HAL open science

Value-Based Core Areas of Trustworthiness in Online Services

Danny S. Guamán, Jose Alamo, Hristina Veljanova, Stefan Reichmann, Anna Haselbacher

To cite this version:

Danny S. Guamán, Jose Alamo, Hristina Veljanova, Stefan Reichmann, Anna Haselbacher. Value-Based Core Areas of Trustworthiness in Online Services. 13th IFIP International Conference on Trust Management (IFIPTM), Jul 2019, Copenhagen, Denmark. pp.81-97, 10.1007/978-3-030-33716-2_7 . hal-03182614

HAL Id: hal-03182614

https://inria.hal.science/hal-03182614v1

Submitted on 26 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Value-based Core Areas of Trustworthiness in Online Services

Danny S. Guamán[1, 2][0000-0003-2794-3079]✉, José M. del Alamo[1][0000-0002-6513-0303], Hristina Veljanova[3], Stefan Reichmann[4] and Anna Haselbacher[3]

[1]Universidad Politécnica de Madrid, Madrid 28040, Spain
[2]Escuela Politécnica Nacional, Quito 170525, Ecuador
[3] University of Graz, Graz 8010, Austria
[4] Graz University of Technology, Graz 8010, Austria
ds.guaman@dit.upm.es, jm.delalamo@upm.es, hristina.veljanova@uni-graz.at, stefan.reichmann@tugraz.at, anna.haselbacher@uni-graz.at

Abstract. In the digital domain, users can be expected to place their trust in online services if they have a reason to believe that, in addition to the functional and quality of service aspects, their rights will be protected and their shared values respected. However, recent studies and surveys suggest that users do not actually trust in online services, one of the reasons being that technology unable to meet their values and address their concerns. To bridge this gap, this work-in-progress paper presents a set of core areas of trustworthiness for online services that have emerged from an interdisciplinary discussion involving a social, ethical, legal and technological perspective while paying due attention to the protection of European fundamental rights and values. It then analyses the manner in which each of these core areas of trustworthiness maps to well-known system properties and (post-compliance) operational requirements.

Keywords: Trustworthiness, Trust, Privacy, Data Protection, Requirements, Ethical, Sociological, Legal, Label, Assurance

1 Introduction

Online services are an inherent component of most organisation processes and individuals' daily activities. Still, besides the numerous benefits they provide, there are several concerns regarding some of their features that can undermine their trustworthiness and this, in turn, users' trust. As indicated by a Eurobarometer survey, the general public's trust in digital applications and services remains quite low. For instance, 63% of the respondents do not trust online businesses [1].On the one hand, the extensive literature on trust offers multiple perspectives, although most of them define interpersonal trust in terms of a relationship between a trustor (i.e. the subject that places trust in a target entity), and the trustee (i.e. the entity that is trusted) [2]. Accordingly, trust forms the basis for allowing a trustee to perform a particular action important to the trustor, regardless of the ability to monitor or control the trustee.

Moving to the digital realm, wherein often there is no personal trustee, trust requires an objective assessment of the system trustworthiness in order to assure that it will perform as expected by the trustor [3].

On the other hand, the literature conceives trustworthiness as a multidimensional construct, as users can expect an online service to perform a diverse set of actions [4]–[6]. These actions have been approached primarily through mature system properties, such as security and dependability [4], focusing on those aspects that protect online services from (malicious) users, but not on those aspects that protect users from (malicious) online services [7]. Trends in users' attitudes suggest that trustworthy services need also to carry out actions to safeguard fundamental rights, such as privacy and autonomy, and to avoid a growing lack of user trust. As suggested by a Eurobarometer survey, the general public's trust in digital applications and online services remains quite low as 72% of Internet users are worried about being asked for a lot of personal data online [1]. Therefore, users can be expected to place their trust in those services if they have a reason to believe that their rights will be protected, and their shared values will be respected.

A sufficient understanding of the concepts of trust and trustworthiness thus needs to be interdisciplinary and include inputs from ethics, law and sociology, addressing concerns regarding users' fundamental rights and values by defining a set of core areas of trustworthiness. Subsequently, these core areas can be (partially) translated into operational requirements to be considered by online services that, in addition to supporting the reliability and dependability of online services, also contribute to engendering trust. The user's trust is not based solely on concrete technical practices for trustworthiness [7] as the core areas of trustworthiness, stemmed from Social Sciences and Humanities (SSH) realm, cannot be simplified and fully achieved only by technical systems, as they themselves are only subsystems of more complex socio-technical systems. This is why novel core areas of trustworthiness emerging from an interdisciplinary perspective can complement and extend the understanding of building, assessing, and providing trustworthy online services.

Towards this end, this paper presents some results of the H2020 TRUESSEC.eu project (https://truessec.eu). Taking into account that to this point labels have generally focused either on security or privacy, TRUESSEC.eu envisions a lightweight trustworthiness labelling scheme which overcomes the limitations of current labels by providing a label that not only contains security and privacy aspects but also goes beyond them. By having a strong focus on European values and fundamental rights, the TRUESSEC.eu label stipulates a set of requirements that make an ICT product and service trustworthy thus directly addressing the issue of how to enhance users' trust in ICT. In this direction, this work-in-progress paper presents a multidisciplinary discussion on identifying a set of core areas of trustworthiness and further analysing how this set could be translated into ICT system properties and detailed operational requirements.

The remainder of this paper is structured as follows. Section 2 introduces the core areas of trustworthiness and how they are approached from the perspective of three SSH disciplines, namely, sociology, law and ethics. Subsequently, Section 3 presents the approach followed to translate the core areas of trustworthiness into operational

requirements. In Section 4, we outline the related works that is relevant for our proposal. Finally, Section 5 provides the conclusion of this paper and outlines future work.

2 Core Areas of Trustworthiness

This section starts by approaching the concept of trust and trustworthiness in the digital realm from a sociological (§2.1), ethical (§2.2) and legal (§2.3) perspective. Subsequently, based on the European values and fundamental rights and interdisciplinary work six common core areas of trustworthiness are defined, namely transparency, privacy, anti-discrimination, autonomy, respect, and protection. These core areas reflect the values that should be considered when developing and evaluating ICT products and services which ought to be trustworthy. Table 1 as well as the following subsections summarize our findings and showcase (a) how each of the three disciplines approach trust and trustworthiness against the background of European values and fundamental rights and (b) the disciplinary understanding of the six core areas. The sociological perspective provides a brief overview over the concept of trust. Additionally, based on survey data collection, the sociological input in Table 1 analyses the core areas from a macro level. The section on ethics presents those aspects that are relevant for trustworthy ICT products and services from a normative point of view. The focus of the legal perspective lies mainly on the European values (Art 2 Treaty on European Union) and the European fundamental rights as stated in the Charter of Fundamental Rights of the European Union as well as the European Convention on Human Rights. For a more detailed description of the core areas interested readers can refer to [8] and [9], [10] for details of the support studies.

2.1 Trust, Trustworthiness, and Social Interaction

Trust is irreducibly social. In many situations, users display "default trust" [11], i.e. trust based on individual assumptions regarding societal expectations pertinent to a given situation. In this way, trust acts as a proxy for cooperation, thereby reducing complexity [12]. Reduction of complexity becomes especially poignant in the wake of industrialization. Industrial and post-industrial division of labour entails, among other things, that users are increasingly forced to rely on expert knowledge which they do not understand. This has to do with a shift in worldview; pre-modern (traditional) societies believe in a universal order that foregrounds individual action. In such a worldview, there is no place for risk. On the other hand, moderns believe in individual autonomy; intentional actions have unintended consequences, which entails individual and collective risks. Trust thus gains traction in modernity: "*The uncertainties and risks modernity entails necessitate a belief in the good intentions of strangers*" [13]; however, "risk" should not, for this reason, be part of a definition of trust, as trust is a fundamental aspect of social interaction, whereas risk is part and parcel of a specifically modern ontology. The widespread use of digital products and services can be conceived in terms of reliance on experts (developers, companies, and lawmakers).

Trust in experts can, therefore, be conceived as a proxy for relying on the technologies themselves.

A lack of trust is the main reason for consumers not to use digital products and services [14]. Contrary to recent theories of e-trust [15], which are overtly behaviourist/cognitive, social theorists stress the fundamentally social nature of trust. Conceptions of e-trust suffer from a cognitive/rationalistic bias that stems from the inability of accounts that ground trust in motivations or morality to apply to artefacts (which have no motivation and hence, no morality). Some authors acknowledge that trust is predicated upon contingency [16]; "trust begins where prediction ends" [11]. Contingency implies expectations. Without expectations, action would be impossible. Social action is directed towards the actions of others. Sociologically, trust is therefore conceived as "a reciprocal orientation and interpretative assumption that is shared, has the social relationship itself as the object, and is symbolized through intentional action" [17]. Georg Simmel observed that there can be love unrequited, but no unrequited trust [18]. "Faithfulness", as Simmel put it, is the only emotion sociological in form; it stems from interacting with others and it is epistemically situated between (complete) knowledge and (complete) ignorance of the other. In any case, it is insufficient to conceive of trust in digital products and services as merely psychological or merely behavioural, because trust necessarily refers to principles of morality, of mutual interests, and to social norms that oblige users to trust and be trustworthy.

2.2 Ethics, Trustworthiness, and ICT

Like any other technology, ICT has introduced numerous benefits to the individual and society, but at the same time, it has also created new ethical concerns and challenges. These concerns and challenges mainly stem from the pervasive and ubiquitous character of ICT. Moreover, ICT is also considered to have a tendency to demote particular values [19]. In that sense, taking values as a starting point for conducting an ethical analysis helps at arriving at a better understanding of the ethical issues related to ICT and paves the way for identifying the central requirements for trustworthy ICT products and services from a normative perspective.

The primary "currency" of today's ICT society is data. This has made *privacy* one of the most pressing issues. Privacy has a normative dimension and can be understood as an individual's claim to exercise control over one's data. Ethical concerns often arise when users lack answers as to activities with their personal data.

Privacy is closely related to the concept of *autonomy* because the former creates the conditions for the exercise of the latter. One way to reinforce autonomy is through informed consent which stands for the possibility of being informed about data processing activities and having the freedom to act upon one's decisions regarding data. Cases, where informed consent lacks, are ethically problematic as they directly undermine the very essence of autonomy.

The most significant concern in the domain of *justice* arises around practices of data-based discrimination and biased-decision making. The concerns pertain to cases where decisions are made based on individual's data that may lead to unjust treatment, bias or exclusion of some users or groups from certain opportunities.

The issues of *responsibility and accountability* play an essential role as well, in particular, due to the possible consequences of ICT. Responsibility and accountability can be observed in a two-fold manner: (a) forward-looking, where responsibility is understood as a duty concerning who should do what, and (b) retrospectively, where the morality of someone's actions is inspected [20].

Security is also one of the leading concepts as it is directly related to privacy, for instance. One way to analyse security issues in an ICT context is as the security of data and systems where data are stored. However, security can also be understood in a much broader sense as freedom from harm and protection of rights and liberties.

Transparency can be considered as the key concept in the ICT discourse as it serves as a means to realising, for instance, privacy, justice, responsibility. Transparency is also even more important due to the extensive informational asymmetry between users and providers of ICT products and services, that is, the lack of clear answers to the question *who* does *what*, *how,* and *why* with individual's personal data.

2.3 Legal Perspective of Trustworthiness in ICT

When two parties decide to establish a legal relationship, its fundament ideally must be mutual trust. We have therefore mapped out the European Union's legal framework regarding ICT, also taking into account the European fundamental rights and values.

Transparency constitutes one main core area of trustworthiness, which is legally assured by information duties. The GDPR's (*General Data Protection Regulation*) severe monetary fines particularly fuel its enforcement. Just like another requirement, transparency does not constitute a stand-alone area but is rather interconnected with others, such as autonomy or anti-discrimination.

While aiming to strengthen user's trust, *privacy* plays an essential role, which is emphasized by the fact that the Right to protection of personal data (Article 8 of the EU Charter of Fundamental Rights - CFR) and respect for private and family life (Art 7 CFR) are the two most referenced CFR in secondary EU legislation regarding ICT.

Legally considered, the core area regarding *justice* means that besides ensuring rights and freedoms to individuals; one must also be provided with effective remedies to effectively enforce the rights (TITLE VI CFR: JUSTICE). From a broader legal understanding justice includes equal treatment of individuals and thus non-discrimination (TITLE III CFR: EQUALITY). Within the ICT context, *anti-discrimination* is the key term to consider, meaning humans must not implement any discriminative features or processes in the online service.

Autonomy constitutes another core area of trustworthiness, legally referring to the individual's guaranteed fundamental freedoms (TITLE II CFR: FREEDOMS; including respect for private and family life and protection of personal data as well as the freedom to conduct business). As it is likely that conflicts will arise between the preserved Freedoms and other guaranteed fundamental rights, the aim is to find a balance between them. Considering ICT, autonomy results in the user's freedom to freely make decisions, thus being respected by the online service provider.

Legally speaking, the requirement of *respect* is referred to as lawfulness and must especially consider consumers. The fact that inside the ICT EU legal framework a usually high number of secondary legal acts can be observed within the area of consumer protection, namely a considerable number of fourteen, supports this view.

From a legal perspective, *security* means protecting individuals from harm, with the utmost fundamental Right to human dignity and life (TITLE I CFR: DIGNITY). In the ICT context, this implies actively providing *protection* to users, by preventing them from harm through fulfilling safety and cybersecurity standards

Table 1. Core areas of trustworthiness (*) The statistical data stem from the Eurobarometer Reports and Summaries and were collected between 2011 and 2017. Details in [10]

Sociological perspective (*)	**Legal perspective**	**Ethical perspective**	**Core areas of trustworthiness**
- Only a minority reads privacy statements (less than a fifth) in general while about 4 out of 10 internet users read the terms and conditions of the online platform. - Over 90% want to be informed if their data ever was lost or stolen, - Users who feel well-informed are more likely to adapt their security behaviour (e.g. changing passwords).	Transparency as in information duties laid down in the GDPR, the Directive on consumer rights or the e-commerce Directive.	Transparency relates to two aspects: i) providing clear and sufficient information about the products and services, and ii) providing information to users regarding activities with their personal data.	**Transparency:** The ICT product or service is provided in line with information duties regarding personal data processing and the product/service itself.
- 72% are concerned about the data collected about them on the Internet. - More than half of internet users are uncomfortable with the use of their personal data for targeted advertising. - General concern about misuse of personal data by corporate entities and public authorities (CMPD).	Privacy as preserving Respect for private life (Art 7 CFR) and the Protection of personal data (Art 8 CFR) in the context of ICT. This includes the GDPR and Directive 2002/58/EC.	Privacy stands for the individual's claim to control the access to and the use of one's personal information. The idea behind it is that people have the claim to determine who knows what about them thus preventing unjustified interferences by others.	**Privacy:** The ICT product or service allows the user to control access to and use of their personal information and it respects the protection of personal data.
- 7 out of 10 are concerned about their personal information being used for other purposes that it was collected for. - Citizens state a negative impact of state surveillance activities on their general trust in ICT.	Lawfulness as in lawful conduct and taking preventative care in accordance with the law, especially when dealing with consumers (Art 38 CFR).	Under the concepts of responsibility and accountability fall the following aspects: i) Attribution of responsibility, ii) Accepting responsibility, and iii) Prevention	**Respect:** ICT products or services are to be provided in accordance with the legitimate expectations related to them.

Table 1. Core areas of trustworthiness, continued.

Sociological perspective	Legal perspective	Ethical perspective	Core areas of trustworthiness
- 20% have changed the default settings of their browser, social network account and so on. - A majority of respondents who use online social networks have tried to change their privacy settings from the default mode. - Two-thirds are concerned about not having complete control over the information they provide online.	Autonomy as preserving freedoms, such as Freedom of thought, conscience and religion (Art 10 CFR), Freedom of expression and information (Art 11 CFR), Freedom to conduct a business (Art 16 CFR) and the Right to (intellectual) property (Art 17 CFR).	Autonomy can be seen as relating to i) capacity for self-determination, i.e. capacity/ability to lead one's life and make decisions based on one's beliefs, values and motives, and ii) possibility (freedom) to act upon one's judgment regarding aspects that affect one's life.	**Autonomy:** The ICT product or service gives users the opportunity to make decisions and respects those decisions. The ICT product or service also respects other parties'/persons' rights and freedoms.
- Concern about targeted advertising and search engine results, which some users expect to be adapted to their needs. However, this is not a majority.	Justice as the remedies against the unjustified use of force by the state, such as the Right to a fair trial (Art 47 CFR) and the Presumption of innocence (Art 48 CFR). This meaning further entails Equality before the law (Art 20 CFR) and Anti-discrimination (Art 21 CFR).	Justice relates to aspects such as: i) anti-bias, ii) fairness, and iii) distributive justice.	**Anti-discrimination:** The ICT product or service does not include any discriminative practices and biases.
- Two thirds to a quarter of EU citizens are concerned about being a victim of cybercrime or that their online personal information is not kept secure by websites or public authorities. - In general, European citizens dislike public authorities having access to their Internet usage data (fear of surveillance).	Security as the protection from harm, such as the Right to liberty and security (Art 6 CFR) as well as the Right to the integrity of the person (Art 3 CFR) and the Right to life (Art 2 CFR).	Security is understood as freedom from (physical, psychological, economic etc.) harm and protection of one's rights, liberties.	**Protection:** ICT products and services are provided in accordance with safety and cybersecurity standards.

3 From Core Areas of Trustworthiness to Operational Requirements

The six core areas (transparency, privacy, anti-discrimination, autonomy, respect, and protection) represent high-level concepts that need to be broken down into a set of requirements in the sense that they relate to more specific, well-known system properties of online services. We acknowledge that the core areas of trustworthiness, stemmed from sociological, legal, and ethical contexts, cannot be simplified and fully achieved only by technical systems, as they themselves are only subsystems of more complex socio-technical systems. However, they, can still contribute to satisfying the aforementioned core areas to certain extents. This section, therefore, presents a translation process consisting of two stages: a mapping of the core areas of trustworthiness to meaningful system properties along with the extent to which these contribute to satisfying the core areas (§3.1) and an operationalization process of the system properties to operational requirements (§3.2).

3.1 Mapping of core areas of Trustworthiness to System Properties

A system property defines a quality or behavioural characteristic of a system that can be evaluated qualitatively or quantitatively. There are numerous system properties enabling trustworthiness already studied in the technical realm (each with a different maturity level), so the knowledge base around them can be leveraged to elicit the specific operational requirements that need to be met and assessed for trustworthy online services. For instance, the S-Cube model considers nine categories of properties or attributes [21], whereas OPTET refines them into 11 categories and 29 sub-attributes [22]. Moreover, Hansen et al. [23] have further divided the notion of privacy into more concrete system properties, and these, in turn, have been broken down them into more concrete requirements for protecting privacy [24]. Our approach does not criticise individual contributions on system properties, but we propose building on these works to build a bridge between the leading SSH requirements and their corresponding operational requirements.

By following a top-down approach proposed in a previous work [25] and backed by interdisciplinary supporting studies (i.e. sociological [10], ethical [9], legal [26], and technological [27]), the core areas of trustworthiness have been mapped into a subset of the more relevant and concrete system properties of online services. When analysing the system properties, several interrelations were observed. Each system property relates to specific Core Areas; however, each system property does not need to address all Core Areas as long as all of them are addressed by a few of the system properties to a satisfactory degree. Table 2 provides an overview of this mapping along with the extent to which a system property contributes to the core areas. A brief rationale for the mapping is elaborated in the following paragraphs; note also that, although the system properties may contribute to multiple core areas to different extents, they are presented within the core area to which they mostly contribute.

Table 2. Core areas and related system properties enabling trustworthiness (●: covers the core area to a high extent; ◐: covers the core area to a medium extent; ◔: covers the core area to a low extent; ○: does not cover the core area)

System property	Core areas of trustworthiness					
	Transparency	Privacy	Anti-discrimination	Autonomy	Respect	Protection
Transparency (Accessibility)	●	●	◔	●	◔	◔
Transparency (Processing of personal data)	●	●	○	◐	◐	○
Intervenability (Consent)	◐	●	○	●	◔	○
Intervenability (Control)	◐	●	◔	●	◐	◔
Unlinkability	○	●	●	◔	○	◔
Explainability	◔	○	●	◐	◔	○
Traceability/Auditability	●	●	●	○	●	◔
Security	◐	○	○	○	◔	●

Transparency. This core area mainly relates to *transparency* including both *accessibility* and *processing of personal data* dimensions and *auditability. Transparency, in terms of accessibility* [28], refers to the form in which information is provided to users. Thus, while information can refer to the system's functionality, usage or quality features, *accessibility* ensures that it has an impact on users' awareness. To this end, an online service must provide information that is easy to find and access as well as easily understandable by users.

Transparency, in terms of the processing of personal data, is one of the three well-known system properties defined in the privacy realm (i.e. transparency, unlinkability, and intervenability [23]). This allows informing users whenever their personal data is processed by providing information on data processing, e.g., the categories of data being collected and used, who access them, the duration for which the data is retained, and the location wherein the data is stored.

Traceability/auditability reflects the capability of a system to generate, collect, and avail the evidence (e.g., records or logs) of a processing instance or any relevant event and, in turn, enable the relative ease of auditing a system. Thus, whereas the target for the above two properties is the user, the target for this property could be a supervisory authority. Increasing transparency requires service providers to clarify how they use and process personal data, and traceability and auditability allow for reconstructing, examining, and using the sequence of events to achieve transparency as well as to demonstrate compliance.

Finally, it can be noted that *transparency,* as a precursor for specific, comprehensive, and understandable information, is a prerequisite for satisfying other core areas. For instance, *autonomy* cannot be ensured if users are unable to understand infor-

mation for decision-making. *Privacy*, in the sense of control over personal data processing, is ineffective if users do not understand information related to the processing of their personal data. Similarly, *auditability* allows for retrospective accountability (which is essential for the *Respect core area)* by providing evidence to help demonstrate compliance, for example, with. a predetermined privacy policy.

Privacy. This core area mainly relates to *transparency, intervenability (control and consent)* and *unlikability*. *Intervenability,* in the sense of *control,* allows providing stakeholders (e.g. data subjects and supervisory authorities) with the means "*to interfere with the ongoing or planned data processing"* [23]. To this end, an online service should keep users in the loop by providing them with accessible means to access and review the accuracy and completeness of their data, as well as update and delete their personal data. Accessibility to control options is essential, i.e. users should be able to exercise them with a reasonable effort.

Intervenability (consent) allows users can give and/or withdraw their consent to the processing of their personal data. Therefore, an online service must be able to provide relevant and sufficient information promptly to enable users to make informed decisions about the use of the service and what information is processed about them. It should be noted that *transparency,* particularly in the sense of *accessibility*, is a precondition for consent, as users are free to give their consent to something they know and understand.

Unlinkability allows for greater (implicit) control over the processing of personal data. The rationale behind this mapping is that by preventing one event from being linked to another, an online service limits the potential impact on the users' privacy. Unlinkability could prevent that online services built to use unique identifiers (e.g., IP address, SIM mobile, or a Wi-fi SSID) from being associated to users and ultimately prevent undesirable profiling based on the actions or data generated by them. Unlinkability is related to *data minimization*, which states that the amount of data processed should be limited to the minimum possible while the consented purpose is achieved.

Anti-discrimination. This core area mainly relates to *unlinkability, explainability* and *auditability.* As already mentioned, *unlinkability* aims at preventing online services from linking personal data within and across domains. This is particularly important, as today's online services can use technologies that have the potential to process large amounts of (personal) data, very often from multiple domains, allowing significant levels of customization and decision making based on criteria such as religion, political affiliation, social status, incomes, and further on.

Explainability [29]–[32] allows decision factors or the decision-making process to be informed to the different stakeholders, allowing them to determine whether online services may have any bias. This property aims at explaining the factors used (or not used) by an online service to make a specific decision rather than a detailed explanation of the system's inner behaviour. For example, if an organization claims that an individual was denied a loan because their income is low, then the online service is expected to consider the individual when his income increases.

Autonomy. This core area mainly relates to *transparency* and *intervenability.* Autonomy is linked to supporting users in delivering informed consent, i.e., the system should provide the means for users to have the possibility to make an informed decision, e.g., on any processing instance regarding their personal data. In this regard, informed consent encompasses two system properties already explained above, i.e. *transparency and intervenability*. *Transparency* ensures that decision-making is supported by comprehensive, accessible, and precise information, while *intervenability* supports decision-making including the possibility of both granting and withdrawing consent.

Respect. This core area mainly relates to *traceability/auditability.* As *respect* implies accountability and liability, this cannot be covered through purely technical measures but requires organizational measures such as the definition of governance, a statement of legal compliance or an appropriate dispute resolution process. Nevertheless, *traceability/auditability* (already explained) is primarily aimed at supporting this core area by providing the evidence to demonstrate compliance with legal requirements.

Protection. This core area mainly relates to *safety*, s*ecurity* (i.e. properties of *confidentiality, integrity,* and *availability*), and *reliability. Safety* ensures the absence of risks that may cause physical injury or damage to the users' well-being, whether direct or indirect, as a result of the damage to the system or its environment [4].

Security, on the other hand, assures online service protection against malicious unauthorized access (*confidentiality*), modification (*integrity*), or use (*availability*) [4]. From these initial properties, other refined properties can be identified, e.g., *authenticity, non-repudiability,* and *control* [33].

It can be noted that the protection-related properties are relevant to the '*privacy'* core area. Privacy has a great value in the trustworthiness discourse; however, it cannot be guaranteed without the existence of a solid security infrastructure that ensures the security capabilities of online services. It can serve as a shield against any security breaches and cases of identity theft, unauthorized access of third parties, etc.

3.2 From System Properties to Operational Requirements

Once the system attributes have been identified, they can be used as a basis for carrying out an operationalization process and deriving a set of more specific operational requirements. When doing that, a challenge arouse with respect to abstaining from turning it into a compliance checklist, as already there is a law regulating this aspect, which ensures that online services act within the legal framework in this regard. This challenge was addressed by distinguishing between compliance and beyond compliance [12]. In this sense, compliance is an important aspect of building trust among citizens. However, very often, it is not sufficient. A measure that could help in filling in this gap is the adoption of a "post-compliance approach". It implies doing more than what the law requires and addressing those aspects that the law does not address in their entirety or to which it does not offer straightforward guidance. Introducing

ethics is one way to do so, as it could provide solutions to questions that the law leaves unanswered. Against this background, the operational requirements elicited stand on this post-compliance level. At the time of writing, a set of 81 operational requirements has been outlined in [34].

Fig. 1 depicts the interrelations among the elements of the operationalization process. The operational requirements define the capabilities that an online service should guarantee in order to satisfy one or more of the aforementioned system properties. They can be used as a precursor to the selection of more concrete (standard) measures or countermeasures that are known as *controls*. Finally, controls are realised by one or more specific *techniques*, which are implementable ways to meet a control.

There are subtle differences and interrelations between the guiding elements of the operationalization process. Although both *operational requirements* and *controls* specify system capabilities required (problem domain) and provided (solution domain), respectively, an *operational requirement* recognises that a capability seldom derives from a single *control* (i.e., fulfilling an *operational requirement* may require multiple *controls*). *Controls,* therefore, are more concrete measures than *operational requirements* and often detail a concrete *technique(s)* to be implemented in a particular context. Accordingly, *operational requirements* can be satisfied to different extents by implementing different *controls* and *techniques*, depending on the needs of the context (e.g. a higher risk scenario may require stronger controls). A *trustworthiness profile* represents a particular set of controls and techniques necessary to meet the operational requirements of an online service to be used in a particular context.

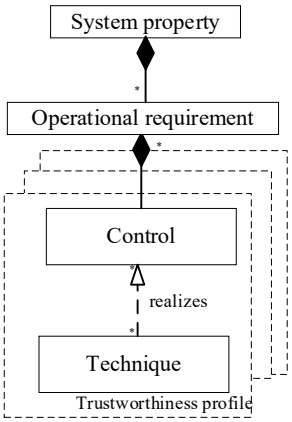

Fig. 1. Guiding elements of the operationalization process [25]

Finally, *does the state of the art support the implementations of the operational requirements related to the core areas of trustworthiness*? Security and privacy task force has provided some inputs regarding some barriers that may prevent existing technologies from satisfying the core areas of trustworthiness (in particular, members

of the IoTUK[1], AIOTI WOG03[2], and IoTMark initiatives have been surveyed[3]). The input received has been crossed with European reports and literature, and the following are highlighted:

- While it should be recognised that the state of the art already provides plenty of *controls* contained in standard catalogues and frameworks for other more mature properties (ECSO presents a syllabus with around 290 standards and certification schemes for the cybersecurity realm [35]) , *controls* related to anti-discrimination or autonomy are scarce and only recently there are some efforts and initiatives to address them. (General)
- Many system properties that enable trustworthiness (e.g. privacy, security, and transparency properties) are fragile with respect to composition, i.e. "*if a system that fulfils a certain property is embedded within or connected to another system, it is hard to assess if that property is preserved*" [36]. The fragility of these properties is becoming even more critical in the current ICT landscape, where an online service is actually a system of systems, being challenging to ensure end-to-end trustworthiness. (General)
- The increasing complexity of the supply chain also increases the difficulty of holding an entity accountable for some action, as "obligations" travels across multiple parties. For instance, it is difficult defining a closed set of technical measures that support the enforcement and auditing of the organizational or legal security obligations due to the cascading effect from the interdependent threats (coming from multiple parties). (General)
- Lack of user-centric assurance mechanisms to inform about trustworthiness-related risks. The information on the functional operation and quality attributes of online services is usually conveyed through assurance mechanism (e.g. third-party certification), and they are mainly oriented to the business market instead of the consumer market. For instance, cybersecurity certification of online services can involve the assessment of hundreds of security controls, so it is impossible for users to identify the level of security offered by online services [7]. It is necessary, therefore, to have alternative user-centric mechanisms, such as ratings, labels and sales, so that users are able to appraise and compare capabilities of different products or services without feeling overwhelmed with technical details. (General)
- Controls (and techniques) to ensure anti-discrimination are expensive, as explaining everything is expensive. This is not a purely technical barrier, but it is related to creating a system that, besides performing complex tasks, must provide an explanation that is a non-trivial engineering task. Thus, as remarked by Doshi-Velez et al. "*requiring explanation all the time may create a financial burden that disadvantages smaller companies; if the decisions are low enough risk, we may not wish to require explanation*" [37]. (Anti-discrimination)

---

[1] https://iotuk.org.uk

[2] https://aioti.eu/working-groups

[3] Appendix B of this document shows the survey carried out in the task force: https://truessec.eu/content/deliverable-52-technical-gap-analysis.

- Conflation about the scope and target audience of transparency mechanisms. Providing transparency about data protection activities has proven to be difficult, with privacy policies being the primary means of informing data subjects. Privacy policies, however, are very complex as users are not familiar with the terminology used by privacy experts, and they do not clearly understand the consequences of accepting the policy because assessing the subsequent risks is not straightforward [28]. Accordingly, only a minority of users read privacy statements. (Transparency)
- One of the essential precursors for informed decision making is to understand what we are agreeing with. Therefore, users cannot exercise their right to autonomy without transparency. However, an issue identified in some of today's online services is that non-expert users cannot connect notices about the processing of their personal data with the risks of consenting to it [38]. (Autonomy and transparency)
- Users exercise their privacy preferences based on the configuration of permissions or access control rules. For instance, in mobile phones *"[some app] should be able to access [some resource]"*. For these mechanisms to be effective, users must be able to exercise them with reasonable effort. However, current online services (e.g. those accessed through mobile apps) require users to manually set one or two hundred on/off options, requiring an amount of time that could overwhelm them. (Privacy)
- Currently, some PETs look like stand-alone solutions that are initiated by users as self-defence measures (e.g. installing web browser add-ons or using anonymity-enhanced browsers), but they are not part of the implementation of the service itself, nor are they considered during the initial design stages. This fact is also observed by ENISA (European Union Agency for Network and Information Security) which claims that "*software development tools for privacy need to be provided to enable the intuitive implementation of privacy properties*" [36]. (Privacy)
- Plenty of privacy techniques and technologies have been proposed in the last years [36], however, one of the key challenges to build a privacy-friendly system "*is the difficulty to decide when a PET may be mature enough to implement it in a system*" [39]. It could lead engineers to try to meet an operational privacy requirement using a low-quality PET. This issue can be found in the literature (e.g., [40]), where some supposedly anonymized data sets may be actually linked to the data subjects' identities. (Privacy)

4 Related Works

Several works have been developed to address different aspects of trustworthiness of online services. Most of them, however, are only approached from a technological perspective addressing only their technical features [22][7], thus failing to adopt a multidisciplinary perspective to address concerns regarding values and fundamental rights. Others, which are also in the technological realm, focus on a particular system property, primarily concentrating on security or safety ([4], [33]) and recently even on privacy ([23], [41], [42]) ,without taking into account other relevant system properties

to meet further core areas of trustworthiness. Thus, the work presented in this paper and, in particular, the criteria of trustworthiness represent a novelty owing to several reasons. First, they are the result of a comprehensive interdisciplinary work comprising ethical, legal, societal and technical aspects. Second, substantial emphasis is particularly placed on the ethical input, as the study conducted by Gibello [43] suggests that it misses existing labels very often or is overshadowed by concerns related to the quality of service or by the dominant legal aspects and the focus on compliance. Third, the criteria of trustworthiness focus on cybersecurity and privacy as the most mature domains regarding certification and labelling (e.g., around 290 cybersecurity standards and certifications schemes are available according to ECSO [35]). However, at the same time, these criteria of trustworthiness go beyond these two domains and cover various aspects of more recently established ones such as transparency, autonomy, and anti-discrimination.

Against this background, the notion that underlies the development of these criteria of trustworthiness, in conjunction with its great focus on the ethical aspects, most closely resembles the framework provided by Luciano Floridi [44]. When referring to digital services, Floridi distinguishes between hard and soft ethics. According to him, the former informs and shapes the law, whereas the latter is *"what we usually have in mind when discussing values, rights, duties, and responsibilities – or, more broadly, what is morally right or wrong and what ought or ought not to be done"*. In this sense, soft ethics operates on the post-compliance level and in this way addresses the issues and aspects that the law does not. This distinction between hard and soft ethics can be valuable in addressing the aforementioned gaps, including the extremely limited focus on ethics or too extensive dominance of legal requirements or service-related concerns in the current labels.

5 Conclusions

In this work, we have presented a set of six high-level core areas of trustworthiness (transparency, privacy, anti-discrimination, autonomy, respect, and protection) which is the result of comprehensive interdisciplinary research, comprising ethical, legal, societal and technical aspects. They complement and extend the state of the art for building and assessing trustworthy online services. Subsequently, these core areas have been translated into well-known properties and turned into (post-compliance) requirements that can be realised and assessed. It should be noted, however, that the requirements described in this paper address ICT products and services in general. This implies that once they are applied to a particular context i.e. particular product or service a certain modification or adjustment can be expected. In this respective, businesses need to recognise the added value of implementing such requirements, which may seem costly in the short run, but actually form an outstanding benefit on the market in the long run.

These contributions pave the way to move towards a lightweight and automated labelling solution for the trustworthiness of ICT products and services. In this context, our future work points in two directions: (i) enable machine-to-machine integration

based on required trustworthiness levels (defined by users through a policy configuration) and trustworthiness levels offered by the labelling subject matters and (ii) a scalable architecture for the automated assessment of elicited operational requirements applied to the mobile ecosystem.

Acknowledgement. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731711. The first author would like to extend thanks to his sponsor Escuela Politécnica Nacional

References

1. T. C. Keighley, "Special Eurobarometer 431: Data Protection Report," 2015.
2. T. Grandison and M. Sloman, "Trust in Internet Applications," *Communications*, pp. 2–16, 2000.
3. M. Taddeo, "Modelling trust in artificial agents, a first step toward the analysis of e-trust," *Minds Mach.*, vol. 20, no. 2, pp. 243–257, 2010.
4. A. Avižienis, J. C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Trans. Dependable Secur. Comput.*, vol. 1, no. 1, pp. 11–33, 2004.
5. N. G. Mohammadi, "Trustworthiness Attributes and Metrics for Engineering Trusted Internet-Based Software Systems," *Cloud Comput. Serv. Sci.*, vol. 1, pp. 165–184, 2012.
6. L. J. Hoffman, K. Lawson-Jenkins, and J. Blum, "Trust beyond security," *Commun. ACM*, vol. 49, no. 7, pp. 94–101, 2006.
7. D. Osterwalder, "Trust Through Evaluation and Certification?," *Soc. Sci. Comput. Rev.*, vol. 19, no. 1, pp. 32–46, 2011.
8. H. Stelzer, et al. "TRUESSEC D4.3: First draft Criteria Catalogue and regulatory recommendations," 2018. [Online]. Available: https://truessec.eu/content/d43-first-draft-criteria-catalogue-and-regulatory-recommendations. [Accessed: 05-Oct-2018]
9. H. Stelzer and H. Veljanova, "TRUESSEC D4.2: Support Study of Ethical Issues," 2017.
10. S. Reichmann and M. Griesbacher, "TRUESSEC D3.1: Assurance and certification of privacy and security of ICT products and services as a question of trust, acceptance and perceived risks across Europe.," 2017.
11. J. D. Lewis and A. J. Weigert, "The social dynamics of trust: Theoretical and empirical research, 1985-2012," *Soc. forces*, vol. 91, no. 1, pp. 25–31, 2012.
12. N. Luhmann, "Trust and power/two works by Niklas Luhmann; with introduction by Gianfranco Poggi." Chichester: John Wiley, 1979.
13. S. Reichmann, "TRUESSEC.eu-European Values and the Digital Single Market from a Sociological Perspec-tive," in *Proceedings of the 21st International Legal Informatics Symposium*, 2018.
14. R. W. H. Bons, R. M. Lee, and R. W. Wagenaar, *Obstacles for the development of open electronic commerce*. Erasmus University, Erasmus University Research Institute for Decision and Information Systems (EURIDIS), 1995.
15. M. Taddeo and L. Floridi, "The case for e-trust," *Ethics Inf. Technol.*, vol. 13, no. 1, pp. 1–3, 2014.
16. P. Sztompka, *Trust: A sociological theory*. Cambridge University Press, 1999.

17. J. D. Lewis and A. J. Weigert, “Social atomism, holism, and trust,” *Sociol. Q.*, vol. 26, no. 4, pp. 455–471, 1985.
18. G. Möllering, “The nature of trust: From Georg Simmel to a theory of expectation, interpretation and suspension,” *Sociology*, vol. 35, no. 2, pp. 403–420, 2001.
19. P. Brey, “Values in technology and disclosive computer ethics,” *Cambridge Handb. Inf. Comput. ethics*, pp. 41–58, 2010.
20. G. Williams, “Responsibility,” *Internet Encycl. Philos.*, 2006.
21. A. Gehlert and A. Metzger, “Quality Reference Model for SBA,” vol. 215483, 2013.
22. N. G. Mohammadi, S. Paulus, M. Bishr, A. Metzger, and H. Koennecke, “An Analysis of Software Quality Attributes and Their Contribution to Trustworthiness,” pp. 542–552, 2015.
23. M. Hansen, M. Jensen, and M. Rost, “Protection goals for privacy engineering,” *Proc. - 2015 IEEE Secur. Priv. Work. SPW 2015*, pp. 159–166, 2015.
24. R. Meis and M. Heisel, “Computer-aided identification and validation of intervenability requirements,” *Inf.*, vol. 8, no. 1, 2017.
25. Y. Martín, J. M. del Alamo, and J. C. Yelmo, “Engineering Privacy Requirements: Valuable Lessons from Another Realm,” *Evol. Secur. Priv. Requir. Eng.*, pp. 19–24, 2015.
26. V. Gibello, “TRUESSEC D4.1: Legal Analysis,” 2017. [Online]. Available: https://truessec.eu/content/deliverable-41-legal-analysis.
27. D. S. Guamán, J. Del Álamo, S. Martin, and J. C. Yelmo, “TRUESSEC D5.1: Technology situation analysis: Current practices and solutions,” 2017. [Online]. Available: https://truessec.eu/content/deliverable-51-technology-situation-analysis-current-practices-and-solutions. [Accessed: 05-Oct-2018].
28. F. Schaub, R. Balebako, and L. F. Cranor, “Designing Effective Privacy Notices and Controls,” *IEEE Internet Comput.*, vol. 21, no. 3, pp. 70–77, 2017.
29. M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier,” 2016.
30. P. Adler *et al.*, “Auditing black-box models for indirect influence,” *Knowl. Inf. Syst.*, vol. 54, no. 1, pp. 95–122, 2018.
31. P.-J. Kindermans *et al.*, “Learning how to explain neural networks: PatternNet and PatternAttribution,” pp. 1–12, 2017.
32. B. Goodman and S. Flaxman, “European Union regulations on algorithmic decision-making and a ‘right to explanation,’” pp. 1–47, 2016.
33. D. Parker, “Our excessively simplistic information security model and how to fix it,” *ISSA J.*, pp. 12–21, 2010.
34. M. Medina *et al.*, “Trust- Enhancing Label launching roadmap,” 2018. [Online]. Available: https://truessec.eu/sites/default/files/evidence/d7.5_european_trust_enhancing_label_launching_roadmap.pdf.
35. ECSO, “State-of-the-Art Syllabus Overview of existing cybersecurity standards and certification schemes,” 2017. [Online]. Available: https://www.ecs-org.eu/documents/publications/5a31129ea8e97.pdf. [Accessed: 05-Oct-2018]
36. G. Danezis *et al.*, *Privacy and Data Protection by Design - from policy to engineering*, no. December. 2015.
37. F. Doshi-Velez *et al.*, “Accountability of AI Under the Law: The Role of Explanation,” pp. 1–15, 2017.
38. A. Felt, E. Ha, S. Egelman, and A. Haney, “Android permissions: User attention, comprehension, and behavior,” *Proc. of SOUPS*, pp. 1–14, 2012.

39. M. Hansen, J.-H. Hoepman, and M. Jensen, *ENISA REPORT: Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies*, no. December. 2015.
40. A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," *Proc. - IEEE Symp. Secur. Priv.*, pp. 111–125, 2008.
41. R. Meis and M. Heisel, "Understanding the Privacy Goal Intervenability," *Trust. Priv. Secur. Digit. Bus.*, vol. 9830, pp. 79–94, 2016.
42. R. Meis, R. Wirtz, and M. Heisel, "A taxonomy of requirements for the privacy goal transparency," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9264, pp. 195–209, 2015.
43. V. Gibello, "Evaluation of existing trustworthiness seals and labels," 2018. [Online]. Available: https://truessec.eu/content/deliverable-71-evaluation-exiting-trustworthiness-seals-and-labels. [Accessed: 05-Oct-2018].
44. L. Floridi, "Soft Ethics: Its Application to the General Data Protection Regulation and Its Dual Advantage," *Philos. Technol.*, vol. 31, no. 2, pp. 163–167, 2018.