes of provision of the service** (Fig. 3.23a) and the very similar symbol for enhancement of the service. By taking into account the comments received in the preceding user test, three alternatives were designed: one icon closely

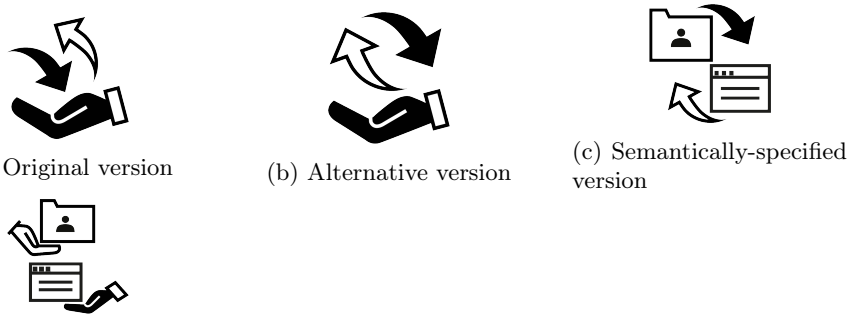


(a) First version

(b) Alternative version

Figure 3.22: The two alternative icons representing the public interest

resembling the original, but with arrows suggesting a circular movement (Fig. 3.23b); one icon that is more semantically specified, with arrows signifying the exchange between personal data and a service, exemplified by a webpage (Fig. 3.23c); and lastly, a similar icon that, instead of the arrows, re-uses the contrast of colors between hands adopted by other symbols pf the set to signify the data subject providing data on the one side and the controller providing the service on the other side (Fig. 3.23d). The preferences are evenly distributed between the two icons that only contain arrows and the two icons that more specifically represent the exchange of data with a service. The two participants selecting the original icon motivated it with the fact that it is more generally representative, while the three people preferring the other circular disposition of the arrows either did not provide any meaningful explanation, or appreciated the order of the two arrows because "we often give before we receive [...] which is often the case when a service is provided" (P5). The icon in Fig. 3.23c received four preferences because of its precise and concrete nature, but was also deemed quite complex, while the fourth choice only received one mark in opposition to the others in which the arrows "are too related with recycling images". In conclusion, the debate around the best manner to represent this concept is still open.



(a) Original version

(b) Alternative version

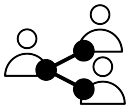
(c) Semantically-specified version

(d) Semantically-specified version

Figure 3.23: The four alternative icons representing the purposes of provision of the service

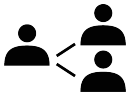
Lastly, three alternative icons for the concept of **data sharing with third**

parties were evaluated: one icon based on the standard symbol of sharing used nowadays on many applications, combined with three parties (Fig. 3.24a); a second icon representing a simplified globe (Fig. 3.24b); and a third icon representing three interconnected parties (Fig. 3.24c). The icon employing the popular symbol of sharing received seven out of ten preferences, precisely because the symbol is "well-known" (p8), "familiar" (P1, P3) and "already in our common understanding" (P2). The alternative representing the globe was chosen in one case because it highlights "the possibility (danger) of data to travel to other (far-away) parties [that] may include a significant loss of control by the user" (P9), while the third icon best represents a "network of people" (P5).



(a) Sharing symbol and three parties

(b) Global reach symbol



(c) Three parties

Figure 3.24: The three alternative icons representing the data sharing with third parties

3.4.6 Discussion of Results

Future Redesign It was expected that one of the best rated icons would have been the one representing storage of data inside of the EU, because of the familiarity with the stars of the European flag. The marks for the icon representing the right to lodge a complaint are satisfactory because the design of the supervisory authority witnessed many iterations. However, some participants pointed out that the fact that there is the possibility to make a complaint is not understandable. For sure, an alternative showing a padlock for encrypted data will need to be explored because, as some people stressed in their answers, it is a more recognizable symbol for the ordinary user than the metaphor of zeros and ones, that is more correct but probably less transparent for people without technical experience.

The icon for the right to object to processing received similar comments to the icon for the right to restriction of processing: the gears have been more easily associated with processing than in the previous experiment, one reason probably being that the participants were more international. However, some comments pointed out that they can be confused with settings. Moreover, a few suggestions made reference to the fact that the objection could be symbolized by the symbol for a stop/alt, because the broken gears might rather suggest that there are difficulties with the processing. It was also expected that the representation for the rights of the data subject could have been easily misunderstood, given the abstractness of the concept and the metaphor behind the diamond. This symbol

was easily linked to something "valuable and important for the user" (P2), but it was also pointed out that it "makes the assumption that people care about their rights" (P7), which is not always the case. Nevertheless, this design choice was deliberately metaphorical and positive to attempt to convey the importance and value that rights can assume for data subjects.

Contrary to the researchers' expectations, the icon for automated decision-making was not well rated. The comments reveal that, although the decision-making process was more easily grasped, the computer symbolizing the automation was not understood. Some participants suggested to explore graphical symbols concerning robots, because more easily associated to automation, or to stress the absence of human intervention.

Directionality of Interpretation Some comments on the graphical symbol representing the concept of profiling also require attention: a few respondents were undecided on whether the puzzle pieces are gathered together to compose the personal data folder or, vice versa, if they are separated from each other in a process of decomposition. Indeed, the image can be read in both senses and icons from the first icon set had received similar comments. However, as it will be discussed in the next chapter, it is hard to show the direction of a process in a static image, whereas movement can be easily conveyed by a gif or any other sort of animated visual.

Black and White Colors The icons for anonymized and pseudonymized data scored badly, but the answers of the respondents reveal that, once that the difference among the colors is noticed, than the metaphor behind the colors is understood: a black user silhouette to identify personal data, a blank silhouette for anonymized data, and a half blank and half white silhouette for pseudonymized data. However, two considerations are necessary: firstly, the meaning of the colors can be grasped only if each icon is shown in combination with at least another icon, as also some respondents hold; secondly, from this derives that such difference might be too subtle to be readily understandable. Many participants also underlined that an additional difficulty had to be ascribed to the complexity of the concept itself.

Usability versus Precision On a general note, the tension between simplicity of design, relevant for usability reasons, and preciseness of representation, important for legal reasons, re-emerged prominently also in this user study. Icon alternatives for the concept of contract, legal obligation and provision of the service were redesigned exactly because in previous experiences the need for more precise representations had emerged. As it will also be discussed in the next chapter, however, this opposition is not easily solvable and can also depend from individual preferences: comments from this last user study show that whereas for some individuals the precision of representation of a concept is of utmost importance, for others simplicity over the complexity of design has to be favored. For example, compare the comments of two respondents that motivate their choice of a

contract icon over the other: whereas P4 motivates her choice as "I prefer the icon with the two silhouettes in it because it conveys the fact that this document (contract) is agreed upon between the two parties. The document without any silhouettes in it could illustrate a set of rules and not necessarily a contract", P10 writes: "I think the simpler the better. The two users are redundant". Hence, more research on this point is needed since this question has re-emerged persistently during every workshop and user study.

One more point that deserves discussion is raised by the representation of the concept of data sharing with third parties. Whereas the two icons displaying three users more closely and literally resembled the underlying concept, and for this reason were preferred by the majority of respondents, the icon displaying a globe focused on the fact that data can be scattered, without exact knowledge of its recipients. In this light, this icon can also, to a certain extent, convey the risks inherited by the data sharing.

Semantic Specifications Another recurrent comment concerned the absence of the folder representing personal data as building blocks in other icons, as in the icon representing the sharing with third parties. Some users would have expected to *see* the application domain, i.e. personal data, in the icons, otherwise the visual elements seemed to them not enough determined, i.e. what is shared with third parties? A similar discussion was opened up in the last user study, with some participants calling for the presence of the folder representing personal data in, for example, the right to rectification because according to them it was otherwise impossible to determine what this right concerned, i.e. the scope of rectification.

Although these comments calling for a semantic specialization in the domain of personal data are reasonable, two considerations are required: firstly, it is the privacy-related context (e.g. the privacy policy) that provides this semantic specialization to the icons and makes it redundant to specify it with visual elements that would make the icon unnecessary crowded. Secondly, for the very functional nature of icons, it is impossible to specify meaning to that extent: as it will be advocated in the next chapter, other elements (like pictograms or comics) are more suitable for an exact and detailed representation of meaning.

3.4.7 Limitations of the Study

Given that it reproduces the same method of the second user study (*see* Section 3.3), also this third user study presents similar limitations concerning the fact that it does not reproduce real-world conditions of icon interpretation, since the participants had the time to scrutinize the icons with attention and the corresponding concept was provided. Nevertheless, this study, like the previous one, only attempts to gather information on people's sense-making process and to offer subjective evaluations about the fitness of correspondence between a symbol and its concepts.

A major limitation constituted by the representativeness of the pool of participants, that were only ten, young, and well-educated, which does not correspond to the majority of Europeans. Although their demographics was expressly selected to be similar to the participants from the previous study, it can be the case that their ratings on the icons excluded in that study would have been different. The fact that the evaluation was not done in the presence of a researcher might as well have influenced the results, because the participants might have felt more free to express negative judgments.

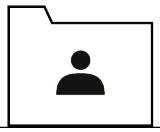
Finally, since this was the third iteration of the icon evaluation, only a selection of icons was shown to the study participants: those that had not received any previous evaluations and those that had been vetted after the last user study. It was attempted to counterweight this limitation by showing the context where the icons would appear, e.g. the privacy policy's section containing the icons of the same conceptual category. However, in the next evaluation phases, it will be fundamental to consider all the icons at the same time, because the interpretation of one icon can be built on and supported by the previous interpretation of a similar icon, as the example of the category of data subjects' rights shows.

In conclusion, the results exposed in these pages only intend to provide a preliminary indication of the more promising icons. Nevertheless, more experimentation is needed and further considerations on this point will be provided in the next chapter.

Chapter 4

The Final DaPIS

According to the results discussed in the previous pages, it can be safely assumed that the icons in Table 4.1 can be adopted. Table 4.2 displays those icons that need some refinements or modification. Table 4.3 presents those icons that need further thought or alternatives, whereas icons of Table 4.4 need further research to determine their effectiveness.

Icon	Corresponding concept
	Personal data
	Data subject
	Right to be Informed
	Right to erasure
	Consent
	Right to withdraw consent

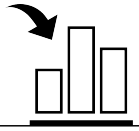
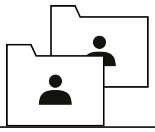
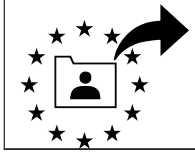
	Right of access
	Enhancement (in combination with other visuals e.g. service enhancement)
	Public interest (but see critics on interest)
	Statistical purposes
	Security purposes
	Research purposes
	Legal
	Copying
	Transfer of data outside of the EU

Table 4.1: Icons that can be safely adopted

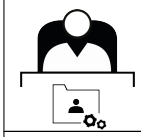
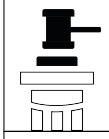
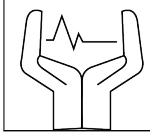
Icon	Corresponding concept	Suggested revision
	Marketing	add a euro symbol
	Controller	it must be more easily distinguishible from the data subject
	Supervisory authority	simplify the part below
	Right to data portability	make movement more evident
	Right to rectification	consider to add a personal data folder to specify its meaning
	Legal bases	erase the column or make it look more like a column
	Vital interest	make it look more like an ECG or provide some context

Table 4.2: Icons that need some minor revisions

Icon	Corresponding concept	Reason
	Gears	more readily associated to settings than processing

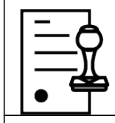
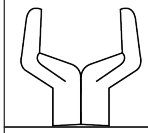
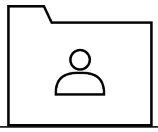
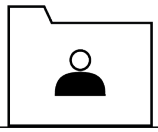
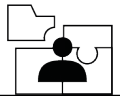
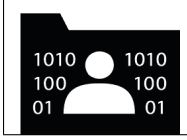
	Service provision (and enhancement)	arrows seem to be too abstract to signify the exchange of data for a service
	Legal obligation	it has been confused with certificate, too general, should be more legal
	Restriction of processing	the color difference might be too subtle and another symbol for limitation might be needed
	Interest	hard to grasp, too open to interpretation
	Purposes	an alternative could be sought

Table 4.3: Icons that need further thought or even need an alternative since they scored badly

Icon	**Corresponding concept**
	Anonymization
	Pseudonymization
	Profiling
	Encryption

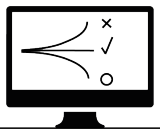
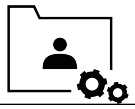
	Automated decision-making
	Processed personal data
	Derived personal data
	Storage of data inside of the EU
	Right to object to processing
	Data subjects' rights
	Right to lodge a complaint to a supervisory authority
	Contract
Missing	Third party

Table 4.4: Icons that need further research

Chapter 5

Conclusions

As other researchers before us [22, 15], we also end with a note on the necessary push towards (a) education of data subjects on privacy-friendly behaviors and data protection topics and (b) towards standardization, since there will always be a margin for individual, i.e. free, interpretation. As of 25 May 2018 the GDPR applies in the EU, therefore such actions would be timely. Whereas researchers from many disciplines (law, semiotics, human-computer interaction, design, computer science, cognitive psychology, philosophy, behavioral economics, etc.) can offer valuable insight for the development and correct user testing methods of data protection icons, it must be a goal of the regulators to find means and resources to carry out data protection education campaigns for European citizens and standardization initiatives that can reach considerable quantities of population. Every segment of society should be possibly included and individuals of each European country should be reached. They should be diverse in terms of age, gender, educational background, profession, technical proficiency, legal knowledge, and privacy awareness. This is a very challenging and ambitious goal, but it is a necessary step to produce icons that can be safely used at the European level.

Nonetheless, it is impossible to produce an icon set that will be considered perfectly representative of data protection concepts, i.e. perfectly semantically transparent. Usability testing can give important insights as for what concerns legibility, but very high rates of ease of recognition will never be reached for unfamiliar concepts or icons, until they will be widely adopted. For this reason, standardization open to versioning is the path to follow: after one icon set has been publicly discussed and adopted, empirical data on its use should be gathered. Subsequent versions will consider and integrate comments on the first version, together with the needs of evolving societies and regulations. Towards this goal, each design choice made about the icons is reported and examined thoroughly in this report: so that public discussion can contribute to their development.

Although visualizations can support the sense-making of legal information, a question remains open: considered the characteristics of data protection con-

cepts, how effective can be the corresponding icons at revealing their meaning? Can icons assume an informative or educational role to a certain extent, i.e. can they convey unknown notions to data subjects? To a certain extent, as some examples from the user studies reveal, they can. However, there are probably different visual means that can be more effective to reach this goal: pictograms, comics, infographics, and videos are obvious candidates. Nevertheless, during one of the workshops for icon creation, an interesting idea was advanced: produce icons that are not immediately understandable by the users but that, because of this reason, invite them to the exploration of their meaning, and thus of the privacy policy. Of course, this could pose issues of interpretation, if subjectivity in lieu of uniformity is valued. However, this kind of visualizations would tackle a different problem: not that of lack of comprehension of legal terms, but namely that of lack of motivation to read them. It is in question, also, whether a privacy policy would be the appropriate *locus* to instill new knowledge into data subjects.

In conclusion, icons will not solve many problems outlined in the introduction: privacy policies as they are now fail to be informative. And it is not solely a matter of information design. Their language is typically vague and ambiguous and the amount of information provided is usually excessive for data subjects (although suitable for those monitoring organizations' data practices such as regulators, supervisory authorities, and advocates [30]). Experience and empirical research show that users are desensitized by too many consent requests and that providing long and tedious privacy policies while the user is carrying out another task is only deemed as a nuisance: if a user is buying a train ticket, he does not want to be educated about all the possible collection of his personal information. Behavioral insights can be used either to favor the data subject or to deliberately obscure information and choices [28]. Data processing activities have a high level of complexity, therefore it is unrealistic to expect that placing icons on privacy notices will alleviate this burden of explaining and understanding such intricacy.

The GDPR will provide a strong push towards compliance with transparency principle, though. Interdisciplinary studies can shed light on many overseen aspects of information transparency to empower data subjects to be more in control of their flow of data. New ways of communicating data practices are spreading, although they constitute only a minority [23]. More research on methods for creating and evaluating legal icons is needed.

Chapter 6

Future Directions

Notwithstanding the two studies carried out on DaPIS, a number of questions remain open and will need further research. These matters are presented in the following.

6.1 Effects of Training

Since one of the main obstacles to ease of recognition is the lack of familiarity with the icon or with its referent, the effect of **training** should be expected to determine positive outcomes and easier recall. For instance, after the first comprehension test, a brief explanation of the symbol can be provided [53]. A second test, after a short amount of time, can be administered to the same subjects to evaluate whether their recognition rate increases over time whilst the false alarm rate decreases, as it is expected after a first exposure. Learning ease can be determined also through more test reiterations and would arguably mirror more closely the actual sense-making of icons in a privacy policy by the data subject than a one-time only recognition test, especially within the view of affirming a standardized icon set. Similarly to symbols on GUIs, it is to be expected that as the icons' use spread and users will gain familiarity with them and their underlying meanings, they will be able to recognize them more readily and do less mistakes, until they will be able to appear and be effortlessly understood in isolation, i.e. without textual labels or explanations, as on IoT devices or during online transactions (e.g. while carrying out a different tasks).

6.2 Number of icons

The icon set produced during the research described in these pages sums up to 42 icons. Although not all of the icons are expected to appear in a privacy policy at the same time and the ontology-based approach has hindered the proliferation of one icon for each individual, research must be devoted to the cognitive overload that such a numerous icon set might cause, especially for first

exposures. If the number of icons is too high to be easily digested and managed, the risk of feeling overwhelmed might arise, which would be counterproductive to the very goals of DaPIS. Furthermore, whereas icons could prove beneficial in first trials because they can attract reader's attention or even curiosity (acting against habituation effects), the effect of habituation over time must also be researched. In this respect, it would be probably useful also to experiment whether the selection of a limited number of icons to be displayed would be more meaningful than the entire icon language. In particular, two scenarios can be envisioned. In the first one, only risky practices or practices that would have a significant effect on the individual are presented in a visual manner, such as the transfer of data outside of the EU and the existence of automated decision-making. The second scenario is shaped around the arising possibility of customization. Indeed, many studies show that expectations, needs, and fears around privacy depend on individual characteristics, thus vary greatly. If, as proposed by the draft ePrivacy Regulation, browsers will directly manage our preferences, the possibility of being shown only the icons that matter to us could become a reality. Artificial intelligence could also play a role in this sense (see [27, **?**, 12]).

6.3 Degree of discriminability

Another dimension that must still be researched is the extent to which each icon is discernible from the others of the set, which is a crucial index for ease of recognition. Identifiability is indeed a relevant dimension: the icons are part of a set and the less they overlap in terms of similarity, the more they will be memorisable. In other words, their design should be sufficiently consistent to identify them as a family of icons, but also sufficiently distinctive to make each element easily distinguishable from the others. For instance, the difference between data subject and controller was not sufficiently relevant in the second user study. This could be carried out as a hit rate task with multiple answers where all the icons under the same class (e.g. rights) or having similar meanings (e.g. consent, right to withdraw consent, etc) are displayed.

6.4 Alternative creations and alternative choice

Given the results of our tests, notwithstanding the inescapable arbitrariness that some icons present, alternative graphical symbols can be considered for the same referent. Indeed, as participatory design methods presume (see Sect. 2.2), there is no fixed solution for a given challenge and in the context of icon design, creativity and whit can create more than one alternatives, provided that a robust methodology is followed. The best candidate can than be elected by a pool of users, even considering their individual characteristics. We invite the readers of this report to submit their alternative proposals for those icons that were poorly rated.

6.5 User Experience

Finally, the user experience should also be considered to determine the users' perception and interaction with an artefact, such as a visualized privacy policy. The extreme length of traditional privacy policeis makes the data subject feel helpless and frustrated, whereas a comprehensible and navigable text can trigger positive emotions, such as satisfaction [44].

6.6 Final Evaluation of the Icons in Context

A subsequent, necessary step to test the effectiveness of DaPIS would be to test the icons in a real context, as anticipated in Chapter 3). This means, for instance, that it is necessary to address questions about how users will make sense of the icons.

It is unrealistic to expect that data subjects will attentively read every line of a privacy policy everytime that they encounter one. Unless the data subject is deciding whether to enter into a contract with a service provider[1], it is more realistic that she will look for specific pieces of information in a limited time span, i.e. she will exercise her strategic reading. Most of the times, the data subject will be asked to agree to the conditions set forth in a privacy policy while executing a different task, e.g. while buying a flight ticket. In this case, the expectation that users will carefully read the privacy policy is unreasonable and the obligation to read the terms will be experienced more as a nuisance (an interruption from the primary task: buying a ticket rapidly and efficiently) than a legal guarantee. It is in such cases, however, that a compact array of icons that summarizes the data practices, or at least the most significant or risky ones, could prove helpful. In this case, the icons would be interpreted as stand-alone elements to offer "in an easily visible, intelligible, and clearly legible manner a meaningful overview of the intended processing" (Art. 12.7 GDPR).

Another realistic set where the interpretation of DaPIS can be tested would be an online, interactive interface where the icons complement the text and act as navigation cues. In this context and by reproducing realistic user tasks, an association task between symbol and referent to assess the icons' usability would be meaningful if carried out as an information finding task. I.e., given a specific privacy policy, how would a user make sense of it? Which elements would she use? Where would she look for specific information items? On which icon would she click to open and expand the relevant section, provided that the policy is interactive and organized in meaningful paragraphs? Such a setting for the usability test would provide higher ecological validity to the study and probably determine higher icons' recognition rates.

The icon set is currently undergoing a final redesign to include the results of the second user test.

[1] And even in this case, nowadays contracts with online service providers or online retailers are concluded with a quick tick of a box or a click on a button (clickwrap agreements).

Appendices

Appendix A

An Example of Visual Privacy Policy

Privacy Policy

Date of application: 25/05/2018

General information

This notice (the "Privacy Policy") is intended to inform you about our practices regarding the collection, use and disclosure of personal information that you may provide via this website or our mobile applications (the "Platforms").

In this notice, the following roles are described:

The controller: .....

.....

The user (or data subject):.....

.....

How do we process your data?

We encrypt your data:.....

.....

We make automated decisions about you: .....

We anonymize your data:.....

.....

We profile you:.....

.....

We pseudonymize your data:.....

.....

Why do we process your data?

On which legal bases?

For security purposes:	Legitimate interest of the controller
For marketing purposes:	Contract
For the provision of the service:	Consent
For purposes of service enhancement:	Legal obligation
For research purposes:.....	Vital interest
For statistical purposes:	Public interest

Where do we keep your data?

We transfer your data outside of the EU:
.....

What are your rights?

You have the right to be informed:
.....

You have the right to data portability:
.....

You have the right to access your data:
.....

You have the right to withdraw your consent:.....

You have the right to rectify your data:
.....

You have the right to object to the processing of your data:.....
.....

You have the right to restrict the processing of your data: .....
.....

You have the right to lodge a complaint to a supervisory authority: .....
.....

You have the right to erase your data:
.....

Appendix B

The First Layer of a Multi-layered approach

CLIXBUS

In order to complete the payment, please enter your credit card details:

Cardholder name

Card number

Expiration date / CVC

☐ I accept the Privacy Policy. A summary of the main points:

Data controller:
Clixbus SpA
Via Roma 2
40121 Bologna
privacy@clixbus.it

Why do we process your personal data?

To offer you our service: to print your name on the ticket and to process your payment

Where do we process your personal data?

We transfer your data outside the EU: we store and process your data in the US

Your rights:

You have the right to be informed about how we use your personal data

You have the right to access the personal data we hold about you

You have the right to correct your personal data, if it is inaccurate

You have the right to erase your personal data. We may keep some of your personal data in specific cases

You have the right to object to the processing of your personal data for marketing purposes

You have the right to restrict the processing of your personal data. This right only applies in specific cases

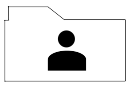
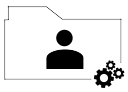
You have the right to data portability: you can receive and/or have us transfer to another data controller that concern you and that you provided us

You have the right to lodge a complaint to a supervisory authority if you believe that your rights have been infringed

Appendix C

The first DaPIS: Personal Data Types, Processing Operations, and Agents' Roles

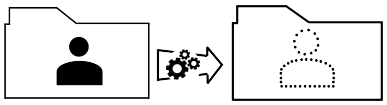
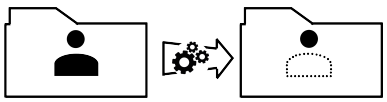
C.1 Personal Data Types

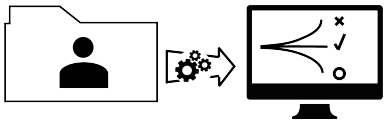
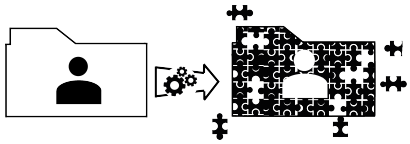
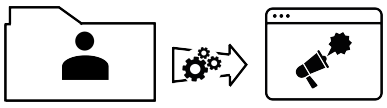
Icon	Description	Legal ref.
	**Original personal data** Definition: it is the personal data provided by the data subject, either directly or observed from her behaviors. Rational behind the choice: Typically, folders contain data and this symbol is widespread on graphical user interfaces, whereas the user's silhouette signifies the data subject.	[4]
	**Processed personal data** Definition: it is the personal data after they have been processed, thus after they have been stored, organised, structured, modified,combined, etc. Rational behind the choice: Gears is a common symbol for (mechanical) transformation or processing. Used as denominator, it indicates that the data contained in the folder has been processed.	[4]

	**Derived personal data** Definition: it is the inferred and derived data generated by the controller from the analysis of the original data. Rational behind the choice: Inferred and derived data is not provided by the data subject. It originates from other data and tells something more on the data subject, hence the pluses that enter into the folder and add novel information to it.	[4]

Table C.1: Icons, respective definitions, and rational behind the visual choice for the icons of the class "personal data"

C.2 Processing Operations

Icon and Description	Legal ref.
 **Anonymization** Definition: it is the process that strips personal data of sufficient elements such that the data subject can no longer be identified. Rational behind the choice: This icon, as the ones reported below in this table, shows a process: the personal data, on the left, are processed (represented by the arrow with gears) and become anonymous. Whereas in the icon on the left the silhouette is black to identify a specific user, it becomes blank and dotted to signify that data was striped of identifiable elements.	Rec. 26 [16] ; [5]
	Art 4.5 [16]

Pseudonymization Definition: it is the process through which personal data can no longer be attributed to a specific data subject without the use of additional information (provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person). Rational behind the choice: This icon is based on the symbol for anonymous data. The silhouette is not completely blank (=‘pseudo’) because it is possible to re-identify the data subject, by retrieving the information that had been separated.	
 **Automated decision-making** Definition: it is the ability to make decisions by technological means without human involvement. Solely automated decision-making, including profiling, produces legal effects or significantly affects the data subject. Rational behind the choice: The three options stand for possible decisions that can be taken. The absence of a human, replaced by a computer, represents the fact that the decisions are taken automatically.	Art 22.1 [16]; [6]
 **Profiling** Definition: it is any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviors, location or movements. Rational behind the choice: Many pieces of a puzzle are combined together to compose the profile of a data subject.	Art. 4.4 [16]
	[18]

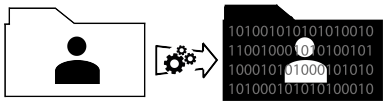
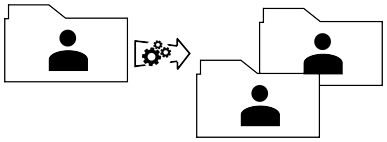
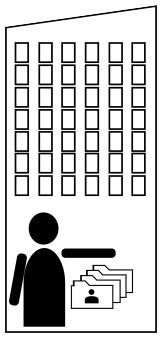
Direct marketing Definition: The communication by whatever means of any advertising or marketing material, which is carried out by the Direct Marketer itself or on its behalf and which is directed to particular individuals. Rational behind the choice: The bullhorn stands for a tool that amplifies the advertisement slogans, represented by the speech balloon. The web interface exemplifies the usual place of display of advertisements (typically online).	
 **Encryption** Definition: Encryption is a mathematical function using a secret value - the key - which encodes data so that only users with access to that key can read the information. Rational behind the choice: The binary code exemplifies a digital transformation of the personal data into encrypted data that can not be read by anybody.	Art 34.3(a) [16]; [33]
 **Copying** Definition: It is the act of making a copy of a certain data. Rational behind the choice: Two personal data folders are exactly reproduced.	Art. 15(3), 15(4) [16]
 **Transfer of personal data to third countries** Definition: It is the transfer of personal data which are undergoing processing or are intended for processing to a third country. Rational behind the choice: The stars in circle are the emblematic symbol of the EU, whereas the arrow signifies the movement of the personal data outside of the European borders.	Art. 44 [16]

Table C.2: Icons, respective definitions, and rational behind the visual choice for the icons of the class "processing operations"

C.3 Agents' Roles

Icon	Description	Legal ref.
	**Data subject** Definition: an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person; Rational behind the choice: The data subject can be generally identified with the user of a certain service, e.g. social media. Thus the silhouette of a user widely adopted on many applications to locate one's own profile's information can easily represent the data subject. To reinforce this idea, 'you' was added to establish a direct connection with the reader.	Art. 4.1 [16]
	**Controller** Definition: it is the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data. Rational behind the choice: The controller decides the destiny of the gathered personal data, represented by the folders. For such reason, this role is symbolized by a user with one raised arm, that exercises her decision-making on the personal data. Typically, the controller is a representative of an organization, such as a company, hence the building.	Art. 4(7) [16]

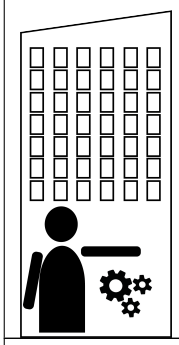
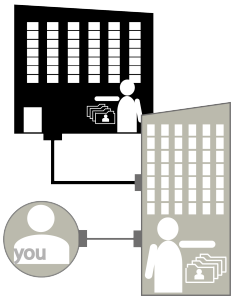
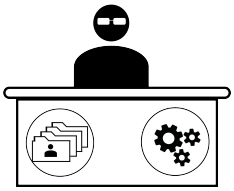
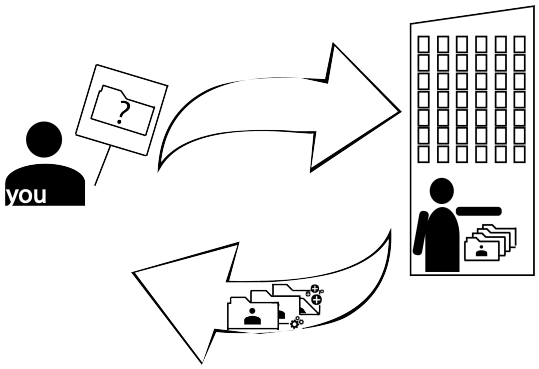
	**Processor** Definition: it is the natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller Rational behind the choice: The processor's icon has a similar structure to the controller's icon, but gears are displayed under its control because it carries out the processing operations.	Art. 4(8) [16]
	**Third party** Definition: it is the natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, is authorized to process personal data. Rational behind the choice: The third party receives the data subject's data through a controller, hence this transfer is symbolized by the cable that connects the data subject with the controller (direct transfer) and the controller with the third party (indirect transfer). The first and second parties (data subject and controller) are grayed out so that the third party can stand out.	Art. 4(10) [16]
	**Supervisory authority** Definition: it is an independent public authority which is established by a Member State to be responsible for monitoring the application of this Regulation, in order to protect the fundamental rights and freedoms of natural persons in relation to processing and to facilitate the free flow of personal data within the Union. Rational behind the choice: Representing this concept as a judge would have been misleading and inherently wrong, thus the authority is sitting at a massive desk and has reading glasses to carry out analyses with the goal of ensuring that a balance between the interests of data subjects (symbolized by the data folder) and controllers/processors (symbolized by the processing gears) is respected.	Art. 4(21) [16]

Table C.3: Icons, respective definitions, and rational behind the visual choice for the icons of the class "agents' roles"

C.4 Right of Access and Right to Data Portability

Icon and Description	Legal ref.
 **Right of access** Definition: The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data [...]. Rational behind the choice: This concept is represented as a narrative, where the data subject holds a sign on which appears a folder combined with a question mark that symbolizes the request of knowing which kind of personal data the controller has about her. The controller sends back personal data to the data subject: not only what she provided, but also the processed and inferred personal data.	Art. 15(1) [16]

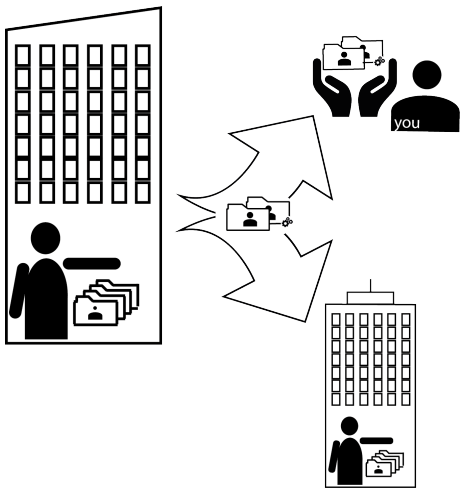
 **Right to data portability** Definition: The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided[...] [T]he data subject shall have the right to have the personal data transmitted directly from one controller to another, where technically feasible. Rational behind the choice: This concept is represented as a narrative, where the controller has two options: either she sends the (original and processed) data directly into the hands of the data subject, or to a different controller.	Art. 20 [16]

Table C.4: Icons, respective definitions, and rational behind the visual choice for the the right of access and the right to data portability

Appendix D

The second DaPIS: Data Subjects' Rights, Legal Bases, and Processing Purposes

The icons are displayedin Tables D.1, D.2, and D.3,, according to the class they belong to . An English translation of the simplified definitions provided to the workshops' participants is also shown, together with the reasons behind each iconographical choice.

D.1 Data Subjects' Rights

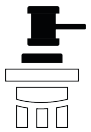
Icon	Description	Legal ref.
	**Data subject's rights** Simplified definition: these are the rights of those (data subjects) that have provided their personal data to an organization (company, e.g. Google or institution, e.g. tax office). Rational behind the choice: the hand means "holding", with metaphorical extension "being in control" or "have the power over" to indicate the possibility granted by a right to its holder. It is an iconographical choice in common with all the other data subjects' rights, whose meaning is specified by the object above. The diamond symbolizes a value, something precious that confers some kind of power to the data subject.	Ch. 3 [16]

	**Right to be informed** Simplified definition: data subjects have the right to know who does what with their data, how, and why. Rational behind the choice: the "i" is an internationally recognized symbol for information. For the hand, see above.	Art. 12, 13,14 [16]
	**Right to rectification** Simplified definition: data subjects have the right to ask the data about them to be corrected or updated in inaccurate and complemented if incomplete. Rational behind the choice: the pencil is a widespread symbol for editing in software applications: it erases incorrect data and rewrites them correctly. For the hand, see above.	Art. 16 [16]
	**Right to erasure ('Right to be forgotten')**: Simplified definition: In some cases, data subjects have the possibility to ask for their data to be erased. Rational behind the choice: the bin is a popular symbol for erasure in software applications. For the hand, see above.	Art. 17 [16]
	**Right of access** Simplified definition: data subjects have the right to know if someone owns data about them and to obtain a copy of it. Rational behind the choice: the folder with a user's silhouette is symbol of personal data, whilst the magnifying lens on the user indicates scrutiny of a specific person's data. For the hand, see above.	Art. 15 [16]
	**Right to withdraw your consent**: Simplified definition: data subjects have the right to revoke the consent on their data processing that they had previously given Rational behind the choice: the cross ("x") and the tick ("v") derive from the representation of consent (see legal basis). The arrow goes from the tick to the cross to signal the transformation from approval/acceptance to disagreement/disapproval. For the hand, see above.	Art. 13(2c) [16]
	**Right to data portability** Simplified definition: data subjects have the right to receive a copy of their data collected by a service provider A and transfer it to a service provider B. They can also ask for direct transfer from A to B. For the hand, see above. Rational behind the choice: the data folder, representing the personal data, takes the shape of a bag with handles to carry it around.	Art. 20 [16]

	**Right to restriction of processing**: Simplified definition: data subjects have the right to ask their data to be processed exclusively for certain purposes. For the hand, see above. Rational behind the choice: the gears represent processing activities, as in other icons. The processing goes on, but only partially: half of the gears continue to work and thus are black, whereas the other half is deactivated.	Art. 18 [16]
	**Right to object to processing**: Simplified definition: data subjects have the right to ask a service to stop processing their data for a certain purpose. For the hand, see above. Rational behind the choice: the gears represent processing activities. If broken, gears stop working. Two versions were produced and it was then in the test phase determined the preferred one.	Art. 21 [16]
	**Right to lodge a complaint to a supervisory authority** Simplified definition: data subjects have the right to file a complaint with a supervisory authority for data protection, if they think that their data is processed unlawfully. Rational behind the choice: see supervisory authority. For the hand, see above.	Art. 13(2d) [16]

Table D.1: Icons, respective simplified definitions, and rational behind the visual choice for the icons of the class "data subjects' rights"

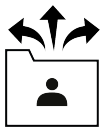
D.2 Legal Bases for Processing

Icon	Description	Legal ref.
	**Legal Basis** Simplified definition: It is the reason why data is processed and must be provided according to the law for the processing to be lawful. Rational behind the choice: the capital symbolizes the bases that bears the law, represented by a gavel.	Art. 6 [16]

	**Consent** Simplified definition: It is the expression of the data subject's willingness to have her data processed. Rational behind the choice: the cross ("x") represents a disagreement, whereas the tick ("v") represents an agreement. The slash conveys the idea of possibilities of an equal choice between the two, whilst the cross is expressly positioned before the tick to stress the chance to not consent, which is not usually the case, whereas the GDPR stresses the fact that consent must be freely given and informed.	Art. 6(1a), Art. 4(11) [16]
	**Contract** Simplified definition: It is an agreement that establishes a legal relationship between two parties. Rational behind the choice: The contract is usually represented as a written agreement that must be signed (hence the "x") by two parties: the data subject and the controller .	Art. 6(1b) [16]
	**Legal Obligation** Simplified definition: It is the duty to carry out what the laws says. Rational behind the choice: The law is represented as an official act, which is here signified by a stamped document with a stamp.	Art. 6(1c) [16]
	**Vital Interest** Simplified definition: A matter of life and death. Rational behind the choice: The two joint hands stand for protection or care, with metaphorical extension for someone's interest. It is an iconographical choice in common with all the other interests, whose meaning is specified by the object between them. The electrocardiogram is an established visual convention to indicate life, as opposite to a flat tracing that means death.	Art. 6(1d) [16]
	**Public Interest** Simplified definition: It is the interest of a community, as opposed to the interest of a private. Rational behind the choice: The community is represented as three users that are blanked out, meaning that their identity is not relevant, as opposed to a specific user, which is black. For the hands' meaning, see above.	Art. 6(1e) [16]
	**Legitimate Interest** Simplified definition: It is a reason that justifies the controller's processing and that prevails on the data subject's rights. Rational behind the choice: The controller is represented as a business man. For the hands' meaning, see above.	Art. 6(1f) [16]

Table D.2: Icons, respective simplified definitions, and rational behind the visual choice for the icon for the class "legal bases for processing"

D.3 Processing Purposes

Icon	Description	Legal ref.
	**Processing Purposes** Simplified definition: the are the reasons why data is collected and processed. Without a purpose, the processing is unlawful. Rational behind the choice: this icon is a superclass of the individual purposes' classes and its iconography must be imagined together with the other purposes and the privacy policy's layout. The arrows symbolizes a direction (=a purpose): the personal data move towards a specific purpose, where the arrow lands.	Ch. 3 [16]
	**Statistical Purposes** Simplified definition: Personal data (e.g. age, gender, personal characteristics) of a certain user can be processed to carry out statistical studies on the population that the user represents. Rational behind the choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The bar graph is a typical figure to represent statistical data.	Rec. 162 [16]
	**Purposes of Information Security** Simplified definition: Personal data can be processed to ensure that the network can resist to events that can compromise its security. Rational behind the choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The shield is a common graphical symbol for security used on antivirus software and alike.	Rec. 49 [16]
	**Research Purposes** Simplified definition: Personal data can be collected and processed to carry out scientific research (e.g. medical research) Rational behind the choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The microscope is a typical and iconographical symbol for science and research.	Recc. 159, 160 [16]

	**Purposes of Provision of the Service** Simplified definition: Personal data can be processed to provide a service (e.g. Google Maps asks for user's location to provide directions). Rational behind the choice: The black arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The white, complementary arrow going in the opposite direction and departing from a black hand with a white cuff (=the controller's hand) symbolizes the service. The two arrows taken together signify the exchange of personal data for a certain service.	
	**Purposes of Service Enhancement** Simplified definition: Personal data can be processed to enhance the functioning of a service (e.g. the navigation on a website). Rational behind the choice: Same as in provision of the service. The additional star/spark (which resembles a plus on purpose) signifies the enhancement in other digital contexts (e.g. videogames).	
	**Marketing purposes** Simplified definition: Personal data can be processed to send advertising material. Rational behind the choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The bullhorn stands for a tool that amplifies the advertisement slogans, represented by the speech balloon.	[18]

Table D.3: Icons, respective simplified definitions, and rational behind the visual choice for the icons of the class "Processing purposes"

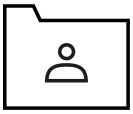
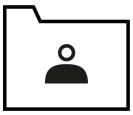
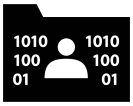
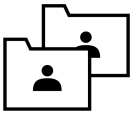
Appendix E

The Final DaPIS

The following tables provide the icons for each data protection concept, alongside its (simplified) definition that was provided to the participants of the workshops and of the user studies and the reasons behind the iconographical choice. The last column gives indication about the legal reference from which the concept was extracted.

Icon	Description	Legal reference
	**Data subject** Definition: an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person; Rational behind the iconographical choice: The data subject can be generally identified with the user of a certain service, e.g. social media. Thus the silhouette of a user widely adopted on many applications to locate one's own profile's information can easily represent the data subject. To reinforce this idea, 'you' was added to establish a direct connection with the reader.	Art. 4.1 GDPR
	**Controller** Definition: it is the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data. Rational behind the iconographical choice: The controller has been one among the most debated icons and has been redesigned multiple times, based on users' feedback. Whereas in the beginning the controller was represented as a tall building (i.e. a company), and then as a man inside the building, for the sake of usability the last design iteration has given as result a business man. Indeed, it needs to combined with other elements to signify more complex notions (*see* contract, vital interest)	Art. 4.7 GDPR
	**Supervisory authority** Definition: it is an independent public authority which is established by a Member State to be responsible for monitoring the application of this Regulation, in order to protect the fundamental rights and freedoms of natural persons in relation to processing and to facilitate the free flow of personal data within the Union. Rational behind the iconographical choice: Representing this concept as a judge would have been misleading and inherently wrong, thus the authority is sitting on an armchair at a massive desk. The colour is white to distinguish it from the user and also because the user testing revealed that a black user in this context was interpreted in a negative sense (e.g. a villain)	Art. 4.21 GDPR

TABLE 1: AGENTS AND ROLES

Icon	Description	Legal reference
	**Processing operation** Definition: processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction. Rational behind the iconographical choice: This icon shows the starting point of a process: the personal data, on the left, undergoes a process represented by the arrow with gears. The result is one of the icons illustrated below in this table.	Art 4.2 GDPR
	**Anonymization** Definition: it is the process that strips personal data of sufficient elements such that the data subject can no longer be identified. Rational behind the iconographical choice: The personal data, after processing, become anonymous. Whereas the icon for personal data shows a black user's silhouette to identify a specific user, here the silhouette becomes blank to signify that the data was striped of identifiable elements.	Rec. 26 GDPR; WP29, *Opinion 05/2014 on Anonymisation techniques,* 2014
	**Pseudonymization** Definition: it is the process through which personal data can no longer be attributed to a specific data subject without the use of additional information (provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person). Rational behind the iconographical choice: This icon is based on the symbol for anonymous data. The silhouette is not completely blank (='pseudo') because it is possible to re-identify the data subject, through retrieval of the information.	Art 4.5 GDPR
	**Encryption** Definition: Encryption is a mathematical function using a secret value - the key - which encodes data so that only users with access to that key can read the information. Rational behind the iconographical choice: The binary code exemplifies a digital transformation of the personal data into encrypted data that can not be read by anybody.	Art 34.3(a) GDPR; ICO Encryption (https://ico.org.uk/for-organisations/guide-to-data-protection/encryption/), 2016
	**Automated decision-making*** Definition: it is the ability to make decisions by technological means without human involvement. Solely automated decision-making, including profiling, produces legal effects or significantly affects the data subject. Rational behind the iconographical choice: The three options stand for possible decisions that can be taken. The absence of a human, replaced by a computer, represents the fact that the decisions are taken automatically.	Art 22.1 GDPR; WP29, *Guidelines on Automated Decision-making and Profiling for the Purposes of Regulation 2016/679 17/EN,* 2018
	**Copying** Definition: It is the act of making a copy of a certain data. Rational behind the iconographical choice: Two personal data folders are reproduced in an exact way.	Art. 15.3, 15.4 GDPR

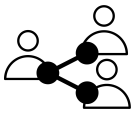
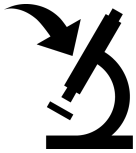
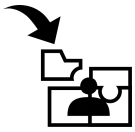
	**Sharing of personal data with third parties** Definition: It is the action of sharing personal data with a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data. Rational behind the iconographical choice: the icon combines the ubiquitous sign for sharing, present especially on social media, with three users that are blanked out to convey the meaning that often the identity of the third parties is unknown	Art. 4.10 GDPR; in many consent forms
	**Transfer of personal data to third countries** Definition: It is the transfer of personal data which are undergoing processing or are intended for processing to a third country. Rational behind the iconographical choice: The stars in circle are the emblematic symbol of the EU, whereas the arrow signifies the movement of the personal data outside of the European borders.	Art. 44 GDPR
	**Storage of personal data in the EU** Definition: It is the opposite of the transfer of personal data to a third country. Rational behind the iconographical choice: The stars in circle are the emblematic symbol of the EU, whit a personal data folder places in the middle, i.e. inside of the European borders.	Recurrent concept in privacy policies

TABLE 2: PROCESSING OPERATIONS

Icon	Description	Legal reference
	**Processing Purposes** Simplified definition: these are the reasons why data is collected and processed. Without a purpose, the processing is unlawful. Rational behind the iconographical choice: this icon is a superclass of the individual purposes' classes and its iconography must be imagined together with the other purposes and the privacy policy's layout. The arrows symbolize a direction (i.e. a purpose): personal data move towards a specific purpose, where the arrow lands (see following icons).	Art. 6.1. GDPR
	**Research Purposes** Simplified definition: Personal data can be collected and processed to carry out scientific research (e.g. medical research) Rational behind the iconographical choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The microscope is a typical and iconographical symbol for science and research.	Recc. 159, 160 GDPR
	**Purposes of Information Security** Simplified definition: Personal data can be processed to ensure that the network can resist to events that can compromise its security. Rational behind the iconographical choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The shield is a common graphical symbol for security used on antivirus software and alike.	Rec. 49 GDPR

	**Statistical Purposes** Simplified definition: Personal data (e.g. age, gender, personal characteristics) of a certain user can be processed to carry out statistical studies on the population that the user represents. Rational behind the iconographical choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The bar graph is a typical figure to represent statistical data.	Rec. 162 GDPR
	**Marketing purposes** Simplified definition: Personal data can be processed to send advertising material. Rational behind the iconographical choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The euro symbol inside the speech balloon recalls advertisement.	Fedma, *Code of Practice for the Use of Personal Data*, 1998
	**Profiling** Definition: it is any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviors, location or movements. Rational behind the iconographical choice: The arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). Many pieces of a puzzle are combined together to compose the a personal data folder i.e. the profile of a data subject.	Art. 4.4 GDPR
	**Purposes of Provision of the Service - alternative A*** Simplified definition: Personal data can be processed to provide a service (e.g. Google Maps asks for user's location to provide directions). Rational behind the iconographical choice: The black arrow metaphorically stands for the point of arrival of the processing purpose (see superclass' icon). The white, complementary arrow going in the opposite direction and departing from a black hand with a white cuff (=the controller's hand, as opposed to the data subject's white hand in the rights icons) symbolizes the service. The two arrows considered together signify the exchange of personal data for a certain service.	Recurrent purpose in privacy policies
	**Purposes of Provision of the Service - alternative B*** Simplified definition: Personal data can be processed to provide a service (e.g. Google Maps asks for user's location to provide directions). Rational behind the iconographical choice: Since the alternative A received much criticism in the second user study because not representative of the concept, this alternative provides a more literal and semantically specified visualization: the user provides personal data in exchange of a service, exemplified by a webpage. The two arrows recall the exchange.	Recurrent purpose in privacy policies
	**Purposes of Service Enhancement - alternative A*** Simplified definition: Personal data can be processed to enhance the functioning of a service (e.g. the navigation on a website). Rational behind the iconographical choice: Same as in provision of the service (alt. A). The additional sparkling signifies enhancement in other digital contexts (e.g. videogames).	Recurrent purpose in privacy policies

	**Purposes of Service Enhancement - alternative B*** Simplified definition: Personal data can be processed to enhance the functioning of a service (e.g. the navigation on a website). Rational behind the iconographical choice: Same as in provision of the service (alt. B). The additional sparkling star signifies enhancement in other digital contexts (e.g. videogames).	Recurrent purpose in privacy policies

TABLE 3: PURPOSES OF PROCESSING

Icon	Description	Legal reference
	**Legal Basis** Simplified definition: It is the reason why data is processed and must be provided according to the law for the processing to be lawful. Rational behind the iconographical choice: the capital symbolizes the basis that bears the law, represented by a gavel.	Art. 6 GDPR
	**Consent** Simplified definition: It is the expression of the data subject's willingness to have her data processed. Rational behind the iconographical choice: the cross ("x") represents a disagreement, whereas the tick ("v") represents an agreement. The slash conveys the idea of possibilities of an equal choice between the two, whilst the cross is expressly positioned before the tick to stress the option of refusing one own's consent in line with the GDPR, which is not usually the case.	Art. 4.11 GDPR
	**Contract** Simplified definition: It is an agreement that establishes a legal relationship between two parties. Rational behind the iconographical choice: The contract is typically represented as a written agreement that must be signed (hence the "x") by two parties: the data subject and the controller .	Art. 6.1(b) GDPR
	**Legal Obligation - alternative A*** Simplified definition: It is the duty to carry out what the laws says. Rational behind the iconographical choice: The law is represented as an official act, which is here signified by a stamped document with a stamp and a pointing hand imposed from above to recall the obligation. A previous icon design without the hand was deemed to similar to a certification and not enough legally defined.	Art. 6.1(c) GDPR
	**Legal Obligation - alternative B*** Simplified definition: It is the duty to carry out what the laws says. Rational behind the iconographical choice: Same as alternative A, but with a simplified layout for usability reasons. If the stamp is necessary to provide enough semantically defined details is still an open question.	Art. 6.1(c) GDPR

	**Vital Interest** Simplified definition: A matter of life and death. Rational behind the iconographical choice: The two joint hands stand for protection or care, with metaphorical extension for someone's interest. It is an iconographical choice in common with all the other interests, whose meaning is specified by the object between them. The electrocardiogram in the heart is an established visual convention to indicate life, as opposite to a flat tracing that means death.	Art. 6.1(d) GDPR
	**Public Interest** Simplified definition: It is the interest of a community, as opposed to the interest of a private. Rational behind the iconographical choice: The community is represented as three users that are blanked out, meaning that their identity is not relevant, as opposed to a specific user, which is black. For the hands' meaning, see above.	Art. 6.1(e) GDPR
	**Legitimate Interest** Simplified definition: It is a reason that justifies the controller's processing and that prevails on the data subject's rights. Rational behind the iconographical choice: The controller is represented as a business man. For the hands' meaning, see above.	Art. 6.1(f) GDPR

TABLE 4: LEGAL BASES FOR PROCESSING

Icon	Description	Legal Reference
	**Data subject's rights** Simplified definition: these are the rights of those that have provided their personal data (i.e. the data subjects) to an organisation (i.e. a company, e.g. Google, or an institution, e.g. tax office). Rational behind the iconographical choice: the hand means "holding", with metaphorical extension "being in control" or "have the power over" to indicate the possibility granted by a right to its holder. It is an iconographical choice in common with all the other data subjects' rights, whose meaning is specified by the element above it. The diamond symbolises a value, stressing that rights are precious and confer a certain power to the data subject.	Ch. 3 GDPR
	**Right to be informed** Simplified definition: data subjects have the right to know who does what with their data, how, and why. Rational behind the iconographical choice: the "i" is an internationally recognized symbol for information. For the hand, see above.	Art. 12, 13,14 GDPR

	**Right to rectification** Simplified definition: data subjects have the right to ask the data about them to be corrected or updated in inaccurate and complemented if incomplete. Rational behind the iconographical choice: the pencil is a widespread symbol for editing in software applications: it erases incorrect data and rewrites them correctly. For the hand, see above.	Art. 16 GDPR
	**Right to erasure ('Right to be forgotten')** Simplified definition: In some cases, data subjects have the possibility to ask for their data to be erased. Rational behind the iconographical choice: the bin is a popular symbol for erasure in software applications. For the hand, see above.	Art. 17 GDPR
	**Right of access** Simplified definition: data subjects have the right to know if someone owns data about them and to obtain a copy. Rational behind the iconographical choice: the folder with a user's silhouette is symbol of personal data, whilst the magnifying lens on the user indicates scrutiny of a specific person's data. For the hand, see above. It is the redesign of a literal representation of this concept.	Art. 15 GDPR
	**Right to withdraw your consent** Simplified definition: data subjects have the right to revoke the consent on their data processing that they had previously given Rational behind the iconographical choice: the cross ("x") and the tick ("v") derive from the representation of consent (*see* consent as legal basis). The arrow stands for the transformation from approval/acceptance to disagreement/disapproval. For the hand, see above.	Art. 13.2(c) GDPR
	**Right to data portability** Simplified definition: data subjects have the right to receive a copy of their data collected by a service provider A and transfer it to a service provider B. They can also ask for direct transfer from A to B. Rational behind the iconographical choice: the data folder, representing the personal data, takes the shape of a bag with handles to carry it around. For the hand, see above.	Art. 20 GDPR
	**Right to restriction of processing*** Simplified definition: data subjects have the right to ask their data to be processed exclusively for certain purposes. Rational behind the iconographical choice: the gears represent processing activities, as in other icons. The processing goes on, but only partially: half of the gears continue to work and thus are black, whereas the other half is deactivated. For the hand, see above.	Art. 18 GDPR
	**Right to object to processing*** Simplified definition: data subjects have the right to ask a service to stop processing their data for a certain purpose. Rational behind the iconographical choice: the gears represent processing activities. If broken, gears stop working. For the hand, see above.	Art. 21 GDPR

	**Right to lodge a complaint to a supervisory authority** Simplified definition: data subjects have the right to file a complaint with a supervisory authority for data protection, if they think that their data is processed unlawfully. Rational behind the iconographical choice: see icon for supervisory authority. For the hand, see above.	Art. 13.2(d) GDPR

TABLE 5: RIGHTS OF THE DATA SUBJECT

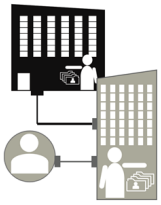
* For the icons with this symbol, research about possible alternatives that can better convey the meaning should be carried out, because in the user studies no definitive consensus was reached.

Appendix F

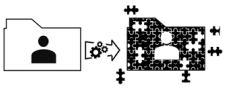
Excerpts from the First User Study

TASK 1

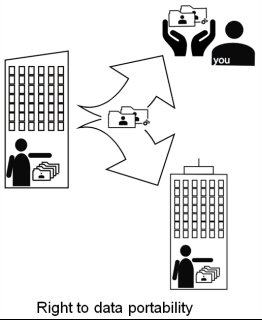
Transfer of personal data to third countries



Third party



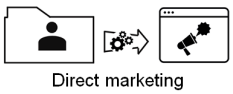
Profiling



Right to data portability

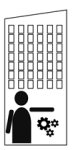
Automated decision-making

Data subject



Direct marketing

Derived personal data



Data processor

When your data is divided from you. Now you can be identified only with extra data.

This is the information about you that is transformed in some way.

This is who can receive your data, but is different from the data controller or data processor.

This is the person to whom personal data refer.

When your data is written in such a way that only authorized people can understand it.

When computers make decisions about you based on your data.

When your data is sent outside of the European Union.

This is who monitors if the law on data protection is applied and protects your rights.

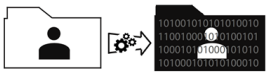
This is who collects your data and decides how your data can be processed.

You have the right to know if someone has information about you. You also have the right to receive a copy of that information.

This is the information about you that is derived from other data or by programs.

Supervisory authority

This is the information about you that is collected from you.

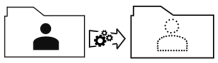


Encryption

When your data is divided from you. Now you cannot be identified.

Original personal data

When your interests, your behaviour, or your skills are predicted based on your data.



Anonymization

This is who carries out operations on your data, on account of the data controller.

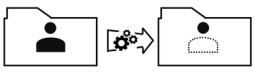
Processed personal data

When your data is used to send you advertising.

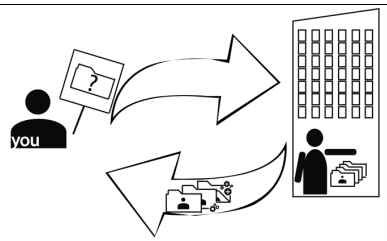
Copying

You have the right to receive data collected about you in a format that supports re-use. You also have the right to ask the transfer of that data to another data controller.

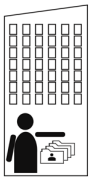
When a copy of your data is made.



Pseudonymization



Right of access



Data controller

Appendix G

Excerpts from the Second User Study

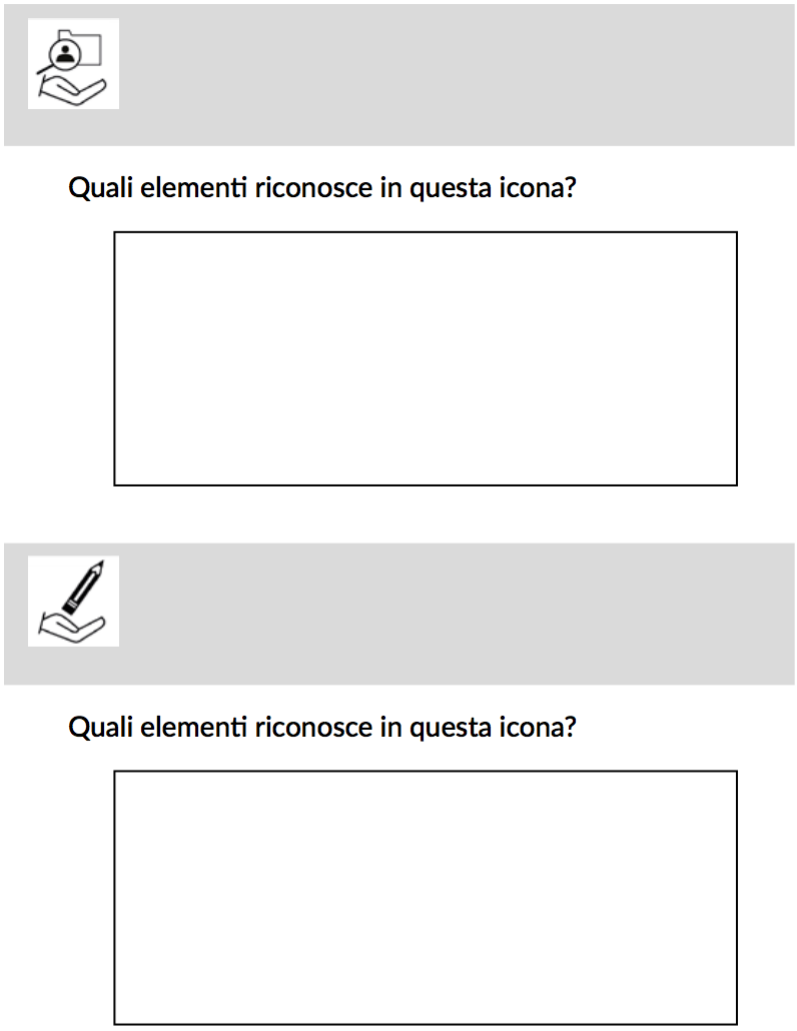


Figure G.1: An example of the legibility task (task 1) in the second user study (see Sec. 3.3). The English translation would be: "Which elements do you recognize in this icon?"

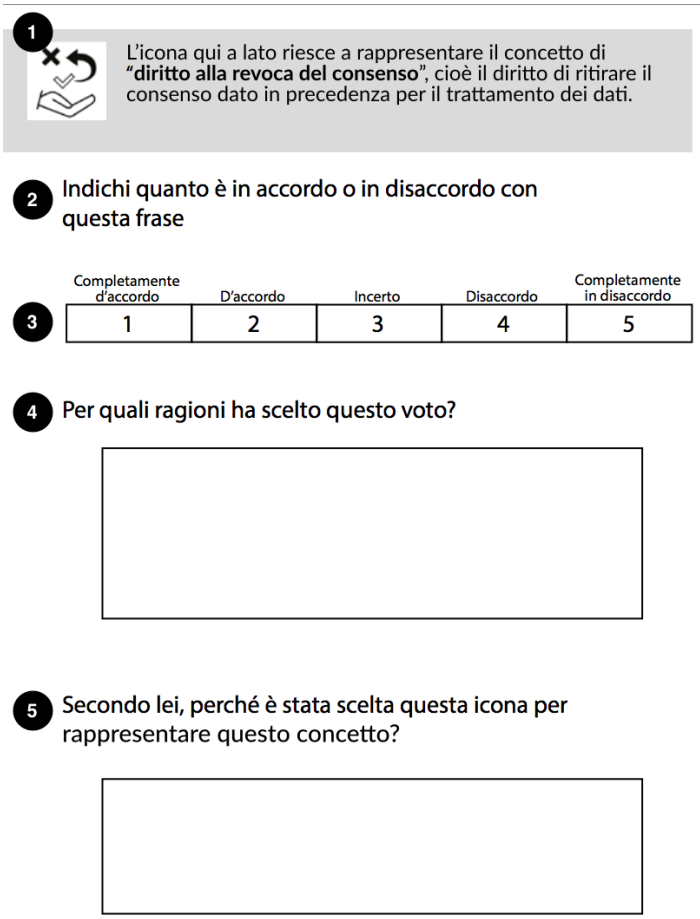


Figure G.2: An example from the second user study (see Sec. 3.3) of the task about ease of understanding (task 2) and the task about the alignment between designers' intentions and users' mental models (task 3). (1) displays the icon and provides the corresponding label and definition, e.g. English translation:"The icon on the side is able to represent the concept of 'right to withdraw your consent', namely the right to revoke the consent on data processing that was previously given."; (2) English translation: "Specify the extent to which you agree or disagree with this statement"; (3) 5 points Likert scale; (4) English translation: "Why have you chosen this mark?"; (5) English translation: "According to you, why was this icon chosen to represent this concept?"

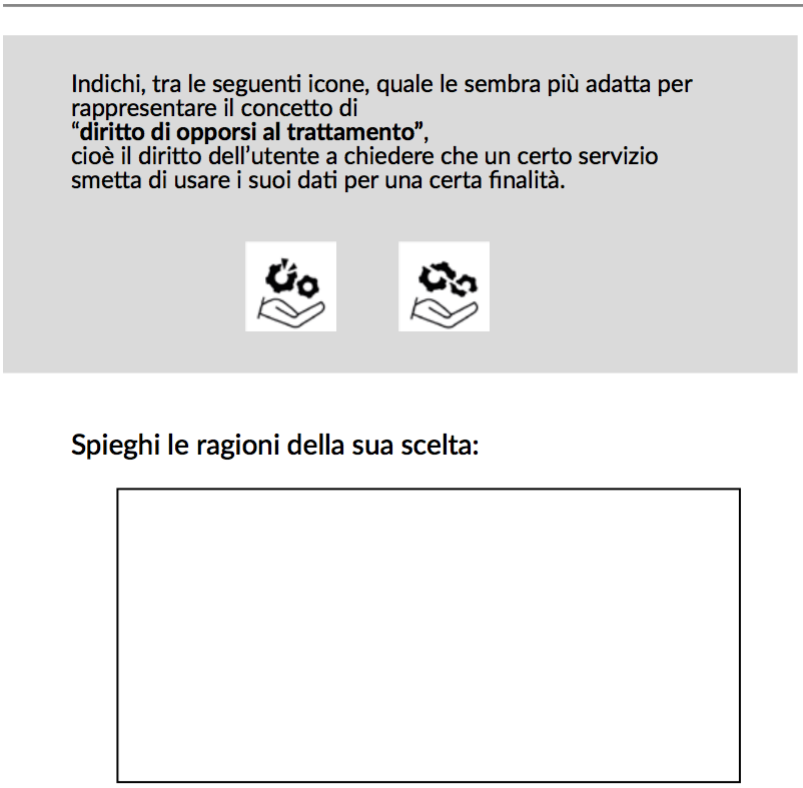
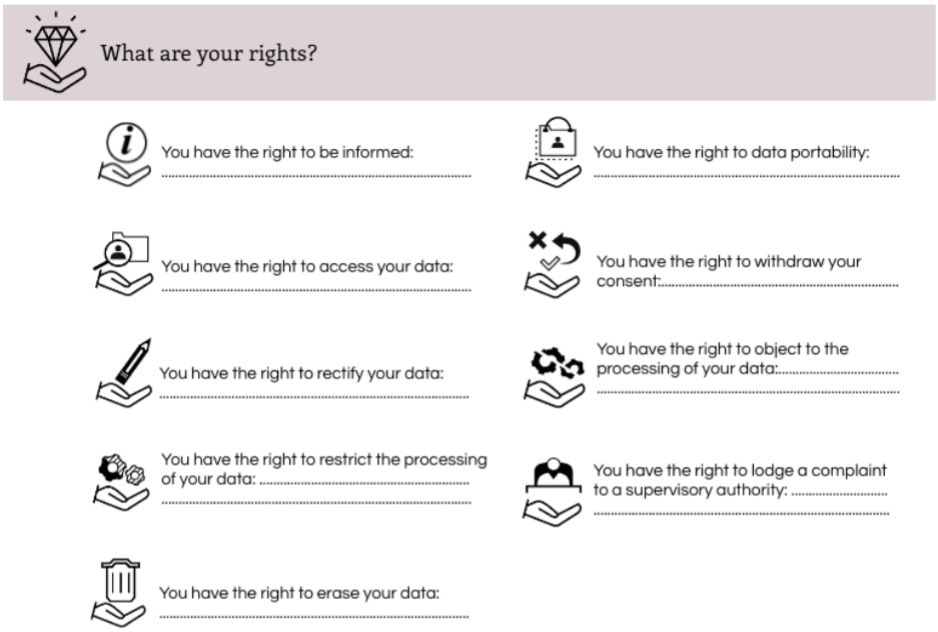


Figure G.3: Alternative choice between two icons representing the concept of "right to object to processing" (task 4), with space to provide reasons for the choice

Appendix H

Excerpts from the Third User Study

This is the section of the privacy policy where you would find the next icons you will see:



Now we will ask you to rate some icons that you could find in this section of the privacy policy.

Question 7: "Rights of the data subjects"

Look at the icon below:

Figure H.1: An excerpt from the third user study, that was carried out online. The questions evaluate the fitness of correspondance between icon and definition, the reasons and the possibility to align users' and designers' mental models

Bibliography

[1] http://www.cirsfid.unibo.it/en

[2] Seminario law & design for privacy, http://www.cirsfid.unibo.it/eventi/seminario-law-design-for-privacy-sala-kelsen

[3] Seminario law & design for privacy 2, http://www.cirsfid.unibo.it/eventi/seminario-law-design-for-privacy-2-sala-kelsen

[4] Abrams, M.: The origins of personal data and its implications for governance (2014)

[5] Article 29 Data Protection Working Party: Opinion 05/2014 on anonymisation techniques (April 2014)

[6] Article 29 Data Protection Working Party: Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679 17/en wp251rev.01 (February 2018)

[7] Article 29 Data Protection Working Party: Guidelines on transparency under regulation 2016/679, 17/EN WP260 rev.01 (April 2018)

[8] Article 29 DPWP: Opinion 15/2011 on the definition of consent wp187. Online (2011)

[9] Ben-Shahar, O., Schneider, C.E.: More than you wanted to know: The Failure of Mandated Disclosure. Princeton University Press (2014)

[10] Berger-Walliser, G., Barton, T.D., Haapio, H.: From visualization to legal design: A collaborative and creative process. American Business Law Journal 54(2), pp. 347–392 (2017)

[11] Boehme-Nessler, V.: Pictorial law: modern law and the power of pictures. Springer Science & Business Media (2010)

[12] Busch, C.: Implementing personalized law: Personalized disclosures in consumer law and privacy law. University of Chicago Law Review (2018)

[13] Caire, P., Genon, N., Heymans, P., Moody, D.L.: Visual notation design 2.0: Towards user comprehensible requirements engineering notations. In: Requirements Engineering Conference (RE), 2013 21st IEEE International. pp. 115–124. IEEE (2013)

[14] Dewar, R.: Visual information for everyday use: Design and research perspectives, chap. Design and evaluation of public information symbols, pp. 285–303. Taylor and Francis London (1999)

[15] Esayas, S., Mahler, T., McGillivray, K.: Is a picture worth a thousand terms? visualising contract terms and data protection requirements for cloud computing users. In: International Conference on Web Engineering. pp. 39–56. Springer (2016)

[16] European Parliament, Council of European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). O.J. L 119, 4.5.2016, p. 1–88 (2016)

[17] European Parliament. Committee on Civil Liberties, Justice and Home Affairs: Draft report on the proposal for a regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation) [com(2012)0011 - c7-0025/2012 - 2012/0011 (cod)] (21 November 2013)

[18] FEDMA: Code of practice for the use of personal data (1998)

[19] Fischer-H´ubner, S., W´astlund, E., Zwingelberg, H.: Ui prototypes: Policy administration and presentation version 2, deliverable d4.3.2 of the ec fp7 project primelife. Tech. rep., PrimeLife Consortium (2010)

[20] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with dolce. Knowledge engineering and knowledge management: Ontologies and the semantic Web pp. 223–233 (2002)

[21] Gangemi, A., Presutti, V.: Ontology design patterns. In: Handbook on ontologies, pp. 221–243. Springer (2009)

[22] Graf, C., Hochleitner, C., Wolkerstorfer, P., Angulo, J., Fischer-Hübner, S., Wästlund, E.: Final hci research project. Tech. rep., PrimeLife Consortium (2011)

[23] Haapio, H., Hagan, M., Palmirani, M., Rossi, A.: Legal design patterns for privacy. In: Schweighofer, E., et al. (eds.) Data Protection / LegalTech Proceedings of the 21st International Legal Informatics Symposium IRIS 2018. pp. pp. 445–450. Editions Weblaw (2018)

[24] Haapio, H., Passera, S.: Contracts as interfaces: Exploring visual representation patterns in contract design. In: M. J. Katz, R.D., Bommarito, M. (eds.) Legal Informatics. UK: Cambridge University Press (Forthcoming)

[25] Hagan, M.: Legal design lab: a new generation of legal services, http://www.legaltechdesign.com

[26] Hansen, M.: Putting privacy pictograms into practice-a european perspective. GI Jahrestagung 154 (2009)

[27] Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K.G., Aberer, K.: Polisis: Automated analysis and presentation of privacy policies using deep learning. arXiv preprint arXiv:1802.02561 (2018)

[28] Hartzog, W.: Privacy's Blueprint: The Battle to Control the Design of New Technologies. Harvard University Press (2018)

[29] Hicks, J.: The icon handbook. Five Simple Steps (2011)

[30] Hintze, M.: In defense of the long privacy statement. Maryland Law Review 76(4), 1044–1085 (2017)

[31] Holtz, L.E., Nocun, K., Hansen, M.: Towards displaying privacy information with icons. In: Jan, C., Bruno, C., Simone, F.H., Ronald, L., Giovanni, R. (eds.) IFIP PrimeLife International Summer School on Privacy and Identity Management for Life. pp. 338–348. Springer (2010)

[32] Iannella, R., Finden, A.: Privacy awareness: Icons and expression for social networks. In: Proceedings of the 8th Virtual Goods Workshop and the 6th ODRL Workshop. pp. 1–15 (2010)

[33] ICO: Encryption (2016), https://ico.org.uk/for-organisations/guide-to-data-protection/encryption/

[34] Institute), E.E.T.S.: Human factors (hf); framework for the development, evaluation and selection of graphical symbols. eg 201 379 v1.1.1 (1998-12). Online at: http://www.etsi.org/ (1998)

[35] Isherwood, S.J., McDougall, S.J., Curry, M.B.: Icon identification in context: The changing role of icon characteristics with user experience. Human Factors: The Journal of the Human Factors and Ergonomics Society 49(3), 465–476 (2007)

[36] Legal Design Lab: Design workshop for eu general data protection regulation. http://www.legaltechdesign.com/design-workshop-for-eu-general-data-protection-regulation/ (July 2017)

[37] Malamed, C.: Visual language for designers: principles for creating graphics that people understand. Rockport Publishers (2009)

[38] Mangen, A., Walgermo, B.R., Brønnick, K.: Reading linear texts on paper versus computer screen: Effects on reading comprehension. International journal of educational research 58, 61–68 (2013)

[39] Mondschein, C.F.: Some iconoclastic thoughts on the effectiveness of simplified notices and icons for informing individuals as proposed in article 12(1) and (7) gdpr. European Data Protection Law Review 2(4), 507–520 (2016)

[40] Moody, D.: The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Transactions on Software Engineering 35(6), 756–779 (2009)

[41] Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L.: PrOnto: Privacy Ontology for Legal Reasoning. In: Accepted at eGovis 2018, the 7th International Conference on Electronic Government and the Information Systems Perspective (2018)

[42] Palmirani, M., Rossi, A., Martoni, M., Hagan, M.: A methodological framework to design a machine-readable privacy icon set. In: Schweighofer, E., et al. (eds.) Data Protection / LegalTech Proceedings of the 21st International Legal Informatics Symposium IRIS 2018. pp. pp. 451 – 454. Editions Weblaw, Wien (2018)

[43] Palmirani, M., Vitali, F.: Legislative XML: Principles and technical tools. Inter-American Development Bank (2012)

[44] Passera, S.: Enhancing contract usability and user experience through visualization-an experimental evaluation. In: 16th International Conference on Information Visualisation. pp. 376–382. IEEE (2012)

[45] Passera, S.: Beyond the wall of text: How information design can make contracts user-friendly. In: International Conference of Design, User Experience, and Usability. pp. 341–352. Springer (2015)

[46] Pettersson, J.S.: A brief evaluation of icons in the first reading of the european parliament on com (2012) 0011,. Privacy and Identity Management for the Future Internet in the Age of Globalisation. Privacy and Identity 2014. IFIP Advances in Information and Communication Technology 457 (2014)

[47] Reidenberg, J.R., Breaux, T., Cranor, L.F., French, B., Grannis, A., Graves, J.T., Liu, F., McDonald, A., Norton, T.B., Ramanath, R.: Disagreeable privacy policies: Mismatches between meaning and users' understanding. Berkeley Tech. LJ 30, 39 (2015)

[48] Richter, A., Courage, M.L.: Comparing electronic and paper storybooks for preschoolers: attention, engagement, and recall. Journal of Applied Developmental Psychology 48, 92–102 (2017)

[49] Rossi, A., Palmirani, M.: From words to images through legal visualization. In: AI Approaches to the Complexity of Legal Systems. Springer (Forthcoming)

[50] Schuler, D., Namioka, A.: Participatory design: Principles and practices. CRC Press (1993)

[51] Wiedenbeck, S.: The use of icons and labels in an end user application program: an empirical study of learning and retention. Behaviour & Information Technology 18(2), 68–82 (1999)

[52] Wogalter, M.S., Conzola, V.C., Smith-Jackson, T.L.: Research-based guidelines for warning design and evaluation. Applied ergonomics 33(3), 219–230 (2002)

[53] Wogalter, M.S., Silver, N.C., Leonard, S.D., Zaikina, H.: Handbook of warnings, chap. Warning symbols, pp. 159–176. Lawrence Erlbaum Associates Mahwah, NJ (2006)

[54] Wolff, J.S., Wogalter, M.S.: Comprehension of pictorial symbols: Effects of context and test method. Human factors 40(2), 173–186 (1998)