# Towards A Collective Awareness Platform for Privacy Concerns and Expectations

Giorgos Flouris[1], Theodore Patkos[1], Ioannis Chrysakis[1], Ioulia Konstantinou[2], Nikolay Nikolov[3], Panagiotis Papadakos[1], Jeremy Pitt[4], Dumitru Roman[3], Alexandru Stan[5], Chrysostomos Zeginis[1]

[1] ICS-FORTH, N. Plastira 100, P.O. Box 1385, GR-70013, Heraklion, Greece
{fgeo, patkos, hrysakis, papadako}@ics.forth.gr
[2] Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium
ioulia.konstantinou@vub.ac.be
[3] SINTEF, Forskningsveien 1a, 0373 Oslo, Norway
{nikolay.nikolov, dumitru.roman}@sintef.no
[4] Imperial College London, South Kensington Campus, London SW7 2AZ, UK
j.pitt@imperial.ac.uk
[5] IN2 Digital Innovations GmbH, Auf dem Hasenbank 23a, Lindau, Germany
as@in-two.com

**Abstract.** In an increasingly instrumented and inter-connected digital world, citizens generate vast amounts of data, much of it being valuable and a significant part of it being personal. However, controlling who can collect it, limiting what they can do with it, and determining how best to protect it, remain deeply undecided issues. This paper proposes *CAPrice*, a socio-technical solution based on collective awareness and informed consent, whereby data collection and use by digital products are driven by the expectations and needs of the consumers themselves, through a collaborative participatory process and the configuration of collective privacy norms. The proposed solution relies on a new innovation model that complements existing top-down approaches to data protection, which mainly rely on technical or legal provisions. Ultimately, the CAPrice ecosystem will strengthen the trust bond between service developers and users, encouraging innovation and empowering the individuals to promote their privacy expectations as a quantifiable, community-generated request.

**Keywords:** Collective Awareness Platforms, Collaborative Platforms, Collaborative Design, Privacy, Digital Social Innovation, Crowdsourcing, Terms of Service, Privacy Expectations.

## 1 Introduction

Privacy and anonymity in the digital world are becoming increasingly difficult to achieve. While we recognize the dramatic progress brought about by Information and Communication Technology (ICT) in almost every aspect of our everyday life, we realize that, in the process, we handed over privacy management to businesses and

corporations that are primarily driven by a profit motive, making our personal data vulnerable to exploitation in ways that are harmful to us. As society in general acknowledges that privacy preservation is essential in human relations, democracy, independence and reputation, nowadays it is openly stated that businesses often offer digital products and services that are inconsistent with consumer values[1]. Yet, for a variety of reasons, the more pronounced being limited awareness of the involved risks, we tolerate and give our consent to untrustworthy software to collect, store and process our data, having limited or no evidence as to how this sensitive information will be protected, who has access to it, or even what the intended purpose is.

The need to forge sound laws to regulate business policies for data protection is judged necessary by many stakeholders in the digital market. Europe, in particular, is pioneering such efforts by recently enacting a new, reformed data protection regulation[2] and by constantly updating its e-Privacy-related directives[3].

Legal frameworks alone are not always effective, as exemplified by the many digital products caught not only breaching national or European laws, but even violating their own privacy policies. The Norwegian Consumer Council (NCC), for example, has been revealing a multitude of such cases[4], having filed a series of complaints for apps that violated both Norwegian and European laws[5]. Similar stories about digital products that have clear discrepancies between their terms and what actually happens when consumers use them reach frequently the press, even regarding products whose main task is to offer a trusted and safer online experience[6].

At the same time, the ease with which we often give our consent to the processing of our data not only hinders the efficacy of legal regulations, but also makes it difficult for technical countermeasures to achieve a broad, society-wide impact to consumers privacy protection. The industry seems to lack incentives to adopt a more privacy-respecting attitude; the much debated Do Not Track[7] policy proposal is a characteristic example: despite its adoption by all main browsers, most web sites ignore it, having no significant reason to do otherwise [3].

Our limited understanding of the potency of digital services and the low degree of awareness on the privacy risks involved help accentuate the problem. The situation is sustained, and implicitly supported, by the current scheme with General Terms and Conditions, Terms of Service, Privacy Policy or End-User License Agreement docu-

---

[1] http://webfoundation.org/2017/03/web-turns-28-letter/

[2] Regulation 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 - GDPR), L 119/14.5.2016

[3] https://ec.europa.eu/digital-single-market/en/proposal-eprivacy-regulation

[4] https://www.forbrukerradet.no/appfail-en/

[5] http://www.forbrukerradet.no/side/norwegian-consumer-council-files-complaint-against-tinder-for-breaching-european-law, http://www.forbrukerradet.no/side/happn-shares-user-data-in-violation-of-its-own-terms/

[6] https://www.ndr.de/nachrichten/netzwelt/Nackt-im-Netz-Millionen-Nutzer-ausgespaeht,nacktimnetz100.html,

[7] http://donottrack.us/

ments (collectively referred to as *ToS* in this paper), which represent the most direct means a consumer has to understand how his/her personal data are handled. A recent study by NCC showed that just reading the ToS for apps on a typical smartphone would take more than 24 hours[8]. Considering their scope, length and complexity, it comes as no surprise that the average consumer is not investing sufficient time to study ToS before agreeing to them, thus unintentionally granting permission to apps to access and process a wealth of personal information.

With the number of privacy violations growing though, it is becoming obvious that the contrasting views between what consumers want and what firms offer can hurt the industry in the long run. As privacy concerns crystallize in public perception, small businesses will be the first to experience the consequences of consumers turning their back on privacy-suspicious products[9]. Furthermore, recent studies provide evidence that privacy policy is interlinked with innovation policy and consequently has impact for innovation and economic growth [4]. A collaborative scheme, built on trust relations, can benefit all involved parties (consumers, app developers, service providers). Within such a scheme, *data protection and privacy will not be seen as barriers to business growth, but as a competitive advantage and an innovation opportunity*. The ensuing competition will provide opportunities for start-ups to enter the market, as well as for established firms to improve their market share by appropriately adjusting the privacy-related characteristics of their digital products/services, all for the benefit of the end-user (consumer) of these products/services.

The main thesis motivating this paper is that technical solutions and solid legal regulations are necessary but not fully sufficient for accomplishing a paradigm shift towards a new data economy. In addition, we firmly believe that *data protection can be powered by the society itself*. By mobilizing consumers to become active players in digital marketplaces and by developing socio-technical tools to harness our collective power, the adoption of the technical and regulatory frameworks will become more effective and ubiquitous, and the market will act with responsiveness. As stated in [14], to protect privacy adequately, society needs awareness, but also consensus about privacy protecting measures and processes that generate social norms, with which service providers will voluntarily comply because it is profit maximizing.

This paper proposes *CAPrice, a suite of mechanisms to facilitate community interaction, enabling the explicit declaration of consumers' privacy expectations of the various digital products*. Through a combination of socio-technical methods, such as community-generated design contractualism, crowd sourcing and a knowledge commons approach to privacy policy, the outcome is a new innovation model that will allow consumers to collectively and collaboratively express their concerns, and developers to adopt more privacy-friendly practices and respond to the needs of consumers with novel products and services. To support this aim, a community is being

---

[8] http://www.forbrukerradet.no/side/the-consumer-council-and-friends-read-app-terms-for-32-hours/ http://www.forbrukerradet.no/side/250000-words-of-app-terms-and-conditions/

[9] http://www.bloomberg.com/bw/articles/2013-03-05/why-mobile-apps-privacy-policies-are-so-important
https://www.cognizant.com/whitepapers/the-business-value-of-trust-codex1951.pdf

formed that wishes to support actions towards the vision discussed in this paper (details are given below). The current paper describes the long-term vision of the CAPrice idea, as well as the current results of applying this vision in practice.

In Section 2 we describe the theoretical framework upon which or work is based, whereas Section 3 describes the complete vision associated with CAPrice. The current progress of CAPrice is described in Section 4. We conclude in Section 5. An earlier version of this idea appears in [10].

## 2 Theoretical Framework

Against the current landscape in the digital world, the protection of privacy is not just the result of applying legal and technical requirements. It seems to be also connected with the idea of the personal privacy expectations of each individual, an expectation that also depends strongly on the context in which the user is interacting (e.g., media sharing sites, social networks, apps), the social status of the user (gender, marital status, age, employment, etc.), and, of course, his/her personality and privacy sensitivity. Digital awareness has become a key issue and, consequently, citizens are another link in the chain of protecting their own privacy. In this regard, the improvement of the individual's empowerment may be the missing link in the implementation of a comprehensive and effective global strategy for the protection of privacy in the digital age. This empowerment, achieved through collaboration, crowd sourcing and collaborative open innovation management, is the main focal point of CAPrice. Before describing the software tools that will enable and facilitate this collaboration, we analyze here the main theoretical principles associated with CAPrice.

Collective Awareness Platforms for Sustainability and Social Innovation (CAPS) is a research initiative endorsed and supported by the European Commission, aiming to explore new solutions at the confluence of social networks, knowledge networks and networks of things [1]. Officially, CAPS is an initiative that "pioneers new models to create awareness of emerging sustainability challenges and of the role that each and every one of us can play to ease them through collective action"[10]. It aims at designing online platforms for creating social awareness and for allowing collective solutions to emerge through the interaction among participants, exploiting the hyper-connectivity characteristic of the digital society. Several projects associated with this initiative have been funded[11] and have already produced (or will produce soon) important results showing how collective action can support and enhance many different facets of human activity.

CAPrice leverages this idea towards *creating a community centered around privacy that will both contribute to, and benefit from, the improved, community-wide awareness on privacy*. More precisely, CAPrice complements top-down efforts by creating a community including consumers, industrial stakeholders, decision-makers and the general public. This community will engage in a *multi-directional communi-*

[10] https://ec.europa.eu/digital-single-market/en/collective-awareness

[11] https://capssi.eu/caps-projects/

*cation*, aided by software tools that will help promote awareness and cooperation among different stakeholders, towards the mutual benefit of everyone. Unlike other initiatives, in which a group of experts tries to inform other users on the privacy-related dangers of certain actions or products, we try to break this asymmetry: every person in the CAPrice community can, potentially, play the role of both the "teacher" and the "student", or both the "informer" and the "informed".

In fact, mechanisms for specifying the intended use of information have been suggested in the past (e.g., P3P [7]), but never achieved wide acceptance. The bottom-up participatory innovation paradigm of CAPS offers the means to achieve a more substantial impact, but society-wide participation and engagement are key aspects for its success. The most important difficulty that most "young" CAPS face is how to reach a critical size above which payoff for the platform (however defined) becomes positive. In order to overcome this initial threshold effect [2], a multi-dimensional strategy is needed to promote user engagement and foster social innovation.

*Design contractualism* is the idea that developers make legal, moral or ethical decisions and then (a) embed these decisions in the code itself and (b) make those decisions manifest to the user. Part (a) is achieved by encoding rules of order for appropriate behavior in computational logic as above, so the second critical innovation is to make those rules manifest to the different actors in the system. Since we are dealing with a knowledge commons, one approach is to extend an idea from the Creative Commons[12]. For example, Creative Commons supports six different licenses in three layers, each of which constitutes a norm, as it serves to coordinate expectations. However, one can imagine a user group operating under one license, but reaching a point where they would prefer to operate under a different license: the question is how to agree changes in licensing arrangements. CAPrice proposes a similar approach through the annotation of ToS documents.

Privacy protection and management, as well as information sensitivity, are inherently user-centered, thus it cannot be claimed that a given set of norms for a given app is suitable for all users and contexts. In the CAPrice model, we encourage debates for norm generation that will allow the identification of groups of people sharing common opinions. Once this happens, a separate debate per group can specify the corresponding fit-for-purpose norms.

These guidelines can be adapted to enhance the privacy policies of diverse digital services. Apps for mobile devices, for instance, specify the groups of capabilities or information (permission groups - PGs) that they need access to. Many platforms operate on a take-it-or-leave-it style, leaving a lot of aspects inadequately supported; in particular, developers are not required to explain why they need access to the requested PGs and what they plan to do with the respective data. CAPrice expands the current scheme with support for explanation generation and justification modeling: for each PG that some app requests access to, the justification can comprise a set of aspects denoting why the app developer needs this PG and a set of aspects denoting the user benefits. Our proposed solutions intend to facilitate discussion about the privacy

[12] https://creativecommons.org/

scope of apps with regards to data access, and enable users become aware and understand how their data is manipulated, as well as to express their privacy expectations.

## 3 Methodology

Our limited understanding as consumers of the capabilities of digital technologies in collecting and processing our personal data, and our inability to easily request guarantees for data protection or to prevent collection and sharing, lead us to adopt behaviors in our digital interactions that would seem unreasonable in the physical world.[13] For the time being, as the current data economy has obvious benefits for both firms and individuals, it comes as no surprise that we seem to feel comfortable with, or at least tolerate, the existing situation. Nevertheless, the protection of privacy in the digital world is becoming a vital societal problem and many stakeholders world-wide ring the bell for appropriate action. Inevitably, as privacy concerns solidify in public perception, the implications of consumers' suspicions towards digital products will eventually hurt the industry, especially the smaller players.

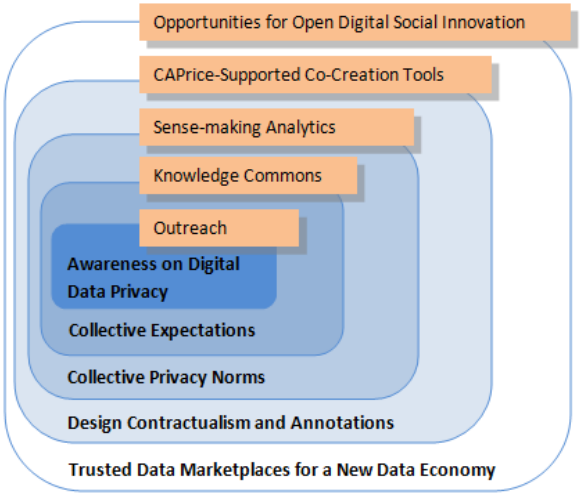


**Fig. 1.** A layered approach to Social Innovation for privacy

Unfortunately, the protection of personal data is not "a few clicks away" for the average consumer; changing application configurations, installing technological countermeasures, even reading the privacy policies and understanding the risks, is a needlessly difficult task, especially for consumers who have grown accustomed to quick-and-easy interactions with technology, or for those with a low level of technological competence. *Our intention is to offer solutions that will make privacy-friendly digital interactions for the consumer as easy to accomplish as it currently is to neglect caring about privacy protection*.

Accomplishing this goal requires a paradigm shift in the way we understand and experience technology, which cannot occur overnight, but needs a methodical approach that will steadily empower passive consumers of digital products to understand the value of their data and take control of it. In this effort, policy makers and ICT tools will offer indispensable leverage; yet, a key step for achieving effective impact

[13] The following video is instructive: https://www.youtube.com/watch?v=xYZtHIPktQg

will be to convince developers that they have many benefits to reap in the new trusted data economy, by seeing privacy and data protection as a competitive advantage, rather than a barrier. CAPrice is a solution that will enable consumers to express their privacy expectations and desires about digital products, while offering innovation opportunities for developers who are willing to listen and respond to their needs. Our approach for contributing towards this paradigm shift will happen along the socio-technical actions and innovations shown in **Fig. 1** and explained next.

### 3.1 Awareness on Digital Data Privacy

The first vital step is to approach individuals from different social and demographic groups who share similar values regarding privacy, and make them aware of the privacy risks that are hidden in the careless use of digital technology. Although digital privacy protection is included in the agenda of many organizations and institutions, in order to achieve a society-wide paradigm shift, it is important to create a global community of citizens that not only subsumes the already established groups, but expands to consumers who never before considered the protection of privacy a key concern of their daily interaction with technology.

Towards this end, we initiated an attempt to create a grassroots community of privacy-aware consumers. Securing participation in virtual online communities is not trivial, and simply bringing together individuals who share similar goals or purposes is not sufficient. To successfully foster and sustain engagement in the CAPrice virtual community, we followed the well-known 3-stages process [8], described below.

First, we need to identify and understand the needs of community members that create the intrinsic motivations for participation. As the numbers from our social channels indicate (see Section 4.1), real stories about smart toys, baby monitors, mHealth apps, even about future autonomous cars can have dramatic effect in driving awareness of diverse audiences, compared to other material.

Second, member participation must be promoted, by highlighting the value of collective actions, by creating enjoyable experiences or by encouraging content creation, among others. In fact, similar community creation attempts in other domains showed that any grassroots community is prone to lose interest, unless a vibrant, self-motivated group of users exists in its core to make it sustainable and to help establish self-definition[14]. In our case, this group is called the *CAPrice Privacy Ambassadors*, a group of individuals with specific technical and social skills, who have taken over the task of engaging citizens in this effort (see Section 4.2).

Finally, the third stage is to sustain member engagement by motivating cooperation, enabling members not only to meet specific needs, but also to co-create value for themselves and the community. We have designed a number of ICT tools to foster cooperation among ordinary consumers, researchers, privacy-enthusiasts, hackers, as well as general-purpose digital-product developers. On top of these tools, a rewards

[14] http://www.scp-centre.org/wp-content/uploads/2016/05/Final_Report_CATALYST.compressed-2.pdf

program will incentivize participation, driving user engagement and supporting reputation mechanisms to assure members that their contributions are recognized.

It is important to repeat here that this is different from top-down efforts, where awareness is achieved through a group of experts. Our aim is to complement such efforts by creating a community including consumers, everyday people, industrial stakeholders and decision-makers, who will engage in a multi-directional communication that will help promote awareness and cooperation among different stakeholders, towards the mutual benefit of everyone. Also, CAPrice differs from technical solutions to privacy (such as, e.g., the PlusPrivacy tool[15]), whose aim are to ensure that digital interactions respect the privacy preferences of the user, by imposing such preferences at a technical (software or hardware) level.

Awareness corresponds to the first, innermost layer shown in **Fig. 1**. So far, we have been quite successful in growing our community; details regarding community creation and sustainability can be found in Sections 4.1, 4.2.

### 3.2 Collective Expectations

The second step (**Fig. 1**) is to *capture consumers' expectations* regarding the privacy policies of the different digital products they ordinarily use. This is achieved by allowing consumers to explicitly state their own expectations and treating these expectations as a common-pool resource. By enabling users to specify which access permissions they find reasonable for products of a given category and which they consider too intrusive, we aim at generating shared content that will be directly exploited by many stakeholders, ranging from simple consumers and developers, to policy makers, even to social scientists that will attempt to interpret the dynamics of the community and their stance towards privacy. Towards this, we are in the process of creating a global repository of human-readable and machine-processable privacy-related content (consumers' expectations, annotated ToS, application ratings, and others) in the form of a semantic privacy wiki (see Sections 4.4 and 4.5).

With the generation of citizens' collective intelligence about privacy expectations in the form of measurable data, the accent is not only on the peer pressure that can be used to drive more privacy-respecting practices by developers, but also on the realization by consumers that expressing privacy needs and requesting solutions is not exclusively a top-down process, but can also be accomplished by each individual user uniting her or his voice to that of other members of the community.

### 3.3 Collective Privacy Norms

The basic position of CAPrice is based on the acknowledgement that, when it comes to privacy, one solution that can serve all needs is not feasible. Within the privacy protection boundaries set by legal regulations, one should listen to the plurality of opinions issued by consumers regarding the level of privacy space they wish to have, which leads to different privacy needs and expectations. Identifying these differences

[15] https://plusprivacy.com/

is of course beneficial for innovative developers who can design flexible services that adapt to the various needs. But this is even more critical for building a society that respects and supports the different trends, and where policy makers can recognize and act upon the dynamics behind the contradicting mindsets of citizens.

One of the key points of our approach is related to the *identification of collective privacy norms* (3rd layer in **Fig. 1**). In contrast to legal regulations, which apply ubiquitously, social norms are more flexible: they can be contradicting, as different attitudes may be considered "ordinary" by different people; they are more dynamic, being easily adapted to societal trends; and they have no geographical restrictions. On the other hand, law and policy making require a thorough understanding of a situation before being issued to guarantee just treatment; however, this reduces their adaptability and makes them unable to confront the astonishing speed with which ICT progresses. And there is always the risk that the country our data go to does not have the desired level of protection (although this problem is being mitigated, at least in Europe, by the introduction of European regulations such as the GDPR). We argue that collective privacy norms that exist inside the boundaries of regulations, despite being less stringent and reliable than legal regulations, can be equally powerful to control market dynamics if appropriately supported.

The aggregation and analysis of consumers' expectations into collective norms that will conceptualize the stance of citizens towards privacy products introduces certain challenges. First, the result should be measurable, to enable developers to weigh their profit-loss trade-off, but also semantically rich, to allow for meaningful interpretations of the data. Otherwise, the industry will find no incentive to adopt a different attitude towards privacy protection, as has happened many times in the past already.

In addition, the privacy principles underlying these norms need to be simple and comprehensive, in order to clearly capture the intuition of consumers and to secure society-wide coverage. We consider for our approach the experience of other initiatives that try to model users' preferences about privacy settings, relying on principles such as transparency and minimization of use.

Simplicity is key for users' comprehension, so we base our approach on a what/who/why/how/how-long scheme, i.e.: what data are being collected and processed; who is collecting or has access to the data (data controller/processor); why are the data collected or processed; how are they published; and for how long are they stored and processed. This is in close compliance with Opinion 02/2013 of the EU Article 29 Working Party on apps for smart devices[16] that provides, among others, the smallest set of recommendations that developers should follow in their privacy policies.

### 3.4 Design Contractualism and Annotation

It is well-understood that the contribution of users in isolation towards a common goal and the aggregation of their data is only half-way towards achieving the collective

[16] http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf

intelligence needed to address a societal problem. What is also imperative is the *participation of users in co-creation processes* that will empower them to collectively generate new ideas and decide collective actions. This co-creation process, which also fosters group awareness and understanding of the problems at hand, requires well-structured deliberation and discussion tools that can support goal-driven exchange of opinions, and where conclusion making is equally important to the identification of the different trends in the dialogue.


**Fig. 2** Visual Cues for Terms of Service Documents (Taken from https://disconnect.me/icons)

In CAPrice, we reuse and extend tools with proven impact, focusing on generating bottom-up solutions on privacy, and incorporating for the first time the consumer's point of view, following the ideas of design contractualism (4th layer in **Fig. 1**). In fact, our advanced notion of design contractualism goes two steps further.

Firstly, because instead of making legal or ethical decisions, designers and developers construct a legal or ethical decision space, and enable the point in that space to be selected by the users. This is the basis of algorithmic self-governance [9], whereby those affected by a set of rules (of an embedded, socio-technical, data-processing system) also participate in the selection, modification and application of those rules.

Secondly, we advance design contractualism by not just encoding this decision space in the software, but crucially also in the interface. This user-centric approach to governance modeling entails the use of visualizations to ensure that the commonly agreed privacy principles are manifested by visually identifiable and interpretable means. Using visual cues, such as the ones shown in **Fig. 2**, CAPrice intends to employ crowdsourcing techniques that will augment privacy policy documents with annotations easy for consumers to check and understand. Appropriate ICT means and personalization algorithms will hide the complexity of the task for users who decide to offer annotation services to the community, and, implicitly, to the general public.

### 3.5 Trusted Data Marketplaces for a New Data Economy

The ultimate objective of this stepwise approach (outermost layer in **Fig. 1**) is to contribute towards *a new marketplace*, where the interactions between consumers and developers are based on trust relations. By associating consumers with their privacy expectations, while providing the technological means for developers to exploit this information for undertaking novel, more privacy-friendly and respectful to consumers practices, we aim towards creating the substrate for developing new ICT tools and services. This will allow the provision of added-value services on top of the open architecture of CAPrice, and will lead to new and innovative privacy-enhancing applications. The engagement of consumers will overcome the problems faced by purely legal or purely technical solutions, creating a novel data economy for developers.

Of course, the legal and technical aspects are also necessary to ensure trust among all involved parties. Existing data marketplaces are essentially centralized systems,

where participants (data providers and consumers) have to trust a third party, the data marketplace provider/operator, with managing their data. Typically, access to data on a marketplace is governed by a set of privacy policies, often rather vague, unclear, and difficult to understand, leaving data providers with little control over their data. The guarantees that current data marketplace players receive give them little confidence that data recipients will treat the received data in the way they promise.


**Fig. 3** The Best Practice Lifecycle of CAPrice

In order to ensure trust, a data marketplace must be transparent with all stakeholders. Transparency is a fundamental principle in data protection and highlighted in the GDPR. This means that the participants in a data exchange should have knowledge about what data are shared and what operations are done over the data, and be in agreement that the data can be used for that purpose.

In CAPrice, we make steps towards offering a starting point for developers to adopt more privacy-respecting practices. In particular, we leverage emerging technological concepts, such as smart contracts and blockchains, and incorporate them into a trusted data marketplace, thereby enabling the processing of data with "by-design" trust and transparency. Smart contracts are self-executing contractual states, stored on the blockchain, and represent computer programs that can automatically execute the terms of a contract. Blockchain, as a decentralized technology, provides security, anonymity and data integrity. An example of a reference architecture for trusted data marketplaces was proposed in [11] where more details are provided on how such emerging technologies can be combined to achieve more trust and transparency in data sharing.

A trusted data marketplace caters to the interests of both application providers and their end-users. Application providers will have the opportunity to develop applications which technically guarantee their end-users' privacy, thus making them more attractive and competitive. End-users, on the other hand, will benefit from the fact that any system based on the trusted marketplace will provide transparency and unbreakable assurances that the promises of the data consumer will be kept.

### 3.6 The Best Practice Lifecycle of CAPrice

To summarize, the CAPrice Best Practice Lifecycle (**Fig. 3**) aims at maximum impact through three conceptual phases. The first phase is *awareness*: only through awareness can people understand the problem and start considering solutions. The second step is *action*: in the context of CAPrice, action consists in participating in the collaborative process of annotating ToS documents, stating privacy concerns, creating and

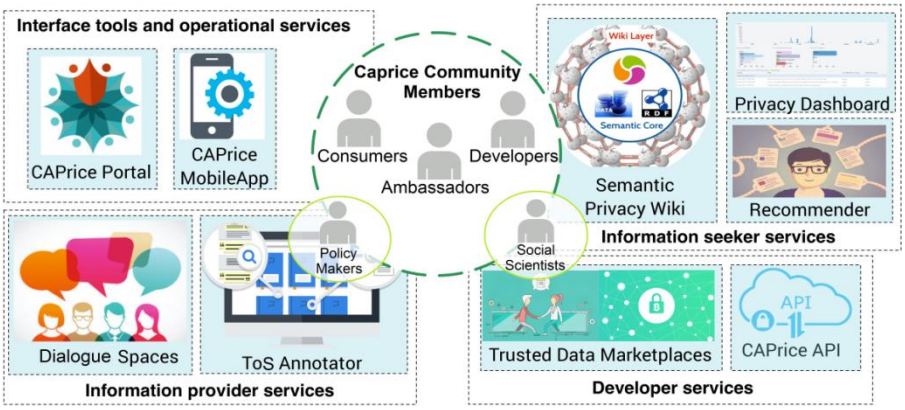


**Fig. 4.** The CAPrice Ecosystem

configuring collective privacy norms, and participating in the co-creation process. The third step is the exploitation of the acquired knowledge through *crowd sourced activities*. In this respect, a series of tools will allow the users to better implement the second step (action), and also other relevant stakeholders (policy makers, developers, legislators) understand better the needs of the public in order to contribute towards making digital products and services more transparent and privacy friendly.

## 4 CAPrice: Architecture and State of Development

The high-level overview of the CAPrice ecosystem is depicted in **Fig. 4**. According to the purpose of use, the members of the CAPrice community will be offered different groups of services, from user interfaces and services for information seekers to services for developers and information providers. These are briefly described below.

Harnessing the power of crowdsourcing tools and methodologies to collect, organize, annotate and simplify this knowledge can achieve immediate results and produce valuable content. At the heart of the CAPrice ecosystem lies the *CAPrice Semantic Privacy Wiki*, an open repository containing, among others, privacy-related information regarding digital products. The repository combines the benefits of semantic technologies with the collaborative editing capabilities associated with wikis, offering a set of functionalities that go beyond simple wiki-style catalogue for ToS: it enables the user to express privacy preferences about each product or category of products, it permits developers to explain their policies and automatically access the underlying data, it offers a public place for experts to post findings about products, and others.

The Semantic Privacy Wiki will be populated with information from the *ToS annotator* (see Section 4.4) and the *Dialogue Spaces* which will facilitate structure discussions and the creation of privacy norms.

The content of the Semantic Privacy Wiki is leveraged by all the other CAPrice services. In particular, the *Information Seeker Services* enables the user to understand better the privacy policies of popular applications, and provides application recom-

mendations that can satisfy the needs of the user while being as compatible as possible with the user's privacy expectations. The functionalities are exposed through appropriate UIs, accessible through the Web (*CAPrice portal*), while being also mobile-friendly (through the *CAPrice Mobile App*). Last but not least, we provide a set of *Developer Services* to allow external developers to improve or enhance the CAPrice functionality by providing new services or by improving existing ones.

As already mentioned, the current maturity level of these tools varies, as the development of the CAPrice platform is work in progress. As a result, some of these tools are still at the planning stage (e.g., Dialogue Spaces, CAPrice API), whereas others have progressed to the implementation phase with varying levels of progress (e.g., Semantic Privacy Wiki, Recommender, Privacy Dashboard, CAPrice Portal, ToS Annotator). In the rest of this section, we give details on the most important activities currently undertaken towards developing the CAPrice platform, including a short presentation of the most mature tools and the results of applying these tools in practice (where available).

### 4.1 Communication Channels Towards A Grassroots Community

CAPrice is a holistic solution towards improved privacy awareness. Even though we have not yet implemented the CAPrice solution to its full extent, some early efforts have led to the implementation of a series of *communication channels* (consisting of a frequently-updated website and social media accounts for improving our engagement and penetration potential), and to the creation of the initial core of the CAPrice ecosystem, including an *Ambassadors' Group* and the *CAPrice Community*.

The CAPrice website[17] offers multiple ways to users to provide feedback, mention their personal stories, and express their opinion. We have been active in continuously providing information regarding the latest policies on privacy of digital apps and services. In addition, the CAPrice website acts as a digital privacy portal presenting privacy leaks, potential solutions and multimedia content regarding digital privacy, focusing on privacy concerns and data protection issues that arise daily. To achieve this, we follow relevant sites, scientific reports/papers, news by privacy experts and hackers, and we publish relevant articles to the CAPrice website. Moreover, there are a lot of short videos/animations in the privacy related section that can be used by teachers and parents to inform kids in a visual and more entertaining way about privacy issues.

The relevant content is also shared through the CAPrice social media accounts (Facebook[18], Twitter[19], Youtube[20]) and is used to gather feedback or interact with users upon relevant posts or issues for our active online community. The content we share is not technical and is intended to the general public. Special focus is given in news concerning toys for kids, student apps and other subjects that, from experience, seem to attract the most attention, in order to ensure that the interest level of CAPrice

[17] https://www.caprice-community.net

[18] https://www.facebook.com/CapriceCommunity/

[19] https://twitter.com/CapriceSociety

[20] https://www.youtube.com/watch?v=4L8gOfU9MXg

community members remains high. The use of social media accounts is a key tool towards maximizing the community outreach and achieving optimal results. Social media are very popular in children and teenagers, which are critical age groups for achieving real, time-enduring change in privacy-related practices.

The aforementioned communication channels have contributed to the creation of the CAPrice community. Although the CAPrice community is by no means a sizable virtual community (yet), the initial statistics show not only prominent indications that the critical mass needed to make the community self-sustainable can indeed be reached, but also that the topic of digital privacy has become a key concern for the average consumer, despite the fact that the current scheme of interacting with digital technology shows otherwise. At the time of writing, the website had 114 unique visitors per day on average and 171 email subscribers, while the Facebook page had 569 likes, the Twitter account 232 followers, and the Youtube video had been viewed 1483 times. An indicator for the impact of this effort is the fact that our tweets overall have earned around 14000 impressions over the last 3 months (May 29, 2018 to August 26, 2018) while the pinned tweet earned 4946 impressions with 63 engagements. Furthermore, the latest 30 posts that have been published in our Facebook page during the same period have earned 9253 reaches and 381 reactions.

### 4.2 The CAPrice Privacy Ambassadors

Perhaps the most challenging part when transferring a socio-technical solution from paper to practice is to achieve the right balance between communities and technology. This is one of the most emphasized lessons learned by almost all past collective intelligence initiatives. Indeed, practice shows that for any established community to grow or for any new community to obtain substance, a group of highly committed and internally motivated individuals needs to be at its core. These individuals support and energize the whole community and maintain the social processes within; they initiate action, generate ideas, and motivate others. Members of the core, which is usually only a small fraction of the community, are characterized by both specific psychological traits (engagement, motivation and charisma), as well as specific structural positions in the social network [13].

Within the CAPrice ecosystem, these members are the *CAPrice Privacy Ambassadors*[21]. The group of Ambassadors is an evolving entity that has a specific role in the entire lifecycle of our initiative. Our intention is to exert only minor control over this group's dynamics, fuelling it with the proper means to help it obtain self-definition, but still leaving the necessary flexibility required to grow in size and adapt to the community's evolving needs.

The key role of the Ambassadors in our effort led us to start contacting and securing the support of the first Ambassadors as one of our first tasks. Currently, the CAPrice Privacy Ambassadors group is a core group of high-profile privacy enthusiasts from Europe and around the world. The founding members were carefully selected to combine three profile characteristics: privacy consciousness, more than average

[21] https://www.caprice-community.net/privacy-community/

knowledge about digital technology, and confirmed desire to motivate society into adopting a more privacy-aware behavior. Currently, CAPrice has employed 20 ambassadors with various characteristics and expertise, ranging from academics to lawyers, developers and entrepreneurs.

### 4.3 Improving Engagement Through the CAPrice Game

To keep our community active, and also to help them become more aware of privacy-related issues, we created the *CAPrice Game*[22], a simple, interactive mobile quiz game that tests the knowledge of kids, parents and teachers regarding the privacy of popular digital apps. The CAPrice game is available through the Kahoot platform[23] and requires only network access and a teacher/manager to control the whole game. This game contains a lot of fun features (music and sound effects, scoreboard to show the current top-scoring players, extra points for correct sequential answers and awards for the top-3 players) in order to increase motivation and engagement. Furthermore, it offers a single-player and a multiplayer mode. The game is highly configurable and scalable to include more questions or request relevant feedback from the players. The results of the game could be saved and exported in various formats in order to gain more knowledge by drawing conclusions about users' privacy expectations and by paying attention to the correct answers.

The CAPrice Game can be easily modified to include more questions or request relevant feedback from the players and can be played in the English or Greek language. We have already tested it to high schools that have visited the Institute of Computer Science at FORTH, and it was also demonstrated at the TEDxUniversityOfCrete conference[24].

### 4.4 Annotating Terms of Service Documents (ToS Annotator)

To cope with the complexity of ToS documents, there are efforts along the following two directions: a) formal privacy policy languages readable by machines that can be used by both the users and the services for describing their privacy expectations, concerns and policies, and b) through annotating the ToS with privacy related information.

A lot of work is currently conducted along the direction of enriching and annotating privacy policies with privacy related information (e.g., specifically designed tags embracing different privacy concerns like data collection, data retainment, etc.). Such tags can be pinned in ToS either by privacy experts or through machine learning algorithms. Unfortunately, although experts are able to provide accurate annotations, the task of annotating the available ToS in the huge and dynamic Internet/Web environment is possibly a Sisyphean one for the limited number of privacy experts. On the other hand, the current machine learning approaches are only able to annotate ToS

[22] https://www.caprice-community.net/game
[23] https://www.kahoot.com
[24] http://tedxuniversityofcrete.com/

segments with the correct but general privacy concern categories, while they are not able to identify more fine-grained information related with the specific values for this category [6], [12].

In CAPrice, we put forward another alternative for annotating ToS that revolves around the wisdom of the crowds. Since the problem of privacy awareness is a social issue, we believe that users should be active producers and reviewers of privacy related content, and not just consumers. Towards this, we have designed and developed *a crowd sourced platform for engaging users in the annotation of privacy policies* [5]. Our aim is to provide to the CAPrice community and all interested users a reference open-source and public platform for the creation, review and evaluation of privacy policy annotations. We already implemented a first pilot version to test various interaction modes for non-expert users and to verify that the content created can be of high quality. Our initial comparative results conducted over the only available expert based OPP-115 ToS privacy annotated collection[25] from the Usable Privacy project, show that the crowd-sourced privacy policy annotations, cooperatively created and reviewed in our platform, are of high importance and quality, comparable in most cases to the annotations created by the expert users [5].

### 4.5 Interacting with CAPrice Data

We are implementing the first release of an *open semantic repository* that constitutes the core of the CAPrice ecosystem and will store a multitude of privacy-related information. Through this tool, all visitors will be able to find information about digital products, such as the requested access policies or the related ToS documents. Facilities are being developed to assist exploration on various axes, e.g., by categorizing products based on their type (smartphone apps, smart products), their purpose (entertainment, weather, travel), the community rating (highly trusted, suspicious), etc.

The system also allows CAPrice members to specify their own expectations and views regarding the privacy policy of each product, e.g., how comfortable they feel about the privacy requests of a particular product, under which conditions they would grant access, and others. We are designing a set of visual cues to help users in expressing their expectations, without overwhelming them with questionnaires and textboxes (see, e.g., **Fig. 5**).

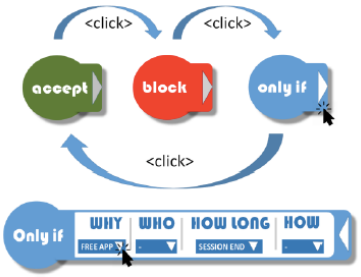


**Fig. 5.** Multi-button for expressing privacy expectations on a specific data access request

Finally, developers will also be able to add input, specifying their access policies and justifying them as appropriate. Note that this latter input is not necessary to ensure a smooth operation of our platform; due to the collaborative nature of CAPrice, simple or expert users can provide relevant information, although of course the active involvement of developers will also be

[25] https://usableprivacy.org/static/data/OPP-115_v1_0.zip

encouraged and supported, in order to help them build a more privacy-sensitive profile.

Our current implementation of the semantic repository stores the aforementioned data in RDF format, the standard Semantic Web language for semantically enriched content. This format allows posing expressive queries that enable more sophisticated forms of automated information seeking and analyis, while also permitting the interconnection of the content with other datasets, following the Linked Open Data paradigm and enhancing the interoperability of the ecosystem. Our current version of the repository uses the open source Blazegraph triple store and currently contains around 2.8M triples concerning information about 241K applications on 55 categories that were automatically extracted from the Android Play Store, our starting point for the first release of the platform.[26]

On top of the repository, we are currently developing a graphical user interface that will be the frontend of the CAPrice portal, along with the first version of the Recommender and the Dashboard. The Recommender uses SPARQL, a standard query language for RDF data, that suggests -among others- similar smart products of comparable quality (based on the Android scoring system), but with fewer (or more compatible with the user's preferences) permission requests, or higher privacy-related rating by CAPrice users. The Dashboard, on the other hand, will aggregate data, in order to extract norms and trends with respect to the CAPrice users' expectations and will visualize analytics in various forms.

There are other implementation tasks that are pending in order to materialize the CAPrice ecosystem shown in **Fig. 4**, but of higher priority is the creation of the engagement and reputation mechanism discussed earlier, that will motivate and reward community members in the generation of new content, while helping them iron out contributions of limited value.

## 5 Conclusions

In this paper, we presented CAPrice, a *socio-technical solution based on collective awareness and informed consent that allows better engagement and awareness of the average consumer towards (digital) privacy*. Our approach aims to make the gains of adopting a more privacy-respecting attitude obvious and measurable, both for consumers, and for service/software providers, while also allowing decision makers and social scientists understand better the consumer needs. This way, the collective pressure of citizens, combined with market forces, will lead to synergies, healthy competition and attitude change for all involved stakeholders.

[26] The endpoint can be accessed from here (using "caprice" as the namespace and "http//caprice/" as the named graph): http://bit.ly/2z3k9jt. The Blazegraph rest API can be found here: https://wiki.blazegraph.com/wiki/index.php/REST_API

## 6 Acknowledgements

The authors thank the following individuals for contributions in earlier versions of this work: G. Baroutas, A. Dimitriadis, K. Doerr, G. Ioannidis, Y. Marketakis, N. Minadakis, G.M. Moen, F. Myrstad, A.K. Ravna, Y. Rousakis, M. Titorencu. The work of N. Nikolov and D. Roman was partly funded by the H2020 projects euBusinessGraph (#732003), EW-Shopp (#732590), and TheyBuyForYou (#780247).

## References

1. Arniani, M., Badii, A., De Liddo, A., Georgi, S., Passani, A., Piccolo, L., Teli, M.: Collective Awareness Platform for Sustainability and Social Innovation: An Introduction, Brussels, EC, CAPS (2014)
2. Bagnoli, F., Guazzini, A., Pacini, G., Stavrakakis, I., Kokolaki, E., Theodorakopoulos, G.: Cognitive structure of collective awareness platforms. In: IEEE 8th International Conference on Self-Adaptive and Self-Organizing Systems Workshops (2014)
3. Carrascosa, J.M., Mikians, J., Cuevas, R., Erramilli, V., Laoutaris, N.: I always feel like somebody's watching me: measuring online behavioural advertising. In: 11th International Conference on Emerging Networking Experiments and Technologies (2015)
4. Goldfarb, A., Tucker C.: Privacy and Innovation, Innovation Policy and the Economy, University of Chicago Press, vol. 12(1), pages 65-90 (2012)
5. Hompis, G.: CAPPA: A Collective Awareness Platform for Privacy Policy Annotations. MSc Thesis, University of Crete (2018)
6. Liu, F., Ramanath, R., Sadeh, N., Smith, N.A.: A step towards usable privacy policy: automatic alignment of privacy statements. In: 25th International Conference on Computational Linguistics (2014)
7. Olurin, M., Adams, C., Logrippo, L.: Platform for privacy preferences (P3P): current status and future directions. In: 10th Conference on Privacy, Security and Trust (2012)
8. Porter, C.E., Donthu, N., MacElroy, W.H., Wydra, D.: How to foster and sustain engagement in virtual communities. California management review 53.4: pages 80-110 (2011)
9. Pitt, J., Diaconescu, A.: Interactive self-governance and value-sensitive design for self-organising socio-technical systems. In: 1st International Workshop on Foundations and Applications of Self* Systems (2016)
10. Patkos, T., Flouris, G., Papadakos, P., Bikakis, A., Casanovas, P., Gonzalez-Conejero, J., Figueroa, R.V., Hunter, A., Idir, G., Ioannidis, G., Kacprzyk-Murawska, M., Nowak, A., Pitt, J., Plexousakis, D., Rychwalska, A., Stan, A.: Privacy-by-norms privacy expectations in online interactions. In: 9th International Conference on Self-Adaptive and Self-Organizing Systems (2015)
11. Roman, D., Gatti, S.: Towards a reference architecture for trusted data marketplaces: the credit scoring perspective. In: 2nd International Conference on Open and Big Data (2016)
12. Sathyendra, K. M., Wilson, S., Schaub, F., Zimmeck, S., Sadeh, N.: Identifying the provision of choices in privacy policy text. In: Empirical Methods in Natural Language Processing (2017)
13. Schroer, J., Hertel, G.: Voluntary engagement in an open web-based encyclopedia: Wikipedians and why they do it. Media Psychology, **12**(1), 96-120 (2009)
14. Sloan, R., Warner, R.: Unauthorized Access: The Crisis in Online Privacy and Security. CRC Press, Inc., Boca Raton, FL, USA, 1st edition (2013)