Multidisciplinary | Rapid Review | Open Access Journal

Date of publication MMMM DD, YYYY, date of current version January 2020.

Digital Object Identifier DOI HERE

Towards an Accountable Web of Personal Information: the Web-of-Receipts

VITOR JESUS[1]

[1]Birmingham City University, School of Computing and Digital Technology, Birmingham, United Kingdom (e-mail: vitor.jesus@bcu.ac.uk)

ABSTRACT Consent is a corner stone in any Privacy practice or public policy. Much beyond a simple "accept" button, we show in this paper that obtaining and demonstrating valid Consent can be a complex matter since it is a multifaceted problem. This is important for both Organisations and Users. As shown in recent cases, not only cannot an individual prove what they accepted at any point in time, but also organisations are struggling with proving such consent was obtained leading to inefficiencies and non-compliance. To a large extent, this problem has not obtained sufficient visibility and research effort. In this paper, we review the current state of Consent and tie it to a problem of Accountability. We argue for a different approach to how the Web of Personal Information operates: the need of an accountable Web in the form of Personal Data Receipts which are able to protect both individuals and organisation. We call this evolution the *Web-of-Receipts*: online actions, from registration to real-time usage, is preceded by valid consent and is auditable (for Users) and demonstrable (for Organisations) at any moment by using secure protocols and locally stored artefacts such as Receipts. The key contribution of this paper is to elaborate on this unique perspective, present proof-of-concept results and lay out a research agenda.

INDEX TERMS Privacy, Consent, Accountability, Web-of-Receipts, Personal Data Receipts

I. INTRODUCTION

CONSENT is perhaps the heart of any Privacy policy or practice. In a recent case in the United Kingdom (at least), a social network was reported to be sending paper letters on behalf of their users without their knowledge. Anecdotally, individuals would only be aware of those letters when other individuals, who received the letters, told them. When journalists confronted the business behind the social network with the fact, they argued that a large notice was displayed to the users when creating their profile on the social network. It was also claimed that users could not finish creating an account without actively accepting it. Several users were interviewed and denied ever encountering such prominently displayed notice.

The following are possible explanations:

- the business is simply not being honest, and consent never existed
- at the time the individuals signed-up, consent was valid but, later, the business changed it to include different terms
- the consent request existed but was part of a long text or otherwise was not clearly informed
- the user interface misled the individuals, did not work as intended on their devices, or its underlying logic tricked the user into accepting conditions they were not aware of
- the affected individuals had poor digital skills and, despite the prominent notice, they still missed it
- the individuals are simply not being honest and valid consent did exist

Such a simple example shows that the current Web has a weak notion of accountability when it comes to personal information. It highlights the importance of obtaining valid Consent particularly when *demonstrating consent* is needed after Personal Information (PI) has been exchanged. In a nutshell, the challenge is how Users and Services can prove that Consent, at a particular time, was obtained and it was *valid*.

The solution, as of today, is to involve Regulatory Agencies such as, in the United Kingdom, the Information Commissioner Office (ICO). The ICO advises that Consent must be *recorded* [1]: "the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data." To this aim, the Controller "must have

an effective audit trail", a "master copy of the document or data capture form containing the consent statement in use at that time" and "If consent was given online, your records should include the data submitted as well as a timestamp to link it to the relevant version of the data capture form." This is, of course, fully aligned with the EU General Data Protection Regulation (EU/GDPR). Other regulations, such as the upcoming California Consumer Protection Act, adopt a similar idea of consent validity.

In this sense, log files or similar artefacts held by participants is commonly, but widely erroneously, thought to be sufficient. As the example of the social network we introduced earlier shows, local records (held either by the User or Service) are not authoritative as they can easily be doctored, manipulated or simply forged. The key fact is that only a lengthy forensic investigation, and later decision by a judge, could give any hope of resolving the dispute. The mentioned case of the Social Network highlights the fact that it is virtually impossible to prove, at least in simple terms and beyond non-repudiation, where the fault lies.

Having indisputable evidence of what-how-when Consent was obtained is an open problem and is our key focus on this paper. A key difficulty in Consent is that it is a multidisciplinary problem. It stems from the Public Policy and Regulatory space; then it is detailed into legal writing, often difficult to comprehensively capture all scenarios. Finally, it is executed and enforced using technical means (such as software). As such, a wide range of professionals and communities needs to be brought together. When bringing the Web into the equation, and as argued in this paper, the problem can become close to intractable unless a clear framework exists. One should stress that this problem exists for every party: Users cannot hold Organisations accountable, the Organisations cannot easily demonstrate compliance, and Regulators find too costly to investigate disputes.

We adopt a unique position in this paper by putting the focus on Accountability and the participants directly. Figure 1 illustrates what we mean by an accountable Web of Personal Information (PI). Upon any exchange of personal data, both Individual and Controller (the entity managing personal data about an Individual), receive an unforgeable and non-repudiable receipt of the transaction. This is inspired in normal payments when shopping. On a dispute, each will have undisputable evidence of the PI-transaction details. Such receipts have been called "Personal Data Receipts" (PDR) [2], a terminology we will keep.

This paper reviews the overall problem of Consent in online data sharing and proposes a concept we call the *Web-of-Receipts* (WoR). We propose that the Web must be accountable for all interactions involving personal information. We will argue that transactions involving Personal Data must always come with a *receipt* that Individuals and Organisations keep in their possession. A key issue is non-repudiation: for such receipts to be valid and beyond any possibility of dispute, a peer-to-peer trusted protocol is proposed. Considering that a WoR requires widespread adoption, we also propose

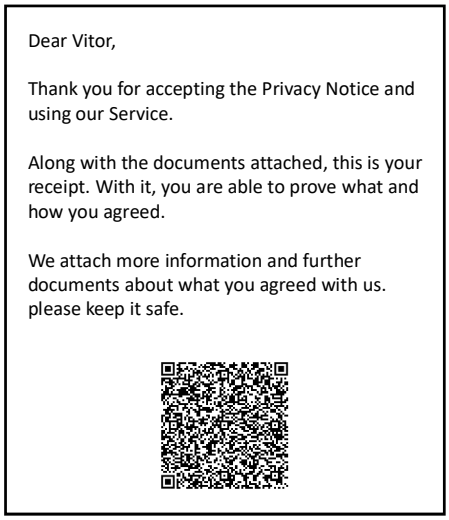

FIGURE 1. A mock-up of a receipt for Personal Information

an evolutionary architecture based on trusted Third-Parties, some components of which, in some form, already exist. As we will argue, more work and research are necessary to firmly establish a WoR.

Section II defines and reviews the components of online Consent. Section III reviews the related work on how Consent can be managed and we argue that this paper presents a unique perspective. Section IV presents our architecture of a WoR and, in particular, our secure protocol which has non-repudiation at its centre. Since our architecture presumes direct support of both the Individual side and the Web side, which requires sufficient adoption, we propose an evolutionary architecture in section IV-D. In section V we present a research agenda on Consent and the Web-of-Receipts. Section VI concludes our paper.

II. CONSENT

Consent is mainly a legal notion. From a technical perspective, the Consent problem should be drawn from the underlying regulations, legislation and public policies. The paradigmatic case is EU/GDPR in force since May of 2018. Other jurisdictions are expected to set similar precedents such as the California Consumer Privacy Act, the UK's Data Protection Act of 2018 and even the 2011's Korean Personal Information Protection Act.

EU/GDPR clearly establishes the need and validity conditions of Consent. It reads *Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data* (GDPR Recital 32) [3]. The essential requirements of *valid* consent are

- The *act* must be affirmative, freely given and with unambiguous manifestation. In other words, the Individual must be able to express consent out of pure volition and without coercion, e.g., there should be no pre-checked

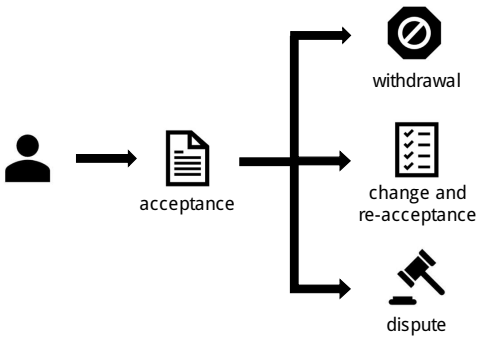

FIGURE 2. Simplified Lifecycle of Consent.

boxes or small fonts.

- The process by which consent is obtained must be driven by *knowledge and voluntariness* [4]. Consent must thus be specific, informed and clear. The information presented to the Individual must be easy to handle and must lead to perfect understanding of what is being consented with. Examples include no long legal texts or using legal terminology outside common knowledge.

Furthermore, and central to this paper, consent must be *demonstrable* (GDPR art. 7): "the controller shall be able to demonstrate that the data subject has consented". To note that this requirement does not affect, directly, Individuals. We shall add that all parties must be able to prove Consent was validly taken, in a form satisfying non-repudiation, at any point in time. This element is where we see the centre of accountability of a Web-of-Receipts. We now explore in more detail each of the above requirements.

1) The Lifecycle of Consent

In Figure 2 we show the essential stages a User goes through when accepting a Privacy Notice. It should be noted that this is a highly simplified diagram. For example, it does not show the complex case of involving sharing personal information with Third Parties which raises complex problems about *revocation* or *re-consenting*, as we discuss later in the paper. We will use the terms *Controller*, *Organisation* or *(online) Service* interchangeably.

As a first step, the User must be exposed to a Privacy Notice. The notice should be clear, specific and unambiguous. This requires that the Web converts from the current state of often long legal texts to human-friendly interfaces. In an interesting title, Obar called this state of play *the biggest lie on the Internet* [5]. To give a paradigmatic example, it has been shown that Users accepted free Wi-Fi in exchange for ridiculous conditions such as giving their first-born child away.

Usability is a clear challenge at this stage. The way such complex information is presented is nearly always controlled by the Organisation which uses this fact in their favour. Some research exists on this topic such as visualisation of Privacy Notices [6] [7] [8], including using comics [9], but the topic needs to be refreshed given the dramatic changes in the landscape in the recent years. A recent example is the Consent Request (CoRe) user interface [10] which presents a user interface to visually structure and manage consent across its lifecycle. The Usable Privacy Policy Project [11] is another initiative which uses natural language processing to help users make informed choices.

Once consent is given, there are three broad paths. The first one is the user withdrawing consent, either wholly or in part. For example, the user may not be comfortable with certain types of personal data and withdraw consent on those specific classes of data at a later stage. The second path is the Controller making changes to the Notice which require the User to re-accept the new Terms.

The third path is the case where, over a previously agreed Notice, there is a disagreement between the user and the controller on how personal data was collected or used. This path likely involves, as it currently stands, authorities. This is the case of the social network we have been using as an example: Users and Service disagree on what was consented with and a Third Party may need to be involved. It is mostly for this last case where we see the value of an accountable Web which is based on architectures centred on non-repudiation.

2) Personal Data Receipts

A Personal Data Receipt (PDR), similar to a conventional paper receipt, is the storable artefact that both User and Service store in order to possess evidence satisfying non-repudiation of who, what and how was agreed. The work done at the Kantara Initiative is an excellent example with their specification on Consent Receipts [2] which has been implemented by a number of organisations. The specification creates a JSON format for a Receipt. In its current form, and under active development, it contains the following generic fields: the collection method, the jurisdiction, PI categories and purposes. To note that Kantara is the main proponent of User-Managed Access (UMA) [12] which creates a technology that enables users to manage their personal data with fine granularity.

A key component for a functional receipt is clear terminology. Similar to the general state of art, terminology is not consensual as there is no established guidance or standard. Jointly with the draft specification ISO/IEC 1st DIS 29184 ("Online privacy notice and consent"), it is expected that the ISO/IEC TS 2043 "Privacy technologies" newly created working group at ISO SC27 (overseeing the ISO/IEC 27000 series) will help by providing guidance on Consent recording.

It should also be noted that, when exchanging proof of acceptance, both on the User and the Controller sides, one should avoid falling into the trap of a legal text which users will mostly ignore. As such, the ideal format of a PDR is one that is machine-readable. Vocabularies and ontologies of Consent are a promising research direction such as the one

proposed by Pandit and Lizar [13]. Beyond the clarity which stems from a machine-readable format suitable for display in human-friendly visualisations, ontologies also open the door to machine-reasoning [14] [11] [15] and, for example, establishing relationships that may not be clear at first glance to a human – hence, supporting key requirements of valid consent such as unambiguity and clarity.

III. RELATED WORK

An early project looking at the problem of Consent was EnCoRe [16] (United Kingdom, 2008-2011). Beyond gathering ideas from different communities [17], a promising recommendation was to wrap personal data and the associated consent with metadata (sticky policies [18]) which establishes the conditions under which data can be used. This is of paramount importance for many cases, including those within the health community which can see projects (and therefore consent) last for decades and, therefore, need re-consenting often under changing circumstances (dynamic consent [19]). It should be, in fact, said that, to a large extent, the health research community is a pioneer in handling Consent [20] [21] [22] [23] albeit mostly from an ethical and administrative perspectives.

A natural step after attaching policies to personal data is to use Ontologies to model Consent: once a robust model exists (which may need formal policy languages such as YAPPL [24]), automatic reasoning can be done over it up to providing autonomous verification of consistency and compliance. On one hand, the formalisation of the ontology rules and the wider semantic models are, for the most part at least, directly mappable from legal requirements. This approach justifies the work of Fatema et al. [25], Bartolini et al [26], among others. To give simple an example, Consent has intrinsically a validity time frame and has a context which is specific to the use-case. On the other hand, once an ontology is established, reasoning can be applied which has the potential of bringing forward non-trivial relationships – for the compliance case, this may have wide cost savings and flag unintended usage of personal data. A particularly comprehensive approach using Ontologies for Consent is the EU-SPECIAL project [27]. It directly addresses the requirements of transparency and compliance of data processors. Their architecture combines several elements of which a Semantic model is at the heart. In order to develop a comprehensive model, a Data Privacy Vocabulary (DPV) was co-developed which is now W3C DPV [28]. Overall, the strategy is to capture metadata about Personal Information and related events which will be automatically checked for compliance using the developed semantic model. EU-SPECIAL also offers a user-interface where Users can visualise and adjust their permissions which helps with the Usability element of Consent.

A current research pattern is to decouple Consent management from the overall data handling which could evolve to a Privacy-as-a-Service model. The EU project OPERANDO [29] demonstrates this approach by creating an infrastructure where independent Third Parties can take Compliance requirements outside organisations. Personal data is stored in an external place and users selectively release their data to organisations on an as-needed basis. This includes reconciliating multiple data sources such as the cases of federations discussed by Ulbright and Pallas [30]. Furthermore, such parties could be in an ideal position to handle the complex problem of Consent delegation to third-parties [31] when personal data is shared with data brokers, often without the User's knowledge.

Decoupling Consent from the individual data paths is also used in the case of Internet-of-Things (IoT). This is perhaps the only possible approach: IoT applications can be extremely challenging for Consent because devices may not even have a screen or buttons, and Users cannot always use an accessory device such as a mobile phone [32]. Particularly when there are multiple devices which collaboratively define the service, an overarching consent needs to be given by aggregating and consolidating different data sources and devices. As such, the typical approach is to define new dedicated elements in a typical IoT architecture such as Agents [33], Privacy proxies on the IoT supporting network [34] or by creating a specific, network discoverable, service for Privacy disclosure (such as the Privacy Assistant of CMU [35]). This is made more challenging by the fact that fact that many IoT applications are dynamic, have lose connectivity graphs and weak (if any) device identity. An example is Vehicular Networks [36] where, if cars are to communicate between each other, there may be no time to contact the network infrastructure and, instead, Consent needs to solely rely on locally generated identities and related cryptographic material such as Hierarchical Identity-based Signatures [37].

Nevertheless, there will always be many cases where Consent will be handled locally, either because Organisations do not want any third party involved or want to keep complexity low. Pro-active and local approaches to Compliance and Consent management will therefore be always a need. The on-going EU SMOOTH project [38] plans on automating compliance by combining machine learning (such as text mining) and software analysis (such as mobile apps). The DEFeND project [39] approaches the problem of enforcing Consent by creating a set of tools organisations can use to create and monitor contracts, along with providing notifications. A key component, similar to Project SPECIAL, is the detection of inconsistencies between Consent and usage of data. This problem is also being tackled by BPR4GDPR [40] which looks at compliance from a workflow and business process perspective.

An interesting technique to measure, at scale, the quality of privacy practices by organisations is to use crowdsourcing: Users themselves rate Organisations which then can be used to build a rating service. The EU Privacy Flag project [41] combines user tools (mobile, web and IoT) with a rating service where Organisations can be rated by users and/or third parties (including collaboratively). Users further share information among themselves. With such empowered crowdsourcing, the expectation is that organisations have now in-

centives to improve their privacy practices. This information, over time, can feed a risk-assessment engine which has the potential to evolve to a global risk index that anyone can use to quickly have some awareness of how trustworthy (or even compliant) an organisation is when handling personal data.

A similar approach is taken by EU project TYPES [42] which is still in progress. It focuses on the specific case of online advertising but it is nevertheless important because of its scale. The project will develop means for users to better understand and control their privacy expectations, including means to report violations. Similar to Privacy Flag, one important mechanism of detecting violations is by designing a crowdsensing platform. Equally important when looking at how incentives inter-play in Privacy, TYPES also designs a valuation mechanism for personal data.

User-empowering tools are universally seen as essential both when visualising the reputation of an organisation and to have some control over the lifecycle of their own personal data. Web-based dashboards are a practical requirement when it comes to allowing users to see history of the flows of their personal data – see projects PoSeID-ON [43] and VisiOn [44]. In particular, the EU VisiOn project designs a set of visualisations for Users to control their privacy settings and better understand and monitor how their personal data is used. VisiOn primarily focusses on Public Organisations, but it seems it can be extended to any other. A key concept developed is that of Privacy Level Agreement which can be directly, and easily (using the same visualisations), measured by users. The project creates also a JSON format for data handling organisational processes [45] which could seemingly be extended to generate Personal Data Receipts. If a secure protocol supporting non-repudiation is added, similar to the one we propose here, the idea of a Web-of-Receipts becomes closer.

Traceability of personal data and immutability of records are equally important and are immediate building blocks for non-repudiation. In general, the problem of keeping trusted records (such as logs) is difficult to tackle [46]. As said, this is also a key motivation to our design of a user-centred secure consent protocol and the wider idea of a Web-of-Receipts. Smart Contracts (generic code running on a blockchain), are seen as a promising technology given its inherent properties of write-once-read-many and immutability of storage of information. A number of proposals are exploring this technology, from PoSeID-ON to the medical community as seen in CrowdMed [47]. Furthermore, given its distributed properties, blockchains are also well poised to facilitate solutions in scenarios where multiple parties is involved. Bhaskaran et al [48] proposes to manage consent with blockchains in financial services in order to hide the identity of the remaining parties. Consentio [49] uses a blockchain to support their Consent architecture albeit with the requirement of a permissioned blockchain which can, arguably, reduce the strength of the accountability requirements. Blockchains to manage consent in IoT is also a promising direction as seen in the ADvoCATE architecture [50]. It is able to consolidate multiple distributed points of consent over a blockchain similarly to CoMaFeds [30]. Overall, even if introducing their own research problems, blockchains provide a framework so that data generators and data consumers can meet at a pseudo-centralised point. Such a meeting point can also be used to support informed and trusted consent.

A. CONTRIBUTIONS

The reviewed approaches to managing Consent shed a light on the complexity of the problem and each explores, partly, solutions to the problem. Whereas the reviewed work can be, broadly speaking, broken into either addressing Compliance or mitigating (typically by automating) the complexity of the different dimensions, we argue, however, that true Accountability can only be reached if both User and Services are in direct control and at the same time. For example, whereas Semantic tools, or platforms such as SMOOTH, can help Organisations better manage, internally, Consent, and detect unexpected uses of Personal Data, the User is still largely unprotected and unable to prove misuse and hold Services accountable. As another example, the use of sticky policies (as in the EnCORE project), or blockchains for immutable storage of consent information, are both promising directions. However, they do not directly address the problem of "who manages the manager". Projects that empower the User with self-management tools (such as Privacy Flag or SPECIAL), or create crowd-sourced communities, must be seen as bringing great benefit, yet, they still do not provide the level of accountability this paper is aiming at.

Furthermore, the more complex and comprehensive a solution is, the higher the need of an evolutionary architecture. We anticipate this to be a challenge in itself. In practice, some of these architectures may require high levels of adoption (both organisationally and technically) and seem to either work only when fully implemented. In other words, their effectiveness is potentially reduced if only a subset of it is implemented. In this sense, we argue that a transition phase needs to be explicitly proposed from the beginning.

On one hand, our approach is complementary to the reviewed ones and several components can, and should, be reused in our approach to a Web-of-Receipts. On the other hand, however, our primary focus and the essential contribution of this paper, is *accountability* of Organisations and Users. We see this as going further than simple Compliance with local regulations. In other words, we do not focus on *how* Organisations deliver their privacy practices but, instead, we focus on how both the User and Organisation can challenge each other without room for dispute.

We argue this is a unique and novel starting point that drastically impacts (1) the set of incentives driving how Organisations manage Consent (and User's behaviours in the long term) and (2) reverses the power relationship from solely Organisations to a shared, "one on one", responsibility. From this unique perspective, we are also able to design an evolutionary architecture that, as will be described, shares several commonalities with other proposals – for example,

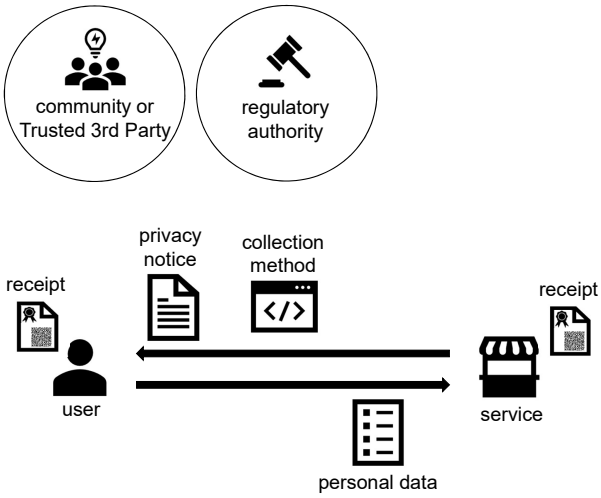

FIGURE 3. Components of the Web-of-Receipts.

with platforms that use crowdsourcing in order to rate the Privacy practices of an Organisation.

IV. THE WEB-OF-RECEIPTS

Having introduced the key concepts and reviewed related work, we now informally present our concept of a Web-of-Receipts (WoR). In the next subsection we formalise the problem and elaborate on our secure protocol supporting non-repudiation for the generation of receipts. Figure 3 depicts the essential scenario.

We use the familiar scenario of a User creating an account on a web Service such as a social network. A first action from the User is to review the Privacy Notice and related documents. On agreement, the User proceeds to create an account using a combination of an HTML form and JavaScript. The personal data is then sent over HTTPS.

When the user submits all the Personal data and Consent, the Service and User will run a secure protocol which has non-repudiation at its heart. Both parties will generate mutual evidence of who, what and how agreement (and, thus, Consent) was obtained. The result of this protocol is a Personal Data Receipt which both User and Service will hold. Upon any dispute in the future, both receipts will have indisputable evidence of what happened at this stage. An authority, as depicted in Figure 3, will not have to start a lengthy and time-consuming investigation as the receipts will hold all the necessary information. Resolving the dispute is only a matter of requesting both parties' receipts and comparing both. If a party, either the User or Service, refuses to produce the receipt, this could be an indication of dishonesty or deliberate non-compliance.

Figure 3 also shows a Trusted Third Party (TTP) which can simply be a community of Users that the User sending personal data trusts. We will discuss the envisioned role of this TTP in a later section. In essence, this entity will review, on behalf of the User, the Privacy Notice and act as an anchor of trust when disputes arise along with providing essential support to a transitional state between the current Web and a Web-of-Receipts.

A. THREAT MODEL

The problem we discuss is mutual for both User and Service. Both need to prove the agreement they claim has been accepted was, indeed, the one mutually agreed to. The following lists key threats in our Trust model:

- *Service tricks the user into accepting an unexpected Notice*. In other words, Consent is not valid as it is doctored to the particular User against expectations. For example, it contains specific and, perhaps, obscure conditions that the User is unaware of.

- *Service shows the expected Privacy Notice but invisibly modifies immediately after technically obtaining consent*. This involves how Consent is recorded such as in a webpage form with JavaScript code. It is trivial, with current Web technologies, to display the user a document that is then invisibly manipulated without the User being aware of.

- *Dishonest user claims Consent for a different Notice*. A User reports having seen a different Notice or not being shown certain parts. We assume it is ultimately up to the Service to make sure the User is fully aware of all elements of the Notice which includes the user interface. For example, the user's device must be able to display the page properly and this check should be done programmatically as much as possible.

B. FORMALISING THE PROBLEM

In this section we advance a technical, but simple, proposal for a Web-of-Receipts. It should be noted that this is just a representative approach as several aspects are not accounted for as later discussed in Section V. The simplest scenario is the case of a User U directly engaging with a Service S with no other party being involved throughout the Consent lifecycle. U and S will generate an agreement $\mathbf{A}_u = \{D_u, L, T\}$ where D_u is the user's personal data, L is the collection logic (such as the JavaScript code involved in Consent form using HTML) and T is the terms in the Privacy Notice. Notice that L, for the simple Web case is straightforward: considering HTTP/HTML is stateless, and the scenario in discussion is a web form for personal data, L is effectively the HTML and JavaScript source files. For the simple web case, this completely captures the collection logic. We will discuss other cases, not as trivial, at a later section.

Our approach to a secure non-repudiation protocol is inspired in the Fair-Exchange problem [51] [52]: in the simplest formulation, two parties, who do not trust each other, are to exchange one item each in a fashion such that either both receive or none receives. Figure 4 shows a signalling diagram of our protocol.

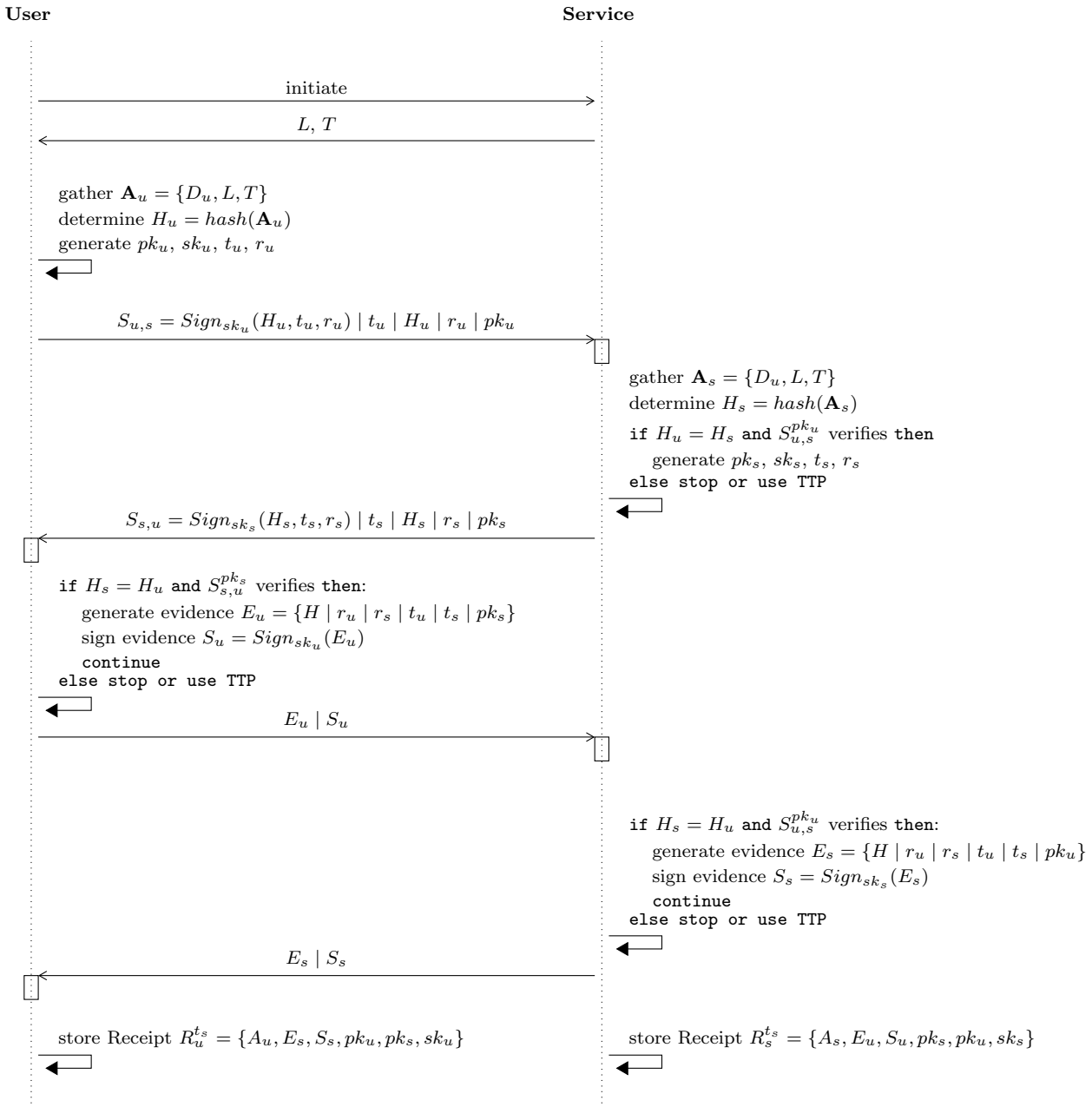

FIGURE 4. Protocol for the generation of receipts.

The protocol work as follows.

- U starts by gathering and locally storing Agreement $\mathbf{A}_u$. U then generates a pair of public/private keys, pk_u and sk_u. This key pair is, effectively, U's identity as one is assuming the generic case of an anonymous User.
- U applies a one-way, collision resistant, function to produce $H_u = hash(\mathbf{A}_u)$.
- U then sends message $M_u = \{\mathbf{A}_u, Enc_{sk_u}(H_u, t_u, r_u), t_u, H_u, r_u, pk_u\}$. This message consists of the agreement $\mathbf{A}_u$, its hash H_u, a timestamp t_u and a sufficiently long random nonce r_u. The role of r_u is that, along with another that S will send, r_s, it provides (with high likelihood) a unique identifier of the particular context of this agreement.
- The message is further encrypted with sk_u (the secret private key) and is sent along with the corresponding public key pk_u so S can verify.
- S runs the same procedure with the necessary changes.
- U sends evidence $E_u = \{Enc_{pk_s}(H, r_u, r_s, t_u, t_s)\}$ and attaches its signature $S_u = Sign_{sk_u}(E_u)$. Together with the agreement gathered at the first step, S generates

$\mathbf{R}_s^t = \{\mathbf{A_s}, E_u, S_u\}$ which is the receipt that S must hold as proof of consent.
- S executes a similar procedure which allows U to generate $\mathbf{R}_u^t = \{\mathbf{A_u}, E_s, S_s\}$ as its receipt from S
- Both parties now hold a receipt of agreement, either $\mathbf{R}_u^t$ or $\mathbf{R}_s^t$.

C. A PROOF-OF-CONCEPT

Figure 5 and Figure 6 demonstrate this approach[1]. We implemented the client-side (a normal web browser) code in standard JavaScript, which virtually all web browsers can run, and JavaScript for Nodejs for the backend server (as if it was the Service). When the user clicks the Submit button, the protocol is run in the background and the web form will not progress until the protocol completes. When it does, both parties will be in the possession of a receipt. The receipt of the user (Figure 5-right) holds the Personal Data sent, the Privacy Notice accepted and the HTML and JavaScript that was used to collect the Personal Data. The Server keeps similar objects. The User has further access to a convenient QR code holding the signatures of the documents included in the download. The QR code is not necessary but is included as an illustration of the usability aspect. If any of the parties refuse to run the protocol, or abruptly terminate the protocol, both are prompted to what to do next and are offered the option to fall back to the current unaccountable web operation.

In Figure 7 we show results of the latency that our protocol added to the webpage and the browser used. Both the webserver and the browser were running on the same computer (a modern Intel i5 laptop). As such, the latency is mostly caused by the processing of messages and the cryptographic operations. As seen, the latency is around 1.5 seconds for an arbitrary file size (HTML, JavaScript files and the Privacy Notice) up to 100kB. For a slow webpage, such as an account creation, this is typically not problematic. However, when considering real-time and fast responding pages, a 1-second latency is undesirable. We will further discuss this point later.

D. EVOLVING TO A WEB-OF-RECEIPTS

An architecture as just proposed will only be meaningful if both Users and online Services support it at large. Whereas on the User side it could be reasonable to think adoption could be very fast, given the incentives, on the Service side it can take a long while given the complexity of aligning current practices with our architecture and the effort of integrating Personal Data Receipts with their technologies.

We therefore propose an evolutionary architecture by using a Trusted Third-Party (TTP) which takes over the protocol if either party, User or Service, do not complete the protocol. As seen in Figure 3, we call this a Consent Manager (CM). The CM can offer *a proof of submission* in case one of the parties do not comply with the protocol. Even though it is not authoritative, it brings a much better level of assurance, especially on checking what the user has accepted such as the specific Privacy Notice. Details on this variation of the protocol can be found on [52].

In Figure 8 we illustrate how we envision a Consent architecture which supports evolution between the current Web and a WoR. It should be noted that this is mainly a functional architecture as most components are part of a research agenda. Our architecture is designed to be transitional by not being dependent on peer-to-peer support from Services or Users. It can still provide assurances such as inspecting policies on behalf of users (to verify Privacy expectations) or to resolve disputes (typically in an automated way).

As in the peer-to-peer case of Figure 3, both User and Service run a Consent protocol with similar properties as the one we suggested before. The primary outcome of the Consent Protocol is a trusted Personal Data Receipt. Both types of parties will use a *wallet* to keep receipts since it is envisioned that online activity will generate a great number of them. A wallet, similar to normal money wallets, is defined as the tool where receipts can be stored, retrieved and managed.

Identity is a necessary component, but we do not strictly require a strong identity such as based on official documents. Our notion of Identity can cope with a range of schemes: full anonymity, an identifier just for receipts, identity based on verifiable credentials, or identity based on official documents. Again, a different Third Party could also be an identity provider (of any form, as depicted in Figure 8)-top-left) and facilitate Consent without sharing it with the Online Services.

The Consent Manager element takes on the following roles:

- *receipt generator* – This is a function to allow Users to retrieve, in the worst case, a receipt of *unilateral* submission of personal data. This is the case where a Service does not comply with our Consent framework, but the User still wants to proceed. The CM will collect and store the components of the Agreement (as explained before): collection logic, personal data, privacy notice. In essence, the Consent Protocol runs between the User and CM.
- *real-time monitoring* – This is a function to identify any changes to a previously accepted Agreement on behalf of the User with or without the Service collaboration. If the Service cooperates by registering itself, the CM can still monitor any changes on behalf of the Service and issue notifications to the User only if in that case. This has the potential of greatly increasing usability. A simple application, that could have dramatic implications in Web usability, is the current situation with cookies: The User only has to accept once, and then upon changes, without revealing any personal data.
- *registration and attestation* – Services register themselves in the CM for the real-time monitoring service and present some sort of Identity.
- *dispute* – The CM cannot directly replace an Authority or Regulator but is able to act as one in first instance for dispute resolution. This function merely compares re-

[1]The source-code is open and freely available on request.

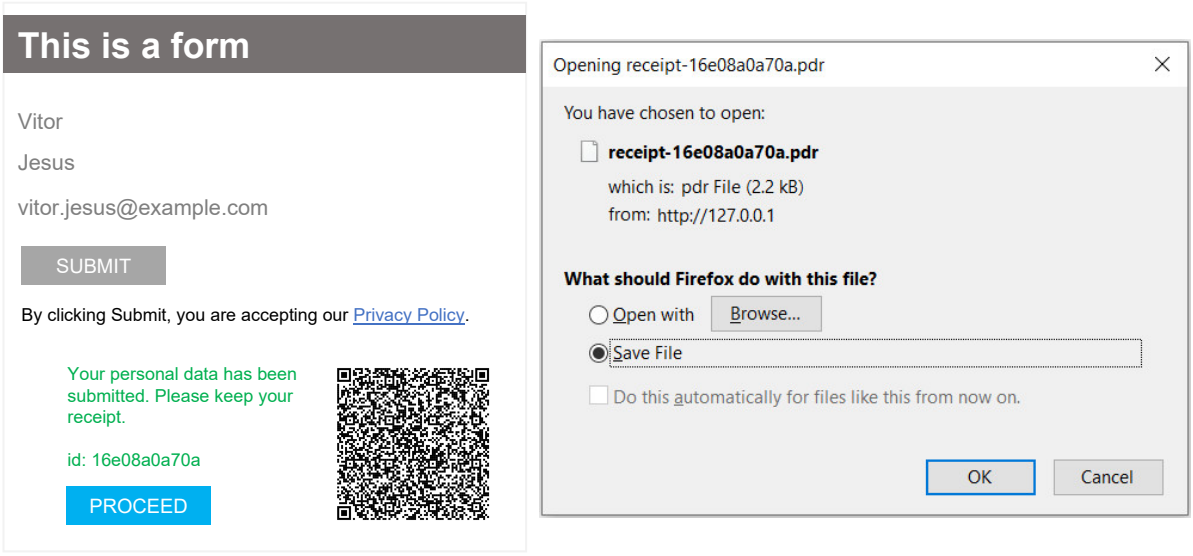

FIGURE 5. A proof-of-concept - user interface.

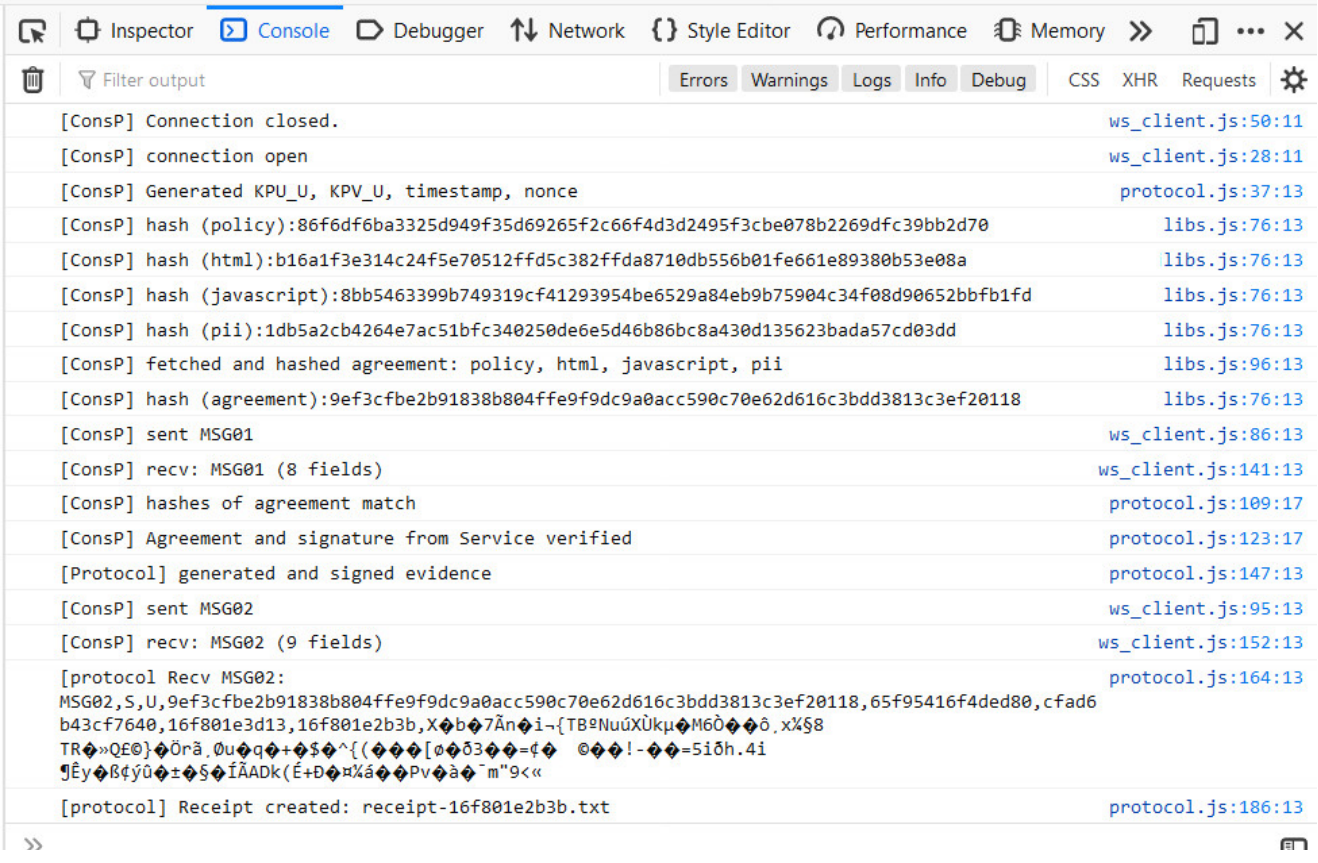

FIGURE 6. JavaScript running on the User's browser implementing the protocol

ceipts previously generated and determines, in case of a dispute, who has failed to deliver the Agreement. Since this is a simple validation and comparison of receipts, this function has the potential of greatly streamlining dispute processes (e.g., by automating) rather than going through a National Agency or Regulator. In case there is

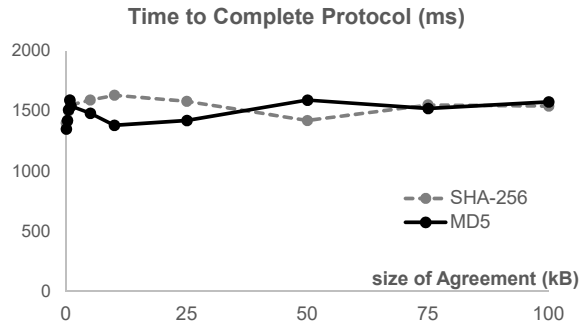

FIGURE 7. Latency incurred with our browser-based protocol.

no receipt available (e.g., either User or Service refused to run the Consent protocol), dispute resolution is likely to require an authority.

Finally, we note that Consent delegation needs to be supported as depicted in Figure 8-top-right. A primary case is when personal data is shared (mostly uncontrollably nowadays) to third parties. Receipts can be generated to cover the case of delegation upon sharing on behalf of the User, perhaps by attaching a policy to the data as suggested by the EnCoRe project [16].

To note that the idea of outsourcing Consent to a third party already exists in some form. Services such as Open Consent[2], Privacy Spy[3] or the Open Rights Group's Data Rights Finder tool[4] can be combined to serve this function to some extent. Whereas they are focused on watching over changes in Policies and aiding in Consent (such as Open Consent), others focus on reviewing Privacy Policies in the hopes Users make a better-informed decision.

V. THE WEB-OF-RECEIPTS: A RESEARCH AGENDA

Up to here we have shown the different aspects that Consent entails. Even so, the scenarios presented suffer from simplicity. In this section we present a research agenda to tackle this problem. It is also evident that the problem of Consent is multi-disciplinary as it involves the social sciences, technology and the law. Even within the technology component, it requires a multi-disciplinary vision in order to bring an accountable Web-of-Receipts to light.

The challenges ahead are numerous. To start with, we need an established *terminology* and vocabulary. Just a simple Privacy Notice can have different designations (such as Terms of Service) which may even be imposed by law depending on the jurisdiction; furthermore, different matters such as commercial terms are often mixed with privacy considerations. Having a precise meaning of each component is essential. As previously said, agreeing on a vocabulary can enable an approach to Consent based on ideas from the Semantic Web.

[2]http://OpenConsent.com
[3]http://PrivacySpy.org
[4]https://www.datarightsfinder.org/

Usability is a key challenge with two examples coming at the forefront. The onboarding of Users is the moment when the user is exposed to a Privacy Notice and requested to act. Whereas in a Web or mobile setting this is not too challenging, since a graphical interface can be used, other domains such as devices in the Internet-of-Things, dramatically lack such features. In essence, Consent needs to be decoupled from the device to a second channel. Another example of usability is a clear interface to manage the consent lifecycle such as when the User needs to revoke or re-confirm Consent.

Whereas we are focused on Personal Information in online scenarios, the problem of Consent exists in other areas such as medical research or IoT, as discussed. These fields require a different approach as risks can be far more severe than simply mismanagement of personal data. For example, a particular problem that does not typically exist with Personal Data is a "break-the-glass" exception. With medical data, the User may need to share their data but be unable to (for example, in case of an accident). When justified, Consent may need to be administratively waived but nevertheless recorded in order to raise accountability.

Storage of Personal data Receipts is also challenging given the scale. The notion of a Receipt Wallet comes as a starting point, similarly to a cloud storage location but fit to hold receipts. In an ideal scenario, the User should self-manage receipts; however, it can be technically challenging for the majority of users. A Third-Party could exist to aid the user in managing the envisioned vast amount of receipts which Users would only notice in case a dispute arises. Furthermore, note that a third-party wallet service does not necessarily have to see Personal Data in itself. The extent that this is applicable needs more research.

Reviewing Privacy Notices is and will be, for the foreseeable future, a problem. To put it bluntly, Users will hardly ever review long Notices so means must exist to (1) establish the reputation of an organisation in terms of Privacy and (2) automate the review of Notices. Once again, we see, at least temporarily, Third-Parties aiding Users with this by using crowd-based annotations and rankings.

The non-repudiation protocol, used to generate and attest the receipts, needs development and, ideally, standardisation. Techniques to facilitate adoption (perhaps in a Consent-as-a-Service model) should be developed so that organisations can quickly support the protocols.

The receipts in themselves need well-defined *syntax and semantics*. However, they should be flexible enough to accommodate new types of data and operations instead of standardising the data categories themselves which, given the necessary slow pace of standards bodies, would quickly be outdated.

Dynamic applications pose a particular challenge. In a full deployment of a Web-of-Receipts, a receipt should be generated not only when a User creates an online account but for all activity which involves Personal Data. Clearly this is, per se, hardly manageable so techniques to minimise data while providing accountability need to be developed.

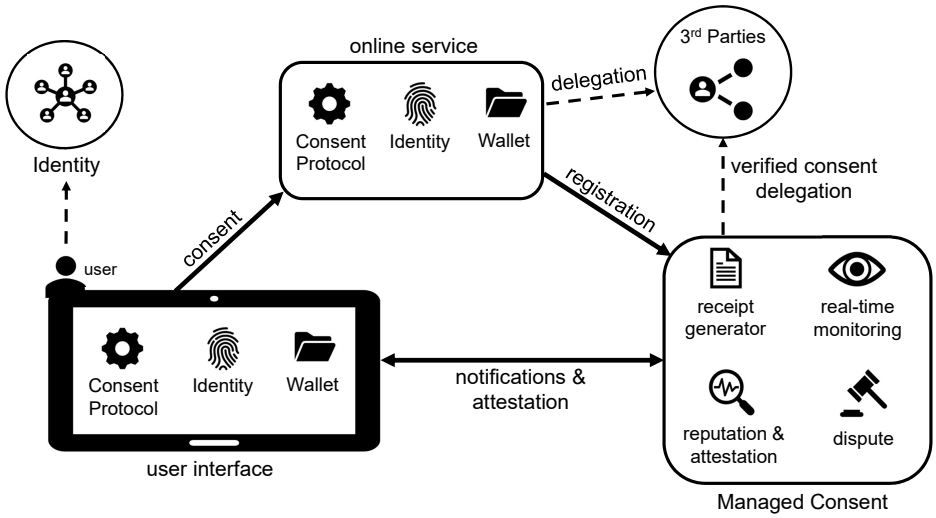

FIGURE 8. Evolutionary Architecture for Consent Management.

Furthermore, considering our threat model in section IV-A, if the application collecting the user data is closed source, new techniques need to be developed to capture that logic beyond simply storing the source or executable code. A possibility is to use software attestation techniques to verify that the application the User is using is indeed what the Data Controller is claiming. This raises challenges in itself such as, for example, requiring the code to be signed. Finally, whereas in our example we used HTTP, and fairly safely assumed full statelessness, dynamic applications such as mobile ones will not be. Capturing the context of the transactions can be challenging.

Data sharing with *Third-Parties* is a further complication particularly when *revocation* of Consent is involved. We prefer to call this problem *delegation of consent*. When the interaction is between the User and the Service directly, the agreement is clear and valid. If the Service shares the data, the obtained Consent assurances can be immediately lost. More research is needed on how to manage the delegation of Consent while embedding secure, non-repudiation protocols.

VI. CONCLUSIONS

We presented our concept of an auditable and accountable Web of Personal Information, which we call the Web-of-Receipts. Beyond valid Consent, all activity involving Personal Information should come with a (Personal Data) Receipt. Such level of accountability can dramatically increase openness, transparency, accountability and trust in today's vastly unregulated Web of Personal Information. We suggest a generic architecture and framework and show that an ecosystem of Third Parties can be used as an evolutionary architecture. Non-repudiation of receipts is further discussed as a central element for which we present a proposal.

Our analysis of the related work and results of our proof-of-concept suggest that a WoR is mainly feasible. Nevertheless, we identify a number of research challenges which need more research or test implementations with field trials. In particular, we are currently developing work on three directions. First, we are working on making available our proof-of-concept Consent protocol to anyone, both Users and Services, in the form of a browser extension. Second, we are creating an open-source prototype of a generic Consent Manager. Finally, but equally important, we aim at developing mechanisms to hold accountable Personal Data aggregators.

ACKNOWLEDGMENTS

We would like to thank Mark Lizar, of Open Consent, and Joss Langford, of Coelition, for their insights.

REFERENCES

[1] The ICO, "How should we obtain, record and manage consent?", online: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/how-should-we-obtain-record-and-manage-consent/#how4 (accessed 26th November 2019)

[2] Kantara Initiative, "Consent Receipt Specification", online: https://kantarainitiative.org/confluence/display/infosharing/Consent+Receipt+Specification (accessed 27th November 2019)

[3] European Parliament and of the council, EU Regulation 2016/679, "General Data Protection Regulation", 27th April 2016

[4] Nancy S. Kim, "Consentability: Consent and Its Limits", *Cambridge University Press*, February 2019

[5] J. A. Obar, "The biggest lie on the internet", *SSRN Electronic Journal*, January 2016

[6] S. Esayas, T. Mahler, K. McGillivray, "Is a Picture Worth a Thousand Terms? Visualising Contract Terms and Data Protection Requirements for Cloud Computing Users", *In: Casteleyn S., Dolog P., Pautasso C. (eds), Current Trends in Web Engineering. ICWE 2016. Lecture Notes in Computer Science*, vol. 9881, Springer

[7] K. Ghazinour, M. Majedi, K. Barker, "A model for privacy policy visualization", *33rd Annual IEEE International Computer Software and Applications Conference (COMPSAC'09)*, vol. 2, 2009

[8] K. Ghazinour, T. Albalawi, "A Usability Study on the Privacy Policy Visualization Model," *14th IEEE Intl Conf on Dependable, Autonomic and Secure Computing (DASC)*, Auckland, 2016

[9] M. Tabassum, A. Alqhatani, M. Aldossari, H. R. Lipford., "Increasing User Attention with a Comic-based Policy", *In Proceedings of the 2018 CHI*

Conference on Human Factors in Computing Systems (CHI '18), ACM, New York, NY, USA

[10] O. Drozd, S. Kirrane, "I Agree: Customize Your Personal Data Processing with the CoRe User Interface", *16th Intl Conf on Trust, Privacy and Security in Digital Business (TrustBus)*, Linz, Austria, 2019

[11] N. Sadeh et al., "The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About", *Tech. report CMU-ISR-13-119*, December 2013

[12] E. Maler, "Extending the Power of Consent with User-Managed Access: A Standard Architecture for Asynchronous, Centralizable, Internet-Scalable Consent", *IEEE Security and Privacy Workshops*, San Jose, CA, 2015

[13] H. J. Pandit, M. Lizar, "OPN: Open Notice Receipt Schema.", *SEMANTICS*, Germany, September 2019

[14] M. Palmirani, M. Martoni, A. Rossi, C. Bartolini, L. Robaldo, "PrOnto: Privacy Ontology for Legal Reasoning", *7th International Conference EGOVIS*, Germany, September 2018

[15] W.B. Tesfay, et al., "I Read but Don't Agree: Privacy Policy Benchmarking using Machine Learning and the EU GDPR", *In Companion Proceedings of the Intl World Wide Web Conference 2018 (WWW '18)*, Republic and Canton of Geneva, CHE, pp. 163–166

[16] M.C. Mont, V. Sharma, S. Pearson, "EnCoRe: Dynamic Consent, Policy Enforcement and Accountable Information", HP Laboratories, report HPL-2012-36

[17] E.A. Whitley, N. Kanellopoulou, "Privacy and informed consent in online interactions: Evidence from expert focus groups.", *International Conference on Information Systems (ICIS)*, Association for Information Systems, 2012

[18] M.C. Mont, S. Pearson, P. Bramhall, "Towards accountable management of identity and privacy: sticky policies and enforceable tracing services", *14th International Workshop on Database and Expert Systems Applications*, Prague, Czech Republic, 2003

[19] J. Kaye, at al., "Dynamic consent: a patient interface for twenty-first century research networks", *Eur J Hum Genet*, vol. 23, pp. 141–146, 2015

[20] M.R. Asghar, et al., "A Review of Privacy and Consent Management in Healthcare: A Focus on Emerging Data Sources", *518-522. 10.1109/eScience.2017.84*

[21] Y. Bo, D. Wijesekera, P.C.G. Costa, "Informed Consent in Electronic Medical Record Systems", *Healthcare Ethics and Training: Concepts, Methodologies, Tools, and Applications*, IGI Global, p1029-1049, 2017

[22] B. M. Welch, et al., "Teleconsent: A novel approach to obtain informed consent for research", *Contemporary Clinical Trials Communications*, vol. 3, pp. 74-79, 2016

[23] I. Budin-Ljosne, H.J.A. Teare, J. Kaye, "Dynamic Consent: a potential solution to some of the challenges of modern biomedical research", *BMC Med Ethics*, vol. 18, issue 4, 2017

[24] M.R. Ulbricht, F. Pallas, "YaPPL - A Lightweight Privacy Preference Language for Legally Sufficient and Automated Consent Provision in IoT Scenarios", *In: Garcia-Alfaro J., Herrera-Joancomarti J., Livraga G., Rios R. (eds) Data Privacy Management, Cryptocurrencies and Blockchain Technology. DPM 2018, CBT 2018. Lecture Notes in Computer Science*, vol. 11025, Springer

[25] K. Fatema, et al., "Compliance through Informed Consent: Semantic Based Consent Permission and Data Management Model", *Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology*, PrivOn, 2017

[26] C. Bartolini, R. Muthuri, C. Santos, "Using Ontologies to Model Data Protection Requirements in Workflows", *In Otake M., Kurahashi S., Ota Y., Satoh K., Bekki D. (eds) New Frontiers in Artificial Intelligence, JSAI-isAI 2015. Lecture Notes in Computer Science*, vol. 10091, Springer

[27] S. Kirrane, et al., "A Scalable Consent, Transparency and Compliance Architecture", *The Semantic Web, ESWC 2018 Satellite Events*, Springer

[28] Data Privacy Vocabulary, online: https://www.w3.org/ns/dpv, W3C (accessed 16th December 2019)

[29] The OPERANDO project. online: http://www.operando.eu (accessed 16th December 2019)

[30] M. Ulbricht, F. Pallas, "CoMaFeDS: Consent Management for Federated Data Sources," *IEEE International Conference on Cloud Engineering Workshop (IC2EW)*, Berlin, 2016

[31] E. Politou, E. Alepis, C. Patsakis, "Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions", *Journal of Cybersecurity*, vol. 4, issue 1, 2018

[32] S. Cha, M. Chuang, K. Yeh, Z. Huang, C. Su, "A User-Friendly Privacy Framework for Users to Achieve Consents With Nearby BLE Devices", *in IEEE Access*, vol. 6, pp. 20779-20787, 2018

[33] R. Neisse et al, "An agent-based framework for Informed Consent in the internet of things", *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, Milan, 2015

[34] V. Morel, M. Cunche, D. Le Metayer, "A Generic Information and Consent Framework for the IoT", *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, Rotorua, New Zealand, 2019

[35] A. Das, M. Degeling, D. Smullen, N. Sadeh, "Personalized Privacy Assistants for the Internet of Things: Providing Users with Notice and Choice", *in IEEE Pervasive Computing*, vol. 17, no. 03, pp. 35-46, 2018

[36] R. Akalu, "Privacy, consent and vehicular ad hoc networks (VANETs)", *Computer Law & Security Review*, vol. 34, issue 1, 2018, Pages 37-46

[37] M. Laurent et al, "Authenticated and Privacy-Preserving Consent Management in the Internet of Things", *The 10th International Conference on Ambient Systems, Networks and Technologies (ANT)*, April 29 - May 2, 2019, Leuven, Belgium

[38] SMOOTH Project, online: https://smoothplatform.eu (accessed 16th December 2019)

[39] L. Piras, et al, "DEFeND Architecture: a Privacy by Design Platform for GDPR Compliance", *16th Intl Conf on Trust, Privacy and Security in Digital Business (TrustBus)*, Linz, Austria, 2019

[40] G. Lioudakis et al, "Facilitating GDPR compliance: the H2020 BPR4GDPR approach", *1st Workshop on Trust and Privacy Aspects of Smart Information Environments (TPSIE)*, September 2019, Trondheim, Norway

[41] S. Ziegler, I. Chochliouros, "Privacy Flag - collective privacy protection scheme based on structured distributed risk assessment", *IEEE 2nd World Forum on Internet of Things (WF-IoT)*, Milan, 2015

[42] The TYPES project, online: http://www.types-project.eu (accessed 16th December 2019)

[43] A. Bagnato, "The POSEID-ON Project", *Workshop on Privacy Challenges in Public and Private Organizations, IFIP Summer School on Privacy and Identity Management*, August 2019, Brugg, Switzerland

[44] V. Diamantopoulou, M. Pavlidis, "Visual Privacy Management in User Centric Open Environments", *11th International Conference on Research Challenges in Information Science (RCIS)*, Brighton, 2017

[45] V. Diamantopoulou, et al., "Supporting Privacy by Design Using Privacy Process Patterns", *In: De Capitani di Vimercati S., Martinelli F. (eds) ICT Systems Security and Privacy Protection. SEC 2017. IFIP Advances in Information and Communication Technology*, vol. 502, Springer

[46] P. Bonatti, S. Kirrane, A. Polleres, R. Wenning, "Transparent Personal Data Processing: The Road Ahead", *In: Tonetta S., Schoitsch E., Bitsch F. (eds) Computer Safety, Reliability, and Security. SAFECOMP 2017. Lecture Notes in Computer Science*, vol. 10489, Springer

[47] M. Shah, et al., "CrowdMed: A Blockchain-Based Approach to Consent Management for Health Data Sharing", *In: Chen H., Zeng D., Yan X., Xing C. (eds) Smart Health. ICSH 2019. Lecture Notes in Computer Science*, vol. 11924, Springer

[48] K. Bhaskaran et al., "Double-Blind Consent-Driven Data Sharing on Blockchain", *2018 IEEE International Conference on Cloud Engineering (IC2E)*, Orlando, FL, 2018

[49] R.R. Agarwal, et al., "Consentio: Managing Consent to Data Access using Permissioned Blockchains", *arXiv:1910.07110 [cs.DC]*

[50] K. Rantos, et al., "ADvoCATE: A Consent Management Platform for Personal Data Processing in the IoT Using Blockchain Technology", *In: Lanet JL., Toma C. (eds) Innovative Security Solutions for Information Technology and Communications. SECITC 2018. Lecture Notes in Computer Science*, vol. 11359, Springer

[51] J. Onieva, J. Zhou, J. Lopez, "Multi-party non-repudiation: A survey", *ACM Computing Surveys*, vol. 41, no. 1, December 2008

[52] V. Jesus, S. Mustare, "I Did Not Accept That: Demonstrating Consent in Online Collection of Personal Data", *16th Intl Conf on Trust, Privacy and Security in Digital Business (TrustBus)*, Linz, Austria, 2019

VITOR JESUS is a Lecturer with Birmingham City University and has 20 years of professional experience, split between Industry and Academia. He holds a BSc in Physics, a MSc and PhD in Computer Science and industry certifications in Cyber Security and Data Privacy. He has held positions with different companies, from start-ups to large organisations. He has authored a number of papers, has been in the review panel of several conferences, was a visiting scholar in different institutions, has worked in a number of international projects and an active contributor to standards and best-practices organisations in Cyber Security and Privacy.

His current research and teaching interests are in Trust problems such as those involved in CyberSecurity and Privacy. We live in a world where CyberSecurity is becoming a basic need just like physical safety is; or its absence can bring down a business, a community or a country. We are also moving towards a world where there is no Privacy by default and everyone's Identity and Data will eventually be stolen and used without permission. His research aims at bringing control back to users by looking at technologies and solutions, such as Blockchains, Secure Networks or Artificial Intelligence, whether it is in Cars, the Internet, Medical devices, Enterprises, Factories or Cities.