Human TECHNOLOGY

ISSN: 1795-6889

https://ht.csr-pub.eu Volume 20(2), September 2024, 183-223

THE CHIMERA OF CONTROL: SELF-SOVEREIGN IDENTITY, DATA CONTROL, AND USER PERCEPTIONS

Nicholas Martin
Fraunhofer Institute for Systems and Innovation Research ISI
Germany
ORCID 0000-0002-8739-0165

Frederik M. Metzger
Fraunhofer Institute for Systems and Innovation Research ISI
Germany
ORCID 0000-0002-3082-7047

Abstract: *This paper analyses the claim that Self-Sovereign Identity technology (SSI) gives users greater control over their data and identity than established digital identity systems, and studies empirically how users view this claim and its impact on adoption decisions. We argue that the claim is objectively false. SSI does not offer users greater control over their data, though in combination with laws like the GDPR, certain add-on features to SSI might have mildly privacy-/control-enhancing effects. Absent GDPR-like laws, however, SSI threatens to turn into a "disclosure machine" where users are forced to give up more data than they would likely have (or be able to) with extant identity solutions. SSI attempts to solve political-institutional problems through technology/architecture, but cannot eliminate the power imbalances at the heart of users' lack of control. Presented with an SSI system, we find that most users do not seem to intuitively conclude that it gives them "control" over their data. A minority does however reach this conclusion, and for these people it seems to be an important factor driving adoption decisions.*

Keywords: *Self-Sovereign Identity, SSI, privacy, data protection, institutions, GDPR.*

DOI: https://doi.org/10.14254/1795-6889.2024.20-2.1

INTRODUCTION

Self-sovereign identity solutions (SSI) have attracted much attention. In the Horizon 2020-funding period, European Union (EU) funding for SSI-related research projects ran to at least ~€45 million.[1] Member states too are funding SSI-related projects with large sums.[2] Numerous start-ups are developing SSI-style identity solutions. Major business groups have expressed strong support for SSI (e.g. Deutscher Sparkassen- und Giroverband et al. 2021). International organisations like the World Food Program have also explored SSI for working with refugees.

Advocates of SSI claim that it offers several benefits compared to traditional digital identity solutions, including greater security and usability (Kubach et al. 2020), greater personal privacy (Schardong and Custódio 2022), allowing users "to fully own and manage [their] digital identity" (Mühle et al. 2018, p. 81), and stimulating the digital economy (Echikson 2020). SSI proponents have even claimed that it could help solve the problem that lack of identity documentation poses for undocumented migrants and/or support economic inclusion (Lim 2020, Wang and De Filippi 2020).

The key benefit of SSI that its proponents reference, though, is that it would put users *in control* of their data and identity. This greater control over one's own data and identity is what supposedly fundamentally distinguishes SSI from other identity solutions – hence its name "*self-sovereign* identity".

Surveys indicate that the idea of getting more "control over one's data" is very popular among users (PWC 2021, Spiekermann and Korunovska 2017). Probably for this reason, business interests have adopted the language of giving users "control over their data" to promote SSI (e.g. Schorlemer 2022).

This paper analyses whether the claim that SSI gives users meaningful control over (i) their data and (ii) their identity, is factually correct. We argue that while SSI advocates' claims are not without grains of truth, they are greatly overblown. SSI can offer users small increases in control of their data, but only if a suitable legal framework is in place, like the GDPR. Where such a legal framework is absent, SSI solutions on the contrary regularly risk compelling users to disclose more data than they otherwise would.

The paper then investigates users' perceptions and responses to the claim that SSI gives them "control", asking two empirical research questions:

> RQ 1: Irrespective of the *objective* truth of the claim that SSI gives users control, do users *subjectively* conclude that SSI *will* put them in control of their data, if the basic functionality and data flows in an SSI system are explained to them?[3]
>
> RQ 2: In as far as users do conclude that an SSI system gives them "control", how important is this for driving adoption?

To our knowledge, the scholarly literature has so far largely failed to critically analyse the claim that SSI gives users "control" over their data or identity. A growing body of work exists on SSI systems (e.g. Strüker et al. 2021, Sedlmeir 2021, Richter and Anke 2021, Liu et al. 2020, Lim 2020, Kubach et al. 2020, Hempel and Anke 2023), but this largely focuses on technical aspects, and occasionally on business models. It and tends to assert, rather thannot critically interrogate, SSI's capacity to give users "control". An exception is a recent Master's thesis (Doesburg 2023), which notes that SSI can encourage "oversharing" of data and

proposes technical measures to ameliorate this problem.[4] But by focusing on technical measures, it falls into the same as trap as, we will argue, SSI advocates have fallen into, namely trying to solve what are fundamentally political-economic challenges through technical fixes.

It is from activists, rather than academics, that the problem of user control in SSI has received greater attention. Thus a policy submission on the reform of the EU's eIDAS regulation (*Regulation on electronic identification and trust services*) from epicenter.works, a digital rights NGO, develops a similar analysis of the potential for abuse of digital wallets as that advanced here. However, it is focused on specific issues with eIDAS and does not attempt a systematic analysis of the concepts of "controlling" one's data or "identity". Instead, it mainly concentrates on proposing alterations to specific parts of the legal text (epicenter.works 2022).[5]

To our knowledge, only the consulting company PWC has investigated to what extent users value SSI's promise to give them "control" of their data. PWC surveyed 2,000 German consumers on their views about digital wallets, which are a component of most SSI systems (see below). PWC found that for 92% of respondents, it is "very" or "somewhat" important to be able to decide themselves about "which data they disclose". For 90%, "sovereignty over storage and use of data", is a "very" or "somewhat" important product feature of digital-wallet systems (PWC 2021). Beyond these two data points, however, the PWC survey provides no further information on users' perceptions and responses to actual SSI systems. Moreover, it does not analyse the data statistically. We are not aware of any other studies that have collected and explored empirical data on how (potential) users of SSI solutions view these questions.

The issues addressed in this paper matter for several reasons. If SSI does not in fact provide users with meaningfully greater control over their data than other identity systems, then advocates promoting SSI on this basis at minimum risk overselling these systems, if not actively and dangerously misleading users. As Doesburg (2023) warns, users may be particularly prone to "oversharing" data if they (falsely) believe that SSI is especially privacy friendly. Misleading advertising risks luring them into a false sense of security.

Users' general cynicism about tech company claims might mitigate the risk posed by misleading claims. An arguably greater danger is that SSI developers and policy makers come to believe their own hype, about SSI solving a series of social-political problems, and doing so by means of a clever technological architecture. If SSI does *not* in fact do so, this would leave the real problems unaddressed. Worse, it risks exporting technological solutions dependent for their safety on specific legal frameworks (e.g. the EU General Data Protection Regulation [GDPR]) to countries where this framework is lacking and the tech solution in question may in fact exacerbate rather than reduce risks, while claiming that the technology has solved these risks.

As we will see, this is precisely the case: Absent legal-institutional safeguards like the GDPR, structural power imbalances between users and service providers can transform the architecture of SSI from a (mildly) privacy-*enhancing* technology into a "disclosure machine", where users are forced to give up more data about themselves than they would likely have to (or be able to) with either centralised or federated identities. This problem emerges particularly clearly with refugees, a potential use case for SSI systems highlighted by several prior authors (Wang and De Filippi 2020, Cheesman and Slavin 2021), but it also

exists for other users. The reason for this is that SSI solutions do not and cannot eliminate the larger imbalances of power in the (digital) economy and the state, that are at the heart of users' lack of control over their data and identity. SSI proposes a technological and architectural solution to what are ultimately political-economic and institutional problems. Accordingly it falls short. For this reason – contrary to some SSI advocates' claims – SSI also cannot solve the political problem facing most undocumented migrants, that no state is willing to issue them with valuable identity rights.

Finally, if this paper's conclusion that the "control" SSI grants to users is largely chimerical is correct, it is to be expected that users will soon see through this promotional rhetoric. As we will see, this is supported by the results for RQ 1. That would be bad for SSI – which arguably does offer users real economic and usability benefits – and for the general level of trust in society. If consumers come to feel misled about something as fundamental and intimate as their control over their own data and identity, this may have more damaging consequences than if they feel they have been sold an overhyped golf club or other random consumer gadget. The good news for SSI advocates is that the findings with regard to RQ 2 suggest that exaggerated claims about control are not needed for driving adoption.

METHOD AND STRUCTURE

The paper has two parts. Part One investigates the claim that SSI provides users with "control" over data and identity. It first provides a conceptual and literature-based analysis of the concept of having "control" over one's data and identity. To analyse this concept, the paper focuses on how it is operationalised in the GDPR. The GDPR offers a useful lens because safeguarding individuals' control over their data (informational self-determination) is one of its core objectives and it represents the most developed attempt to operationalise this concept in practice for the digital world. It also constitutes the central legal framework within which SSI solutions, at least in Europe, have to work and be designed for.

Next, the paper describes how SSI solutions try to provide users with "control" by technological and architectural means. SSI is compared to two established types of identity system: centralised and federated identity management systems. This analysis is based on the extant literature describing these different identity systems, plus in-depth discussions with SSI developers from the Spanish research organisation GRADIANT.[6]

On this basis, the paper shows why the promise of SSI largely falls short, with the (distinctly modest) improvements in user "control" that SSI solutions may sometimes be able to achieve dependant entirely on the existence of legal protections for users – not the technology as such (as SSI advocates tend to claim). The paper then turns to the problem of situations where these legal protections are absent. For this, it re-analyses a case study presented in Wang and De Filippi (2020) to show how under these circumstances, SSI risks turning into a "disclosure machine" that weakens, not strengthens, user control. Summary conclusions with regard to SSI's claim to provide users with "control" are then presented.

Part Two investigates empirically how users perceive and respond to the claim that SSI gives them "control" (RQ 1 and RQ 2). To answer these research questions, data from a unique user survey (n = 740 and a subset thereof) is analysed. A detailed description of the survey instrument, data and analytical method are presented in Part Two.

PART ONE: "CONTROL" IN SSI SYSTEMS

Data, Identity and Control

Control over One's Data

Since the German Constitutional Court's 1983 ruling establishing a right to "informational self-determination", the idea that users ("data subjects") should "control" their data has become increasingly central to European debates about data policy and data protection. The Court ruled that individuals should in principle be able "to determine for themselves when, how, and to what extent information about them is communicated to others." (Lazaro and Le Métayer 2015: p. 6; also Veil 2018). This idea, that individuals should have the power to decide over the use and disclosure of their data (Gonzalez Fuster and Gellert 2012), has since become a cornerstone of European data protection law and the EU General Data Protection Regulation (GDPR) in particular. As Clifford and Ausloos (2017) note, "control" here must be understood broadly, as providing "a robust architecture of control that actively ensures individual autonomy."

The GDPR tries to implement this through two mechanisms (Vogiatzoglou and Valcke 2022). Firstly, it equips the individual with a battery of so-called "data subject rights", vis-à-vis data controllers (e.g. online service providers). Users have rights to information and access about the use of their data (Art. 12–15 GDPR); to rectification of erroneous data, erasure ("to be forgotten"), and to restrict or object to processing of their data (Art. 16–18, 21), among others.

The second mechanism by which the GDPR tries to safeguard the individual's control over their data is by specifying the legal bases under which personal data can be legitimately processed (Art. 6). In particular, it defines the conditions under which the data subject may be said to have "consented" to processing (Art. 7). This is crucial, as consent remains one of the most common legal bases for processing. However, consent may often provide a particularly flimsy safeguard, or means of control, for the individual. This flimsiness is due to how, in the digital world, nominal "consent" can and often is obtained through power imbalances (take-it-or-leave-it offers) and even manipulation. To prevent this, the GDPR specifies that "consent" must be "freely" and demonstrably given. Information about *what* is being consented to, must be presented in an "intelligible and easily accessible form, using clear and plain language". Data subjects must be able to withdraw consent at any time and this must be "easy" (Art. 7).

These stipulations show that while conceptually, the notion of having "control" over one's data is simple (to be able to determine who gets it and what they can do with it), operationally it is anything but. To have "control" while also using data-based services requires one to be able to meaningfully exercise and enforce multiple rights/capabilities vis-à-vis data processors.

The question of the effective exercise of rights/capabilities points to the weaknesses inherent in consent as a mechanism for regulating user/service provider-relationships in the digital economy. Often this relationship rests on large, even extreme, power imbalances (Lazaro and Le Métayer 2015). For many of the socially or economically most valuable services in the digital economy, few substitutes exist. Absent legal safeguards, the user is thus

left with the choice to "consent" to whatever terms the service provider stipulates, or forego the service. Depending on the value and uniqueness of the service, though, that may not be an option for the user. Furthermore, a large information asymmetry often exists between user and service provider regarding what data are being collected, and how and to what ends they are being processed.[7] Under these circumstances – which are omnipresent in the digital world – "consent" as an expression of meaningful control over data processing is likely to be an illusion. While the GDPR's insistence on consent being "informed" and "freely given" is laudable, it is questionable to what extent even this regulation has been able to do much to level the underlying power imbalance and make "consent" meaningful. This has important implications for SSI.

Control over One's Identity

Identity, and "control" over it, are trickier concepts. As Zwitter et al. (2020) discuss, there is a longstanding debate in philosophy about how to understand identity. Philosophical naturalists believe that identity is inherently "tied to the properties of an object or a person". It is "whole" and "distinguishable" from others (ibid.). In this view, identity is in effect a kind of object or "thing". By contrast, philosophical constructivists understand identity as being primarily relational. It is not so much an object as a status that arises out of the relationship of a person or an object to other persons, objects or institutions (ibid.).

This paper does not aim to resolve these philosophical debates. Arguably, "identity" has both relational/constructivist and objectivist/naturalist dimensions. This dual nature of identity – naturalist/objectivist *and* constructivist/relational, "thing" and "status" – is evident in digital and in legal-political identities. It is these that are mainly of interest here, as SSI is a type of digital identity and can also potentially serve as a digital version of offline legal-political identity documents like passports or identity cards, or at least depends on them in order to offer SSI users much of value, a point discussed further below.

Structurally, both digital identities and legal-political identities consist of:

- An *identifier* – such as a username, account number, tax number or state identity card number – that uniquely identifies the entity (the person, user, machine, organisation, …) and distinguishes it from other entities (Dunphy and Petitcolas 2018);
- *Attributes*, that is, pieces of information that describe the entity, such as their real name, age or citizenship, etc. (Dunphy and Petitcolas 2018). Attributes can also include rights or entitlements (in other words, status), such as the permission to access a certain IT system (e.g. an email inbox) or to receive a certain payment (e.g. child allowance).[8]

In as far as digital and legal-political identities are data (the identifier and attributes) encased in a digital system (e.g. an app) or a physical artefact (e.g. a card), they are "things" and can thus arguably be, in some sense, "controlled" by their holder (user). It is likely no coincidence that the SSI discourse tends to talk about identity in these objectivist/naturalist terms.[9] Practically, "control" may mean something like being able to decide how and where to use one's identity, whom to disclose attributes to, whether and how long those to whom attributes have been disclosed to may store this information, and for what purposes they may process it. It might also mean being able to decide who can add further attributes to one's

identity. This is broadly similar to having "control over one's data" (attributes are data).[10] Of course, as with data, the "choice" over whether to disclose one's identity to someone may often be largely chimerical. If I want to buy alcohol, I *will* have to show my ID.

But the practical social, economic and legal-political value of identity – which is what people are usually most interested in – is strictly relational. Identities, or rather their attributes (e.g. citizenship, status as a single parent, holder of university degree, …) usually only create value for the person who holds them because the identity/attribute is recognised by others, who agree to act on this identity/attribute in ways beneficial to its holder (e.g. grant them the right to vote, to pass a border, get benefits or a job, or access to an email system). An identity/attribute that is not recognised by others is – in social, economic or legal-political terms – largely worthless, however expressive of the holder's most deeply held beliefs it may be.

It is much less clear what having "control" over this identity-as-relationship/status might mean. The most plausible reading of "control" here would be that it means to have some degree of control or power *over the relationship*. It is not immediately obvious what this would mean practically, but it could be something like being able to *oblige* counterparties to act towards one in certain ways (e.g. grant access to an email system, pay benefits).[11] It is hard to see how any identity management technology can do that.

The Promise of Self-Sovereign Identity: Solving for Control Technologically

To understand how SSI works and tries to exploit technology and architecture to give users control over their data/identity, it is helpful to contrast it to two established, widely used identity management approaches: centralised and federated identity management systems.

In *centralised identity management systems*,[12] the user is issued with an identity by the system administrator. The system administrator retains ultimate control over how the user uses the identity and can, at least within their system, track user behaviour. Amazon is a typical example of this. To use Amazon's service (i.e., purchase books, films, etc.), the user must create an account (i.e. a digital identity) with Amazon (Figure 1, left). Amazon "controls" this account/identity, in the sense that the company can define the terms under which it may be used, monitor user behaviour on the Amazon site (and possibly also on other websites if the user stays logged in to their account),[13] and revoke the account and identity at will, subject to law. Another example of centralised identity systems are employees' user accounts/digital identities with their employers' IT systems (computer, corporate email, server access etc.). Again – subject to data protection and employment law – the employer can define the terms under which the employee may use their account/identity and the corporate IT systems, monitor user behaviour within the corporate systems, and revoke the identity. Centralised identity systems are often system-specific, like in these two examples: neither Amazon nor most employers offer users an identity that can be used with other service providers (e.g. to authenticate to an online banking service).

In *federated identity management systems*,[14] also known as Single Sign-On (SSO), the user signs up for an identity with a central identity service provider, such as Google or Facebook, and can then use the identity thus created (e.g. their Gmail address and password) to sign up for, log in to and use the services offered by other ("federated") service providers

(e.g. a music streaming service) that accept the SSO provider's identity system (e.g. "sign up with Gmail") (Figure 1, right). Federated systems give the SSO provider (e.g. Google) control over the user's identity in the sense that the provider can set the terms and conditions of use, revoke the user's identity (thus locking her out of all services she uses the SSO identity for), and can track the user across all services she accesses with the SSO identity.[15]

A number of different approaches to *self-sovereign identity (SSI)* exist. There is no single definition of the term (Kubach et al. 2020, Lim 2020). Generally though, SSI approaches share a common conceptual reference point in the work of Christopher Allen (2016), whose "ten principles of self-sovereign identity" serve as a kind of foundational text for the community. More importantly, SSI systems generally rely on the same technical building blocks and foresee similar architectures and ecosystems built around several common key roles (Strüker et al. 2021, Lim 2020, Kubach et al. 2020, W3C 2024).[16]

The key technical building blocks are *digital wallets* (generally to be held on the user's smartphone or other devices), *verifiable credentials* (VCs) and *decentralised identifiers* (DIDs). World Wide Web Consortium (W3C) standards exist for VCs (draft recommendation) and DIDs (W3C 2022, 2024; Strüker et al. 2021; Richter and Anke 2021).

Verifiable credentials (VCs) are tamper-proof digital certificates that are issued and cryptographically signed by a trusted party (e.g. a university issuing a digital degree certificate, an online service provider issuing a credential to authenticate a registered user). VCs, in effect, contain attributes about a user (e.g. name, university degree class). Users store the VCs issued to them in the *digital wallet* on their smartphone or other device (Kubach et al. 2020, W3C 2024).[17]

A major strength of VCs is that – provided their issuer is trusted – the information contained in them can be accepted as true because they are tamper-proof and cryptographically signed by the issuer. Their validity can thus be proven automatically, with no need to contact the issuer to confirm a credential. This offers some – arguably mostly minor – privacy benefits to the user (see below), but also substantial efficiency gains for users and service providers, as the verification of identity attributes no longer requires lengthy "manual" communication between service provider and issuer (Kubach et al. 2020, Richter and Anke 2021, W3C 2024).

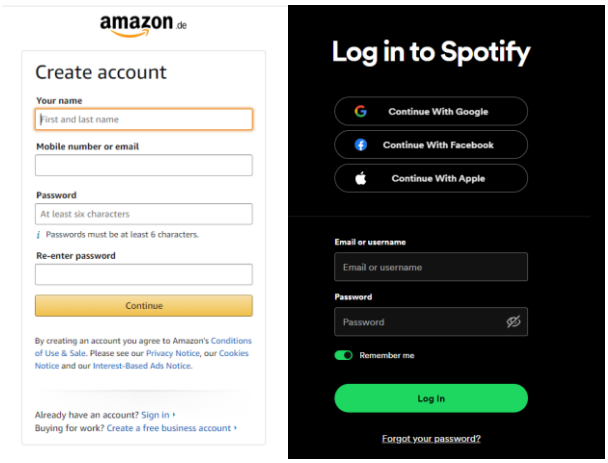

Figure 1. Registration interface in a centralised identity system (l); Log-in interface with SSO Identities (r)

A further strength of VCs is that they can be exchanged through encrypted peer-to-peer channels. For this, *Decentralised Identifiers* (DIDs) are used. DIDs identify parties to an exchange to each other (e.g. user and service provider, who exchange information/attributes). They allow the two parties to recognise each other. Importantly, the DIDs allow for this exchange to take place via an encrypted peer-to-peer channel, that can be observed only by the parties to the exchange (Richter and Anke 2021).[18] It is common to create a unique, temporary DID channel for each interaction with each service provider.

The communication between user and service provider in an SSI system using DIDs is thus fundamentally different to that in an SSO system. With SSO, the communication is tripartite and intermediated by the SSO provider, who can observe the user and, in theory, even block her from using her identity and accessing the service. With SSI, using a DID channel, the communication between user and service provider is strictly bilateral. No third party can observe or interfere with the communication. (It is however possible for service providers to collude with each other and/or VC issuers *after* the communication, to further profile and track users [W3C 2024].) Thus, the key roles in the SSI ecosystem are:

- the *issuer* of credentials (VCs)
- the *user* (sometimes called "holder"), who receives VCs from the issue and stores them in her digital wallet
- the *service provider* (often called "relying party" or "verifier")

When she accesses a service or engages in some other transaction with the service provider, the *user* presents VCs to the *service provider* via a DID channel. The *service provider* then "verifies" the VCs and "relies on" the *issuer* for their trustworthiness.[19]

Note that service providers can simultaneously also be issuers. For example, in an SSI ecosystem, signing up to an online service provider (e.g. creating an account with an online bank) could involve presenting the service provider (the bank) with a state-issued VC of one's identity card to prove one's identity, whereupon the bank issues one with a further VC, with which one can authenticate (log in) to one's new account. When one next logs in, one then just presents this VC.

The Chimera of Control

SSI advocates point to three mechanisms through which SSI supposedly gives users control over their data and identity. Firstly, users must consent to any disclosure of their VCs. This is architecturally hardwired into SSI, as only the users control their digital wallets with the VCs. Secondly, identity providers are disintermediated. There is no SSO or other identity provider, who stands between the user and the service provider and can revoke the user's access to and use of her identity and attributes. Rather, these are all contained in the digital wallet and are directly presented by the user to the service provider, via peer-to-peer DID channels. By the same token, there is no SSO or other identity provider who can directly monitor the user's behaviour. Likewise, because the VCs' authenticity is verified automatically by the service provider, without direct involvement of the issuer, the issuer too is unable to monitor the user's behaviour (unless issuers and service providers collude to do so.)

Thirdly, it is in principle possible to equip SSI systems with additional features to enhance user privacy; especially Zero Knowledge Proofs (ZKPs) and Wallet Histories (Hempel and Anke 2023, W3C 2024). ZKPs enable attributes about an entity to be proven without disclosing the underlying information (e.g. that a user is older than 18, but not her actual age). Thus, the amount of data that needs to be provided about a user can be minimised. Wallet Histories give users a more complete overview of what data (VCs) they have disclosed to which service providers, for what purposes. That makes it easier for users to exercise GDPR rights, e.g. to request that service providers delete their data.

All three mechanisms provide users with some potential influence over their data and, in this sense, their identity (if "identity" is treated in objectivist terms: none offer control over identity-as-relationship). Unfortunately, this added influence is far from the kind of "sovereign control" that the rhetoric of SSI advocates frequently implies. On the contrary, at best SSI provides users with some technical aids to help them operationalise GDPR rights and via ZKPs, the GDPR's data minimisation requirement.

The fundamental reason for this lies in the highly unequal political-economic (power) relationship between users and service providers in the digital economy, which SSI is unable to address. This is exemplified by the problem of consent.

Consent and Control over the Information Contained in VCs

As noted, in an SSI ecosystem, users store their data as VCs in their digital wallets. Service providers must ask users for permission (consent) to access and process these. Users can then decide whether they wish to grant service providers this consent. SSI advocates cite this as a key instance of users exercising "control" over their data (e.g. Schardong and Custódio 2022). Superficially, this is indeed the case – users are technically free to refuse consent.

But there are three problems with this. For one, this is hardly very different to how registration processes *already* work in centralised or federated identity systems. Recall the example of onboarding to Amazon (or most other online service providers): as part of the onboarding process, Amazon (the online service provider) requests various pieces of user data and spells out – in lengthy legal documents[20] – under what terms and conditions it will store and process this; terms and conditions the user must consent to in order to use the service. This is no different if she uses an SSI system to onboard and provides her data to Amazon via VCs.[21]

Secondly, while VCs remain in some sense in the user's possession, the *information they contain* (i.e., the user's data) do not. Rather, when the user presents her VCs to the service provider, the service provider reads out the information (data) contained in the VCs. Provided the user "consents", there is no technical barrier to stop the service from storing and further processing these data, including tracking the user across external data and devices, and in cooperation with other third parties.[22] Once the information is entered into the service provider's systems, it is beyond the user's control.

Thirdly, while, at least in Europe, the user does need to consent, as indicated above, consent is only meaningful if the user has real alternatives and a full understanding of what they are consenting to. Only then can the act of consent represent the exercise of some form of genuine "control" or self-determination. In the case of Amazon, most goods sold there can likely be obtained through other online or offline sellers. In this sense, the user at least has

alternatives. But it is likely that most alternative online sellers would present users with comparable or even more invasive data requests, couched in equally long and complex legal documents.[23] The alternatives may thus present little real choice. Conversely, making purchases offline requires the user to pay a steep opportunity cost (time, effort).

Making online purchases may be a trivial example, as offline alternatives exist and access to Amazon or other e-commerce sites is rarely existential. But access to other services is more so. Take banking, social networks or messenger services. It is all but impossible to participate in a modern economy without a bank account. Given the central role of social networks and messenger services to communication flows today, it is difficult in many situations to refuse their use, if one wants to participate in society. Any theoretical "choice" users may have about whether to consent to their data requests is thus largely chimerical. To "consent" is less the sign of users exercising "control" than facing take-it-or-leave it choices where "leave it" is not an option. SSI does not change this.

Disintermediation of (some) Identity Providers

SSI does disintermediate SSO providers. This is a net gain for user privacy and user data "control", in as far as it reduces the number of parties monitoring user behaviour and the number of parties to whose data collection users must "consent" to. Likewise, it aims to remove certificate issuers from the verification process. However, these privacy gains are arguably minor: Google and Facebook, the main SSO providers, have many other means to surveil users (Christl and Spiekermann 2016), and users must still consent to data processing by service providers. While the *aim* is to remove certificate issuers from the verification process, this is ultimately up to the actors in the ecosystem: there are few technical barriers to stop service providers and issuers colluding to track users (though the W3C standard appeals to them not to do so).

It is unclear whether SSI would also disintermediate *centralised* identity providers, though this seems unlikely. Centralised identity providers are usually also service providers, and issue an identity as part of, and in order to, provide the service. They would likely do this also in an SSI ecosystem. For instance, Amazon would presumably still want its users to create accounts with unique identities for use of its services. While no SSO-type intermediary would exist to potentially block users from using an identity, service providers and certificate issuers remain free to revoke identity credentials they have issued, e.g. if users violate terms of service.

More fundamentally, as discussed, the *value* of an identity/identity attribute depends on it being recognised by others. Generally, the most valuable identities and attributes are those issued by powerful states or institutions that command unique resources (e.g. famous universities with degree-granting power). These identities/attributes offer their holder the greatest panoply of rights and benefits. SSI does not change this. While users may be able to use some SSI systems to issue themselves with their own identity certificate, such a self-issued identity certificate is likely without value. On the contrary, to access most valuable services, SSI users are all but certain to remain dependent on identity credentials (VCs) issued by states and other powerful institutions.

Wallet Histories and Zero Knowledge Proofs

Wallet Histories and ZKPs do offer users modest gains in data control and privacy – provided service providers act on user requests enabled by Histories (e.g. to delete data) and accept ZKPs rather than demanding the underlying data. But nothing about SSI (or Histories and ZKPs) forces service providers to do so. What *can* force service providers to do so is not technology, but law – for instance the GDPR data subject rights, whose exercise Wallet Histories can facilitate, and the GDPR's data minimisation principle (Art. 5(1)(d)) and privacy by design and default stipulation (Art. 25), which could be read as *mandating* ZKPs once these are technologically mature.

In summary, the three mechanisms by themselves do not significantly increase users' control over their data and their identity-as-thing, while they cannot in any way increase users' power or control over the identity relations that generate actual value from identity (identity-as-relationship/status). Wallet Histories and ZKPs arguably can somewhat increase users' privacy and control, but this depends on institutions (e.g., laws) actually forcing their use. In Europe, the GDPR may ultimately do this. That raises the question of what effects SSI may have in environments where institutions such as the GDPR do not exist.

SSI as Disclosure Machine?

A key benefit that SSI systems offer users is the digital wallet and the ability to store all their identity credentials and attributes as VCs in one easily accessible place. This allows users to conduct digital transactions faster and with less friction than with existing systems, where credentials continue to be stored over a wide range of systems (desktop, smartphone, paper box files) and in many different formats (pdfs, jpegs, hard copy,…) , often distributed over multiple entities (user, employer, health insurance, …). Moreover, unlike the pdfs and jpegs that users often still rely on today, VCs are hard to falsify and can be quickly and easily verified.

This is a key reason why business interests support SSI. It promises to reduce costs and increase service consumption as friction is reduced, and to provide more reliable data about users, VCs being tamper-proof and verifiable (Anke and Hempel 2023).

But absent legal safeguards like the GDPR, these characteristics also threaten to turn SSI into a disclosure machine. Faced with wallets brimming with interesting data that can be made available in seconds via a few clicks, businesses and state institutions will be sorely tempted to demand all manner of additional data disclosure as a condition of service access, well beyond what is strictly necessary to provide the service. Facing take-it-or-leave-it choices about service access, users will likely find it hard to refuse. We call this "data extortion". With terms of service commonly packed into pages of dense legalese, users will often struggle to even understand what is being demanded of them. Far from securing user privacy and data control, SSI would thus make users *more* transparent and *reduce* their control over their data.

In Europe, the GDPR's data minimisation principle, among other stipulations, provides some safeguards against this. But in many other contexts institutions like the GDPR are

absent. SSI's potential to turn into a disclosure machine under these circumstances is underscored by Wang and De Filippi's (2020) description of the UN World Food Program's (WFP) "Building Blocks" program. As they describe it, Building Blocks sought to harness blockchain, biometric recognition and public key infrastructure for disbursal and management of Syrian refugees' humanitarian assistance entitlements (food, health, cash, education etc.) in Jordan. The system conforms broadly to the SSI idea, albeit with some architectural adjustments to fit the circumstances of a refugee camp. For instance, it did not require users (refugees) to possess smartphones or similar devices. Rather, each refugee received a unique identifier anchored to their biometric data plus associated public/private keys, and an individual account or "blockchain wallet" (Wang and De Filippi 2020) on the chain, in which information about their entitlements and transactions conducted was stored.[24] The initial aim of the system was to streamline cash disbursal to refugees and save banking fees (Microsoft Sway n.d.). However, as the account in Wang and De Filippi (2020) and statements from the WFP make clear, operators quickly recognised the system's potential to give them a much more systematic and comprehensive, "unified" view of the refugees: the ~106,000 refugees in Jordan were served by another 45 different NGOs and aid agencies aside from the WFP, each with their own data systems. This led to a "fragmented view of the people served", duplicated efforts and inequities (Wang and De Filippi 2020). A solution was seen in collecting all entitlements as well as other relevant information (e.g. status as a nursing mother) in the refugees' blockchain wallets, thus providing the agencies with a "unified view" of the refugees, to ensure equity of provisions and avoid duplications. In other words, the potential of digital wallets to render the refugees highly transparent and the functional benefits this could provide to operators were quickly recognised, and steps taken to explore implementation options.

It is unclear to what extent other aid providers, beyond the WFP, were ultimately given access to Building Blocks and refugee entitlements collected in such a comprehensive way. It should also be noted that important measures were implemented in Building Blocks to avoid disclosing excessive information to actors within the network.[25] The point is not that Building Blocks was "evil" or that rendering refugees more transparent to aid agencies is unjustifiable or would lead to disaster. Ensuring equity and *adequacy* of aid provision is often likely to be more important than adhering to a privacy absolutism. Rather, the case illustrates how the capability of the digital wallets, that are at the heart of SSI, to serve as collection point for all of an individual's relevant identity information immediately leads to covetousness for ever more of this data among service providers and other digital actors. As discussed above, left to themselves, users will rarely if ever be in a position to resist such demands.

Interim Conclusion: "Control" in SSI Systems

We thus conclude that the claim made by SSI advocates that SSI gives users meaningful control over their (i) data and (ii) their identity, is factually wrong. Unless additional functions (Wallet Histories and ZKPs) are included, SSI solutions do not provide any significant increase in user's control "over their data" than existing solutions. Even with Wallet Histories and ZKPs, SSI can unfold its limited increase in data control only through the interplay with powerful supportive legal structures. Absent these, SSI risks turning into a

"disclosure machine" wherein users become *more* transparent and more vulnerable to "data extortion" by service providers than they often are already. The reason for this is that users' lack of control over their data derives from structural power imbalances that SSI neither can nor tries to fix.

The claim that SSI gives users control "over their identity" is harder to evaluate, as it is less clear what it means. If "identity" refers to the underlying data (attributes, VCs), then the above applies. If "identity" refers to users' social, legal-political or economic status (identity as relationship), then too the claims for SSI necessarily fall short. The reason is again its inability to alter the power imbalances in the relationship between users, issuers of credential and grantors of benefits.

PART TWO: USER PERCEPTIONS

Method

To explore users' views of SSI solutions we included several pertinent questions in a larger online survey about the use and adoption of digital identity technologies, conducted as part of the IMPULSE project.[26] The survey ran from February to June 2023. The survey link was initially disseminated through the personal and professional networks of the researchers and staff members from the IMPULSE consortium's 15 partner institutions in their countries (Germany, Spain, Italy, Denmark, Iceland, Bulgaria, Finland and France). Respondents were asked to pass the survey on further in their own personal networks (convenience sampling). This dissemination method was chosen as financial constraints prevented the use of e.g. a survey company. To incentivise participation, a donation of €2 to charitable organisations was promised per completed survey, up to a total of €500.

There were 740 valid responses (205 to the German-language instrument, 192 to the Spanish, 116 to the Italian; and 62, 60, 42, 39 and 24 to the Danish, Icelandic, Bulgarian, Finnish and French, respectively). Respondents' age ranged from 18 to 82 years (average 43.4; median 41) and were disproportionately well-educated (79% had or were currently in university education) and affluent (by income, 53% belonged to the top-20% of income earners; 16% to the top 60-80%), and 93% white. Conversely, the gender split was almost exactly 50-50 (49.5% male, 49.7% female, rest diverse or n.a.). 57.4% of respondents live in small or medium-sized towns; 25.5% in major cities, 17% in villages. Evidently, the survey population is not representative of the general population, as was to be expected given the dissemination strategy. The implications thereof for the external validity of the following analysis are discussed in closing.

The main objective of the survey was to understand user acceptance decisions regarding the IMPULSE digital identity system, an SSI-type digital identity system developed in the eponymous project. To this end, demographic, attitudinal and experiential data were collected (among others: age, gender, education, income; privacy concerns, technology openness; experience with digital identities, authentication technology preferences; willingness to use IMPULSE, reasons for non-use, and perceptions of the system). The full survey instrument is provided in the Appendix. The survey items were based on the extant literature and theories about user acceptance of digital technologies and identity systems. A full analysis of the

determinants of user acceptance is presented in Martin and Metzger (under review). Of interest here is the information that the survey provides about whether users are likely to believe that SSI systems give them "control" over their data, if such a system is described to them (RQ 1), and if so, how important this belief is for driving adoption (RQ 2).

To collect data on this, survey respondents were shown a five and a half minute long animated video explaining IMPULSE (described below), and then asked whether they "would use IMPULSE instead of the digital identity (log-in) systems [they] currently use (like username/password, smartcard, PIN, etc.), if IMPULSE were available?" Answers were collected on a Likert scale running from 1 ("certainly not") to 5 ("certainly yes").

Several questions later, respondents were given 18 different words and phrases and asked to choose all that "you feel describe IMPULSE". These included positive and negative characterisations (e.g., "convenient", "saves time", "complicated", "surveillance"). Of interest here are the following three: "IMPULSE gives me control over my data", "With IMPULSE, I can decide who gets my data", and "privacy-friendly". These items were interspersed in the middle and towards the end of the entire set of descriptors, with various other descriptors in between.

Finally, again two questions later, they were given six statements describing different aspects of "having control over one's data" and then asked to choose the three most important ones in their view.[27] The following statements were presented to them, in the order below:

- Online service providers have to ask for my consent before collecting or using my data.
- Online service providers must provide a short, easy-to-understand privacy policy explaining how they use my data.
- Online service providers cannot demand more data from me than necessary to make the service work.
- Online service providers cannot refuse me their service unless I allow them to use my data for advertising.
- Online service providers cannot use website lay-out and other tricks that manipulate me to give them extra data or data-use permissions.
- Online service providers must delete my data if I ask them to.

The logic behind these different questions and their order of appearance was as follows. Asking whether respondents would switch to IMPULSE is self-explanatory. We formulated the question so as to create a direct comparison between IMPULSE and respondents' current digital identity systems. This creates a harder test of IMPULSE true adoption potential, since this primes respondents to think about pros/cons compared to their current system.

The descriptors "IMPULSE gives me control over my data" and "With IMPULSE, I can decide who gets my data" present respondents with the standard claim SSI advocates make about these systems, formulated in two slightly different ways. The first phrase is more abstract than the second and arguably rhetorically slightly stronger (IMPULSE *gives control*), but semantically they are near-identical: to have control over one's data is to get to decide

who gets it, and vice-versa. The more general descriptor "privacy-friendly" gives respondents an alternative item to tick if they want to comment positively about the system's privacy qualities, thus reducing the risk that the two "data control" items get "abused" for general commentary on privacy. To avoid these three items priming respondents to assume that IMPULSE is particularly privacy friendly, the negative descriptors "surveillance", "creepy", "weird" and "dangerous" were also shown. The description of the IMPULSE system presented to respondents in the video was careful to explain the data flows in the system but not make any positive or negative statement about its privacy qualities or "data control".

Finally, the dimensions of data control, from which respondents were asked to pick the three most important, include both the relatively weak and formalistic requirements to ask users for consent and provide a privacy policy, and more substantive controls: deletion of data upon request, protection from manipulation, use of data for a purpose unrelated to service provision, i.e. advertising, and data extortion: demanding more data than necessary for service provision – one of the key threats SSI systems pose (cf. Part I). Of interest here is whether respondents associate data control primarily with the weaker and more formalistic dimensions consent and privacy policy (which SSI advocates regularly reference as means by which SSI operationalises data control for users), or whether they see the other, more substantive dimensions – that SSI does little to operationalise and can even actively threaten ("disclosure machine") – as critical. The data control-dimensions question was asked last, to avoid priming respondents when they answered the foregoing questions.

The IMPULSE System

IMPULSE consists of three core components: A smartphone application for the user, an enterprise module for the online service provider (installed in her IT systems), and a blockchain[28]. As a typical SSI system, IMPULSE uses decentralised identifiers (DIDs) and blockchain for communication between user and service provider; verifiable credentials (VCs) to represent the user's data and identity, and a digital wallet in the app on the user's smartphone to store the VCs in. Authentication works via facial recognition, but this is incidental.[29]

The process by which the system works and the user onboards to an online service (e.g. an online bank, a state agency, or an e-commerce site) with IMPULSE is shown in Figure 2: The service provider stores his DID and trusted issuer information ("verification information") on the blockchain (Step 0 in Figure 2 left). When the user accesses the online service (e.g. via a "Sign Up with IMPULSE"-link), the IMPULSE app accesses this verification information to check that the service provider has been registered as a trusted party. This happens automatically in the background.

The first step on the user onboarding journey with IMPULSE is that she is presented with the service provider's data use/privacy policy. She has to consent to this to be able to onboard to and use the online service. This done, she is asked to take a photo of the front and back of her state ID card or passport, and a selfie (Step 1). These are uploaded to the IMPULSE enterprise module installed in the service provider's IT system via an encrypted p2p DID channel. Prior to upload, a biometric module integrated in the IMPULSE app creates a biometric profile of the user from the selfie and stores this profile in the user's

smartphone. This profile is used when she authenticates to the online service next time. The profile is *not* shared with the enterprise module. The enterprise module (on the online service provider side) of IMPULSE then uses artificial intelligence to check that the selfie and the user's picture on her ID card match and verifies the authenticity of the ID documents. This user verification can be fully automated, but for compliance with EU eIDAS regulation, IMPULSE foresees a human in the loop (a worker on the online service provider side who manually confirms that the photos match). Next, data that the online service requires are automatically extracted from the ID card photos (e.g. name, data of birth, address). If additional data that is *not* contained on the ID card is needed (e.g. bank details), then the user enters these manually. The data are encoded in the verifiable credential that the service provider creates for the user (to be read out later by the service provider when the VC is presented during authentication) and/or stored directly by the service provider in their systems (Step 2). With verification of the user completed, all photos are erased and the service provider creates the VC for the user (Step 3). The VC is sent to the user, to be stored in the IMPULSE app digital wallet on her smartphone (Step 4). The user is now onboarded to the online service provider (Step 5).

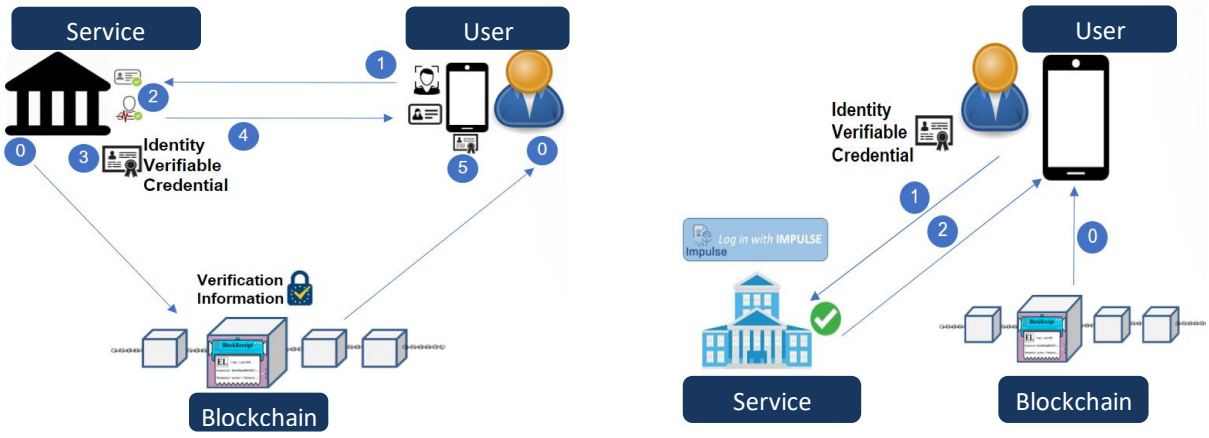

Figure 2. User onboarding to IMPULSE (left) and authentication after completed onboarding (right)

When she next wishes to authenticate to the service to use it, she goes to the service provider's webpage, chooses "log in with IMPULSE", and the IMPULSE app on her smartphone again checks the service provider's verification information from the blockchain (Step 0 in Figure 2 right). To access the verifiable credential that she will present to the service for authentication, she takes a selfie, which is checked via facial recognition against the biometric profile stored on her Smartphone. She can now present the VC to the service provider via a DID channel (Step 1). The service provider "verifies" the VC and authenticates the user (Step 2). She can now use the service. Data required by the service provider is read out from the VC or retrieved from service provider's IT systems. If additional data that the user has not yet provided to service provider subsequently becomes necessary, she can supply this in several ways. If she already has the data in the form of a further VC, perhaps issued to her by a different entity (e.g. a VC of her university diploma), she can simply provide this VC to the service provider, for him to read and extract the information from. Alternatively, if she

does not yet have the data in VC form, she can enter it manually (or in some other form, e.g. as a JPEG) to the service provider. Depending on the context, the service provider may again turn this into a VC for her to store in her digital wallet – and potentially share with other service providers. (She could of course refuse to provide the additional data, at the cost of foregoing the feature that it is needed for or even giving up using this service provider entirely, if he makes the new data request a precondition for further use.)

While IMPULSE's use of facial and document recognition is distinctive, the basic architecture and data flows are typical of SSI systems: Users hold their data in the form of VCs – issued to them by service providers or other third parties – in their digital wallet, and provide the VCs to service providers who read and/or extract data from them in the context of service provision via DID channels. We can therefore treat survey respondents' reactions to IMPULSE as at least indicative for how users might view other SSI systems, too.

Such a DID/VC/digital wallet-based system provides the user with greater convenience and security compared to many existing processes. Important data (e.g. one's legal identity, as attested by one's state ID card) can be easily and securely provided to service providers and other entities. The data is in a format that can be very quickly and automatically verified as correct. With IMPULSE, the use of facial recognition in place of passwords or PINs provides additional convenience and security. Together with the document recognition function, this allows even onboarding to sensitive services – which conventionally would require in-person or video-chat sign up – to be done swiftly and conveniently from home.

As discussed in Part I, what such a system does *not* however do, is to provide users with meaningful "control" over their data. But would users recognise that too? To understand this, in the survey, we described the onboarding and authentication processes and basic data flows in IMPULSE to survey respondents by means of an animated video, narrated in the respondents' languages (Spanish, Bulgarian, German etc.). Online banking was used as an illustrative example use case.[30]

The video explained that to use internet services like online banking one needed a "digital identity", often a username/password and associated data like one's address, and that one usually created this digital identity when signing up for the service. It noted that people often chose simple, insecure passwords, and introduced IMPULSE as a novel smartphone-based digital identity system that instead used facial recognition. It outlined the main steps in the onboarding and authentication journeys: accessing the service provider's website, taking the photos of ID card and selfie, AI-based analysis of the photos by IMPULSE to verify the user and create the verifiable credential (called "digital identity" in the film to avoid tech jargon), deletion of the photos and secure, encrypted storage of the VC ("digital identity") on the user's smartphone. The film was careful to avoid broaching the topic of "controlling one's own data". It noted that for the service provider to make the service (e.g. banking) available, the user might have to provide data stored on their phone (i.e. VCs) or additional data, and noted that "the service provider may sometimes keep this data, if you agree, or if they have to for legal or technical reasons." The issue of data extortion discussed above (having to provide more data than necessary for the service, or be refused service) was not raised.

Results for Research Question 1: User Perceptions of "Control"

Figure 3 shows the percentages of respondents who (1) stated that they were "concerned" or "very concerned" about online privacy[31], (2) who chose "privacy friendly" as a description of IMPULSE, (3) who chose the descriptor "IMPULSE gives me control over my data", (4) who chose the descriptor "with IMPULSE I can decide who gets my data", and (5) who chose the descriptor "convenient" for IMPULSE. Of all descriptors, "convenient" was the one most frequently chosen.

As we can see, the sample population was quite concerned about online privacy with 60.7% choosing 4 or 5 on the Likert scale. But only 28.5% believed IMPULSE was privacy friendly[32]. Even fewer believed that the system would "give [them] control over their data" (13.2%) or felt that "with IMPULSE [they] could decide who gets [their] data" (14.2%). By contrast, over 37% thought it was "convenient".

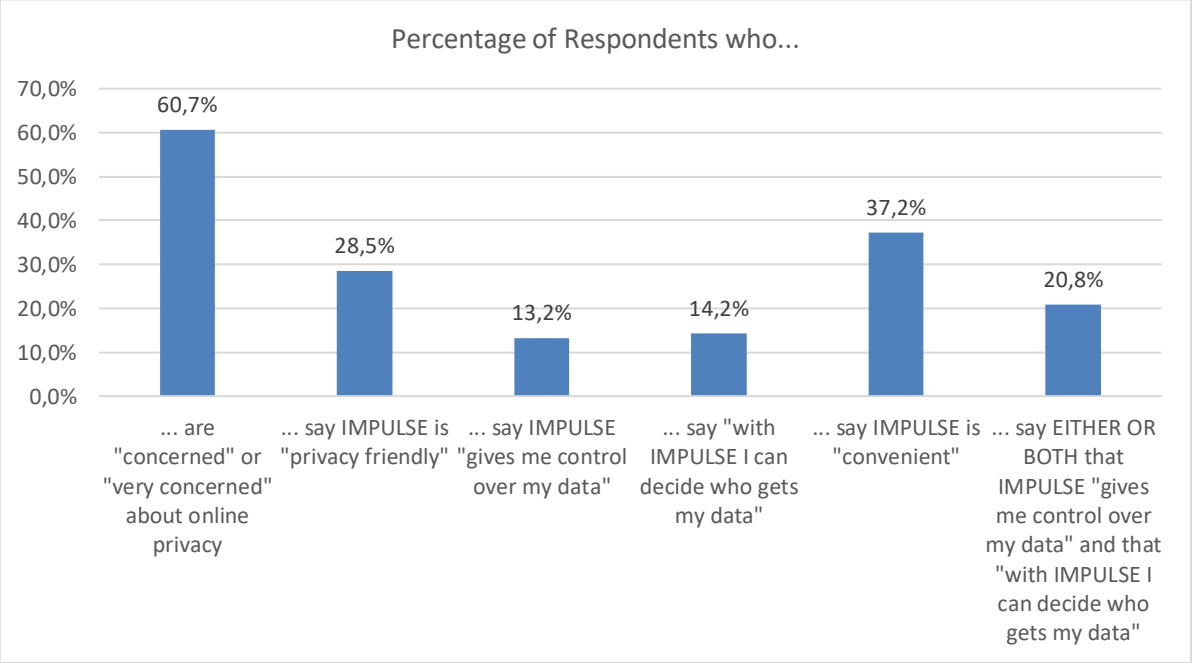

Figure 3. User perceptions of IMPULSE

The two descriptors, "gives me control" and "I can decide who gets my data", are semantically near-identical. Presumably they capture the same underlying phenomenon/belief structure, and respondents who ticked one ought logically to tick the other, too. In practice, though, only ~32% of the respondents who chose one of the two descriptors also chose the other (Figure 4). The rest chose only one of the two. If we add these two groups together, the share of respondents who seem to think that IMPULSE in some way gives them control or decision-making capability over their data rises to 20.8% of the total sample population. (Figure 3). However, the fact that nearly 70% of the respondents who picked either of these two semantically identical descriptors failed to also pick the other one, suggests that most

struggled with the concept of "having control"/"getting to decide" over their data. Indeed, only 6.6% of the total sample population picked both.

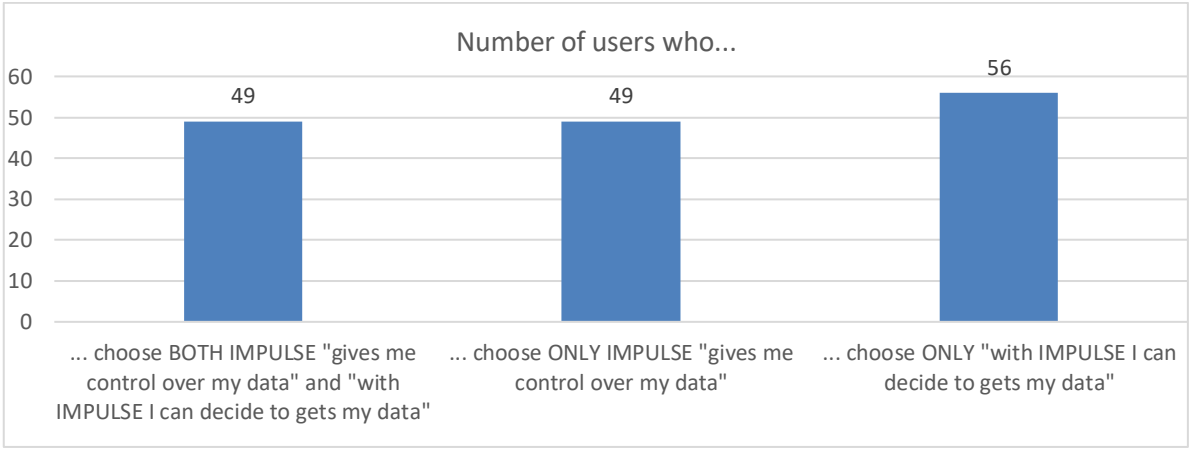

Figure 4. User choices of "control"/"decide"-descriptors

This is also supported by how respondents answered the question about different dimensions of having "control over data" (Figure 5). Three items – that service providers must ask for consent, cannot demand more data than necessary to make the service work (data extortion), and must delete data upon request – are chosen as among the "most important" by relatively large shares of respondents, viz. 69%, 66% and 59%. Conversely, only 31% to 37% choose any of the other three items (no manipulation to get data, privacy policy, no use of data for advertising as a condition of service access).

Yet only 22% chose *all three* of the "popular" items (consent, no data extortion, deletion upon request) (Figure 5). The other 18 possible three-item combinations (e.g. consent, no data extortion, no manipulation, etc.) are each chosen by just ~3% to ~11% of respondents (not shown here). This suggests that while some items clearly resonate with far more respondents than others, overall the respondents struggle to form a systematic and encompassing concept of "having control over their data". In turn, this suggests that even those respondents who selected either or both of the "control over my data"/"get to decide who gets my data"-descriptors, likely did *not* have a well-developed, rigorous concept of what such control would actually entail or require. On the contrary, rather strikingly 70.4% of the respondents who selected the "control"-descriptor, and 63.8% of those who chose the "decide"-descriptor, *also* chose "service providers cannot demand more data from me than necessary" as one of their three most important dimensions of data control. But as we have seen, this is precisely *not* ensured by IMPULSE, or any other SSI system.

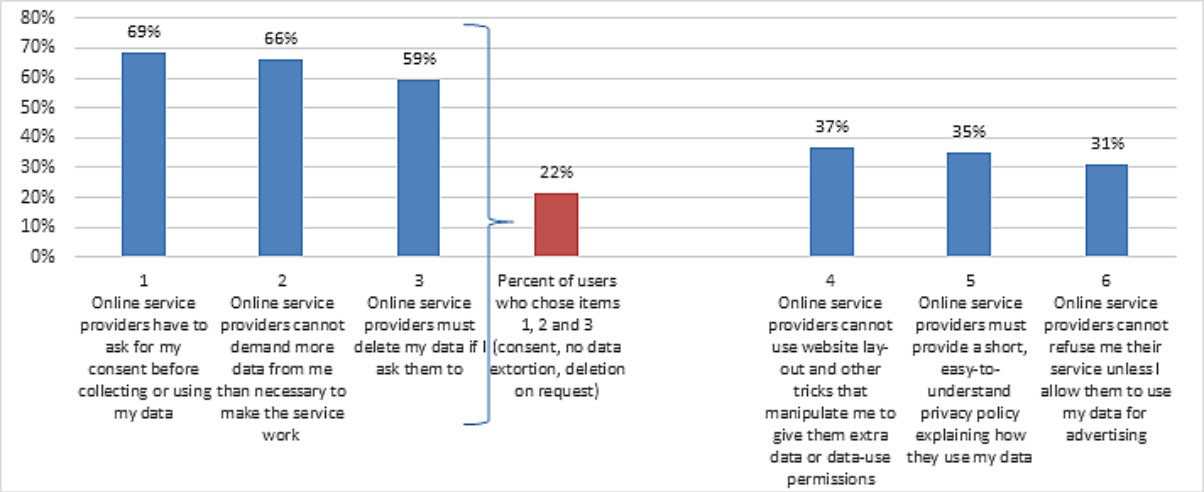

Figure 5. Percentage of users who regard selected dimensions of "data control" as among the 3 "most important" dimensions

This raises the question of who the 20.8% of respondents are who seem to believe that IMPULSE gives them "control over their data". Figure 6 compares them to the entire survey (sample) population on several demographic variables. What stands out is their *lack* of differentiation. Whether in regard to age, gender distribution, income levels, place of abode, technology openness and privacy concern, the subpopulation of respondents who chose either the "control"- or "decide"-descriptor is virtually indistinguishable from the general population: they are on average a little over 40 years old[33], about 50-50 male/female, highly educated with ~80% having completed or currently in university, and earn above-average incomes. Over half live in small towns; about a quarter each in major cities or villages. Most are open or very open to new technology, but also concerned or very concerned about online privacy. There is more national variation, with respondents from Germany, Denmark and Spain ticking the two descriptors at about the rate of the sample population average, Italians and Icelandics somewhat less, and responses from Finland and France swinging wildly (likely also due to their smaller sample sizes) (data shown in Appendix). However, lacking strong theoretical reasons to expect national divergences here, we put this down to random variation.

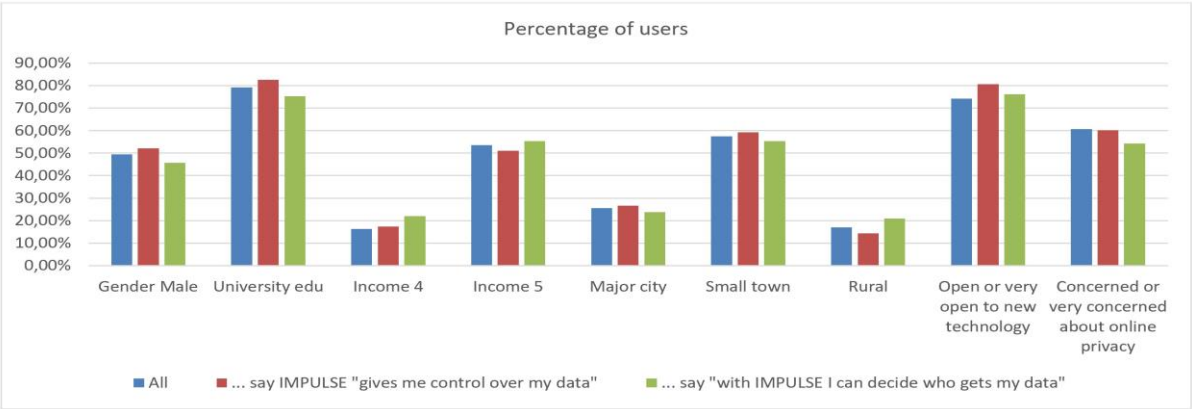

Figure 6. Demographic and attitudinal characteristics of different respondent (sub) populations

Interim Conclusion: Findings related to Research Question 1

Regarding RQ 1 – whether users, if given a description of how an SSI system works, are likely to conclude that it will put them in control of their data – we thus conclude: Most users are *not* likely to conclude that SSI will put them in control of their data. In our survey, at a maximum 20.8% of respondents seem to have concluded something along these lines about IMPULSE. Only 6.6% seem to have concluded that IMPULSE gave them control of their data, and had a sufficiently rigorous understanding of this notion to tick *both* descriptors. (For further discussion of external validity see below.)

Furthermore, the survey suggests that users are unlikely to have a systematic, broadly shared understandings of what having "control over their data" means, and may struggle to apply what notion they do have in this regard, to assessing SSI solutions in a rigorous fashion. This is unsurprising, given the abstract nature of the concept and the sophisticated understanding of digital technology and ecosystems required to apply it rigorously. That in turn however implies that users may be quite suggestible: as they lack a clear idea of what "control over their data" means, it could be fairly easy to persuade them, at least temporarily, that SSI (or some other technology) gives them this. I.e. while the number of users who by themselves are likely to conclude that an SSI solution would offer them "control" is low, a professional marketing campaign may well be able to change that dramatically, at least temporarily.

Thirdly, no specific demographic or attitudinal characteristics distinguishing those who seem prone to concluding that SSI gives them "control over their data" could be identified.

Results for Research Question 2: Effect on Adoption

To understand whether the belief that an SSI solution provides "control over data" materially affects adoption decisions, we conducted Ordinary Least Squares regression analysis using the R software package.

The results are shown in Table 1. The *dependent variable* in all models was respondents' self-reported likelihood of switching to the IMPULSE system, should the system become available. The item was measured on a Likert scale from 1 to 5 (from "certainly not" to "certainly yes").

The explanatory variables of interest here are:

- *ControlPerception:* an ordinal variable with values 0, 1, 2, that captures whether respondents ticked none, one or both of the "control over/get to decide who gets my data"-descriptors.
- *ControlDescriptor:* a dummy variable with value 1 if respondents ticked the "control over my data"-descriptor, and value 0 if not.
- *DecisionDescriptor:* a dummy variable with value 1 if respondents ticked the "get to decide who gets my data"-descriptor, and value 0 if not.

The following control variables were included in different models:

- *Age*: a continuous variable running from 18 to 82, capturing the respondents' age.

- *Gender_Male:* a dummy variable with value 1 if the respondent is female, and 2 if male (Respondents who reported their gender as "diverse" were excluded, due to their very small number).
- Education: a categorical variable with value 1 if the respondent has completed high school (but not attended university), and value 2 if the respondent has completed or is attending university. The value 0 is given if the respondent has not completed high school, and is the reference category. R reports the coefficients for each value that the variable takes separately in the regression table (except the reference category). Thus the table includes *Edu_HighSchool* (when the education variable has the value 1) and *Edu_University* (educational variable value of 2).
- Income*:* a categorical variable with values 1 to 5, corresponding to the income quintile (based on EuroStat data) that the respondent's self-reported income is in. (I.e. value = 2 if the respondent's income situates them in the second quintile; value = 3 for the third income quintile, etc.). The reference category is the first quintile. R again reports the coefficients for each value that the variable takes on separately (except for the reference category), viz. *Income_Q2* for second quintile, *Income_Q3* for third quintile, etc.
- Ethnicity: a categorical variable with value = 1 if the respondent self-reported to be "white" (reference category), value = 2 if the respondent self-reported to belong to an ethnic minority, and = 3 if he or she refused to report their ethnic identity. The reference category is 1; R reports the coefficients for Ethnicity = 2 (*Ethnic_Minority*) and = 3 (*Ethnicity_Refuse*) separately.
- *TechOpenness:* an ordinal variable running from 1 to 5, capturing the respondent's self-reported enjoyment of new technology.[34]
- *PrivacyConcern:* an ordinal variable running from 1 to 5, capturing the respondent's self-reported level of concern over their online privacy
- *Usability:* an ordinal variable running from 0 to 5, with 1 point added for each of the descriptors "convenient", "easy to use", "makes signing up for services easier", "makes logging in easier" that the respondent ticked.
- *Country*: a categorical variable with country coding for Bulgaria (reference category), Denmark, Finland, France, Germany, Iceland, Italy, Spain.
- An interaction term between the *country* variable and *PrivacyConcerns*.

In further models we also added interaction terms between the country variable and income quintiles, and further controls for digital skills and place of abode (rural village, major city). For simplicity of presentation these are not included here as they were consistently not significant and had no impact on the significance or effect size of the three "data control"-variables that are of interest here. A detailed analysis of all factors determining adoption decisions is in preparation (Martin and Metzger u.r.). Here we concentrate on the impact of the three "data control"-variables.

Model (1) in Table 1 includes all control variables except for Country, and the Country-*PrivacyConcerns* interaction. The *ControlDescriptor* and *DecisionDescriptor* are both significant at the $p<.001$ level, with a positive sign and similar coefficient sizes, a pattern that continues in the following models. This indicates that the two variables are indeed capturing

the same underlying phenomenon/belief structure in respondents. Significant at the p<.001 level, too, are *Usability* and *TechOpenness*. *PrivacyConcerns* is negative but *not* significant. The other control variables are likewise not or only weakly significant, with the exception of *EthnicityRefuse*.[36] (For further discussion see Martin and Metzger u.r.). In Model (2) we add the Country variable. *PrivacyConcerns* now becomes significant at the p<.001, with a negative sign. The other variables remain largely unchanged – including *ControlDescriptor* and *DecisionDescriptor*. The Country controls are all negatively signed and mostly significant at the p<.001 level.[37]

Model (3) includes the Country-*PrivacyConcerns* interaction. This turns *PrivacyConcerns* positive (!) albeit robbing it of significance. *ControlDescriptor* and *DecisionDescriptor* remain significant at the p<.001 level (as do *Usability* and *TechOpenness*), with similar coefficient sizes as before. In Model (4) we use *ControlPerception* instead of *ControlDescriptor* and *DecisionDescriptor*; it too is positive at the p<.001 level and coefficient size is similar to that of the two Descriptor variables in the previous models.

In summary, the "data control"-variables are significant at the p<.001 level across all models, with a consistently positive effect. Notably this is robust to the inclusion of country dummies. While these cause the significance, coefficient size and, with interaction effects, even direction of *PrivacyConcerns* to bounce around considerably; these remain largely unchanged for the "data control"-variables. This is striking, as one might expect "data control" to have some relationship to privacy attitudes and these variables therefore to move together or affect each other. Evidently though, this is not the case. Moreover, while the swings of *PrivacyConcerns* suggest that how privacy is understood – at least with respect to IMPULSE adoption – has a strong country-specific dimension, this is not the case for the control variables. Finally, several other factors are also important for respondents' adoption decisions, in particular perceptions of usability and their personal openness and interest in new technology.

Table 1. Results of the Regression Analysis

	Dependent variable: *Self-reported likelihood of adopting IMPULSE*			
	Model 1	Model 2	Model 3	Model 4
Age	0.006*	0.008**	0.009***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)
Gender_Male	-0.072	-0.095	-0.078	-0.078
	(0.086)	(0.084)	(0.084)	(0.084)
Edu_HighSchool	-0.401**	-0.353*	-0.358*	-0.357*
	(0.193)	(0.193)	(0.196)	(0.196)
Edu_University	-0.231	-0.193	-0.187	-0.186
	(0.160)	(0.161)	(0.163)	(0.162)
Income_Q2	0.391*	0.255	0.281	0.281
	(0.228)	(0.224)	(0.228)	(0.227)
Income_Q3	0.070	-0.049	-0.074	-0.074
	(0.207)	(0.207)	(0.209)	(0.209)
Income_Q4	0.079	-0.083	-0.092	-0.093

	(0.188)	(0.189)	(0.192)	(0.191)
Income_Q5	0.228	0.088	0.080	0.080
	(0.171)	(0.172)	(0.175)	(0.174)
Ethnic_Minority	-0.137	-0.158	-0.082	-0.082
	(0.236)	(0.236)	(0.247)	(0.246)
Ethnicity_Refuse	-0.616**	-0.502**	-0.486**	-0.486**
	(0.239)	(0.235)	(0.236)	(0.235)
TechOpennness	0.217***	0.190***	0.188***	0.188***
	(0.044)	(0.044)	(0.045)	(0.045)
PrivacyConcern	-0.032	-0.118***	0.244	0.245
	(0.041)	(0.043)	(0.229)	(0.228)
Usability	0.335***	0.310***	0.298***	0.298***
	(0.027)	(0.028)	(0.028)	(0.028)
ControlDescriptor	0.388***	0.362***	0.374***	
	(0.134)	(0.133)	(0.135)	
DecisionDescriptor	0.346***	0.351***	0.348***	
	(0.132)	(0.130)	(0.130)	
ControlPerception				0.361***
				(0.072)
Country_Denmark		-0.990***	0.865	0.870
		(0.242)	(1.121)	(1.119)
Country_Finland		-0.684***	1.847	1.849
		(0.256)	(1.253)	(1.252)
Country_France		-0.780***	1.428	1.423
		(0.286)	(1.503)	(1.501)
Country_Germany		-0.843***	0.928	0.929
		(0.206)	(1.105)	(1.104)
Country_Iceland		-0.865***	0.809	0.812
		(0.235)	(1.159)	(1.157)
Country_Italy		-0.538**	1.433	1.437
		(0.214)	(1.168)	(1.167)
Country_Spain		-0.325	0.859	0.865
		(0.201)	(1.120)	(1.118)
Denmark_PrivacyCon			-0.431*	-0.432*
			(0.256)	(0.256)
Finland_PrivacyCon			-0.614**	-0.615**
			(0.295)	(0.295)
France_PrivacyCon			-0.501	-0.498
			(0.344)	(0.343)
Germany_PrivacyCon			-0.397	-0.398
			(0.244)	(0.244)
Iceland_PrivacyCon			-0.366	-0.366

			(0.259)	(0.258)
Italy_PrivacyCon			-0.445*	-0.446*
			(0.259)	(0.258)
Spain_PrivacyCon			-0.237	-0.239
			(0.244)	(0.244)
Constant	1.884***	2.956***	1.298	1.294
	(0.338)	(0.410)	(1.100)	(1.099)
Observations	635	635	635	635
R^2	0.312	0.353	0.361	0.361
Adjusted R^2	0.295	0.329	0.330	0.331
Residual Std. Error	1.034 (df = 619)	1.009 (df = 612)	1.008 (df = 605)	1.007 (df = 606)
F Statistic	18.688*** (df = 15; 619)	15.153*** (df = 22; 612)	11.772*** (df = 29; 605)	12.212*** (df = 28; 606)

Note. *p<0.1; **p<0.05; ***p<0.01; standard errors in parentheses.

Interim Conclusion: Findings related to Research Question 2

With regard to RQ 2 – in as far as users believe that SSI gives them "control" over their data, how important is this for driving adoption? – we can thus conclude that such a perception may well have a significant positive impact on adoption. While this is highly plausible, given extant survey evidence that users highly value control over their data (PWC 2021, Spiekermann and Korunovska 2017), the regression results must be interpreted cautiously. At the point in the survey where respondents were asked about adopting IMPULSE, the issue of data control had not yet been raised. Thus it remains unclear whether these respondents really had already formed an idea about IMPULSE giving them control over their data, or whether this notion only came to them in the subsequent items, when it was explicitly raised (which would suggest that the "data control"-variables in the regression are picking up some other underlying factor).

CONCLUDING DISCUSSION

Conclusions and Implications

This paper concludes: Firstly, the claim that SSI systems provide users with "control" over their data and/or identity, and does so via its technology and architecture, is incorrect. On the contrary, SSI by itself not only provides users little more control than established identity technologies do, it is even liable to render them more vulnerable to excessive demands for data; i.e. leave them less in control. If Wallet Histories and ZKPs, which are optional and in the case of ZKPs, technologically immature features, do get built into SSI systems, this *could* mildly improve users' data privacy and control over their data. However, Wallet Histories and ZKPs depend on a conducive legal-institutional framework (e.g. the GDPR) to have this effect. By contrast, without an appropriate legal framework, SSI is liable to turn into a disclosure machine for user data.

Secondly, when presented with the workings and data flows of an SSI system, most users do not seem to conclude that it provides them with a particular degree of "control" over their data. A minority of users however does seem to come to this belief. It remains unclear who these users are or why exactly they reach this interpretation. Indeed, many users seem to struggle to form and apply a clear and coherent idea of what "having control over their data" would mean – unsurprisingly so, given that this is a rather abstract concept that requires a good understanding of digital technologies and ecosystems. However, this means that users may also be quite suggestible in this regard, not least as people seem to find the idea of having "control" over their data appealing.

Consistent with this is the third finding, namely that perceptions that an SSI system gives users' "control" over their data, are significantly positively correlated with the decision to adopt the system.

These findings have several implications. Firstly, marketing campaigns pushing the message that SSI gives users "control" may well boost adoption, at least in the short-term. Given that this claim is objectively incorrect, it is however to be feared that a backlash may follow, too. This would be obviously detrimental to SSI, but also to trust in society in general: users and citizens may well prove less forgiving – and less forgetful – of a broken promise to give them "control" over something as personal and sensitive as their data, than they perhaps are of an overhyped tech gadget or golf club.

These points do not imply a rejection of the identity solutions being developed under the SSI label. On the contrary, the VCs, DIDs and digital wallets offer real benefits for users and service providers. However, the benefits lie in their greater usability, reduced friction and costs and greater data security compared to existing solutions and customary processes. A second implication is thus that SSI developers should focus their development and promotional efforts on these advantages, and avoid the unconvincing and misleading rhetoric about giving users "control". Usability has regularly proven tremendously important for driving adoption (and our own regression results consistently find that usability is significantly predictive of a positive adoption decision), suggesting that this is promising strategy. Thirdly, this may imply the need for a new term for this class of identity solutions, since they hardly make the user "self-sovereign" in any meaningful sense. Fortunately, a good term for them already exists: user-centric identity solutions.

A final and broader implication of this research is the critical importance of locating the design and the analysis of technical systems (whether SSI systems or some other system) within the specific legal-institutional environment that they are to be used in. Depending on the legal-institutional context, the same system may produce quite different effects. Technology design and analysis that ignores these realities is liable to generate unintended and unrecognised consequences, not all of which may be desirable. Conversely, design and analysis that is sensitive to the diverging and idiosyncratic effects different legal-institutional contexts can have on technology use, and the design requirements this implies, can create genuinely beneficial technological solutions.

Limitations and Future Research

This study has limitations. Most importantly, the survey population is not representative, but disproportionately well-educated, well-off and white. However, it is not clear what effect this is likely to have had on the questions of interest here. Well-educated respondents may be harder to deceive than less well-educated ones, implying that a more representative survey might have found higher numbers agreeing that the IMPULSE system gave them "control". However, even well-educated respondents might often lack the specialised knowledge about digital technologies and ecosystems required to accurately assess this question, which would moderate the effect of our respondents being better educated. It is notable that on a number of measures, the attitudes and behaviours related to digital identity systems revealed by our survey population, was similar to the attitudes and behaviour of the general population, as found in several representative surveys.[35] This strengthens our confidence in our specific results. In any case, the larger implications of our argument – that the claim that SSI gives users "control" is untrue, and that marketing SSI on the basis of this claim highly inadvisable – remain unaffected by this possible bias.

A further limitation is that with regard to the "gives me control over my data"/"I get to decide who gets my data"-descriptors, our survey instrument could not distinguish between those respondents who did *not* tick these because they consciously disagreed with these statements, and those who did not tick them because they were unsure or confused. Length constraints on the survey instrument prevented more detailed questions on this point. However, as the key question of interest was what share of respondents consciously *agreed* with these statements, this limitation seems less critical, if still unfortunate. Further research could address this issue by disambiguating and also adding more breadth to this question area. For instance, a number of reflective items, forming latent constructs, could better grasp the nuances underlying agreement or disagreement.

Finally, as already noted, the issue of "data control" was only raised after respondents had been asked whether they would adopt IMPULSE. While it is intuitively plausible that the belief that the system gives one "control" should increase adoption likelihood, it is possible that the significantly positive relationship between ticking the "control"/"decision"-descriptors and the adopting IMPULSE does *not* in fact reflect respondents – at the time of being asked about adoption – having already formed an idea that IMPULSE gives them control and *therefore* wanting to adopt it, but rather some underlying factor. A more systematic test of this relationship would therefore be desirable, for instance by splitting a population sample in two and giving them different information about the system (e.g. one that brands the system as "giving control" and one that skips the topic). However, given the appeal that "data control" has to users, it is likely that such tests would find substantively the same relationship.

More broadly, this paper points to the value of future research that would investigate in more detail how ordinary users understand and make sense of terms like "control", "data sovereignty" or also "data protection" and "privacy friendly" and other data-governance terms. While not the main focus of this research, our paper provides evidence that users may often struggle to form a clear – let alone a coherent – understanding of what these terms mean. How ordinary users understand these terms may well also diverge – perhaps markedly – from how they are used in the specialist discourses of law, policy, technology and

academia, where they were originally coined. This would be quite problematic. These and related terms are the key concepts in which data governance is discussed in the specialist legal, policy and technology communities. Rapid digitisation of (almost) all areas of life means that data governance is increasingly relevant to people's everyday lives. But if the terms that data governance is discussed in – and with which products are advertised to users – are unintelligible or mean quite different (and maybe idiosyncratic) things to lay users than to the specialists, the stage is set for serious miscommunication. Understanding better how ordinary users in fact make sense of these terms is the first step to resolve and avoid such confusion.

ENDNOTES

1. Calculated by summing the total funding for projects contained in the https://cordis.europa.eu/ data base that referred to "self-sovereign identity" in their project description or referenced the term in their key words. No effort was made to disentangle how much of this funding went to SSI development in a narrow sense, and how much to (broadly) related ancillary activities. Accordingly, the headline number should be treated as indicative only.
2. For example, in Germany the Economics Ministry funding program "Schaufenster Sichere Digitale Identitäten" has reportedly provided some €45 million to four projects (Heeger 2020), at least two of which describe their objective as developing SSI solutions (the projects SDIKA [https://www.sdika.de/] and ID-Union [https://idunion.org/]).
3. RQ 1 focuses only on whether users believe SSI would put them "in control over their data", not whether they believe it would give them "control over their identity." The reason for this is twofold. Firstly, as we will see shortly, in many – but not all – regards, to have "control over one's data" and to have "control over one's identity" are identical statements, and the SSI discourse tends to treat them as synonymous. Secondly, while the notion of "having control over one's identity" can be understood in a way that is *not* synonymous with having control over one's data, to construe "having control over one's identity" in this other way turns it into an (even more) complex and abstract notion than "having control over one's data" is, that would be difficult to implement in a quantitative survey.
4. We are grateful to Henk Marsman for alerting us to this thesis. (Private communication from Henk Marsman, 12 February 2024.)
5. We are grateful to Henk Marsman for alerting us to the epicenter.works submission. (Private communication from Henk Marsman, 14 January 2024.)
6. GRADIANT was coordinator of the EU Horizon 2020 IMPULSE project, which developed a prototype SSI solution and studied its impact in the context of public services. The authors were too members of IMPULSE project, and this study was funded as part of the IMPULSE project (see acknowledgements).
7. This asymmetry derives from the sheer complexity of modern data processing and data value chains. Combined with most people's limited attention for what is, *prima facie*, an excruciatingly boring topic, this makes it difficult for even the best intentioned data processor to communicate to their users what is really going on. Of course, many data processors are not well-intentioned, and exploit the possibilities offered by impenetrable legalese and/or manipulative technology design (e.g. "dark patterns").
8. Digital identity systems generally also include some security features, like passwords, PINs or public/private keys.
9. For example, Allen (2016), in what is a foundational document of the SSI movement, speaks about placing identity "back under [the users'] control". Users should be "the rulers of their own identity"

and “the ultimate authority on their identity. They should always be able to refer to it, update it, even hide it.” In the same vein, Sedlmeir et al. (2021) describe a digital identity as “something that a subject has”. Similar language can be found in Dunphy und Petitcolas 2018, Lim 2020, Liu et al. 2020, Bazharnova et al. 2019, Richter and Anke 2021 and Kubach et al. 2020 and references quoted therein.

10. See again Sedlmeir et al. (2021): “A digital identity consists of attributes that can be revoked, deleted, transferred, or exchanged”
11. Understanding the main value of identity to lie in the status/entitlements it confers also has implications for how to think about the plight of undocumented migrants (a use case for SSI frequently referenced in the literature): arguably, their problem is not so much a lack of identity documents, as the unwillingness of states to issue them with new and valuable legal-political status and rights. (Something the states could of course do at any time, if they were willing.)
12. This paragraph is based on Strüker et al. (2021), Sedlmeir et al. (2021), Lim (2020), Kubach et al. (2020)
13. Amazon, like many other internet service providers, may also be able to monitor users and other visitors to the Amazon website through other technologies, like browser cookies, irrespective of whether the visitor has an Amazon account or is logged in.
14. This paragraph is based on Strüker et al. (2021), Lim (2020), Kubach et al. (2020.)
15. Again it is worth remembering that both Google and Facebook, the two most widely used SSO systems, have many further ways of tracking users’ behaviour across the internet, irrespective of whether they use their Gmail or Facebook log-ins to access other services (Christl and Spiekermann 2016).
16. We are indebted to Xavier Martinez and Jaime Loureiro from GRADIANT for talking the time explain SSI architectures, their building blocks, and data flows within them, to us in detail. All remaining errors, needless to say, are our own, as well as the larger argument about SSI, and should not be construed as representing the views of GRADIANT.
17. For a detailed technical discussion of VCs see the W3C draft standard (W3C 2024).
18. For a detailed technical discussion of DIDs see the W3C standard (W3C 2022).
19. Typical transactions could be, for example, presenting one’s state identity card (or information derived from it) in VC form to a bank to prove one’s identity in order to open an account or for a financial transaction, presenting certificates (e.g. one’s child’s birth certificate) to state agencies to access social services, or one’s educational credentials to a prospective employer
20. The (German language) data protection notice on the Amazon.de site including the additional note on Cookies is 13 pages (5026 words) long, and the general contractual terms and conditions 18 pages (6965 words). The additional note on programmatic advertising could not be opened. (Accessed 19 May 2023)
21. Using SSI would though likely be more user-friendly than the current manual data entry, and could provide Amazon with more trustworthy data about the user, if the VC was based on e.g. her state identity card. Among other things, this could reduce the scope for online fraud.
22. See the extensive discussion of privacy risks with VCs in the W3C draft standard for Verifiable Credentials (W3C 2024).
23. The privacy policy and related documents on cookies and interest-based ads of Abebooks.com, for instance, runs to 4348 words; the privacy policy of Alibaba.com to 5560 words. (Accessed 19 May 2023)
24. A detailed account of how the “Building Blocks” system functions is provided in Wang and De Filippi 2020.
25. For instance, merchants from whom refugees made purchases via the Building Blocks system, as well as the payments processors, could not learn the refugees’ identities or account balances. The

biometric data, identifiers, and underlying identity information were stored separately (Wang and De Filippi 2020).

26. IMPULSE – Identity Management in PubLic SErvices (www.impulse-h2020.eu) was an EU Horizon 2020-funded project (European Commission Grant Agreement 101004459) that ran from 02/2021 to 01/2024 and aimed to develop a smartphone-based SSI-style identity solution, trialling this in six pilots mostly in local public administrations across Europe. The authors, and the Fraunhofer Society, participated as partners in the project, responsible mainly for social-political and economic impact assessments, standardisation and roadmapping. Fraunhofer and the authors have no commercial or financial interest in the IMPULSE solution.
27. The exact formulation of the task was as follows: "For many people, it is very important to have 'control over their data'. Having 'control over your data' has many dimensions. Below is a list of some dimensions. Please choose the three most important in your view."
28. IMPULSE uses the European Blockchain Services Infrastructure (EBSI) blockchain (https://ec.europa.eu/digital-building-blocks/sites/display/EBSI)
29. For a detailed technical description of the IMPULSE system and its building blocks see Deliverables D2.10V2, D5.1, D5.2 and D5.4 of the IMPULSE project (Martínez et al. (2023a), Martínez et al. (2023b), Martínez et al. (2022)). The following description is based on and partly taken from these documents, as well as from extensive discussions with the IMPULSE developers from GRADIANT, who are also the lead authors of these Deliverables.
30. It can be viewed at https://www.youtube.com/@impulse_EU. Online banking was chosen as example because it is increasingly common, so respondents were likely to be familiar with it.
31. The question asked "How concerned are you about your privacy when you access services on the internet?" with responses collected on a Likert Scale from 1 ("Not concerned at all") to 5 ("Very concerned"). We have coded "4" on the Likert scale as "concerned" (Note that in the survey instrument, only "1" and "5" were explicitly described as signifying "not concerned at all" [1] and "very concerned" [5] respectively).
32. It is worth noting that respondents who described themselves as "concerned" (Likert 4) or "very concerned" (Likert 5) about online privacy were slightly *more* likely to describe IMPULSE as privacy friendly than those who gave themselves lower "privacy concern" scores.
33. The average age of the survey population is 43.4 years; that of the respondents who chose the descriptor "gives me control over my data" is 41.5, of the respondents who chose "with IMPULSE, I can decide who gets my data" is 43.9.
34. The survey item asked respondents where they stood with regard to the statement "I enjoy trying out new technologies" (Likert scale 1 ("strongly disagree") to 5 ("strongly agree")).
35. Attitudes to using multiple digital identities, experience with passwords, and use of digital public services by our respondents were notably similar to those found in the Initiative D21 E-Government Monitor, a representative survey conducted annually in Germany, Switzerland and Austria.
36. We interpret this value of the ethnicity variable (refusal to disclose ethnicity) as indicating a particularly intense concern over privacy. Note that unlike *PrivacyConcerns*, the sign and significance of *EthnicityRefuse* remains unaffected by the inclusion of the Country variable. This may be related to the small number of respondents (29 individuals) who refused to give their ethnicity, compared to the much larger number (>440) who reported a hightened level of *PrivacyConcerns*. For full discussion see Martin and Metzger (u.r.).
37. We can interpret this to mean that – other things equal – respondents in Bulgaria (the reference category) are significantly more likely to express intention to adopt the system than respondents elsewhere.

REFERENCES

Allen, C. (2016). The path to self-sovereign identity. https://github.com/WebOfTrustInfo/self-sovereign-identity/blob/master/ThePathToSelf-SovereignIdentity.md

Bazarhanova, A., Lindman, J., Magnusson, J., Nilsson, A., & Chou, E. (2019). Blockchain-based electronic identification: Cross-country comparison of six design choices. In Proceedings of the 27th European Conference on Information Systems (ECIS). https://aisel.aisnet.org/ecis2019_rp/79

Cheesman, M., & Slavin, A. (2021). Self-sovereign identity and forced migration: Slippery terms and the refugee data apparatus. In E. E. Korkmaz (Ed.), Digital identity, virtual borders and social media (pp. 10–32). Edward Elgar Publishing. https://doi.org/10.4337/9781789909159.00006

Christl, W., & Spiekermann, S. (2016). Networks of control: A report on corporate surveillance, digital tracking, big data & privacy. Facultas. https://crackedlabs.org/dl/Christl_Spiekermann_Networks_Of_Control.pdf

Clifford, D., & Ausloos, J. (2017). Data protection and the role of fairness (CiTiP Working Paper). KU Leuven Centre for IT & IP Law.

Deutscher Sparkassen- und Giroverband, Bundesverband deutscher Banken, Bundesverband Deutscher Volksbanken und Raiffeisenbanken, Commerzbank, Deutsche Bank, & ING Deutschland. (2021). Gemeinsames Positionspapier: Selbstsouveräne Identitäten (SSI).

Doesburg, J. (2023). Measures against over-asking in SSI and the Yivi ecosystem [Master's thesis, Radboud University].

Dunphy, P., & Petitcolas, F. A. P. (2018). A first look at identity management schemes on the blockchain. IEEE Security & Privacy, 16(4), 20–29. https://doi.org/10.1109/MSP.2018.3111247

Echikson, W. (2020). Europe's digital identification opportunity. The Centre for European Policy Studies (CEPS).

Epicenter.Works. (2022). eIDAS policy paper. https://epicenter.works/content/eidas-policy-analysis-english

Gonzalez Fuster, G., & Gellert, R. (2012). The fundamental right of data protection in the European Union: In search of an uncharted right. International Review of Law, Computers & Technology, 26(1), 73–82.

Heeger, V. (2020, May 25). Digitale Identitäten: Zehn Ökosysteme für Deutschland. Tagesspiegel Background. https://background.tagesspiegel.de/digitalisierung/digitale-identitaeten-zehn-oekosysteme-fuer-deutschland-1

Hempel, G., & Anke, J. (2023). Privacy management mit self-sovereign identity: Potentiale zur Erhöhung der informellen Selbstbestimmung [Manuscript submitted for publication].

Kubach, M., Schunck, C. H., Sellung, R., & Roßnagel, H. (2020). Self-sovereign and decentralized identity as the future of identity management? In Open identity summit 2020 (pp. 35–47). Gesellschaft für Informatik e.V. https://doi.org/10.18420/ois2020_03

Lazaro, C., & Le Métayer, D. (2015). Control over personal data: True remedy or fairy tale? SCRIPTed, 12(1), 3–34. https://doi.org/10.2966/scrip.120115.3

Lim, J. (2020). Self-sovereign identity: The harmonising of digital identity solutions through distributed ledger technology [Master's thesis, University of Twente]. https://essay.utwente.nl/81862/

Liu, Y., He, D., Obaidat, M. S., Kumar, N., Khan, M. K., & Raymond Choo, K.-K. (2020). Blockchain-based identity management systems: A review. Journal of Network and Computer Applications, 166, Article 102731. https://doi.org/10.1016/j.jnca.2020.102731

Martin, N., & Metzger, F. M. (2023). What determines the acceptance of facial recognition-based digital identity technology? Evidence from a multi-country survey [Manuscript submitted for publication]. Journal of Innovation Management.

Martínez, X., Alonso, J., City of Aarhus, Gijón City Council, Ertzaintza, Municipality of Peshtera, City of Reykjavik, UnionCamere, & InfoCamere. (2023a). D2.10 Implementation of basic system V2. https://www.impulse-h2020.eu/public-deliverables/

Martínez, X., Cuenca, A., Loureiro, J., Núnez, I., Markos, G., Bianchini, A., Nanclares, J. A. A., Dai, K., & Jimenez, A. (2022). D5.1 IMPULSE technology block V1. https://www.impulse-h2020.eu/public-deliverables/

Martínez, X., Loureiro, J., Núnez, I., Markos, G., Bianchini, A., Nanclares, J. A. A., & Dai, K. (2023b). D5.2 IMPULSE technology block V2. https://www.impulse-h2020.eu/public-deliverables/

Martínez, X., Rodriguez, J., & Loureiro, J. (2022). D5.4 IMPULSE wallet V2.

Microsoft Sway. (n.d.). Blockbuster collaboration [Interview with H. Haddad]. https://sway.office.com/03wqpaSpBpokJ3ZR?ref=Link

Mühle, A., Grüner, A., Gayvoronskaya, T., & Meinel, C. (2018). A survey on essential components of a self-sovereign identity. Computer Science Review, 30, 80–86. https://doi.org/10.1016/j.cosrev.2018.10.002

PricewaterhouseCoopers (PWC). (2021). Der Online-Ausweis auf dem Smartphone und die digitale Brieftasche.

Richter, D., & Anke, J. (2021). Exploring potential impacts of self-sovereign identity on smart service systems. Business Information Systems, 1(1), 105–116. https://doi.org/10.52825/bis.v1i.68

Schardong, F., & Custódio, R. (2022). Self-sovereign identity: A systematic review, mapping and taxonomy. Sensors, 22(15), Article 5641. https://doi.org/10.3390/s22155641

Schorlemer, J. von. (2022, January 18). Selbstbestimmte digitale Identitäten sind der Grundbaustein für eine digitale Wirtschaft. bpö – blog politische ökonomie. https://www.blog-bpoe.com/2022/01/18/schorlemer/

Sedlmeir, J., Smethurst, R., Rieger, A., & Fridgen, G. (2021). Digital identities and verifiable credentials. Business & Information Systems Engineering, 63(5), 603–613. https://doi.org/10.1007/s12599-021-00722-y

Spiekermann, S., & Korunovska, J. (2017). Towards a value theory for personal data. Journal of Information Technology, 32(1), 62–84. https://doi.org/10.1057/jit.2016.4

Strüker, J., Urbach, N., Lautenschlager, J., & Ruhland, N. (2021). Self-sovereign identity: Foundations, applications, and potentials of portable digital identities. Fraunhofer-Institut für Angewandte Informationstechnik FIT.

Veil, W. (2018). The GDPR: The emperor's new clothes - On the structural shortcomings of both the old and the new data protection law. Neue Zeitschrift für Verwaltungsrecht, 37(10), 686–696.

Vogiatzoglou, P., & Valcke, P. (2022). Two decades of Article 8 CFR: A critical exploration of the fundamental right to personal data protection in EU law. In P. Valcke & P. Vogiatzoglou (Eds.), Research handbook on EU data protection law (pp. 11–37). Edward Elgar Publishing. https://doi.org/10.4337/9781800371682.00010

W3C World Wide Web Consortium. (2022). Decentralized identifiers (DIDs) v1.0: Core architecture, data model, and representations (W3C Recommendation). https://www.w3.org/TR/did-core/

W3C World Wide Web Consortium. (2024). Verifiable credentials data model v2.0 (W3C Candidate Recommendation Draft). https://www.w3.org/TR/vc-data-model-2.0/

Wang, F., & De Filippi, P. (2020). Self-sovereign identity in a globalized world: Credentials-based identity systems as a driver for economic inclusion. Frontiers in Blockchain, 2, Article 28. https://doi.org/10.3389/fbloc.2019.00028

Zwitter, A. J., Gstrein, O. J., & Yap, E. (2020). Digital identity and the blockchain: Universal identity management and the concept of the "self-sovereign" individual. Frontiers in Blockchain, 3, Article 26. https://doi.org/10.3389/fbloc.2020.00026

ACKNOWLEDGMENTS

This research was supported by the European Commission, Grant Agreement 101004459, as part of the Horizon 2020 Identity Management in PubLic SErvices (IMPULSE) Project (www.impulse-h2020.eu). We thank two anonymous reviewers and the associate editor for valuable comments on previous versions of the manuscript.

Authors' Note

All correspondence should be addressed to
Nicholas Martin
Fraunhofer Institute for System and Innovation Research ISI
Karlsruhe, Germany
nicholas.martin@isi.fraunhofer.de

Human Technology
ISSN 1795-6889
https://ht.csr-pub.eu

Appendix

Distribution by country of survey respondents who say that IMPULSE gives them “control over their data” / lets them “decide who gets their data”

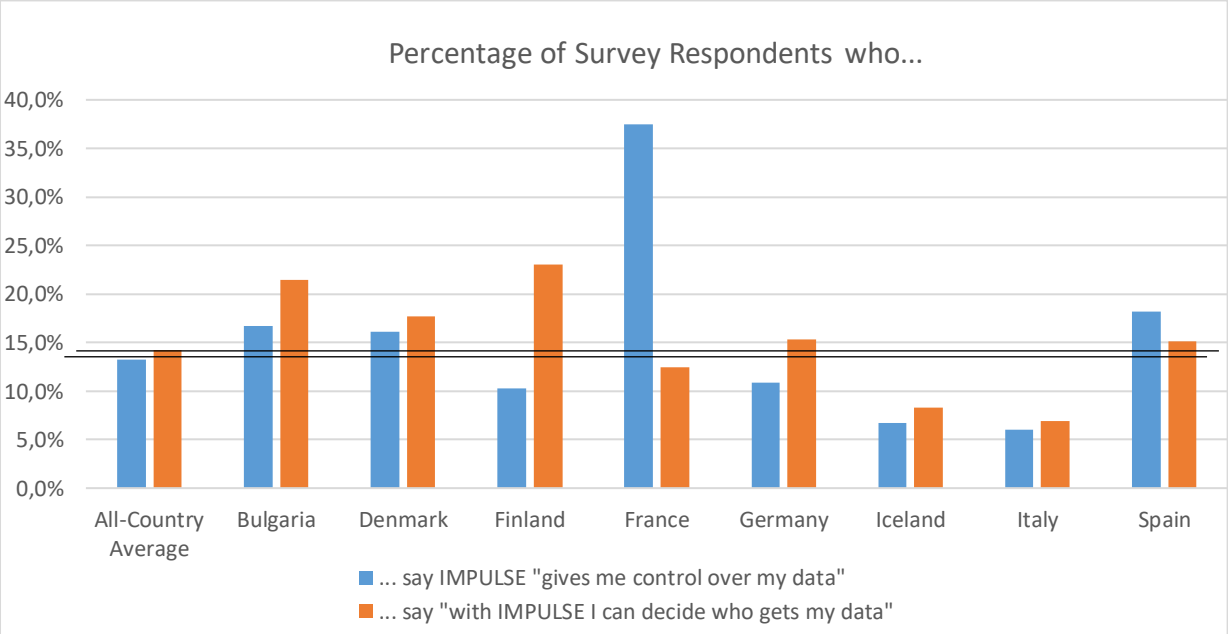

Figure 6. Percentage of survey respondents by country who say that IMPULSE gives them “control over their data” / lets them “decide who gets their data”

Survey Instrument

The text was translated into the languages of the countries surveyed.

Introduction

Thank you for your help!

This survey is conducted by a group of researchers at institutes and universities in Finland, Germany and Austria. Our survey is for a scientific project funded by the European Union. We want to understand what people think about online public services, computers and smartphones, and how they use them. This will help us develop new technology and improve public services online.

By taking the survey, you provide really important input for our research!

The survey is completely anonymous. No information you share can be traced to you, nor can you be traced by any information you provide. You can also withdraw from the survey at any point.

For each completed survey, we will donate 2€ to one of several charities (you can pick!), up to a total of €500 Euros.

Let me thank you again for support

Nicholas Martin, Fraunhofer ISI, Germany

If you have any questions, please feel free to contact me at nicholas.martin@isi.fraunhofer.de

To learn more about our research project, visit https://www.impulse-h2020.eu/

Section 1: Demographics

To understand survey results better, we would first like to collect some background information.

Q1. How old are you?

- Free text box

Q2. What is your gender

- Female
- Male
- Diverse

Q3. What is your highest level of education?

- Completed primary school
- Completed secondary school
- Completed post-secondary school vocational training
- Completed university
- Currently at university
- Currently in post-secondary school vocational training

Q4: What is your annual household income, after tax?
(country-specific quintiles, based on Eurostat data)

- Less than X
- X+1 to XX
- XX+1 to XXX
- XXX+1 to XXXX
- More than XXXX+1
- Prefer not to answer

Q5. What is your citizenship?

- Citizen of [*survey country*]
- Permanent resident of [*survey country*]
- Migrant / non-permanent residency
- Prefer not to answer

Q6. Due to discrimination and for other reasons, minority groups often have distinct views about technology. To help us develop genuinely inclusive technology, we would therefore like to ask you for your ethnicity.

- White
- Minority
- Prefer not to answer

Q7. Do you live in a…

- Rural area or village
- Small or middle-sized town

- Major city [country-specific population number based on size of ~top 5 cities]

Q8. I enjoy trying out new technologies

Strongly Disagree ... Strongly Agree.
1 2 3 4 5

Q9. People have different views on privacy. In general, how concerned are you about your privacy when you access services on the internet?

Very Concerned ... Not Concerned at All
1 2 3 4 5

Section 2: IMPULSE

We are developing a new Log-In system for online services. We would like to get your feedback on it. Please watch the video, which shows the basic elements of our system.

Video here

Qu. 10. Would you use IMPULSE instead of the digital identity (log-in) systems you currently use (like username/password, smartcard, PIN, etc.), if IMPULSE were available?

Certainly not ... Certainly yes
1 2 3 4 5

Qu. 11. *Only for "NOs" (1, 2) at Qu. 9***. Why would you not use IMPULSE instead of the digital identity (log-in) systems you currently use? Tick all that apply**

- I do not want to depend on my Smartphone
- I do not have a Smartphone
- I am worried about what happens if I lose my Smartphone or it is stolen
- I am worried about facial recognition technology
- IMPULSE is too complicated
- I use too few online services to make IMPULSE worthwhile
- I am worried about hackers and identity theft

Qu. 12. For which online services do you think it would be most sensible to use IMPULSE? Please select no more than five

- Online banking
- eHealth (e.g. electronic communication with a doctor to get a prescription instead of going in person)
- Digital vaccination certificate for Covid or other diseases
- Social media
- E-commerce (e.g. Amazon, Airbnb)
- Completing tax returns online
- Registering for social services online
- Email

- Other (free text box)
- None

Qu 13. Please choose all of the following words and phrases that you feel describe IMPULSE.
Convenient ; Easy to use ; Saves time ; Makes signing up for services easier ; Makes logging in easier ; Exciting ; Safe ; privacy-friendly ; interesting ; Unnecessary ; weird ; Complicated ; Creepy ; IMPULSE gives me control over my data ; Dangerous ; Surveillance ; Not useful ; With IMPULSE, I can decide who gets my data ; Other *(free text box)*

Qu 14. Today, many people have multiple digital identities (log-ins), with different usernames/passwords. People have different views and experiences of this. Please tell us which statements you agree with

Strongly Disagree ... Strongly Agree
1 2 3 4 5

- Using multiple digital identities is a hassle, but sensible
- I would prefer to have a single digital ID for all online services and accounts
- I sometimes forget my username/password, or forget which identity to use for a service or account
- I sometimes don't use a service because the sign-up process (when you enter information like your name and address and create a username/password or similar) is too much hassle
- Managing multiple digital identities and ways of identifying myself overtaxes me
- If signing up for services were easier and faster, I would use more services
- If I had only one single, secure digital ID, that I could use for all online services, I would use more services

Qu. 15. With IMPULSE, you create a new digital identity for each online service, and store these identities on your phone, in the IMPULSE App. As we saw in the video, you create these digital identities with photos of your state identity card, but then the photos are deleted. No record of your identity card is kept.

Some European governments have proposed an alternative system: a digital version of your state identity card, that you store on your phone and can use as a single, universal digital identity for all online services – instead of having multiple digital identities like with IMPULSE. Which system would you prefer to use? *Choose One*

- IMPULSE: multiple digital identities stored on your phone but no record of your ID card kept on your phone
- Alternative: digital version of your identity card stored on your phone that serves as a single, universal digital identity

Qu. 16. For many people, it is very important to have "control over their data". Having "control over your data" has many dimensions. Below is a list of some dimensions. *Please choose the three most important in your view.*

- Online service providers have to ask for my consent before collecting or using my data
- Online service providers must provide a short, easy-to-understand privacy policy explaining how they use my data
- Online service providers cannot demand more data from me than necessary to make the service work
- Online service providers cannot refuse me their service unless I allow them to use my data for advertising
- Online service providers cannot use website lay-out and other tricks that manipulate me to give them extra data or data-use permissions
- Online service providers must delete my data if I ask them to

Section 3: Experience with eID and & Digital Experience

Qu 17. Some people are very skilled with computers and smartphones; others are just getting to know them. Please tell us about your skills

Strongly Disagree ... Strongly Agree
1 2 3 4 5

I know how to…

- …use Google or other internet search engines
- …find relevant information on websites of state agencies or the municipal government
- …use email or social media
- …save or store files (documents, music etc.) on my device and retrieve them when I want them
- …use Cloud Applications like Dropbox, iCloud, Google Drive, or SharePoint to store and share documents
- …use online services like online banking, e-government or e-health
- …use Word or Powerpoint or similar applications
- …read a simple computer code and make basic changes to it
- …re-install or update computer programs

Qu. 18. There are many different technologies to give users a digital identity to log in to an online service, computer or smartphone, like username and password, PIN/TAN, Smartcard etc. Please tell us which ones you have used or heard about

	I have used this technology	**I have heard about this technology**	**I have not heared about this technology**
User Name + Password			
SmartCard + PIN-Number			
NemID / MitID *(Denmark only)*			
PIN / TAN			
Fingerprint recognition			
Face recognition			
Voice recognition			
Eye [iris] recognition			
Other [please specify]			

Qu. 19. From the list below, please choose the 3 digital identity (log-in) technologies that you prefer to use for your log-ins. Don't worry if you have not used some of the technologies yet.

- User Name + Password
- SmartCard + PIN-Number
- PIN / TAN
- Fingerprint recognition
- Face recognition
- Voice recognition
- Eye [iris] recognition
- NemID / MitID *(Denmark only)*
- Other [please specify]

Qu. 20. Many online services require you to use a digital identity (log-in), like a username/password, PIN/TAN, Smartcard/PIN or biometric recognition

Please estimate how many *private-sector* online services you regularly use that require a digital identity (log in), like online banking, social media, insurance, Amazon/online shopping, Airbnb, Booking.com, …
Don't worry if you are not 100% sure, just give us your best guess.
Freetext box

Please estimate how many public-sector online services you use that require a digital identity (log in), like completing tax returns online, Covid App, registering a car or your residency, participating on a public discussion forum, registering online for a government service, …
Don't worry if you are not 100% sure, just give us your best guess.
Freetext box

Section 4: Experience with eGovernment

eGovernment means digitising public services, so citizens can do them online: for example complete tax returns, pay municipal bills, register a car or residence, or apply for social benefits.

Q. 21. Have you used eGovernment services (online public services)?

- Yes
- No

Q.22a. *For "Nos"*. Please tell us about why you have not used eGovernment services, and your views about eGovernment in general.

Tick all that apply

- I am not aware of any eGovernment services
- The eGovernment services that exist are not relevant to me
- Going to the public administration in person is faster than using eGoverment
- Doing things in person at the public administration is easier than using eGoverment
- eGovernment feels cold and nonpersonal. I prefer having personal interaction with the civil servants
- I am worried about the security of my data if I use eGovernment
- I would be worried to make a mistake if I use eGovernment
- I would use eGovernment if it was fast, simple and I could be sure my data are safe
- Other Free Textbox

Qu.22b. *Only for "Yes" at Qu.21.* Thinking about your experience with eGovernment, what aspects should be improved the most? Please choose the three most important

- Make signing up for eGovernment (creating a digital identity to use for eGovernment) easier and faster
- Make logging-in to eGovernment services easier and faster
- Make more public services available online
- Make finding information about eGovernment services easier
- Improve the layout of eGovernment websites
- Offer more assistance for using eGovernment (e.g. helplines, chatbots)
- Ensure all eGovernment services can be completed fully online (i.e. no need for any offline steps, like providing physical signatures

End

Many thanks for your time and support!