Identity Agents

Paul Trevithick and Sergey Kucherenko, The Mee Foundation

March 3, 2023. Revised December 5, 2025

Abstract

The internet has evolved to exhibit a power asymmetry between organizations and individuals–an asymmetry that comes at the expense of the autonomy, agency, and privacy of the individual. This asymmetry arises because individuals lack (i) the technical means to create their own digital identities independent of businesses or governments, (ii) practical and convenient means to control their own data, and (iii) effective legal protections to prevent the disclosure of their information to third parties. We present design considerations for, and the architecture of, *identity agents* whose goal is to restore the power balance for individuals.

1 Imbalance

The power imbalance between internet technology users and service providers (businesses and governments) has been recognized for some time. It was described over a decade ago by the World Economic Forum[8]:

> An asymmetry of power exists today between institutions and individuals—created by an imbalance in the amount of information about individuals held by, or that is accessible to, industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information.

We argue that this asymmetry arises because individuals lack (i) the technical means to create their own digital identities independent of businesses or governments, (ii) practical and convenient means to control their own data, and (iii) effective legal protections to prevent the disclosure of their information by service providers to third parties.

Before we discuss solutions that could restore power to individual, we will provide an overview of how personal data is managed and handled on the internet today.

Digital service providers (businesses, organizations of any kind, and governments) interact with their users via websites, local and mobile apps, chatbots that we will refer to simply as *apps*. We will also include in this term another person's identity agent which appears to the first person as an app.

App process personal data three main ways: (i) data related to user interactions is collected and stored in user accounts, (ii) third-party adtech systems collect user data for advertising on these apps, and (iii) third-party payments systems collect and process the user's transactional data.

As a user interacts with an app/site, whatever they type, click, enter, upload is (or can be) collected by the service provider's app. Observations, e.g. the kinds of things they click on, and spend time on, may also be collected. These data include what is called first-party data[1] by commercial service providers as well as data the service provider has acquired from third parties (e.g. data brokers).

Users have very limited control over what is collected and how it is used. At best apps provide a means to review and update selected portions of it via an online form–often in the user's profile. In some cases the app allows the user to download a copy of the data collected about them, although doing so is time-consuming, labor-intensive, and produces dozens of files that the user probably don't know how to use. In some jurisdictions (e.g. GDPR in the EU) the user has the right to rectify and/or erase their data. Unfortunately, in practice these rights remain almost entirely formal and theoretical due to the unmanageable burden placed on the user to exercise them.

The app provider may share collected data (sometimes partially anonymizing it first) with third parties. They may sell it to data brokers[2] who buy data from a variety of sources and then resell it to other aggregators and providers.

Businesses earn ad revenue by implementing tracking using in-app technologies apps (e.g. third-party cookies, transparent pixels, fingerprinting, etc.) and monetize it through integrations with the adtech ecosystem.

Tracking data is behavioral data used to infer traits about the individual (e.g. age-range, income level, and many of other demographic and psychographic traits). Advertisers pay to get their messages (ads, images, videos, text, etc.) in front of cohorts with shared traits (called "audiences") irrespective of which app a member of that cohort is using. Apps sell ad inventory (i.e. ad "slots") to these advertisers. Although some are sold directly, most are sold via ad networks and ad exchanges that take part in a high volume, high-speed real-time auction process called real-time bidding[3]. A complex ecosystem of thousands of adtech firms is involved in the supply chain stretching from advertisers, through ad exchanges, to the apps acting in the role of publishers.

Apps that sell products or services leverage payment gateways that allow the app provider to receive funds from the user (e.g. via a payment card). In most cases this involves

[1] www.salesforce.com/ap/blog/first-party-customer-data/

[2] theconversation.com/its-time-for-third-party-data-brokers-to-emerge-from-the-shadows-94298

[3] en.wikipedia.org/wiki/Real-time_bidding

sending financial data (including identifiers) about the user through financial systems run by banks, credit card associations, and their service providers.

In addition to the privacy risks associated with the flow of payment transactions, some app providers also earn money by selling purchase information to data brokers.

In the data flows just mentioned, the user is relatively powerless over their data. They have little or no visibility into what's being gathered, where it's being shared and how it's used. Users live in what Alicia Solow-Niederman calls an "inference economy"[22] wherein big data and machine learning are used to infer traits that form new kinds of personal information–often more sensitive than the underlying source data. Harm and risk can rarely be evaluated outside a specific situation[21], yet it is useful to list a few representative types of harm. Individuals, are vulnerable to data breaches, they can be spammed by marketers, they are vulnerable to identity theft, they can be exposed to price and/or hiring discrimination, and they can be stalked.

"Debates over privacy are really debates about how power will be allocated in an information society and how much power the humans in that society will get as consumers or citizens."[17] Today, despite significant new regulation, the basic approach to protect privacy hasn't changed since the 1970s. It is often called *notice and consent*. Solove described it using the term *privacy-self management*, as follows:

> [T]he law provides people with a set of rights to enable them to make decisions about how to manage their data. These rights consist primarily of rights to notice, access, and consent regarding the collection, use, and disclosure of personal data. The goal of this bundle of rights is to provide people with control over their personal data, and through this control people can decide for themselves how to weigh the costs and benefits of the collection, use, or disclosure of their information.[20]

Although well-intended, and necessary, *notice and consent* does not provide people with meaningful control over their data.

The U.S. population consistently misunderstands the meaning of the term privacy policy.[4] A majority of Americans believe incorrectly the mere presence of a privacy policy indicates a website will not share information without permission.[4] The problem is well summarized as follows:

> When presented with click-through consent, privacy policies or terms of use statements, most people reflexively select "I agree". An extensive body of aca-

[4] "Privacy policies have been widely adopted and are now commonplace. This kind of transparency is good in theory, but less so in practice since it places the onus of privacy on end users. In general, attempts to improve privacy by helping end users have not worked, since most people don't have the time, expertise, or desire to deal with all the nuances of privacy."[9]

> demic research specifically on privacy and data collection notices demonstrates that members of the public don't read them and might not understand them if they did and that many misinterpret their purpose, assuming that the existence of a privacy policy displayed by way of notice means that the entity collecting the data offers a level of data protection when, in fact, privacy notices do not guarantee privacy. Since the terms offered are typically "take it or leave it", to decline often results in being denied the product or service one seeks, creating a disincentive for consumers to do anything other than accept the terms.[5]

"We agree to all these 'privacy notices' so we must have privacy, right? Notice and choice is thus an elaborate trap, and we're all caught in it."[17]

The most substantive lever for progress has been legislation such as GDPR and CPRA, along with regulatory fines by organizations like the FTC.

In a growing number of jurisdictions, including Europe under GDPR[5] and California under CPRA[6], the person's *data rights*, (e.g. the right to access, rectify and erase their data), are clearly described. In practice, the time and effort required to exercise these rights at each app individually is enormous. The individual must, for example, send written requests to get copies of their data, to have it updated, or to have it deleted (as in the so-called right to be forgotten). Furthermore, even if the individual makes these requests, they have no way of knowing if these requests have been implemented. Until these processes are automated by personal agents, these rights don't meaningfully exist.

Society agrees to supervise the places children inhabit, protect them from environments they should not encounter, and regulate the products they use. As a result, businesses are not permitted to sell tobacco, alcohol, pornography, handguns, certain kinds of fireworks, and other products and services to minors. However, none of this is true online. In the virtual world children are largely unprotected despite being exposed to wide range of potential harms.

Many approaches have been proposed and tried without much success. Existing laws have proven to be insufficient, and industry self-regulation has largely failed. Today there is a renewed global push to protect children's safety through stronger laws and regulations. Although some use other approaches[7], many mandate age verification.[7][10] However, privacy advocates and others have shown that many of the mechanisms for verifying age online weaken anonymity and privacy.[18]

[5] gdpr-info.eu/

[6] thecpra.org/

[7] Such as requiring online services that are likely to be used by young people to default to the highest privacy setting possible for minors, as mandated by California's Age-Appropriate Design Code Act.

1.1 Autonomy

In real life we each have a self that embodies our unique individuality. We "bring" that independent selfness to our interactions with others. However, online "we have no *digital embodiment*."[8] Our identifiers and their associated account data are provided to us by online service providers (e.g. in the form of a Facebook or an Amazon account) and without them, we don't exist. We can't "bring" them anywhere. Anyone who has been banned from a platform, or uses a platform that has been shut down, is sharply reminded that their account and its data exists at the pleasure of that platform. We believe that each of us has an inalienable right to a digital identity that we create and control and that neither a business nor a government can delete, revoke or withdraw.

In theory to autonomously *own* our own digital identity and personal data doesn't require physical possession of it because through technical and legal mechanisms control over it can be achieved irrespective of where the data is stored and by whom. In practice, however, possession tends to shift power to the possessor. Since with few exceptions our personal information is possessed by service provider's apps, power shifts to them. This pattern of what could be called *app-held data* by the *first-parties* we interact with is so common that it's hard to imagine an alternative. Beyond first-parties, our data is also collected and held by *third-parties* (e.g. data brokers) with whom we have no direct interaction. In short, as Johannes Ernst has put it, "everybody has our data ... except us."[9]. Giving individuals possession of their data doesn't mean that it doesn't also exist in many other places, but what it does mean is that *at least* they have it as well.

We lack the ability to communicate (e.g. chat) directly from one person to another without requiring that all parties have accounts on some shared server. With rare exceptions[10], we don't have the ability to do so *peer-to-peer*–i.e. from one person's device to the other person's device. Instead, we're dependent on servers hosted by intermediaries. Further, whereas it is standard practice that the content of messages is encrypted end-to-end, the *metadata* about this content (e.g. who a person communicates with, from where, at what time, how often and from which device, etc.) is in many cases visible to the intermediary server.

1.2 Agency

Individuals lack computational power "on their side" to enable them to easily manage, control and protect their personal information being shared with and collected by apps.

Privacy laws such as the GDPR provide the individual (data subject) formal rights over their personal information regardless of where it is stored. These include the right to

[8] Phil Windley, personal communication, September 2022

[9] reb00ted.org/personaldata/20210620-who-has-my-personal-data/

[10] berty.tech

rectification, access and erasure. Unfortunately, in practice, these rights are largely not actionable because the burden required to exercise them using the provider-side mechanisms is extreme. To be actionable the service provider apps would have to implement APIs on their side, and the individual would have to have software agents to consume these APIs on theirs.

Unfortunately, an individual's account identifiers and associated human data are bound to specific online service providers and can't be moved freely from one to another. In other words they are not *portable*.[11]

In many jurisdictions service providers are required by law to provide individuals with access to their data, but they usually offer this by means of a set of files emailed to the individual as an attachment several hours or days after the request. There are significant problems with implementing portability in this manner. First, it is tedious, manual and slow. Service providers don't support data "export" APIs, so an individual can't use technology to automate the process. Second, the individual ends up with dozens of sets of files (one set from each provider) that are not largely unintelligible to them.

Beyond access and export problems, providers generally don't provide "import" APIs to allow the individual to upload their data. Even if an individual could import their data, it first must be transformed into the format of the recipient, since each provider uses their own format. The result is a lack of portability.

Advocacy groups, including the EFF, are pushing for interoperability as an antidote to corporate concentration. This is good, but they should insist that apps implement import/export APIs that can be leveraged by agents such as identity agents. "A new regime of interoperability can revitalize competition in the space, encourage innovation, and give users more agency over their data..."[2]

2 Related work

Many initiatives seek to address various subsets of the power imbalance we've described. We mention a few of them here.

The lock-in and lack of data portability and interoperability between service providers is being fought using both policy and technical means[3][2].

Work on re-decentralizing the internet includes: redecentralize.org[12], nlnet.nl[13], DWeb

[11] Efforts such as the Data Transfer Institute (https://dtinit.org) are relevant here.
[12] redecentralize.org
[13] nlnet.nl

principles[14], The Web3 Foundation[15], the Decentralized Identity Foundation(DIF)[16], "local-first"[12] software principles[17], ProjectVRM[18], Blue Sky[19], and Berners-Lee's Decentralized Information Group[20].

See also work on *personal agents* [21]–software tools that work (i.e. provide agency and power) "on the individual's side"[22] for, and *exclusively* on behalf of, the person. Personal datastores[23] and the *self-sovereign identity*[16] movement are squarely aimed at addressing our lack of autonomy.

3 Design considerations

In this section, we discuss design considerations for identity agents that promise to restore the power imbalance individuals experience on the internet. We discuss requirements related to ensuring that identity agents enforce the user's privacy data rights, shift control to the individual and are trustworthy.

3.1 Data rights

In a growing number of jurisdictions privacy regulations describe *data rights* to access, correct and delete the personal information about users that is managed by service provider's apps. In practice the user burden of exercising these rights across hundreds of apps is unmanageable without automation on the user's side. Although identity agents provide this automation, this is only half the answer. The other half involves constraining how service providers use the user's data, and requiring them to implement APIs that the identity agents can consume on the user's behalf.

3.2 Convenience

Identity agents must make the individual's life easier. To do so they must *automate* the burdensome tasks related to controlling and managing a person's information and not introduce new friction and effort.

[14] getdweb.net/principles/
[15] web3.foundation/
[16] identity.foundation
[17] inkandswitch.com/local-first/
[18] projectvrm.org/
[19] blueskyweb.xyz/
[20] dig.csail.mit.edu
[21] What Mozilla calls a *user agent* (https://developer.mozilla.org/en-US/docs/Glossary/User_agent)
[22] Project VRM[19] refers to this as "tools for individuals to manage relationships with organizations" to which we would add "...or with other individuals."
[23] Examples of open-source personal datastores include `https://solidproject.org` Decentralized Web Nodes(DWN) is relevant here. For more about personal datastores see `https://wikipedia.org/wiki/Personal_data_service`

3.3 Inclusivity

Identity agents must be affordable by all socio-economic classes, not just those better off. For this reason, solutions that incur monthly hosting costs to the user are disqualified at least for a useful baseline level of functionality. Agents should be available at no charge and run on devices the user already owns, although admittedly there may be additional costs incurred if the user wishes to store extremely large datasets on their device.

3.4 Loyalty to the user

Much of the power asymmetry described in the first section is created by the economic incentives online service providers have to collect and monetize user's personal information. Providers are loyal to their *customers* (e.g. an advertiser) not *users* (individuals). Thus, they do just enough in the user's interest to ensure that users continue to use their services, while collecting and monetizing as much of their user's data as possible.

For a user to trust that their agent works *exclusively* on their behalf, the agent *provider* (i.e. organization that develops and provides it) must not have an economic incentive to be disloyal to the user. This can be ensured by designing identity agents such that the personal information that they process is never accessible to the identity agent provider. Doing so largely eliminates the need to trust the agent provider organization, their security infrastructure, and their processes.

3.5 Open source

The transparency of open-source software can increase the confidence that an identity agent built from this source code is trustworthy. The source code is visible to all and can be reviewed and audited to ensure that the agent is secure, works as expected, and truly works in the person's interest.

3.6 Trustworthiness

Users require the developer organization behind their agent is trustworthy. To achieve this, the organization's financial incentives should be aligned with the user's interests. Providers using nonprofit or similar organizational forms have the important benefit that they have no financial incentive to exploit the user's data.

3.7 Trust Framework

Once data is shared from an agent to a service provider no purely technical means exists to constrain what the provider can do with it. Technical means, for example, can't prevent them from selling it to or sharing it with others. Instead, legal means called trust frameworks, can be employed.

Some providers may be willing to join a trust framework and *license* the user's information received from the user's identity agent. If so, they would agree to the terms of a license agreement which include terms that respect the user's privacy rights. This contract is signed by a trusted organization[24] that represents the community of identity agent users thereby making the processes effortless for individual users. This organization is responsible for enforcement of the contract's terms, again, on behalf of the user. A related approach is called MyTerms[25].

3.8 Human-centricity

The internet is *provider-centric* rather than *human-centric*. The internet includes millions of providers, each offering their own apps. In this provider-centric model each app sees a thin slice of the individual through their direct interactions with them. The burden of managing personal data across these apps falls to the individual.

For the individual the situation is reversed. The user sits at the center of a hub with connections to apps (and relying parties) radiating outwards from them. Even for a single app there is considerable burden for the person to enter and update personal information, payment details, and preferences, and review privacy policies, and set cookie preferences, and so on. Multiplied by often as many as one hundred or more connections, the resulting burden is unbearable.

Tools to manage these chores must sit on the user's side, and work on their behalf across all of their interactions. Technologies of this kind, that empower a person across multiple apps (relying parties), e.g. browsers and password managers, are called are *user-agents* as they act as agents of the user.

3.9 Metacontextuality

Zuckerberg once said that "[h]aving two identities for yourself is an example of a lack of integrity"[11]. However, even if one could force all users of a given platform (e.g. Facebook) to have a single identity[26], this approach is clearly unworkable for a solution that represents the person across multiple, widely varying systems and contexts. People need the freedom to be themselves–selves that are complicated and messy. Our identities vary depending on whom we are interacting. We choose to express different parts of ourselves within different contexts. Not only are the attributes we share different, but the values of one attribute may be different in different contexts.

[24]These kinds of organizations have been variously described in the literature as "data unions," "data coalitions," "Mediators of Individual Data" (MIDs) by Lanier et al.[13], etc.

[25]doc.searls.com/myterms/

[26]Note: *identity* is term we prefer to avoid due to its semantic ambiguity, but this is the word he used.

> "[A]t various times in the same day, virtually every adult can be a friend, a worker, a supervisor, a citizen, a mentor, a student, a musician, a customer, a lover, a child and a parent. Each of these roles demands different behavior and different aspects of our selves, aspects that need not be consistent. We behave, for example, in different ways with loved ones than with those we encounter in commercial or professional settings. Even among our loved ones, we behave very differently (and often show very different sides of ourselves) to our children, our parents, and our sexual partners. But this is not dishonest, nor is it inconsistent. At the very least, it's no more inconsistent than is the complicated nature of having a self. It is human."[17, p122]

Let's look at a person's age as an example. We see that across contexts they might share, their exact chronological age among their close friends, a fictional age to a music recommendation service, no age at all in contexts wherein doing so might cause discrimination against them, or a merely a statement that they exceed the minimum legal drinking age.

In his last public speech[27] Kim Cameron[28] introduced two useful definitions based on archaic English:

- **Selfness**: The sameness of a person or thing at all times or in all circumstances. The condition of being a single individual. The fact that a person or thing is itself and not something else. Individuality, personality.
- **Whoness**: A distinct impression of a single person or thing presented to or perceived by others. A set of characteristics or a description that distinguishes a person or thing from others.

Figure 1 illustrates these concepts and introduces the notion of context.

Using these terms we can say that in everyday life people have one *selfness*, but they have many, context-dependent *whonesses*. Any solution must be meta-contextual–it must embrace and support the complicated, multi-contextual nature of our lives.

3.10 Local-first

In this section we explain the motivation for a local-first architecture.

A user's personal datastore may be on-device or in the cloud. By *on-device* we mean that the individual's datastore and processing is on their own phones, laptops, and/or home servers. By *cloud-based* we mean that the person's datastore and processing lives in the cloud (e.g. on a SOLID[29] pod). The local-first software[30] principles are highly relevant.

[27] www.youtube.com/watch?v=9DExNTY3QAk
[28] en.wikipedia.org/wiki/Kim_Cameron_(computer_scientist)
[29] solidproject.org
[30] www.inkandswitch.com/local-first/

Figure 1: Multiple whoness-contexts around a single selfness

Although there are other points of view, we contend that as long as a relatively large number of people are using the solution, having a personal datastore on-device is more secure than one in the cloud. Even if these alternatives were equivalently secure for a single person, a cloud-based architecture by nature aggregates large numbers of personal datastores at one cloud service provider location, and thereby creates a proportionately larger economic incentive for hackers.

3.10.1 Synchronization

A user may have two or more identity agents running on multiple devices each of which is only intermittently connected to the internet. The user's data needs to be kept consistent across these identity agents and devices, at least eventually. This requires that these identity agents implement data replication and syncing between themselves in a peer-to-peer (P2P) fashion. Unfortunately, pure P2P internet communication between agents remains an unsolved technical challenge and thus intermediate "relay" servers are sometimes required.

These relays have their own privacy, autonomy, trust, and security considerations. We mention a few of them here. Technically competent individuals can host their own relay servers, presuming it is relatively straightforward to build and deploy. Everyone else will

have to rely on an external administrative hosting authority. Ideally they will be available freely or at low cost. In either the self-hosted or external case, the relay needs to be trusted. For this reason its source code should be open. The relay should only handle encrypted data transiently. That is, it should temporarily buffer (encrypted) message data while waiting for the recipient app or agent to come back online.

3.10.2 Backup and recovery

One disadvantage of noncustodial, on-device architectures, as opposed to more conventional cloud-based architectures, is that they are vulnerability to data loss if all the user's devices are lost or damaged simultaneously, and no backups exist. For users with more than one device this kind of catastrophic loss is less likely since data is replicated across their devices and a repaired or replaced device's data can be restored from one of the user's other devices.

Since we envision that a person would use an agent for their entire life, agents must implement backup/restore approaches that would provide recovery from even the worst case disaster scenario. The identity agent's data must be backed up in secure remote storage location(s) and encrypted. The user's private key(s) must also be recoverable. Approaches combining sharding, shared secrets[31], and social recovery have been proposed, although this remains an area of active research.

3.11 Guardianship and delegation

In addition to managing their own personal data, identity agents, need to enable individuals to act as guardians for others (e.g. minors, the elderly and other vulnerable people). To implement this, agents must include the ability for a user to delegate access to portions of their personal data to another user.

4 Identity agents

Identity agents promise to address the power imbalance described in the first section and take into account the design considerations in the second section. Through a combination of technical and legal mechanisms an identity agent gives individuals control over their personal information as they interact with websites, mobile apps, as well as other people's identity agents. The solution combines a legal contract with a trusted, personal agent[32] integrated with a traditional digital wallet[6].

An identity agent is a native software application (e.g., written in Swift on iOS, Kotlin on Android, etc.) that runs on a user's devices (e.g., mobile phone, laptop, etc.). It maintains

[31] en.wikipedia.org/wiki/Shamir%27s_secret_sharing

[32] Similar ideas have been proposed by others. See *personal user agents*, in [5, p24]

a local, private datastore of the user's personal information.

Apps can request personal information from, and provide information to, the user's identity agent using the PDN protocols (see section 4.1). The information may flow between the app and the identity agent in one direction, the other, or in both. The app reads and writes data using the identity agent's data model. This data may include both structured and unstructured data and may include digitally signed documents (e.g. Verifiable Credentials, etc.).

4.1 Private Data Network

The Private Data Network (PDN) is a set of protocols designed to support data sharing between PDN nodes. These nodes may be integrated with a provider's apps or websites, and/or within the user's identity agent if they choose to install and configure one or more of them on their devices.

The PDN includes a trust framework wherein apps authorized by The Mee Foundation implement PDN protocols and to agree to the terms of the PDN License. This license requires that apps abide by certain privacy principles regarding how they handle the individual's data (e.g. requiring explicit consent for collection, processing, storage and sharing of the person's data) as well as implement the PDN protocols. These protocols, combined with the license, enable *private sharing* between the identity agent and the app or site.

Private sharing allows the user to share personal information with confidence that it remains under their control. Following intellectual property law precedents, the user licenses their information to the app rather than transferring a copy of it, and then hoping that the app will treat it with care.

The PDN is described in more detail in section 6.

4.2 Benefits for the individual

4.2.1 Privacy

When an identity agent interacts with apps that are part of the PDN, these apps agree to process the user's personal information under the terms of the PDN License. By default, the app can't sell, transfer or share the user's information without their consent.

If the identity agent includes a browser extension component, it can add the Global Privacy Control[33] field in the HTTP header expressing the user's privacy preference. The extension may also include the ability to delete third-party cookies and other kinds of trackers used in surveillance advertising. Identity agents can participate in new, *private advertising*

[33] globalprivacycontrol.org

networks, that don't rely on cookies, trackers, data brokers, etc. but instead rely on user profiles that are anonymized, and never shared outside the ad network.

4.2.2 Protection of Minors

Minors can be given a special child-oriented identity agent by their guardians which provides the minor with an age-appropriate experience online. The guardian would register their minors on a third-party age verification service. This service would issue an age verification credential which is stored in the minor's agent (and its replicas on other devices if they exist). When the minor uses an app, the agent signals that the minor wishes to have an age appropriate experience.[34] In response, the app requests the age verification credential from the minor's identity agent and adapt its experience accordingly.

4.2.3 Autonomy

Identity agents provide individuals with digital embodiment of themselves. Over time, the identity agent develops rich, context-specific data profiles about them. This embodiment can move autonomously under the user's control between first-party apps. A *local* identity agent can do so independent of any external administrative authority.

Identity agents reduce the user being locked-in to provider apps/sites by supporting data portability. Using the PDN APIs of provider apps, the agent allows the user to automatically retrieve their personal information from one app, and share it with another.

As the usage of agents grows, surveillance free, end-to-end encrypted, peer-to-peer communications can interconnect these users to allow messaging and data sharing. These communications can be designed with minimal reliance on cloud-based relay servers which are often needed to buffer messages to endpoints that are temporarily offline.

4.2.4 Agency

A foundation for Personal AI

Rather than requiring individuals to trust a shared AI-in-the-cloud service with all of their sensitive personal information, a better approach is to have the *Personal AI* algorithms run on the individual's devices. These algorithms read and write personal information to/from the person's identity agent.[35]

Wielding credentials

In real life an individual can, say, present their driver's license to a wine seller to prove that they are of drinking age because the wine seller trusts the license issuer. This interaction is

[34] This signaling is used in the https://ageprotect.org proposal

[35] Iron Man's J.A.R.V.I.S.[36] is an example of this architecture.

privacy-respecting because the presentation interaction is never disclosed to the issuer. This driver's license use-case involves the individual *wielding* a trust credential. Unfortunately, there is no commonplace way to do this online. There's no standard way to be issued a credential, hold it in a digital agent (acting as a digital wallet), and then present it to another party. With a few, domain-specific exceptions (e.g. cryptocurrency), there is no standard online method for an individual to prove something one party states about them, to another party. Digital wallets are emerging to meet this need and this wallet-like capability is included in an identity agent.

Automated data presentation

Apps rely on form filling and other kinds of tedious, manual data entry because individuals lack the ability to *digitally* present personal information about themselves. Individuals must manually re-enter personal information into each app, endlessly repeating themselves. They lack an agent that can automatically present information on their behalf.[37]

This endless repetition is a symptom of the internet's silo-ed architecture wherein each app maintains its own database of personal information. The individual has the hassle of repeated data entry, and the app offers a less-than-optimal user experience. Armed with an identity agent, the user is no longer required to repeat themselves as they move from app to app.

This inability to present ourselves digitally is a contributing factor to the concentration of corporate power on the internet. For example, it's simply easier to buy something from Amazon because so many of us have already entered so much information to them. We have a preferential attachment to Amazon that goes beyond their intrinsic advantages. Continuing with the shopping example, identity agents can represent an individual to any e-commerce website, and thereby provide the same Amazon-like, frictionless user experience that can mitigate corporate concentration and "natural" monopolies.

Password-less login

Identity agents enable the user to log in to apps using a variety of password-less authentication technologies. The identity agent knows who the user is because the user authenticates to it, so the identity agent can represent the user in their interactions with apps, and can do so without revealing correlatable identifiers. This is both private and convenient.

Infer and present ad profiles

An identity agent can generate on-device an ad profile by inferring traits from the user's browsing behavior. The user can review and edit this profile, and may choose to share it with apps that are supported by interest-based *private* advertising technology. This

[37] The credential presentation interaction just mentioned is another example of this.

approach eliminates the need for surveillance by third-parties using cookies and other tracking technologies. It is similar in design to Google's Topics API[38].

Delegation

In the offline world one entity can grant access to some resource to another entity. For example, an individual can give their car keys to a friend, so they can borrow their car. At present, there is no standardized way to do this online. This is especially problematic in healthcare scenarios where a healthcare provider needs access to health-related data about a patient, but the patient is not in a situation where they are able to provide it by themselves and must instead rely on someone else, e.g. a family member to grant the needed permission. In the online world each service provider not only possesses the individual's data, but they manage it in such a way that it is impossible for the individual to delegate rights to it to others.

Content filtering

Social networking platforms have replaced human content editors with algorithmic filters. Individuals may think that they are seeing a balance of content whereas in reality they are trapped in what Pariser called "filter bubbles."[15] Pariser's recommendation is that if platforms are going to be gatekeepers, they need to program a sense of civic responsibility into their algorithms, they need to be transparent about the rules that determine what gets through the filter, and "they need to give user control of their bubble."[14, p66] Identity agents can achieve this.

Account management

The individual carries the burden of maintaining the timeliness and consistency of their account information at hundreds of apps. For example, updating contact or credit card information at each is tedious, time-consuming and encourages the individual to spend more time at sites that already have their information. The relative convenience of shopping on Amazon vs. other e-commerce sites is partly a consequence of the individual not having an easy way to manage and update their personal information at multiple sites–it's just easier to buy things on Amazon because Amazon already has all of their personal information.

5 Identity agent implementation

In a human-centric architecture the user's identity agent is at the center with the user's interactions with multiple apps radiating out from it. "When we put the user at the center, and make them the point of integration, the entire system becomes simpler, more robust, more scalable, and more useful."[1]

[38] developer.chrome.com/en/docs/privacy-sandbox/topics/overview/

This necessitates that an identity agent have the ability to interact with a wide variety of different kinds of apps. To illustrate this point, let's consider the user's interactions with six apps. The first app might need the user's email address, and ask for it in a web form. A second form-filler app (which might be a browser extension integrated with the agent) uses this value to fill in the form. A third app might support password-less sign-in (e.g. using OpenID Connect) that leverages an identity agent acting as the so-called *identity provider*. A fourth might request a digital driver's license credential from the agent–a credential that had presumably been installed into it from a fifth credential-issuing app. Finally, a sixth app could be some other person's identity agent (acting as an app) requesting contact information about the user.

5.1 Self and Contexts

The identity agent represents both the person's single *selfness* and a set of *whonesses*, each used in the context of their interaction with a different relying party.

The selfness of the person is represented by a person entity in a data container called the *self*. The person entity in the self is the point of integration across contexts each of which may use differing identifier namespaces, protocols for communication, and data schemas. The contents of the self entity are secret to the user.

Each context represents a relationship between the agent user and some relying party and is represented by a *context* data container. A directed *correlation* link points from the singleton entity in the self to the individual entities representing the user in each context. The identifier of the entity representing the user may differ by context. To ensure privacy, i.e. to prevent correlation across contexts, only the agent user knows that each of these separate contexts contain representations of them.

We can illustrate these concepts with a simple example. A person named Alice might have a relationship with someone called Robert Fox, another identity agent user, and she may interact with a fictional airline website called Untied Airlines. Figure 2 shows a simplified view of how this is represented:

In our example Alice has two digital relationship contexts called *connections*. The blue circle at the left inside black "Alice" container represents Alice's selfness–her innate sense of being a single individual person. In the context of each relationship Alice may choose to express herself differently. The blue circle within the Robert Fox box represents who she is to the relying party named Robert Fox (her whoness to Robert). The blue circle within the Untied Airlines box represents who she is to this airline relying party. The personal information (attributes, messages, credentials, etc.) about Alice that she chooses to present in each context may differ, and usually does.

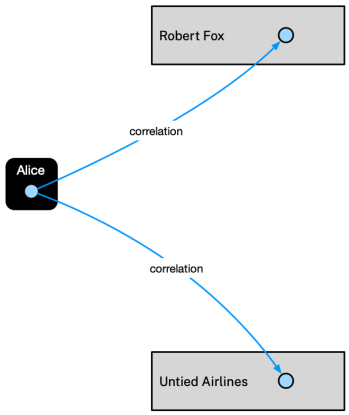

Figure 2: Two facets of Alice in two different contexts

5.2 Functionality

Figure 3 summarizes the functionality of an identity agent. The cells in light and dark green show the progress of implementation work being led by The Mee Foundation.

In addition to authenticating the user to the identity agent, an identity agent performs the following functions:

- **Organize** relationships with relying party's apps into a set of connections.
- **Request** access to a context managed by another app.
- **Grant** access to a context managed by the user.
- **Sync** contexts across user's devices.
- **Delete** all data associated with this set of contexts.
- **Consent** to share data with an app.
- **Edit** data in self-asserted contexts within a connection.
- **View** data in a context (connection).

5.3 Identity agent architecture

The architecture of an identity agent is shown in the center of Figure 4. The state of the identity agents is replicated and synchronized across this pool of identity agents.

Identity Agent Functionality		iOS	Android	Mac/ Win/ Linux
Connection Data Management Functions	**Request** access to a context managed by another app			
	Grant access to a context managed by the user			
	Organize relationships into a set of connections and groups			
	Sync contexts across PDN nodes running on user's devices			
	Delete connection			
	Consent to share data with an app			
	Edit data in self-asserted contexts			
	View data in connection's context			
Other	**Authenticate** the user			

Feasible | Under Development | Implemented

v39

Figure 3: Identity agent Functionality

The terms *self*, *context*, and *connection*, are defined in section 5.5 as part of the Persona data model description.

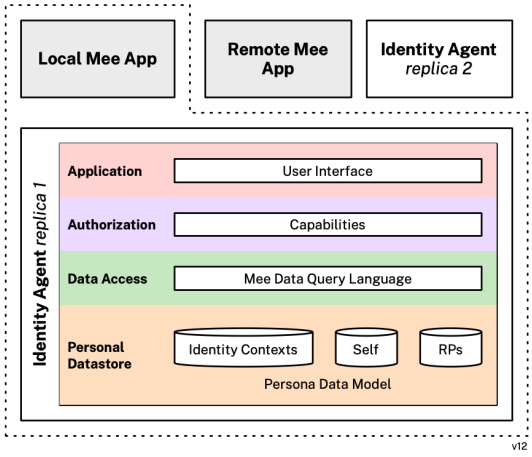

Figure 4: Identity Agent Architecture

5.3.1 Application layer

The Application layer of an identity agent consists of a User Interface (UI) subcomponent and associated business logic.

The UI provides an interface that enables users to manage their data sharing relationships

with apps. Using this UI the user can add and delete connections. Within each connection they can consent to data being shared from their identity agent, see what data is involved in the connection, and in some cases edit attribute values.

5.3.2 Authorization layer

The Authorization Layer manages the granting, verification and revocation of capabilities.

5.3.3 Data access layer

The Mee Data Query Language subcomponent is responsible for management of the user's data whether it is stored locally, replicated on another of the user's identity agents, or managed by a service provider's app. It exposes data contexts in the User Interface where they can be viewed and in some cases edited.

5.3.4 Personal datastore layer

This layer manages local data contexts (some of which may be accessed by Connectors), representations of the Self and the set of relying parties with which the user is connected. Data is encrypted at rest using FIPS-compatible[39] algorithms.

5.4 Data sharing

An identity agent contains an embedded personal datastore, although storage of the user's data may also be distributed among multiple PDN nodes within PDN-compatible apps.

5.5 Persona data model

This section describes the data model used by identity agents to represent personal information. The user's data may be replicated across multiple identity agents on different devices, but we focus here on the logical model, not these replicas.

These Person instances are connected into a directed graph that spans *Context* containers. The singleton Self container holds a single Person node that represents the selfness of person as a single individual. The Self node links to Person nodes in distinct context containers. Each Person contains information about how the user (Self) perceived by the relying party, that is, their *whoness*.

The namespace of the identifier of the Person node in each context varies based on the digital protocol used to interact with the relying party. Since a hypothetical user named Alice's relationship with Robert is via an agent-to-agent protocol, the namespace would

[39] www.nist.gov/standardsgov/compliance-faqs-federal-information-processing-standards-fips

utilize a pseudonymous DID URL[40] identifier, exchanged during an introduction ceremony with Robert. In Alice's relationship with Untied Airlines, her Person identifier is an email address. The graph of Person nodes with identifiers is shown in Figure 5

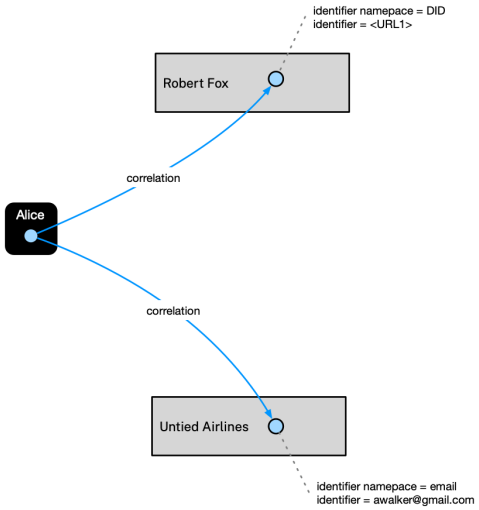

Figure 5: Contexts and Person nodes with identifiers

In the simplified example shown earlier in Figure 2 Alice has two connections, one to a person named Robert Fox (also using an identity agent) and the other to the Untied Airlines.

In each context the Person node has a set of information (attributes, messages, credentials, etc.) defined by the user. We call claims *about* the user (Alice) and are made *by* the user (Alice) *UBU* (User-By-User) claims. For example, Alice might create the UBU claim that her first name is "Alice" and another UBU claim the value of which is her street address. In her relationship with Untied Airlines she might include her first name UBU claim of "Alice", but not choose to include a UBU claim of her home address.

So far we've discussed the claims Alice makes about herself in a context. But in a relationship context the relying party (RP) may make claims about Alice too. We call these claims about the User By the RP or *UBR* claims. For example Unitied Airlines might claim that Alice's frequent-flyer-number is 823-21-5531. To reduce clutter in the diagrams, from now on we will omit the Self and the graph of Person nodes. See Figure 6 for a representation of these UBUs and UBRs.

In the context of Alice's a relationship with a relying party, the RP shares claims it makes

[40] www.w3.org/TR/did-1.0/

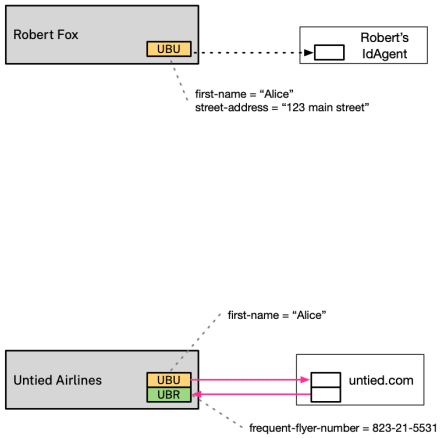

Figure 6: Claims about the user: UBUs and UBRs

about itself/themselves. For example in the context of Alice's connection to Robert Fox, Robert's identity agent shares claims about Robert back to Alice's agent. These claims about RPs made by RPs are called *RBR* claims. An example of an RBR claim is the claim made by Robert Fox that his first name is Robert. To complete our tour of claim types, here is one fourth and last type. These are claims made about the RP by the user are called *RBU*. RBU claims are never shared with the RP. An example of an RBU claim by Alice might be a private note to herself that she met Robert in 2019. See Figure 7.

Now that we've discussed contexts, connections and the four types of claims, we turn to the aggregates of connections called *Collections* and *Groups*.

A collection is a set of connections. In addition to individual connections having UBU claims, the collection itself can have its own UBU claims. The values of the UBU claims that are shared with an RP are computed by taking the value of the claim at the collection level and overriding it with the value of the claim at the connection level if it exists.

In Figure 8 we show Alice as having organized connections into two collections, Friends and Work. The former contains her connections to Bob Fox and Kristin Watson. The latter contains two connections to co-workers, Charlie Smith and Devon Lane and two connections to service providers, Google and Untied Airlines.

A *group* is also comprised of connections, but with different characteristics from a collection. In a collection the RPs to which the agent is connected are unaware of one another. The fact that an RP is a member of a collection is not shared with the RP. A Collection is a set

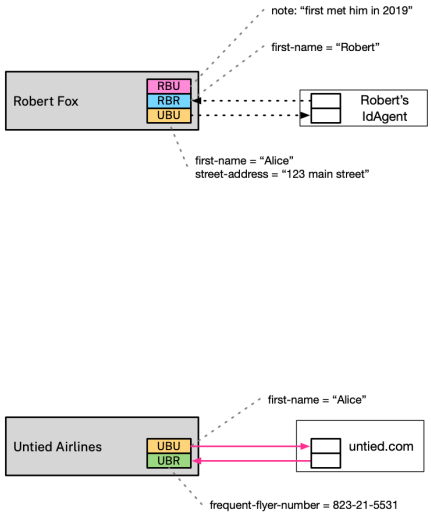

Figure 7: Claims about the RP: RBRs and RBUs

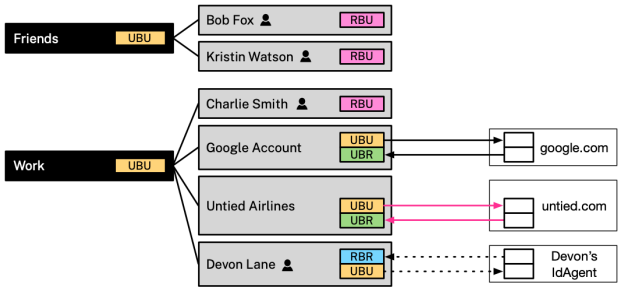

Figure 8: Two Collections

of 1:1 relationships and connections may be moved between collections by the user.

A *group* extends the concept of a collection. A group is a set of connections to RPs. This set itself is replicated to all connections in the group. For example, if Alice has a group with connections to RPs B and C, then the group membership set, (A, B, C), is shared and synchronized to RPs B and C. Thus RP B knows that it is in a group with A and C, and RP C knows it is in a group with A and B. The group also may have a set of *group attributes* which are dynamically and continuously shared with all members of the Group. A group attribute could be a shared calendar, group chat, etc.

In our example, Robert Fox, Kristin A Watson, and Susan Franks have agreed to connect with Alice, join her Bridge Club group, and share claims about themselves using their identity agents. The group attributes of the Bridge Club are shared and synchronized between all four group members. All members of the group see the same shared, updated state of all Group Attributes.

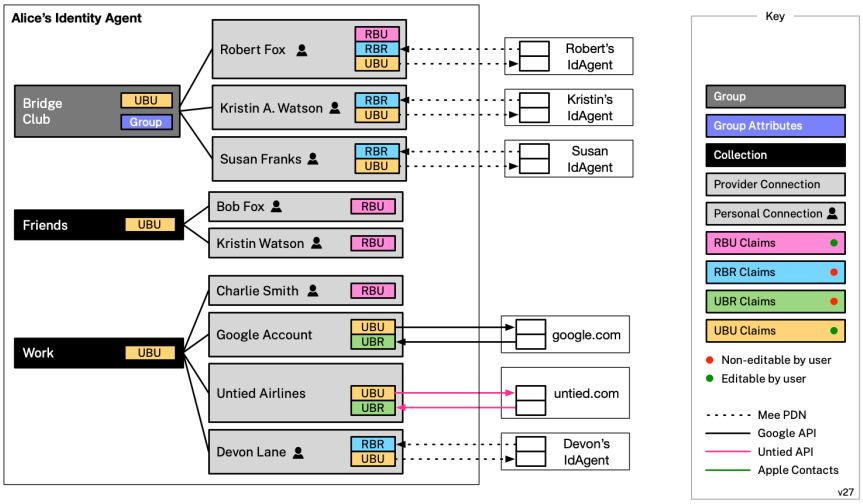

Figure 9: A Group and two Collections

6 Private Data Network

Although identity agents can form connections with many kinds of existing apps using a variety of protocols, we describe here a network of apps and identity agents called the Private Data Network that use a specific set of protocols and adhere to a specific trust framework. These two, taken together, offer identity agent users particularly strong privacy guarantees.

6.1 Private data sharing and the Private Data Network

It is obvious that data held and/or managed by a user's identity agent and stored locally on a device the user owns, is inherently under this user's control. The challenge is that data that a user shares with another party or that is collected by that party in other ways *also* needs to be under the user's control. Unfortunately, it is impossible using solely technical means to remotely control data held by another party. Privacy laws and regulations on the other hand, while intended to provide this control, in practice place such burdens on the user to effectuate this control that it hardly exists. The solution is to combine both legal

(license agreements) and technical means (identity agents and apps on the Private Data Network).

The legal mechanism we propose is the Mee License[41]. The license is a pairwise contract between two parties. The first is the service provider providing an app. The second is an organization that represents the community of identity agent users (e.g. The Mee Foundation). This organization acts as a *Mediator of Individual Data* (MID), a term coined by Lanier et al.[13], that enforces the terms of the license on behalf of the community.

The Mee License imposes obligations on the app provider, among which is the requirement to respect the user's *data rights* to access, correction (editing), and deletion of the information collected and held by them. The Mee License covers information that the user may have shared manually (e.g. by filling in a form, or other kinds of in-app interactions) or shared with them by a person's identity agent. The license requires the provider to implement *data rights* APIs that an identity agent uses to remotely control this app-held data. In this way, we tie the legal (license) and technical means (identity agents and APIs) together.

The Mee License's provisions are intentionally generic. They are designed to meet the needs of the entire community of identity agent users. We expect that other contracts containing more specific provisions will be required to meet the needs of more specialized communities. Each community can amend the license to meet the specifics they require, providing that they do not weaken the license's existing provisions and protections. These specialized communities would organize, govern and operate independent MIDs that enforce their more specialized Mee license-based contracts. These specialized MIDs would enter into agreements with one or more providers which would be held to both the generic terms of the Mee license and the additional, specialized terms.

7 Acknowledgements

Contributors to this paper include Alexey Pepeskul, Kirill Khalitov, Alexander Yuhimenko, Maria Vasuytenko, Vlad Fisher, and Xenya Shatalova.

References

[1] Joe Andrieu. Vrm: The user as point of integration. *joeandrieu.com*, 6 2007. URL: `https://blog.joeandrieu.com/2007/06/14/vrm-the-user-as-point-of-integration/`.

[41] docs.google.com/document/d/13aGk5adoncMxxfl5637NfqP6fl6q_op_1CF50UrJNjg

[2] Bennett Cyphers and Cory Doctorow. Privacy without monopoly: Data protection and interoperability. *EFF*, 2 2021. URL: https://www.eff.org/wp/interoperability-and-privacy.

[3] Cory Doctorow. Competitive compatibility: let's fix the internet, not the tech giants. *Communications of the ACM*, 64:26–29, 2021. URL: https://dl.acm.org/doi/fullHtml/10.1145/3446789.

[4] Nora A Draper and Joseph Turow. The corporate cultivation of digital resignation. *New media and society*, 21:1824–1839, 2019. URL: https://www.cs.cornell.edu/~shmat/courses/cs5436/draper-turow.pdf.

[5] Anne Josephine Flanagan, Jen King, and Sheila Warren. Redesigning data privacy: Reimagining notice & consent for human technology interaction. *World Economic Forum*, 2020. URL: https://www.weforum.org/reports/redesigning-data-privacy-reimagining-notice-consent-for-humantechnology-interaction.

[6] Gordon Graham. Why the world needs an open source digital wallet right now. *The Open Wallet Foundation*, 2023. URL: https://project.linuxfoundation.org/hubfs/LF%20Research/OpenWallet%20Open%20Digital%20Wallet%20-%20Report.pdf?hsLang=en.

[7] Chris Griswold. Protecting children from social media — national affairs. *National Affairs*, 55:3–17, 2023. URL: https://nationalaffairs.com/publications/detail/protecting-children-from-social-media.

[8] William Hoffman. Rethinking personal data: Trust and context in user-centred data ecosystems. *World Economic Forum*, 2014. URL: https://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf.

[9] Jason I Hong. Teaching the fate community about privacy. *Communications of the ACM*, 66:10–11, 2023. URL: https://dl.acm.org/doi/abs/10.1145/3603718.

[10] Lauren Jackson. A driver's license for the internet. *The New York Times*, 7 2023. URL: https://www.nytimes.com/2023/07/03/briefing/age-verification.html.

[11] David Kirkpatrick. *The Facebook effect: The inside story of the company that is connecting the world*. Simon and Schuster, 2011.

[12] Martin Kleppmann, Adam Wiggins, Peter Van Hardenberg, and Mark McGranaghan. Local-first software: you own your data, in spite of the cloud. *Proceedings of the 2019 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*, pages 154–178, 2019. URL: https://dl.acm.org/doi/abs/10.1145/3359591.3359737.

[13] Jaron Lanier and E Glen Weyl. A blueprint for a better digital society. *Harvard Business Review*, 26, 2018. URL: http://eliassi.org/lanier_and_weyl_hbr2018.pdf.

[14] Roger McNamee. *Zucked: Waking up to the Facebook catastrophe*. Penguin, 2020.

[15] Eli Pariser. *Eli Pariser: Beware Online" filter Bubbles"*. TED, 2011.

[16] Alex Preukschat and Drummond Reed. *Self-sovereign identity: decentralized digital identity and verifiable credentials*. Simon and Schuster, 2021.

[17] Neil Richards. *Why privacy matters*. Oxford University Press, 2021.

[18] Emma Roth. Online age verification is coming, and privacy is on the chopping block - the verge. *The Verge*, 2023. URL: https://www.theverge.com/23721306/online-age-verification-privacy-laws-child-safety.

[19] Doc Searls. Vrm is me2b. *Project VRM Blog*, 5 2019. URL: http://blogs.harvard.edu/vrm/2019/05/13/me2b-2/.

[20] Daniel J Solove. Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.*, 126:1880, 2012. URL: https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2093&context=faculty_publications.

[21] Daniel J Solove. Data is what data does: Regulating use, harm, and risk instead of sensitive data. *Harm, and Risk Instead of Sensitive Data (January 11, 2023)*, 2023. URL: Solove,DanielJ.,DataIsWhatDataDoes:RegulatingBasedonHarmandRiskInsteadofSensitiveData(January11,2023).118NorthwesternUniversityLawReview(Forthcoming),GWULegalStudiesResearchPaperNo.2023-22,GWULawSchoolPublicLawResearchPaperNo.2023-22,AvailableatSSRN:https://ssrn.com/abstract=4322198orhttp://dx.doi.org/10.2139/ssrn.4322198.

[22] Alicia Solow-Niederman. Information privacy and the inference economy. *Nw. UL Rev.*, 117:357, 2022. URL: https://heinonline.org/HOL/LandingPage?handle=hein.journals/illlr117&div=18&id=&page=.