GESELLSCHAFT
FÜR INFORMATIK



Heiko Roßnagel, Christian H. Schunck,
Daniela Pöhn (Hrsg.)

Open Identity Summit 2025

22.-23. Mai 2025
Neubiberg, Germany

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Proceedings
Series of the Gesellschaft für Informatik (GI)

Volume P-364

ISSN 2944-7682 (Online)

Volume Editors
Heiko Roßnagel | Christian Schunck
Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering IAO
Nobelstr. 12, D-70569 Stuttgart, Germany
heiko.rossnagel|christian.schunck@iao.fraunhofer.de
Daniela Pöhn
University of the Bundeswehr Munich
Fakultät für Informatik, Institut für Softwaretechnologie
Forschungsinstitut CODE
Cascada Haus, Carl-Wery-Str. 22, 81739 München
daniela.poehn@unibw.de

Series Editorial Board
Andreas Oberweis, KIT Karlsruhe,
(Chairman, andreas.oberweis@kit.edu)
Torsten Brinda, Universität Duisburg-Essen, Germany
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Barbara Hammer, Universität Bielefeld, Germany
Falk Schreiber, Universität Konstanz, Germany
Wolfgang Karl, KIT Karlsruhe, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Heiko Roßnagel, Fraunhofer IAO Stuttgart, Germany
Kurt Schneider, Universität Hannover, Germany
Andreas Thor, HFT Leipzig, Germany
Ingo Timm, Universität Trier, Germany
Karin Vosseberg, Hochschule Bremerhaven, Germany
Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations
Rüdiger Reischuk, Universität Lübeck, Germany
Thematics
Agnes Koschmider, Universität Kiel, Germany
Seminars
Judith Michael, RWTH Aachen, Germany

© Gesellschaft für Informatik, Bonn 2025



This book is licensed under a Creative Commons BY-SA 4.0 licence.

Preface

Welcome to the “Open Identity Summit 2025”, which has been jointly organized by the Special Interest Groups BIOSIG within the German Computer Science Society (Gesellschaft für Informatik), Fraunhofer Institute for Industrial Engineering IAO and the University of the Bundeswehr in Munich.
The international program committee performed a strong review process according to the LNI guidelines with at least three reviews per paper and accepted 50% of the 18 submitted papers as full scientific papers.
Furthermore, the program committee has created a program including selected contributions of strong interest (further conference contributions) for the outlined scope of this conference.
We would like to thank all authors for their contributions and the numerous reviewers for their work in the program committee.

Neubiberg, 5th of May 2025

Heiko Roßnagel
Fraunhofer IAO

Christian H. Schunck
Fraunhofer IAO

Daniela Pöhn
University of the Bundeswehr

Conference Chairs

Heiko Roßnagel, Fraunhofer Institute for Industrial Engineering IAO
Christian H. Schunck, Fraunhofer Institute for Industrial Engineering IAO
Daniela Pöhn, University of the Bundeswehr Munich

Programme Committee

Jürgen Anke, Germany
Franco Arcieri, Italy
Tamas Bisztray, Norway
Arslan Broemme, Germany
Christoph Busch, Germany
Victor-Philipp Busch, Germany
Bushan Chavan, USA
Konrad Degen, Germany
Nicolas Fähnrich, Germany
Daniel Fett, Germany
Lothar Fritsch, Norway
Igor Furgel, Germany
Michael Grabatin, Germany
Jochen Günther, Germany
Marit Hansen, Germany
Michael Heinl, Germany
Olaf Herden, Germany
Gerrit Hornung, Germany
Detlef Houdeau, Germany
Detlef Hühnlein, Germany
Tina Hühnlein, Germany
Ulrike Korte, Germany
Michael Kubach, Germany
Andreas Kühne, Germany
Sebastian Kurowski, Germany
Herbert Leitold, Austria
Luigi Lo Iacono, Germany
Sebastian Mödersheim, Denmark
Alexander Nouak, Germany
René Peinl, Germany
Henrich Pöhls, Germany
Daniela Pöhn, Germany
Alexander Roßnagel, Germany
Heiko Roßnagel, Germany
Mirko Ross, Germany
Martin Schanzenbach, Germany
Christian H. Schunck, Germany
Rachelle Sellung, Germany
Jon Shamah, United Kingdom
Filipe Souza, Portugal
Hermann Strack, Germany
Maurizio Talamo, Italy
Don Thibeau, USA
Karsten Treiber, Germany
Perdri Vieira, Portugal
Samuel Wairimu, Sweden
Tobias Wich, Germany
Thomas Wieland, Germany
Alex Wiesmaier, Germany
Jan Zibuschka, Germany
Frank Zimmermann, Switzerland

Hosts and Partners

BIOSIG – Biometrics and Electronic Signatures (www.biosig.org)
The special interest group “Biometrics and Electronic Signatures” (BIOSIG) within GI e.V. is dedicated to the fundamentals, methods, techniques, processes and implementations used to guarantee the authenticity and integrity of entities.

Contents

Regular Research Papers

Aaron Witzki, K. Valeria Carl, Léon Dankert, Oliver Thomas, Oliver Hinz
What are your intentions with my data? A tool to enhance consumer data sovereignty in digital ecosystems 13

Ignacio Alamillo, Steffen Schwalm
The role of EBSI in eIDAS – how qualified ledger with governmental trust anchor could shape eIDAS ecosystem 29

R.X. Schwartz, Kerri Lemoie, Jeanne Kitchens
Developing A Governance Framework for Learning and Employment Record (LER) Issuer Registries 41

Daniela Pöhn
"Then I clicked something"– Helping Users to Report Security Incidents with Digital Identity Wallets 55

Michael Kubach
Economic Challenges of Cybersecurity Innovations Pursuing an Ecosystems Approach 71

Erran Carmel, Edgar Whitley, Emma Grubb
Tracking the Shift to Decentralised Identity: A Case Study in International Air Travel 85

Lara Eberl, Lisa-Marie Engländer, Caroline Löhle, Dennis Jahnecke, Aylin Baris, Donjeta Seljaci, Schabnam Shamsi, Jochen Günther
Phishing and Identity Manipulation through Audiovisual Channels 101

Eric Biele, Daniel Richter, Ricardo Bochnia, Jürgen Anke
Supporting trust decisions of holders and verifiers: a unified model of trust management in SSI 113

Tina Hühnlein, Detlef Hühnlein, Steffen Schwalm, Carsten Stöcker
Towards the European Business Wallet 127

Further Conference Contributions

Lennart Kiss, Rachelle Sellung
Securing Engagement: Key Takeaways from Onboarding Users into Data Protection Apps 145

Mrudula Arunkumar, Christian H. Schunck, Salma Ben Mamia, Heiko Roßnagel
How can Design Thinking benefit Cybersecurity?: Insights from Don Norman's The Design of Everyday Things 153

Anne Elisabeth Krueger
Making Security Matter: Engaging Non-Security Users Effectively through Empathy 161

Salma Ben Mamia, Christian H. Schunck, Mrudula Arunkumar, Heiko Roßnagel
Security awareness versus secure behaviour: A bibliometric study of the state of research on human factors in IT Security 167

Isaac Henderson Johnson Jeyakumar, Michael Kubach
FidNET: Flexible Infrastructure for Decentralized Trust Establishment in Industrial Supply Chain Networks 175

Awid Vaziry, Luca Vetter, Axel Küpper
eIDAS 2.0: Evaluating the Issuance of Digital University Diplomas 183

Regular Research Papers

What are your intentions with my data? A tool to enhance consumer data sovereignty in digital ecosystems

Aaron Witzki [1], K. Valerie Carl [2], Léon Dankert [3], Oliver Thomas [4], and Oliver Hinz [5]

Abstract: As data collection enhances, consumer data sovereignty becomes increasingly important. Despite initiatives like the GDPR, information asymmetries persist, and privacy policies remain complex or lengthy, often preventing consumers from giving truly informed consent for data sharing. Building on established concepts for more informed consent, we propose a tool—the consent registry—to enable informed consent and strengthen consumer data sovereignty while being easy to implement in a variety of digital ecosystems and applications. Our concept takes three different stakeholder groups into consideration, the consumer, the inquiring institution and the application provider, since often times the consumer is not directly connected to inquiring institutions and instead relies on third-party applications in digital ecosystems. Our tool employs case-to-case consent, describing that consent is valid only for a singular request. This empowers consumers to exercise informed control over data collection and usage while enhancing data availability across institutional borders in digital ecosystems.

Keywords: Data Sharing, Consumer Sovereignty, Privacy Policies, Informed Consent, Privacy Tool, Digital Ecosystem

1 Introduction

As companies today collect an ever-increasing amount of data of their customers, they can create even more detailed personal profiles, enabling them to, e.g., tailor advertisements

[1] Goethe University Frankfurt/Main, Chair of Information Systems and Information Management, Theodor W.-Adorno-Platz 4, 60323 Frankfurt/Main, witzki@wiwi.uni-frankfurt.de,
https://orcid.org/0009-0009-7150-6627
[2] Goethe University Frankfurt/Main, Chair of Information Systems and Information Management, Theodor W.-Adorno-Platz 4, 60323 Frankfurt/Main, kcarl@wiwi.uni-frankfurt.de,
https://orcid.org/0000-0003-4655-1046
[3] German Research Centre for Artificial Intelligence, Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, leon.dankert@dfki.de,
https://orcid.org/0009-0008-7989-8058
[4] German Research Centre for Artificial Intelligence, Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, oliver.thomas@dfki.de,
https://orcid.org/0009-0007-3326-4175
[5] Goethe University Frankfurt/Main, Chair of Information Systems and Information Management, Theodor W.-Adorno-Platz 4, 60323 Frankfurt/Main, ohinz@wiwi.uni-frankfurt.de,
https://orcid.org/0000-0003-4757-0599

doi:10.18420/OID2025_01

and services [VT21]. However, this also leads to growing information asymmetries, as consumers provide increasingly more data, while companies disclose little in an understandable way, particularly regarding how they handle data [Wa20]. Especially in digital ecosystems, consumers often share data with the whole ecosystem without being actually aware of it. For this reason, legislation such as the General Data Protection Declaration (GDPR) represents a significant step in the direction of more data sovereignty. However, it remains far from comprehensive. For example, it often results in lengthy privacy policies that consumers in many cases do not understand or even read at all, leading to uninformed consent [Ba22]. This lack of awareness can diminish consumer data sovereignty despite being an important value, especially in a more digitalized world. In addition, access to data presents a significant challenge, particularly for smaller companies, as there are currently few alternatives available for obtaining data, other than through consumers' acceptance of privacy policies when using the product and collaborating with other firms [JB24]. Consumers can profit from more transparent methods supporting their data sovereignty in the digital world and reducing information asymmetries, while responsibly granting deliberate access to collected data for a wider audience. These measures should empower consumers with a clear understanding of how institutions handle their data [Mi22, Tr23] and give consumers the ability to make informed decisions about data usage, eventually even on a case-to-case-basis. Previous research shows that users might be willing to disclose data in exchange for desired services [Ko17]. However, such cases might not enhance consumer empowerment, as they do not provide sufficient information about data usage, thus not supporting informed disclosure.

We present a novel tool specifically designed for digital ecosystems which we term as the *consent registry*, to enhance the empowerment of consumers. This registry has the ability to enhance consumer data sovereignty by enabling use-case-specific, active and informed consent to data requests. The acceptance of one request is only valid for a single case for one institution. In this way, our tool inherently increases transparency, yet the main focus remains in strengthening consumer data sovereignty. Privacy policies and legislation such as the GDPR lay the groundwork for our solution, yet we aim to enhance them by increasing consumer friendliness and easing the way they can exercise power over data collection and usage. Often times, inquiring institutions present privacy policies unintuitively to consumers today [Ef19], potentially leading to uninformed consent. Previous works investigated possible solutions and more comprehensible representations of privacy policy, e.g., privacy icons [RP20] and one-pager solutions [RS18]. Drawing from such concepts, we present a compact, yet concise, representation of privacy policies to consumers in a dedicated tool to control their data in digital ecosystems, to ultimately enhance consumer data sovereignty in the digital age. In contrast to previously established concepts, such as ID wallets [e.g., PAZ22], our tool only focuses on the user data that is generated in the application, in our use case (energy) consumption data and does not include further identifying information. The goal of our concept is to strengthen consumer data sovereignty by granting the consumers granular decisions regarding their data sharing. We structure the reminder of the paper as follows. Section 2 introduces the concepts of data privacy and sovereignty. Following, section 3 provides the architecture

of the data management tool: the consent registry. Finally, in section 4, we present an outlook of our work and the contribution to consumer data sovereignty in theory.

2 Data privacy and consumer data sovereignty

Data privacy describes a broad concept with the goal to ensure the responsible usage of data [CH24]. It encompasses various aspects [Ca21b] regarding, e.g., data protection (e.g., protection against unauthorized access) [He25] and the ethical use of data [Mi22]. Laws set minimum requirements for data privacy [Ca22a] in various regions around the world: e.g., the Personal Information Protection Law in China [Ca22a] and the GDPR in the European Union [LYH19]. The GDPR, for instance, describes data privacy as a "fundamental right" in recital 1. These frameworks have been established to safeguard fundamental citizen rights in the digital realm [CM22]. Apart from aspects of data protection and compliance with laws, the concept of data privacy also includes more consumer-focused aspects, e.g., transparent communication to the consumer that exceed regulatory minimum requirements [JLL21, Ca22b, CH24].

Data sovereignty, as a concept, centers on the consumer, their personal data and their ownership of their data [Hu21]. The focus of data sovereignty is the empowerment of the consumer in the digital sphere with the goal to ensure the ownership of the personal data [Ca21a], thereby often exceeding regulatory minimum requirements. Data sovereignty includes multiple aspects, like data ownership and control of the data [Hu21]. To ensure data sovereignty, consumers should have control over their data, including who can access it and how different actors utilize it [BC11]. The GDPR defines various aspects of data control in Articles 15-22, for instance the right of access by the data subject (Art. 15 GDPR). To ensure this, transparency practices [Lo24] and informed consent [AKR22] are potential measures. Transparency hereby includes all relevant information regarding the further processing of the data for the consumer, to control the compliance with the stated usage [Lo24], therefore being a crucial prerequisite for an informed consent.

The informed consent is a necessity, since a general acceptance of privacy policies might not be sufficient [Ef19]. Consumers tend to underestimate the consequences of the general acceptance of privacy policies [Ba22]. This could potentially lead to unforeseen and undesirable consequences (e.g., loss of privacy or unknowingly relinquishing rights) for the consumer and may permanently damage their trust [MF16]. Often times, institutions present the privacy policies in a way that does not focus on easy understanding, e.g., through policies that are hard to comprehend for the consumer [CC15] or not of an appropriate length for the service or product used. Another aspect is that, despite lengthy privacy policies, information asymmetries remain [Wa20]. Consumers often lack information about already collected data and to what extent inquiring institutions will actually use the data [Mi22]. This further complicates the informed consent [Ef19] due to an inadequate level of transparency. Despite the right to access and alter collected data (i.e., granted by the GDPR), consumers do not exercise their power often.

Through this lens, it is apparent that regulations alone, e.g., the GDPR, might not be sufficient to ensure consumer data sovereignty in practice. While the regulations include sufficient formal requirements, their implementation often falls short in usability and effectiveness for consumers, hindering consumers from performing their legal rights. For this reason, previous research has proposed various consumer-centered concepts [SE24]. Concepts such as privacy icons [Wi22] or one-pager privacy policies [RS18] arose (partly indicated by the GDPR but not mandatory), to further enhance consumer data sovereignty and enable the informed consent [Ca22b]. However, these concepts face challenges as well. For instance, privacy icons alone might not be sufficient to convey the complexity of certain cases and may vary in effectiveness depending on the context [Gr24]. One-pager solutions for the general consent face the risk of oversimplifying circumstances and may lead to misinterpretations of the consumers [KKS20]. Still, those approaches are valuable measures to strengthen consumer sovereignty beyond legal requirements.

We build upon these concepts to present a tool that supports the informed consent in digital ecosystems. As of now, digital ecosystems mainly employ complicated, lengthy privacy policies and often only require one consent per check mark to share collected data with the whole institution or even ecosystem. Therefore, we aim to develop an easy to implement tool that fosters data sovereignty by easing control over collected data, enhancing data control in digital ecosystems on a case-to-case basis. We consider three distinct stakeholder groups in digital ecosystems: consumers generating data, institutions requesting data, and application providers. Consumers often lack a direct connection with the inquiring institutions and instead rely on third-party applications in digital ecosystems.

3 The consent registry tool to enhance consumer data sovereignty

To facilitate the practical applicability, we present our tool within the context of a digital ecosystem for smart living, since data from personal households is sensitive and requires informed consent for sharing. The tool will be implemented in an already used application to track consumption data, so that the consumers will only face the choice of sharing their data without any further needed configurations. When designing the tool, we tried to ensure self-explainability and ease of use. This application is not independent of requiring institutions, rather the application is part of the same digital ecosystem like requiring institutions, despite the app users not being an active participant of the digital ecosystem yet before introducing our tool. Our consent registry is situated within the context of emerging data ecosystems that prioritize consumer data sovereignty, such as SmartLivingNEXT [Sm25]. We developed the tool based on prior research and stakeholder requirements through dedicated focus groups and discussed the consent registry during several workshops with academics, practitioners and interested persons. We display the general process of the proposed consent registry tool in Figure 1.

To issue such requests, institutions need to fill out predefined request forms, which serve to inform the consumer about, e.g., not only the intended purpose of the data, but also how

long they intend to use the data. Often times consumers are not in direct contact with requesting stakeholders in a digital ecosystem, but use third-party applications and services to record and manage their data in such ecosystems and are therefore part of such ecosystems. Therefore, at first, the requests will pass these third-party application providers. The application providers have the ability to block requests that they deem inappropriate for their user community. If the application providers allow requests, they will be sent to consumers and be presented in a comprehensible format. Only if the consumer decides to accept the request, the data will be sent to the institution.

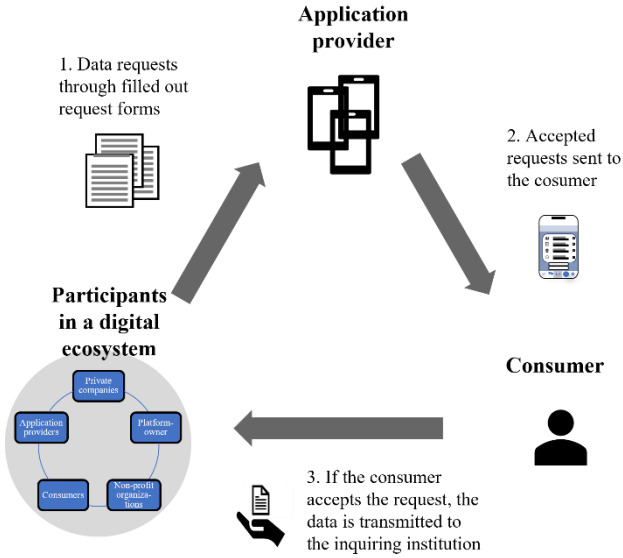


Fig. 1: Process of the consent registry

The consumers need to be aware of the exact goals that institutions follow with their data to give informed consent. Such active consent does not necessarily hinder data collection and usage. Rather, giving informed consent can enhance trust in digital ecosystems and particular participants. Further, entities collecting consumer data are sometimes afraid of legal and security risks when sharing data with external partners. However, employing such an active-consent-approach can allow more external institutions access to collected consumer data and therefore benefit particularly small and medium-sized enterprises as well as start-ups in their initial data access. In this way, all involved stakeholder groups can potentially benefit from such a consent tool: Consumers can easily exercise their power over collected data and deliberately choose with whom to share data for which reasons. Corporate entities in a digital ecosystem can request data from consumers without a direct business relationship and can therefore benefit from a vast amount of data available in such digital ecosystems. And finally, the application provider benefits from a more trusting relationship to their users since they allow them to decide about data sharing

rather than obtaining general consent. Overall, the whole digital ecosystem can benefit from consumer trust on the one side and an enhanced data availability on the other side.

3.1 Request from the inquiring institution

A lack of transparency in privacy policies, e.g., through non-consumer-centered designs [CC15], is a fundamental barrier to achieving informed consent today [Ba22]. Consumers need to have a holistic view of the data activities to be able to give an informed consent. The first part of the consent registry is a standardized form that the inquiring institutions need to completely fill out to ensure a comprehensive request that can be displayed to the consumer (see Fig. 2). We carefully considered the trade-off between comprehensibility and exhaustiveness. For this reason, the form includes formalized, predetermined answers to increase the consumers' comprehensibility and open-text fields for detailed information.

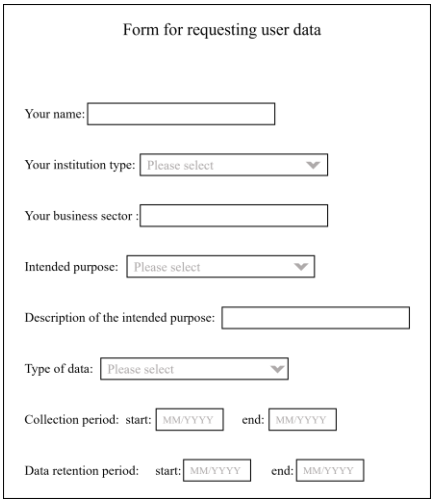


Fig. 2: Inquiring institution's view of the form for requesting consumer data

First, the inquiring institutions need to identify themselves. This is necessary, since only the exact entity that requested the data can use the data. Institutions should not place a request for a whole conglomerate but rather for the particular entity requiring this data. Prior studies show that the requester influences the consumers' willingness to share data [Li17, WGW19]. This might, e.g., be due to familiarity and prior experiences with the inquiring institution, through which the consumers were able to build trust [Ro15, Wi16].

Further, consumers might have particular preferences for sharing data with different institution types. For example, consumers showed a higher willingness to share data for medical research compared to private companies [Mc16]. The inquiring institution needs to specify their type of institution, selecting from a set of predetermined options: (i) (private) research institute; (ii) university; (iii) private company; (iv) non-profit

organization; (v) governmental institution; and (vi) other […]. The indication of different institution types gives consumers, that might not be familiar with specific business sectors, the possibility to gain a more general insight towards the organization.

To further ease the evaluation for consumers, the requester needs to specify their business sector. The institution's name alone might not be sufficient, since consumers might not be familiar with all actors in a digital ecosystem but may have specific preferences based on the business sector. Prior work highlights the importance of the business sector of the inquiring institution for consumers' willingness to share data [Ac21].

Further, the inquiring institution needs to specify the intended use of the requested data in a standardized form. The intended use of the data is crucial, since it significantly affects the consumers' willingness to share data [Ro15]. For instance, prior works have shown, that consumers might be more willing to share data for research purposes than for commercial usage [TLS17]. To facilitate the consumers' understanding, the inquiring institution needs to specify their intended use among predefined options for a first evaluation: (i) research; (ii) product development; (iii) artificial intelligence (AI) training; (iv) personalized marketing; (v) market research; and (vi) other […]. In this way, consumers can form a general opinion towards specific purposes and can potentially easily understand the actual purpose compared to free text fields only.

To further enhance the consumers' understanding of the intended use of data, we added a particular field that details the intended use expanding the overall category of usage and describes in an easy-to-understand way how the inquiring institution will use the data. This additional information should enhance consumers' understanding of the general purpose behind the proposed use of their data.

The type of data is another relevant information for the consumers [Ha17]. For instance, the works of Marxen et al. [MFF24] have shown the importance and influence of the type of requested data on the willingness to share it. For this reason, the inquiring institution needs to define the type of the requested data. The type of data is deeply connected to the use case of the implementation of the consent registry and describes the type of data that is requested by the inquiring institution. For this reason, we deviate from the GDPR definition of categories of (personal) data (Art. 9 GDPR). To facilitate the comprehensibility for the consumers, the inquiring institution can only choose between predetermined options available in the particular digital ecosystem. In our case in the smart living domain (i) electricity consumption, (ii) water consumption and (iii) gas consumption data is available through the intermediary application that will implement the consent registry tool first within an initial case study before it is rolled out across the entire ecosystem, leading to more variety in available data types.

In addition, the inquiring institution needs to determine the collection period of the desired data: They need to state start and end dates of the targeted data. It enables the consumer to gain a clearer understanding of the volume of the requested data, thereby enabling a more informed consent, as the consumer is aware of the exact time period and extent of their data that they will share with the inquiring institution.

Further, the inquiring institution needs to specify the intended retention period of the desired data. The retention period is another crucial information, addressed, e.g., in the GDPR's right to be forgotten [Ko15]. Consumers need to be aware of the period during which the inquiring institution will handle the data, after they give the consent. This constitutes another critical piece of information that allows consumers to better assess the potential consequences of their consent.

3.2 Application provider as intermediary

Within digital ecosystems, individuals frequently do not interact directly with institutions seeking their data, relying instead on third-party platforms and services for data storage and management. For this reason, the consent registry will not immediately forward a request to the consumers. In our concept, the application providers take the role of intermediaries. Previous research has underlined the focal role of application and service providers and their public role. For instance, service providers inherently have a gatekeeping function, to protect their users [TF17]. For this reason, our concept registry enables application providers to accept or decline every request before sending them to their users, thereby preserving their gatekeeping role. The application providers thereby keep the control over which institutions might gain access to user data of their platform, enabling them to protect their users from actors who may not align with core values of the application or actors which the users might perceive negatively, thereby potentially undermining trust between the users and the application providers. Additionally, the application providers are also able to protect their users from overwhelming amounts of requests by restricting too many requests from one inquiring institution. Furthermore, in the current state of our concept, application providers also act as arbitrators, e.g., by verifying whether the description of the intended purpose aligns with the selected intended purpose. A different institution in the ecosystem could potentially perform this function, depending on the implementation of the ecosystem, like a trust anchor that at least sporadically checks whether intended general purposes matches the free text description.

Administering each individual request separately might become an overwhelming and time-consuming task for the application providers, potentially resulting in imprecise management of requests or rejecting them and the consent registry in general. For this reason, the application provider has the option to determine rules about incoming requests when initially implementing the consent registry, thereby mitigating inappropriate requests (e.g., declining requests from private companies or particular intended purposes).

For instance, the application provider is able to predefine the general types of institutions that are allowed to send requests. They might deem certain types of institution inappropriate for requesting data from their users, e.g., permitting research institutions to submit requests while considering private companies unsuitable, as leveraging data for profit could conflict with their fundamental values. Application providers have the ability to prevent institutions they consider inappropriate from requesting data from their users.

Similar to the institution type, application providers have the ability to predetermine specific purposes. For instance, application providers might be opposed to data requests for marketing purposes, since this might affect the user's perception of the (own) application and could ultimately negatively influence trust in the application provider.

Further, application providers can predetermine which types of data institutions could request from their users. Our use case involves three types of usage data, whose confidentiality may be perceived similarly. However, the consent registry is supposed to be adapted to various different use cases with different types of data. For instance, in use cases that include wearable devices, the requested data might include movement profiles or health data. For this reason, the option to exclude requests for specific types of data might be particularly useful for use cases with data of varying sensitivity.

Application providers can determine different rules connected with Boolean operators to specify the use cases for data requests they would generally approve. This aspect of the consent registry limits the number of potentially unnecessary requests for application providers, since too many requests might burden them and ultimately lead to rejecting consent at all for any requests and the consent registry in general. Requests, that the application providers did not exclude still require their explicit approval to be sent to users, if desired, as the application provider may still consider certain requests inappropriate.

3.3 Request for the consumer

Finally, consumers receive data requests that the application providers deemed appropriate and can decide on a case-to-case basis which request they want to accept. To enable an informed consent, we follow two key goals: First, the consumer must receive complete information of the processing of their data. Second, the information needs to be presented to the consumer in a comprehensible format. To achieve the comprehensible format, we utilize the concept of progressive disclosure. Progressive disclosure describes the initial presentation of condensed information to the consumer to enable a quick overview of a certain topic [SW19]. The consumer only receives detailed information upon request. Researchers have applied the concept of progressive disclosure in various fields, including algorithmic transparency [SW19], explainable AI [Mu24] and education [HG18].

Drawing from the concept of progressive disclosure, we opted for the design displayed in Figure 3. We utilize the concept of privacy icons to facilitate the understanding of the key aspects of the request, expanding the icons with an additional description for enhanced clarity. The presented form contains four fields, extracting the data submitted by the inquiring institution via the request form and displaying it in a concise, streamlined format.

At first, the consumer will see four key information patterns of the request: (i) inquiring institution; (ii) intended purpose; (iii) type of data; and (iv) retention period. However, the consumer is able to expand the respective fields and receive additional information on demand, indicated by the plus next to each field (i.e., a progressive disclosure).

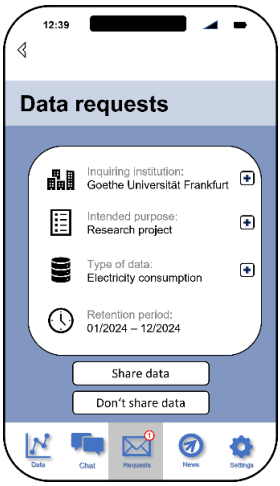
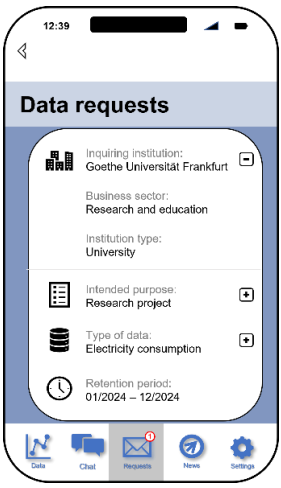


Fig. 3: Consumer's initial view of the request form

In its default state, the request form displays only the name of the institution to the consumer. This information on its own is sufficient to unambiguously identify the inquiring institution. However, if consumers are interested in additional information, they can expand this field and receive further information regarding the inquiring institution like the business sector and institution type (see Fig. 4a). This approach prevents consumers from being overwhelmed with unnecessary information if they are already familiar with the inquiring institution. If consumers are not familiar with the institution, they are able to easily access additional information to form an opinion. This could support and facilitate the decision process if consumers are not familiar with a particular institution but are generally willing to share data with organizations in this sector.

Initially, the consumer will only view the intended usage category of the data in the default form. This information is necessary for the consumers to be aware of the general intended use of the inquiring institution. Since the inquiring institution chooses among predetermined options, the consent registry offers a quick overview of the intended purpose to the consumer. This might facilitate an informed decision-making. However, the consumer is able to extend this field as well, receiving the detailed description of the intended purpose given by the inquiring institution in the request form (see Fig. 4b).

Since the type of the requested data is crucial information to the consumer, the consent registry immediately presents this information, along with the other three fields. The tool further displays information on the collection period of the requested data when the consumer expands this field. This information is crucial and should remain accessible to the consumer. However, to balance comprehensibility and exhaustiveness, we decided to present only the information on the type of data initially.



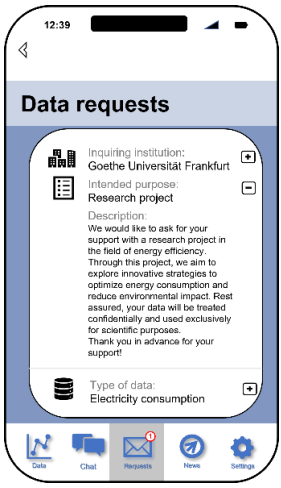


Fig. 4a and Fig. 4b: Expanded fields inquiring institution (4a, left) and intended purpose (4b, right)

The last information that the consent registry initially presents to the consumer is the retention period of the data. This information is another key aspect consumers need to be aware of to enable a comprehensive understanding of the implications of sharing data, thereby supporting the informed consent. It contains implications about the long-term effects of data sharing and should therefore be immediately visible to the consumer. This field, in the current form of the concept registry, does not include additional information and therefore the consumer is not able to expand this field.

Above all, this concept is supposed to enable an informed consent for consumers. On the one hand, it reduces and structures the given information to minimize the cognitive load and facilitate the understanding of the consequences of the consent to share data by employing a standardized form. On the other hand, consumers are able to access additional information, e.g., a comprehensive description of the intended use, in addition to a general classification, therefore allowing a comprehensively informed consent.

However, receiving individual requests could potentially burden consumers. Through the application providers' function as intermediaries, consumers will probably only receive a reduced amount of requests, thereby decreasing the cognitive load again. Future research could explore further approaches to limit the burden of requests for consumers even more, through, for instance, only presenting requests on a daily or weekly basis.

4 Conclusion and Outlook

The goal of our work is to sketch a tool to enhance consumer data sovereignty. Our concept encompasses three stakeholder groups involved in a data request: consumers producing

data, inquiring institutions desiring data access and application providers in a digital ecosystem as intermediaries with a certain responsibility toward their users. Our tool addresses the prevalent information asymmetries. While lawmakers have already addressed this issue with legislative measures, prior investigations indicate that they might not be sufficient to enable an informed consent for consumers [Ef19, Ba22]. While concepts such as privacy icons and one-page solutions aim to empower consumers, they also have notable shortcomings. We address this issue by building on these concepts and combining them with progressive disclosure, offering consumers an initially simplified request while providing the option to access additional information on demand. Our tool facilitates the consumers' exercise of power proposed by legislative measures like the GDPR and thereby helps strengthen consumer data sovereignty. The presented concept is applicable to various different use cases, encompassing various data types across different industries. The developed tool could potentially enhance data handling transparency by presenting the individual requests to the consumer in a more comprehensible format. We specifically positioned our exemplary use case within the context of a digital ecosystem, as such environments in particular may lead to consumers not being aware of who might have access to their data and the potential consequences of such data sharing. Yet, the consent registry can not only be applied in digital ecosystems, but also in single applications as well. The consent registry therefore represents another step towards enabling informed consent and enhancing consumer data sovereignty while also promoting (responsible) data sharing and addressing data silos. Its applicability to different digital ecosystems and use cases should support widespread adoption in practice.

Future research could consider the validation of the consent registry with consumers in real-world scenarios to test its effectiveness and consumer acceptance. We will contribute to this initial evaluation by implementing our proposed concept in a case study and testing the concept with consumers in a digital ecosystem in the context of smart living appliances. In real-world application, we will further develop the tool and implement changes as needed. For instance, if the consumers prefer it, we might implement the possibility to disclose data in bulks, rather than limiting the disclosure to individual datasets. Additionally, we might limit the retention period by only allowing requests with a retention period that does not exceed a certain time. However, the priority of consumer data sovereignty will remain in our tool and thus changes must not compromise this objective. Another important avenue for future research is ensuring that companies adhere to their stated purpose of data usage. The current consent registry does not guarantee this, but addressing it is crucial for fostering user trust.

Acknowledgement

The project COMET is funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK) as part of the technology program "SmartLivingNEXT – Artificial Intelligence for Sustainable Living Environments." SmartLivingNEXT is creating a universal, AI-based ecosystem for the simple and cost-effective development of intelligent and sustainability-oriented smart living services and applications.

Bibliography

[Ac21] Ackermann, K.; Miesler, L.; Mildenberger, T.; Frey, M.; Bearth, A.: Willingness to share data: Contextual determinants of consumers' decisions to share private data with companies. In: Journal of Consumer Behaviour 21, pp. 375–386, 2021

[AKR22] Andreotta, A. J.; Kirkham, N.; Rizzi, M.: AI, big data, and the future of consent. In: AI & SOCIETY 37, pp. 1715–1728

[Ba22] Bahrini, M.; Zargham, N.; Wolff, A.; Kipker, D.K.; Sohr, K.; Malaka, R.: It's Long and Complicated! Enhancing One-Pager Privacy Policies in Smart Home Applications. In: Nordic Human-Computer Interaction Conference '22., pp. 1–13, 2022

[BC11] Bélanger, F.; Crossler, R. E.: Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. In: MIS Quarterly 35, pp. 1017–1041, 2011

[Ca21a] Calzada, I.: Data Co-Operatives through Data Sovereignty. In: Smart Cities 4, 1158-1172, 2021[Ca21b] Carl, K. V.: Corporate Digital Responsibility: Evaluating Privacy and Data Security Activities on Company-level. In: INFORMATIK 2021, pp. 757–771, 2021

[Ca22a] Calzada, I.: Citizens' Data Privacy in China: The State of the Art of the Personal Information Protection Law (PIPL). In: Smart Cities 5, 1129-1150, 2022

[Ca22b] Carl, K. V.: The status-quo of companies' data privacy and security communication: An ethical evaluation and future paths. In: INFORMATIK 22, pp. 195–206, 2022

[CC15] Capistrano, E. P. S.; Chen, J. V.: Information privacy policies: The effects of policy characteristics and online experience. In: Computer Standards & Interfaces 42, pp. 24–31, 2015

[CH24] Carl, K. V.; Hinz, O.: What we already know about corporate digital responsibility in IS research: A review and conceptualization of potential CDR activities. In: Electronic Markets 34, pp. 1-30, 2024

[CM22] Custers, B.; Malgieri, G.: Priceless data: why the EU fundamental right to data protection is at odds with trade in personal data. In: Computer Law & Security Review 45, 2022

[Ef19] Efroni, Z.; Metzger, J.; Mischau, L.; Schirmbeck, M.: Privacy Icons: A Risk-Based Approach to Visualisation of Data Processing. In: European Data Protection Law Review 5, pp. 352–366, 2019

[Gr24] Grafenstein, M.; Kiefaber, I.; Heumüller, J.; Rupp, V.; Graßl, P.; Kolless, O.; Puzst, Z.: Privacy icons as a component of effective transparency and controls under the

GDPR: effective data protection by design based on art. 25 GDPR. In: Computer Law & Security Review 52, 2024

[Ha17] Haeusermann, T.; Greshake, B.; Blasimme, A.; Irdam, D.; Richards, M.; Vayena, E.: Open sharing of genomic data: Who does it and why? In: PLOS ONE 12, 2017

[He25] Hering, F.; Hinz, O.; Pfeiffer, J.; Aalst, W. van der: The Damocles Sword of Cyber Attacks - A Call for Information Systems Security. In: Business & Information Systems Engineering 67, forthcoming, 2025

[HG18] Howard, M. L.; Gaviola, M. L.: Progressive disclosure cases: The design and evaluation of use in multiple therapeutics courses. In: Currents in Pharmacy Teaching and Learning 10, pp. 723–729, 2018

[Hu21] Hummel, P.; Braun, M.; Tretter, M.; Dabrock, P.: Data sovereignty: A review. In: Big Data & Society 8, pp. 1–17, 2021

[JB24] Jøranli, I.; Breunig, K. J.: Unlocking data's potential: navigating the challenges of data-driven innovation for start-ups. In: Measuring Business Excellence 28, pp. 334–346, 2024

[JLL21] Jelovac, D.; Ljubojević, Č.; Ljubojević, L.: HPC in business: the impact of corporate digital responsibility on building digital trust and responsible corporate digital governance. In: Digital Policy, Regulation and Governance 24, pp. 485–497, 2021

[KKS20] Korunovska, J.; Kamleitner, B.; Spiekermann-Hoff, S.: The challenges and impact of privacy policy comprehension. In: Twenty-Eighth European Conference on Information Systems 2020, pp. 1–17, 2020

[Ko17] Kokolais, S.: Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. In: Computers & Security, 64, pp. 122–134, 2017

[Ko15] Korenhof, P.; Ausloos, J.; Szekely, I.; Ambrose, M.; Sartor, G.; Leenes, R.: Timing the Right to Be Forgotten: A Study into "Time" as a Factor in Deciding About Retention or Erasure of Data. In: (Gutwirth, S., Leenes, R., de Hert, P. eds.) Reforming European Data Protection Law, pp.171-201, 2015

[Li17] Li, Y.; Kobsa, A.; Knijnenburg, B.; Nguyen, C.: Cross-Cultural Privacy Prediction. In: Proceedings on Privacy Enhancing Technologies 2017, pp. 113–132, 2017

[Lo24] Long, Y.; Luo, X.; Zhu, Y.; Lee, K. P.; Wang, S. J.: Data Transparency Design in Internet of Things: A Systematic Review. In: International Journal of Human–Computer Interaction 40, pp. 5003–5025, 2024

[LYH19] Li, H.; Yu, L.; He, W.: The Impact of GDPR on Global Technology Development. In: Journal of Global Information Technology Management 22, pp. 1–6, 2019

[Mc16] Mccormack, P.; Kole, A.; Gainotti, S.; Mascalzoni, D.; Molster, C.; Lochmuller, H.; Woods, S.: „You should at least ask“. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. In: European Journal of Human Genetics 24, pp. 1–6, 2016

[MF16] Mittelstadt, B.D., Floridi, L.: The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. In: Science and Engineering Ethics 22, pp. 303–341, 2016

[MFF24] Marxen, H.; Frank, M.; Fridgen, G.: The Role of Gender in Data Sharing for Smart Charging of Electric Vehicles. In: AMCIS 2024 Proceedings, 2024

[Mi22] Mihale-Wilson, C.; Hinz, O.; Aalst, W. van der; Weinhardt, C.: Corporate Digital Responsibility. In: Business & Information Systems Engineering 64, pp. 127-132, 2022

[Mu24] Muralidhar, D.: The Effect of Progressive Disclosure in the Transparency of Explainable Artificial Intelligence Systems. In: 2024 IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 382–383, 2024

[PAZ22] Podgorelec, B.; Alber, L.; Zefferer, T.: What is a (digital) identity wallet? A systematic literature review. In: IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC) 2022, pp. 809–818, 2022.

[Ro15] Roeber, B.; Rehse, O.; Knorrek, R.; Thomsen, B.: Personal data: how context shapes consumers’ data sharing with organizations from various sectors. In: Electronic Markets 25, pp. 95–108, 2015

[RP20] Rossi, A.; Palmirani, M.: Can Visual Design Provide Legal Transparency? The Challenges for Successful Implementation of Icons for Data Protection. In: Design Issues 36, pp. 82–96, 2020

[RS18] Renaud, K.; Shepherd, L. A.: How to Make Privacy Policies both GDPR-Compliant and Usable. In: 2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment, pp. 1–8, 2018

[SE24] Schramm, J.; Eichinger, T.: Towards Building GDPR-Friendly Consent Management Systems on Top of Self-Sovereign Identity Ecosystems. In: (Roßnagel, H.; Schunck, C. H.; Sousa, F. eds.) Open Identity Summit 2024, pp. 93-102, 2024

[Sm25] SmartLivingNEXT,https://smartlivingnext.de/en/, accessed:14/04/2025

[SW19] Springer, A.; Whittaker, S.: Progressive disclosure: empirically motivated approaches to designing effective transparency. In: Proceedings of the 24th International Conference on Intelligent User Interfaces, pp. 107–120, 2019

[TF17] Taddeo, M., Floridi, L.: The Moral Responsibilities of Online Service Providers. In: (Taddeo, M., Floridi, L. eds.) The Responsibilities of Online Service Providers. Law,

Governance and Technology Series, 31, pp. 13-42, 2017.

[TLS17] Thiebes, S.; Lyytinen, K.; Sunyaev, A.: Sharing is About Caring? Motivating and Discouraging Factors in Sharing Individual Genomic Data. In: ICIS 2017 Proceedings, pp. 15, 2017

[Tr23] Trier, M.; Kundisch, D.; Beverungen, D.; Müller, O.; Schryen, G.; Mirbabaie, M.; Trang, S.: Digital Responsibility. In: Business & Information Systems Engineering 65, pp. 463-474, 2023

[VT21] Viktoratos, I.; Tsadiras, A.: Personalized Advertising Computational Techniques: A Systematic Literature Review, Findings, and a Design Framework. In: Information 12, 2021

[Wa20] Waerdt, P. J. van de: Information asymmetries: recognizing the limits of the GDPR on the data-driven market. In: Computer Law & Security Review 38, 2020

[WGW19] Wessels, N.; Gerlach, J.; Wagner, A.: To Sell or not to Sell – Antecedents of Individuals' Willingness-to-Sell Personal Information on Data-Selling Platforms. In: ICIS 2019 Proceedings, 2019

[Wi16] Wieneke, A.; Lehrer, C.; Zeder, R.; Jung, R.: PRIVACY-RELATED DECISION-MAKING IN THE CONTEXT OF WEARABLE USE. In: PACIS 2016 Proceedings, 2016

[Wi22] Windl, M.; Ortloff, A.-M.; Henze, N.; Schwind, V.: Privacy at a Glance: A Process to Learn Modular Privacy Icons During Web Browsing. In: Proceedings of the 2022 Conference on Human Information Interaction and Retrieval '22, pp. 102–112, 2022

The role of EBSI in eIDAS – how qualified ledger with governmental trust anchor could shape eIDAS ecosystem

Ignacio Alamillo[1], Steffen Schwalm[2]

Abstract: eIDAS 2.0 as a legal and technical framework for trustworthy, decentralized identities in conjunction with the EU digital wallet and various trust services introduced Electronic Ledger as own (qualified) trust service. This could lead to a rise in distributed ledger technologies (DLT) in general and European Blockchain Services and Infrastructure (EBSI) in particular. As EBSI is provided by Member States it contains governmental trust anchor per default and so strength the European digital sovereignty if EBSI could be integrated into eIDAS ecosystem

Keywords: eIDAS, EU Digital Wallet, DLT, EBSI, Electronic Ledger

1 Introduction and status of DLT in Europe

Within the framework of the European Blockchain Partnership (EBP), the European Commission established a European DLT-infrastructure provided by the Member States EBSI. This means that the DLT nodes are under the responsibility of the Member States and so ensure a government trust anchor. Since EBSI contains its own governance and technical specifications together with conformance tests for wallets it could solve the trustworthiness issues in DLT but, as it lacks security standards and independent audit processes, the growth of EBSI was limited.

The [eIDAS] establishes as an amendment of eIDAS 1.0 a legal and technical framework for trustworthy decentralized identities with the EU Digital Wallet (EUDIW) and related (qualified) trust services, using or not DLT, on one hand but with a new dedicated Qualified Trust Service Provider (QTSP) for Electronic Ledger on the other hand. Although the term Electronic Ledger in [eIDAS] does not necessarily mean only DLT – even less, blockchain – this regulation seems like a step forward to close the gaps and to enable DLT to be used in regulated environments with typically comprehensive requirements on proven security and legal trust. As several project within the Digital Europe Programme already work to establish decentralized ecosystem based on EBSI as infrastructure provided by Member States and so governmental trust anchor for European companies and citizens the following questions occur. What's the role of DLT within [eIDAS] in general? Which role could EBSI play between EUDI Wallet and qualified trust services?

[1] Logalty C/ de Cantabria, 2 28108 Alcobendas, Spain, ignacio.alamillo@logalty.com
[2] msg systems ag, Repobert-Buerkle.-Str. 1, 85737 Ismaing, Germany, steffen.schwalm@msg.group

doi:10.18420/OID2025_02

The paper describes based on introduction on [eIDAS] and electronic ledger in general (Section 2) as well as EBSI (Section 3) it discuss possible roles EBSI could play within the [eIDAS] ecosystem (Section 4) The paper closes with overview on ongoing standardization and considerations on the future of EBSI in eIDAS ecosystem.

2 Electronic Ledger in eIDAS and other regulations

2.1 eIDAS Regulation

In April 2024 eIDAS 2.0 as amendment of eIDAS 1.0 was published [eIDAS]. As [eIDAS]requires mandatory implementing acts for each component referencing European standards from ETSI or CEN the regulation also creates a much more coherent technical framework than eIDAS 1.0 where only fewer implementing acts were mandatory. The presumable biggest change in [eIDAS] is the requirements for every Member state to provide an EU-Digital Wallet (EUDIW) to its natural entities. The Wallet could be published by member state, under authority of member state or recognized by member state. This makes also private wallet possible under the recognition of a Member State. Any EUDIW will contain a Personal Identification (so called PID for natural or legal entity as wallet holder) based on notified eID scheme on LoA “high” and has to achieve LoA “high” itself. Directly corresponding with the EU-Digital Wallet the new qualified attestation services acc. Art. 45a-e [eIDAS] have to be taken into account. This means that EU-Digital Wallet will contain the core identity currently covered by government eID as well as additional attributes. Beside attestation of attributes, it`s worth to mention that it`s required that the EUDI Wallet can be used for creation of (qualified) signatures/seals by law [eIDAS] but practically needs interaction also with other (qualified) trust services to serve fully digital business processes and so comprehensive digitalization. EUDIW and (qualified) trust services) will be certified by independent Conformity Assessment Body (CAB) which ensures the proven security. In the consequence [eIDAS] crosses digital identity means and (qualified) trust services – they determine each other and for both DLT as infrastructure is possible to use. The core requirements on QTSP like liability, periodical re-certification, reporting obligation on security issues etc. remain in [eIDAS].

2.2 Electronic Ledger in eIDAS

With Section 11 [eIDAS] also introduces (qualified) trust services on Electronic Ledger (Art. 45h following). eIDAS defines Electronic ledger as a sequence of electronic data records, ensuring the integrity of those records and the accuracy of the chronological ordering of those records. Art .45l [eIDAS] defines that “Qualified electronic ledgers shall meet the following requirements:

- they are created and managed by one or more qualified trust service providers.

- they establish the origin of data records in the ledger;
- they ensure the unique sequential chronological ordering of data records in the ledger;
- they record data in such a way that any subsequent change to the data is immediately detectable, ensuring their integrity over time."

This definition is based on definition of the term "ledger" in ISO 22739 [ISO01] which was adopted as CEN EN into European standardization Framework and is so mandatory in terms of European standardization. In order to ensure a coherent European standardization framework as requested by CEN Internal Regulation Part 1 as well as ETSI Directive it`s recommended to follow this definition:

- Ledger 3.54.: information store that keeps records of transactions that are intended to be final, definitive and immutable
- Distributed Ledger: ledger that is shared across a set of distributed ledger technology (DLT) nodes and synchronized between the DLT nodes using a consensus mechanism

This means that DLT is technical only a special kind of ledger and the basic properties like immutability as well as the fact that on ledger records are final per definition are valid for any kind of ledger. The only difference between a ledger and a distributed ledger is the distributed provision. Both kinds of ledger are covered by [eIDAS]. As the regulation is technology neutral it`s possible to use DLT as infrastructure for or by any component or actor within the eIDAS 2.0 ecosystem:

- Infrastructure and comprehensive components like e.g. List of EUDI Wallets (TrustList) acc. Art.5a [eIDAS], List of Relying party acc. Art.5b [eIDAS], TrustList acc. Art. 22 [eIDAS]
- EUDI Wallet and/or (qualified) trust services using electronic ledger
- QTSP for Electronic Ledger.

Since the legal and certification requirements those will be proven by independent conformity assessment body the security and trust in DLT is ensured in any case.

2.3 Electronic Ledger in other regulations and initiatives

Beside the eIDAS regulation also the Data Act [Da23] mentions the Electronic Ledger. Recital 104 states that "To promote the interoperability of tools for the automated execution of data sharing agreements, it is necessary to lay down essential requirements for smart contracts which professionals create for others or integrate in applications that support the implementation of agreements for data sharing." Those agreements can be

implemented as smart contract which could run on an electronic ledger. Similar subjects can also be found in regulations and related initiatives on Digital Product Passport, Supply chain or protection of intellectual property [TR24].

3 Fundamentals on European Blockchain Service Infrastructure

The project, which was set up in 2018, aims to lay the foundation for future ledger based services within the EU and EFTA. The EBSI is maintained and operated by the European Digital Infrastructure Consortium (EDIC). It shall be understood that because of GDPR the EBSI design does not intend to anchor any citizen data on immutable electronic ledgers. EBSI contains:

- comprehensive governance covering onboarding, offboarding core security as well as roles and responsibilities
- trust model including trusted issuer and verifier registry [EBSI04]
- fundamental architecture requirements [EBSI01]
- technical framework covering
    - protocols, formats, issuance requirements, validation and revocation requirements for verifiable credentials or in eIDAS terms attestations of attributes aligned with technical framework [EBSI05] related to eIDAS
    - own DID method incl. Resolver [EBSI02] to combine identities, attestations and transactions
    - interfaces and requirements for anchoring data on ledger and interaction of off-ledger records [EBSI05]
- requirements and conformance testing on wallets [EBSI03]

The EBSI validator nodes are fully operated by Member States. This leads to the determination that EBSI contains governmental trust anchor per default. Any application or additional node always build and rely on those core network from European Member States as shown on the picture below:

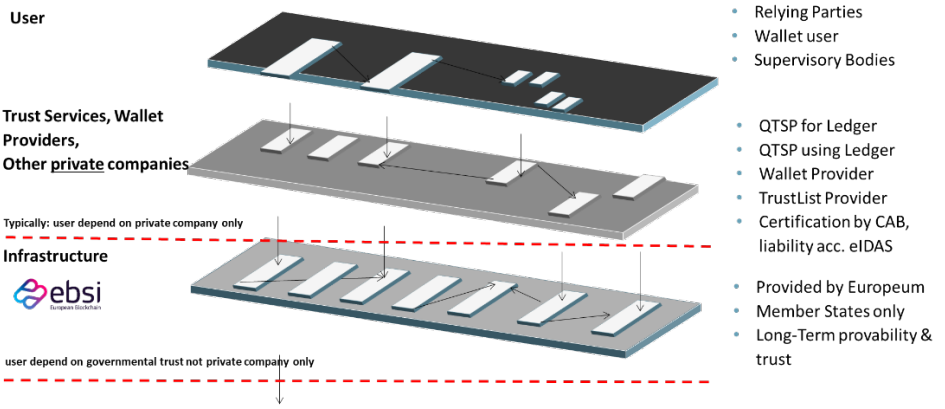


Figure 1: Layers of EBSI

This approach aligns with the decentralized nature of DLT acc. ISO 23257 [ISO02] and is suitable for multi-party cooperation. EBSI on one and it ensures a governmental trust anchor and so clear responsibility on the other hand this approach leads to the question on how such a network might be provided (QTSP for Electronic Ledger) or use (by EUDI Wallet Issuer or QTSP using DLT) by a certain provider. Currently the EBSI network is improved for production [EBSI], used to pilot the EUDI Wallet [DC4EU] and to establish cross-border use cases in combination of digital identities such as Legal PID acc. Art. 5a [eIDAS], attestations of attributes and transactions between companies using the ledger as decentralized PKI and infrastructure for the transactions themselves:

- EBSI VECTOR [VE24]
  - Business Wallets and Legal PID inc. Power of Attorney
  - Education
- TRACE4EU [TR24]
  - Digital Product passport and supply chain acc. European Regulation
  - Evidence and trade of intellectual property and other Attestations

Against this background EBSI could possibly fulfil nearly all roles defined in section 2.2. How this could be done using the EBSI capabilities describes section 4 further.

4 Possible Role of EBSI in eIDAS

4.1 Infrastructure and comprehensive components

As described in section 2.2 an Electronic Ledger can basically support in the provision of Infrastructure and comprehensive components like e.g. inclusion EUDI Wallets in TrustList acc. Art.5a [eIDAS], List of Relying party acc. Art.5b [eIDAS], TrustList acc.

Art. 22 [eIDAS]. In terms of EBSI this would mean: As EBSI already contains a list of conformant wallets same could be done for EUDI Wallet. The issuance of certification of EUDIW could be executed using the trust model. It´s only necessary to define the (Root-)Trusted accreditation organization (RootTAO/TAO), typically the authority issuing the EUDIW in certain member state which grants the allowance for issuance to the authority supervising the conformity assessment and issuing the certificate or endorsement for EUDIW, The whole procedure can be automized using the ledger. Would also mean that any revocation could be done in real-time too so not only in data base but related to transaction in ledger – from the time of revocation the validity of the certain EUDI wallet is deleted too and as it`s done on immutable ledger the revocation can`t be undo, also no cheating from anybody possible as the EBSI is public permissioned so any change publicly provable. Similar approach applies for List of Relying parties and Trust list. In case of List of Relying parties it`s the reuse of existing Trusted Verifier Registry, as the Relying Parties are practically a verifier of certain PID/LPID or e.g. (qualified) attestations of attributes (also if they may use validation services) within the eIDAS. This could mean:

Role acc. eIDAS	**Role acc. EBSI**	**Task**
(Root-)TAO	Organization defining the Registrar	Accreditation of Registrar as Trusted Issuer
Trusted issuer	Registrar	Issuance Relying Party registration certificate Issuance Relying Party access certificate

Table 1. EBSI and Relying Party Registration

According to Implementing Act on Art. 5b the Relying Party Access Certificate " means a certificate for electronic seals or signatures authenticating and validating the wallet-relying party issued by a provider of wallet-relying party access certificates". Even an attestation of attributes can be used for electronic seals and signatures as both can be combined. As the Implementing Act contains the "the obligation for the providers of wallet-relying party access certificates to log all wallet-relying party access certificates they have issued to ensure certificate transparency, in compliance mutatis mutandis with IETF RFC 9162" and EBSI is public permissioned the EBSI contains the certificate transparency by default. Exactly this approach is already implemented within EBSI and any (Root-)TAO, TAO and Trusted Issuer, same as Trusted Verifier (the Relying party) provable via their DID:EBSI which DID Document reference the actual identity stored off-chain. Similar approach applies for a performant alternative to the XML TrustList acc. to [eIDAS] which is in EBSI represented through the Trusted issuer Registry. As shown below the eIDAS Trust Model related to conformity assessment and listing of QTSP can be fully mapped to EBSI.

Role acc. eIDAS	Role acc. EBSI	Task
Supervisory Ministry	Root TAO	Accreditation National Supervisory Body as Trusted Issuer Attestation QTSP in case of appeal of QTSP against decision of National Supervisory Body
National Supervisory Body	Trusted Issuer	Issuance Attestation Trust Service qualification to applicant (possible QTSP) Issuance (or revocation) qualified status of Trust Services Issuer for Attestation to be listed in TrustList
	TAO	Accreditation applicant on QTSP to issue attestation to QTSP to start conformity assessment
Conformity Assessment Body (CAB)	Trusted Issuer	Issuance Conformity Assessment Report

Table 2: EBSI and (qualified) Trust Services (TrustList)

In case the Service identifier defined in ETSI TS 119 612 [ETS02] could be a DID:EBSI the whole procedure so issuance and revocation of qualified status of certain QTSP could be automized using EBSI similar as to issuance and revocation of Relying Party registration and/or access certificates. This would practically mean that EBSI in general and its trusted issuer registry in particular could technically serve the List of Relying parties acc. Art. 5b [eIDAS] or the one for EUDIW resp. replace the existing TrustList acc. Art. 22 eIDAS resp. ETSI TS 119 612 which is currently implemented in XML. As with the new (qualified) trust services like QEAA the number of QTSP will foreseeably increase the current XML may not fulfil the required scalability across Europe. In the final consequence this means that EBSI could be used to build up complete trust chains in order to automize the trust framework of [eIDAS] and ensure the real-time provability of Status of QTSP or EUDI Wallet and so enhance the trustworthiness of conformity assessment reports and needed certificates for conformity assessment.

As EBSI is provided by the Member States it would also solve the requirement of a publicly provable TrustList as defined in Art. 22 and ETSI TS 119 612 [ETS02]. The compilation on level of the European Commission seems not necessary anymore with EBSI due to the fact that all nodes contain all transactions within a public permissioned

ledger per definition [ISO02], [ISO06]. As EBSI is immutable and contains governmental trust anchor per default also any cheating avoided as it´s transparent to anybody especially in case the nodes used for trusted issuer (same applies for trusted verifier registry) provided by (qualified) trust service for ledger.

In summary EBSI could bring the added value of reusing already existing sustainable solution for listing of EUDI Wallets and Relying Parties as well as scalable one for TrustList. Beside this the ledger adds the issuance and revocation in real time including the public transparency in real time as any transaction directly executed and provable which would ensure same security level for all member states instead of possible race to the bottom as registration Relying Party, EUDI Wallet and QSP (incl. conformity assessment) can be done in any Member State. The governmental trust anchor should not be forgotten as this is unique property of EBSI as public infrastructure. The [VE24] Project developed complete mapping of all the 3 options on infrastructure and comprehensive components including an example implementation [AlSc24].

4.2 EUDI Wallet PID and LPID

The already existing Trusted Issuer Registry of EBSI could be reused as solution for a scalable Trust List in [eIDAS]. EUDI and other wallets. As EBSI already contains conformance testing for wallets including list of conformant wallets provable using the ledger the use for EUDI Wallet means only to add the certification requirements for EUDI Wallet and integration interaction eIDAS Trust Model with EBSI one as described in section 4.1. An Electronic Ledger in general and EBSI in particular can be used as decentralized PKI for certain EUDI Wallet for natural and/or legal entities (Enterprise Wallet). PID of Legal entities is already piloted within [EBSIVECTOR]. In case EBSI is used as infrastructure for the EUDIW and to ensure the migration of existing wallets certified against EBSI technical framework it`s required to integrate the EBSI specific requirements into the relevant standards on EUDIW defines by ESI and/or CEN. The reason is that the Conformity Assessment Body will use those standards for certification of the EUDIW. With regard to the obligations to issue and accept an EUDIW for Member States and defined relying parties it´s also recommended to enable EUDIW using EBSI to store not only the PID on LoA “high”, but additionally other Attestation used for personal identification which could be executed using (Q)EAA. In this case an EBSI Wallet which is certified as EUDIW can be used in 2 modes: LoA “high” and LoA “substantial” – recommended especially in case LoA “high” may limit the user-friendliness of the EUDIW. As existing EBSI wallet in most cases fulfil LoA “substantial”, the described approach could ensure an easy migration and growth of the EBSI ecosystem within [eIDAS]. In order to ensure that existing EBSI certified wallets may be used as EUDIW in [eIDAS] the certification framework for EUDIW should contain optional parts, means if certain applicant for EUDIW wants to become EBSI compliant at same time, the CAB shall assess the related conditional requirements too.

4.3 (qualified) Trust Services using EBSI

EBSI could be used as infrastructure for following (qualified) trust services: Especially the QEAA would serve the current use cases of EBSI which are mainly focused on issuance and verification of verifiable credentials as identity attributes such as diploma, social security, digital product passport etc. using EBSI certified wallet which have toe be migrated into the eIDAS ecosystem The aims could be [AlSc24]:

qualified) Trust Service (examples)	**Role of EBSI**
Issuance (qualified) certificates for (qualified) electronic signatures, seals and/or timestamps,	Decentralized PKI
Issuance (qualified) timestamps only	Delivery of DLT native (qualified) timestamps
Validation of (qualified) electronic signatures, seals and/or timestamps	Decentralized PKI
Issuance of (qualified) attestations of Attributes	Decentralized PKI
(qualified) Electronic registered mail/ delivery services	Audit Trail (as piloted e.g. in [TR24]
Issuance of qualified) website certificates	Decentralized PKI
(qualified) Archiving	Tracing and Logging [Ce25b]

Table 3: EBSI as infrastructure for QTSP using ledger

4.4 QTSP for Ledger

For QTSP for ledger EBSI or those part relevant for the certain QTSP would be part of the related conformity assessment executed by certain CAB and subject to the requirements on QTSP acc. Art. 20 ff. and Section 11 [eIDAS]. The ledger could be used for certain use case like supply chain or tokenization but also act as infrastructure (or trust service component acc. ETSI EN 319 403 [ETSI03] to other QTSP especially in case of combination of identity/attestation and transactions like for e.g. supply chain and digital product passports [TR24]. With regards to the specific governance of EBSI after which the EDIC as fundamental operating organization always provides main nodes always the following questions occurs to define the concrete design of the role of EBSI in case of QTSP for Ledger. In order to solve certain challenges like e.g. longevity of cryptographic algorithm EBSI would interact with other trust service like preservation as described in ISO TR 24332 [ISO04].

4.5 Summary

In case EBSI is used as infrastructure for EUDIW and/or QTSP using the EUDIW as well as the (qualified) trust service would be application or services build on top. These applications together with the infrastructure they use form the actual wallet and especially (qualified) trust service as subject of conformity assessment of the CAB. This leads to the conclusion that in case EBSI is used as infrastructure for EUDIW or QTSP using Ledger both will achieve additional trust in comparison to other EUDIW/QTSP not using EBSI as EUDIW/QTSP using Ledger will always rely on governmental trust – exactly the basic infrastructure provided by EDIC. As EBSI can serve all components of eIDAS: Alternative to TrustList, listing EUDIW, Registration Relying Parties, infrastructure for EUDIW, trust services or trust service itself it also enables the combination of digital identities (PID/LPID), attestations and transactions cross-organizations, cross-industry, cross-border in one infrastructure including the signing, timestamping etc. of transactions. It could serve as trusted infrastructure for core European initiatives and regulations like: Digital Product Passport and verification of supply chains [TR24], [UN24], Data Sharing using smart contracts acc. Data Act in industry 4.0, proof and trade of intellectual and other properties, tokenization [Mi23] up to the digital euro which is already combined with [eIDAS]. As EBSI contains governmental trust anchor per default it would ensure also European digital sovereignty on its identity and core services.

5 Standardization and Outlook

The standardization on technical framework on [eIDAS] is mainly done in ETSI and CEN. CEN JTC 19 applies for the standards related to electronic ledger. This contain the Policy & Security Requirements for trust services providing ledger services so basically the requirements relevant for any electronic ledger but also EBSI to be used for the any subject mentioned in section 4 so e.g. registration of Relying Parties, alternative to TrustList, Trust Service Component to other trust services or ledger as trust service itself. Beside this it´s necessary to standardize the security, functional and interoperability requirements of Decentralized Identifiers and related methods with regards to ledger as DID in general and DID:EBSI is exactly one key for the unique property to combine (L-) PID, attestations and transactions in one trustworthy infrastructure. A new Technical Specification in CEN JTC 19 covering these requirements is under development. As CEN standards can be used by CAB for conformity assessment acc. to [EU12] both specifications directly serve the sustainable utilization of EBSI. The standardization of both [eIDAS] related projects is closely aligned with the worldwide framework in ISO Tc 307 to ensure international compatibility. This applies especially to ongoing work on ISO TS 23353 [ISO03] on auditing guidelines and ISO 23042 [ISO05] on decentralized identity management. In order to use EBSI as possible scalable alternative to current XML based TrustList an update of ETSI TS 119 612 [ETSI02] needed.

Electronic Ledger is here to stay and with EBSI Europe has the unique chance to use an

infrastructure based on governmental trust anchor by default to combine identity and transaction in one trustworthy infrastructure. Main requirement is the transformation of EBSI into the [eIDAS] ecosystem. EBSI could adopt all components in [eIDAS] and those mechanism especially in case of trust list (trusted issuer registry), trust list for wallets, list of Relying Parties (Trusted Verifier Registry) or issuance and verification of attestations of attributes already exists it seemingly makes no sense to reinvent the wheel with other solutions for same subject. The necessary standards are under construction so that main challenge seems to be the establishment of well working operating company with related marketable portfolio through EDIC and to ensure the reuse of the results from the ongoing research projects. This needs further analysis and solution development.

Bibliography

[AlSc24] Alamillo, Dr. I., Schwalm S., Stoecker, C., Thiermann, R.: Qualified Ledgers: Bridging the Gap between Blockchain Technology and Legal Compliance. 2024.

[ARF25] Architecture and Refence Framework. https://github.com/eu-digital-identity-wallet/eudi-doc-architecture-and-reference-framework/blob/main/docs/architecture-and-reference-framework-main.md

[Ce25a] CEN TS (WD) Functional and interoperability requirements on Decentralized Identifier (DID). 2025

[Ce25b] CEN TS (WD) Policy and security requirements for trust service providers providing electronic ledger services. 2025

[Ce25c] CEN TS 18170:2025. Policy and Functional requirements for electronic archiving services

[Da23] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (Data Act) (Text with EEA relevance). PE/49/2023/REV/1 OJ L, 2023/2854, 22.12.2023.

[Di24] Digital Credentials for Europe Project https://www.dc4eu.eu/

[EBSI] EBSI, European Blockchain Services Infrastructure, https://ec.europa.eu/digital-building-blocks/sites/display/EBSI/Home

[EBSI] EBSI-NE Project https://www.ebsi-ne.com/

[EBSI01] https://hub.ebsi.eu/apis

[EBSI02] https://hub.ebsi.eu/blockchain

[EBSI03] https://hub.ebsi.eu/conformance

[EBSI04] https://hub.ebsi.eu/vc-framework/trust-model

[EBSI05] https://hub.ebsi.eu/vc-framework

[eIDAS] REGULATION (EU) 2024/1183 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 April 2024 amending Regulation (EU) No 910/2014 as regards

establishing the European Digital Identity Framework

[ETS01] ETSI TS 119 472-1. Electronic Signatures and Trust Infrastructures (ESI); Profiles for Electronic Attestations of Attributes; Part 1: General requirements

[ETS02] ETSI TS 119 612 Electronic Signatures and Infrastructures (ESI); Trusted Lists

[ETS03] ETSI EN 319 403 Electronic Signatures and Infrastructures (ESI); Trust Service Provider Conformity Assessment; Part 1: Requirements for conformity assessment bodies assessing Trust Service Providers

[EU12] REGULATION (EU) No 1025/2012 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 October 2012 on European standardization.

[ISO01] ISO 22739:2024: Blockchain and distributed ledger technologies - Terminology, 2024

[ISO02] ISO23257:2022: Blockchain and distributed ledger technologies – Reference Architecture, 2022

[ISO03] ISO WD TS 23353 Blockchain and distributed ledger technologies — Auditing Guidelines. 2024

[ISO04] ISO TR 24332. Information and documentation — Blockchain and DLT in relation to authoritative records, records systems, and records management. 2025

[ISO05] ISO 23042 (WD). Blockchain and distributed ledger technologies. Reference architecture for DLT-based decentralized identity systems.2025

[ISO06] ISO TS 23635:2022 ISO/TS 23635:2022. Blockchain and distributed ledger technologies — Guidelines for governance

[Mi23] Regulation (EU) 2023/1114 of the European Parliament and of the Council of 31 May 2023 on markets in crypto-assets, and (Text with EEA relevance). PE/54/2022/REV/1. OJ L 150, 9.6.2023, p. 40–205

[RF21] Laurie, B., Messeri, E, Stradling, R.: RFC 9162 Certificate Transparency Version 2.0.

[TR24] TRACE4EU Project https://trace4eu.eu/

[Un24] https://github.com/uncefact/spec-untp

[VE24] EBSI VECTOR project https://www.ebsi-vector.eu/en/

Developing A Governance Framework for Learning and Employment Record (LER) Issuer Registries

R.X. Schwartz [1], Kerri Lemoie [1], and Jeanne Kitchens[2]

Abstract: Learning and employment records (LERs) are increasingly deployed in the context of verifiable credentials, enabling more reliable and faster credential verification. However, there exists a missing link in the trust chain in order to substantiate if the issuer of a verifiable credential is trustworthy. For this reason, LER issuer registries, or lists of trusted issuers, are an increasingly relevant research area. An integral element of issuer registries is the governance framework used, which is essential for promoting trust in the issuer registry and supporting technical adoption. This study develops a governance framework for LER issuer registries as a part of a non-profit/higher education consortium research project, in collaboration with a group of eighty-one advisors from academia, government, private industry, standards bodies, and nonprofits. This advisory group participated in the brainstorming/data collection and iterative refinement of a governance framework using a two-stage process modeled on Mayer et. al's "Integrative Model of Organizational Trust" (ability, benevolence, and integrity perceived trustworthiness factors). The final governance framework consists of 7 categories which comprise a total of 48 guidelines (full governance framework provided as a table within the paper). The impact of this study is that trust in LER issuer registries can be studied and supported using a theoretically-grounded governance framework developed among industry stakeholders, promoting trusted engagement and authority of the issuer registry.

Keywords: Issuer Registry, Trust Registry, Learning and Employment Records, Verifiable Credentials, Trust, Governance

1 Introduction

The 2024-2025 Issuer Registry open-source project, a collaboration between the Digital Credentials Consortium higher education consortium, the Credential Engine non-profit, and a targeted advisory group of 81 subject matter experts/implementers, aims to provide a replicable and scalable open-source issuer registry pilot for the LER (Learning and Employment Record) space. The goal of the project is to enhance trust in Learning and Employment Records by creating a technical and governance pilot for interoperable issuer registries. An issuer registry is a cryptographically supported list of trusted credential issuers, enabling near-instantaneous trusted credential evaluation. Rapid and trusted credential evaluation is a novel functionality compared to traditional credential evaluations for professional licenses and degrees, which may take weeks to months to complete [Bo24].

One objective of the Issuer Registry pilot project is to develop and integrate both the technical and governance needs of an LER issuer registry. It is understood that LER

[1] Massachusetts Institute of Technology, Digital Credentials Consortium,
, https://orcid.org/0000-0002-8072-6420; klemoie@mit.edu, https://orcid.org/0009-0001-4812-2379
[2] Credential Engine, jkitchens@credentialengine.org

doi:10.18420/OID2025_03

issuer registry technical and governance aspects are mutually interdependent [Tr24a]: an appropriate technological approach will only be adopted if the governance of the issuer registry is accepted; and the governance of the issuer registry will be enabled by the capacity of the implementing technologies. Previous academic research has not yet analyzed governance needs in the LER issuer registry space. As such, a stakeholder-informed and theoretically-grounded governance framework was developed in the first phase (governance phase) of the Issuer Registry project to ensure that the eventual issuer registry technology would be deployed in alignment with stakeholder needs.

The aim of the governance phase of the project was to react to stakeholder needs and incorporate them in a best practices framework which answered the question "What is the minimal governance framework for a viable issuer registry?", informed by industry needs and academic research on trust. Capturing the socio-economic and socio-technical needs of project stakeholders first via an open-response exercise and then by a review exercise ensured that the advisory group was able to effectively communicate relevant requirements. Given the need to ensure that the governance framework could be interpreted according to inter-organizational trust, the exercises were conducted with respect to ability, benevolence, and integrity as defined in Mayer et. al's Integrative Model of Organizational Trust [MDS95].

The end result of the governance phase of this project (captured in Tab. 3) is an organized and classified framework of governance best practices for use in Learning and Employment Record issuer registry implementations. The governance best practices, listed out by category (*Registry Overview and Purpose, Governing Parties and Oversight, Privacy Policy and Terms of Use, Issuer Information and Verification, Data Download Format, Technical Standards, Security and Operations*) and further classified by ability, benevolence, and integrity enable implementers, researchers, and users of LER issuer registries to better understand and deploy governance guidelines to establish trust.

2 Background

2.1 Trust Registries and Issuer Registries

Trust registries are secure digital lists run by a trusted entity, which contain information about multiple organizations [Di21]. To join or maintain participation in a trust registry, an organization has to meet certain criteria: an organization is added to a trust registry by the registry operator once these criteria are met. The end result is that organizations can be easily and instantaneously checked to confirm if they have a particular status, often designated by governments, accreditation bodies, or training groups. Trust registries are a part of technology infrastructure in government document signing [Eu25b], public health [Eu25a], university environments [Va24], financial services [Ga23], and for USA interstate digital driver's licenses [Am25]. Various standards have been used to implement trust registries, including [IS21], [ET16], and [He24], among others.

An issuer registry is a specific kind of trust registry which contains organizations which are trusted to issue credentials (as opposed to a verifier registry). In the education domain, an issuer registry may list universities, schools, training courses, or testing centers.

2.2 Verifiable Learning and Employment Records

A key driver for the implementation of issuer registries in the education domain is the increasing scale and scope of educational and employment credentials, also known as Learning and Employment Records, or LERs [Go22]. These credentials, including course records, test results, or diplomas, can also be represented as digital credentials and can be implemented in a verifiable way [Sp25] [1E24]. Individuals or organizations which receive a credential from an issuer may want to confirm if the issuing institution is accredited and by whom. The method by which this verification can be completed is by creating a trust chain between the signed credential, the issuer, and an issuer registry. Currently, this trust chain for educational credentials is rarely made on an automated basis (and is often impossible to complete on an automated basis with paper records): a lack of automated checks may also increase delays and promote fraud [Bo24] [AD11]. Pilot implementations of issuer registries can demonstrate secure and rapid checks for issuer trust, which can serve as an enabling technology for verifiable LERs by completing this trust chain.

2.3 Governance and Trust

Digital governance is a concept which is discussed in various management fields. Relevant to the LER space is previous research that has been completed in the eHealth space, which involves a similar mix of institutional, personal, and governmental actors. [My16] defines governance components for maternal and child health registries to be "purpose specification, legal, fiscal, and operational responsibility, reporting requirements and enforceability, data quality, data security, confidentiality policies, and data access, and public engagement." These areas are relevant in some cases to LER issuer registries but may not be relevant regarding eHealth reporting requirements, confidentiality policies, and public engagement, due to the relatively non-sensitive nature of the data accessible via an LER issuer registry.

Other work on governance in digital platforms [Sc18] has described a classification of relevant trust elements for multi-sided platforms (MSPs, or platforms where users offer goods or services): "Governance structure", "Resources and documentation", "Accessibility and control", "Trust and perceived risk", "Pricing", and "External relationships." However, LER issuer registries do not follow the same economic logic as MSPs. An issuer registry is not necessarily more successful if adapted more widely or if more issuers join the issuer registry: in fact, some LER issuer registries will be more successful if specialized within specific subject domains. Additionally, low-quality offerings on MSPs would be expected to be offered at prices according to supply and demand, while an LER issuer registry may

in fact be provided at no cost and the level of trustworthiness of an issuer in the registry is usually determined by contextual factors, not by the market. Broader considerations of issuer registry design also suggest that not all MSP trust elements from [Sc18] are relevant in the LER issuer registry context (such as platform "modules"; cultivation of the platform ecosystem through APIs; and output evaluation, penalization, or rewards). Additionally, LER issuer registries, given the nature of educational data and government involvement in the university sphere, may be subject to certain legal and regulatory constraints which are not mentioned in [Sc18]. In sum, the components of governance in the LER issuer registry space have not yet been formally defined in the academic literature nor evaluated or designed among stakeholders of LER issuer registries.

Given the novelty of LER issuer registry technology, establishing trust in these issuer registries is of particular academic and practical interest. The relationship between governance and trust is disputed within organizational management literature [PV09], with authors noting both a potential substitutive effect and potential complementary effect [EE12]. We hypothesize that the LER issuer registry space, involving relatively little competition but the need for high standardization between the issuer registry and other stakeholders, is characterized by a complementary effect in which "formal control can contribute to increasing trust when parties agree on formal control to realize a common goal" [EE12]. For this reason, we apply Mayer et. al's Integrative Model of Organizational Trust [MDS95] (ability, benevolence, and integrity) to governance in the LER space, with the aim of identifying how governance criteria for LER issuer registries are related to (and can potentially promote) trust.

3 Methods

Determining and classifying governance guidelines for the LER issuer registry space was completed within a group of 7 higher education consortium/non-profit facilitators and 81 advisory group members, consisting of experts from academia, state and national governments, private companies, standards bodies, and nonprofits. All advisory group members were informed at the project start that results would be public and available under open licenses.

Development of the governance best practices framework occurred in two stages: a data collection stage with 30 advisory members in a live session and 5 higher education consortium/non-profit facilitators, followed by an asynchronous review session open to all 81 advisory members; and an iterative refinement stage reviewed with 21 advisory members in a live session and 5 higher education consortium/non-profit facilitators, followed by an asynchronous review session open to all 81 advisory members.

3.1 Stage 1 (Data collection)

A brainstorming exercise for governance guidelines ("Map which features of a hypothetical organization running an issuer registry contribute to three aspects of trust") was conducted among 30 advisory group members using an online sticky note collaboration tool. First, definitions of ability, benevolence, and integrity from [MDS95] were presented to the advisory group in a live session. Then, a 3-circle collaborative digital Venn diagram mapping ability, benevolence, and integrity was shared with the group, with 7 sticky notes for governance guidelines already added to the diagram by the higher education consortium/non-profit facilitators to spark discussion. The 30 members present in the live session or the 81 members in the larger group added 22 sticky notes to the diagram and rearranged the notes over the following week. Additional notes were recorded by facilitators.

3.2 Stage 2 (Iterative refinement)

The data collected from Stage 1 was sorted into categories and refined by one of the higher education consortium/non-profit facilitators with a background in organizational governance. In this initial step, the goal was to isolate core sticky notes and themes to identify "the minimal governance framework for a viable issuer registry." This facilitator also referenced

Category name	**Description**
Registry Overview and Purpose	Defines the scope, objectives, and operational principles of the issuer registry.
Governing Parties and Oversight	Addresses the roles and responsibilities of the governing bodies and mechanisms for ensuring transparent, effective oversight of an issuer registry.
Privacy Policy and Terms of Use	Establishes the legal and contractual rights and obligations of the issuer registry with respect to data providers and data users.
Issuer Information and Verification	Details the processes and requirements for onboarding, verifying, and maintaining issuer data.
Data Download Format	Details the processes and policies regarding downloading issuer registry data.
Technical Standards	Establishes the technical foundation for an issuer registry based on specific standards used.
Security and Operations	Describes cybersecurity and technical operations information regarding how the issuer registry code is operated and maintained.

Tab. 1: Description of project output (final governance framework) categories for LER issuer registries.

policies from two academic credential networks ([Tr24b], [Ve23]) and previous notes taken during this project's advisory meetings. Several sticky notes were merged or split apart, and the sticky note "Skills Translation" was removed because it was not relevant to the level of the detail of the overall classification.

After the initial categorization, two more higher education consortium/non-profit facilitators refined the secondary classification for ability, benevolence, and integrity for each entry, based on [MDS95]. These labels were reviewed and edited again in a group of 5 higher education consortium/non-profit facilitators. Next, the document was presented to 21 advisory members in a live session for their review, followed by a chance for the larger group of 81 advisory members to review and comment. Within the live session, two more guidelines were added ("Credential holder privacy protections" and "Use/storage of private individuals' data") and 3 individuals made formatting or clarifying comments to the document. Following this process, an additional organizational step was completed by higher education consortium/non-profit facilitators.

4 Results

The final governance framework consists of 48 guidelines divided among 7 categories (Tab. 1). Each category is meant to stand alone, referring to a different facet of the registry implementation. The overall list captures information relevant to advisory group stakeholders in order to make an informed decision about how much to trust an LER issuer registry, assuming an LER issuer registry provides information corresponding to each guideline.

Ability, benevolence, and integrity from [MDS95] were classified among each guideline in each category (Tab. 2), with the highest number of guidelines classified under integrity

Category name	**Ability classif.**	**Benevolence classif.**	**Integrity classif.**	**Total guidelines**
Registry Overview and Purpose	4	4	5	5
Governing Parties and Oversight	5	7	5	8
Privacy Policy and Terms of Use	0	5	5	5
Issuer Information and Verification	10	7	13	14
Data Download Format	4	3	3	4
Technical Standards	7	0	4	7
Security and Operations	5	3	5	5
Totals	**35**	**29**	**40**	**48**

Tab. 2: Distribution of governance framework guidelines for LER issuer registries, with ability/benevolence/integrity definitions from [MDS95]. Classifications may overlap within a category, but all classifications will be less than or equal to the total guidelines per category (rightmost column).

(40 of 48 guidelines), followed by ability (35 of 48 guidelines), followed by benevolence (29 of 48 guidelines). The full governance framework with all categories, guidelines, and classifications is provided below (Tab. 3).

Category name	**Description**		
Registry Overview and Purpose	Defines the scope, objectives, and operational principles of the issuer registry.		
Guidelines and perceived trustworthiness factors	**A**	**B**	**I**
Purpose and scope of the registry.	•		•
Subject matter focus of the registry: Limited to issuers, potentially specific types of issuers.	•	•	•
Funding model and sources.		•	•
Business model and fee structure for issuers.	•	•	•
Use/storage of private individuals' data (storage of private data should ideally not occur, or should only occur on a minimum basis).	•	•	•
Category name	**Description**		
Governing Parties and Oversight	Addresses the roles and responsibilities of the governing bodies and mechanisms for ensuring transparent, effective oversight of an issuer registry.		
Guidelines and perceived trustworthiness factors	**A**	**B**	**I**
Organization: Organizational mission.		•	
Governing body or advisory board structure: Composition and roles.	•	•	•
Governing body or advisory board structure: Processes for decision-making and resolving disputes.	•	•	•
Responsibilities of the governing parties: Oversight of operations, policies, procedures, and performance.	•		
Responsibilities of the governing parties: Communication and transparency policies for governance decisions.		•	•
Responsibilities of the governing parties: General transparency and accountability to stakeholders.		•	•
Dispute resolution mechanisms: Protocols for handling conflicts or challenges related to registry governance.	•	•	•
Communication and reporting: Channels for stakeholder engagement. This could include messaging about updates to the registry, requests for updates from the participants, updates about the organization hosting the registry, etc.	•	•	
Category name	**Description**		
Privacy Policy and Terms of Use	Establishes the legal and contractual rights and obligations of the issuer registry with respect to data providers and data users.		

Guidelines and perceived trustworthiness factors	A	B	I
Limitation of liability: Provisions to hold the issuer registry owner harmless from certain claims or misuse of the registry.		•	•
Misuse of data: Policies and consequences for improper use of registry data.		•	•
Privacy and data protection requirements, including compliance with applicable laws.		•	•
Credential holder privacy protections (preventing “calling home”, etc).		•	•
Intellectual property and licensing considerations: Ownership and usage rights for registry content and data.		•	•

Category name	Description
Issuer Information and Verification	Details the processes and requirements for onboarding, verifying, and maintaining issuer data.

Guidelines and perceived trustworthiness factors	A	B	I
Process for initial issuer data submission: Mandatory and voluntary issuer information (Name, description, location, identifiers, credentials offered, contact information, etc).	•		•
Process for verifying issuer identity and legitimacy: Know Your Customer (KYC) process/other identification process.	•		•
Process for verifying issuer identity and legitimacy: Use of third-party services (if any).			•
Process for reviewing and maintaining issuer information: Initial verification steps and estimated processing time.	•		•
Process for reviewing and maintaining issuer information: Frequency and criteria for periodic reviews and estimated processing time.			•
Trust and reputation policies: Defining trust levels or scales based on criteria met (e.g., evidence-based ratings).	•	•	•
Trust and reputation policies: Addressing negative attestations or endorsements of issuers (e.g., “bad reviews”).	•	•	•
Issuer planned/emergency key rotation and/or key compromise policies (including response, audit, and notification policies).	•		•
Retention and archival policies: longevity of issuer data for long-term verification.	•		•
Policies for organizations that close, merge, or leave the registry.	•	•	•
Good-faith policies for payment defaults or non-renewal.		•	•
Policies for issuer removal: Voluntary exit or non-compliance.		•	•
Documentation for issuers.	•	•	•
Issuer support contact.	•	•	

Category name	**Description**		
Data Download Format	Details the processes and policies regarding downloading issuer registry data.		
Guidelines and perceived trustworthiness factors	**A**	**B**	**I**
Machine-readable and human-readable issuer registry retrieval methods.	•		•
Public vs. private registry access.	•	•	•
Documentation for users.	•	•	•
User support contact.	•	•	

Category name	**Description**		
Technical Standards	Establishes the technical foundation for an issuer registry based on specific standards used.		
Guidelines and perceived trustworthiness factors	**A**	**B**	**I**
DID URL/traditional URL/other identifiers used.	•		•
Machine-readable credential formats supported.	•		•
Issuer registry standard/standards used.	•		•
Cryptographic signing mechanisms: how external applications can confirm record integrity.	•		
Verification infrastructure: Mechanisms for validating records against the issuer's cryptographic signature.	•		
Verification infrastructure: Mechanisms for validating records against the issuer registry's cryptographic signature.	•		
Support for tamper-proof storage of records (e.g., append-only logs, blockchain).	•		•

Category name	**Description**		
Security and Operations	Describes cybersecurity and technical operations information regarding how the issuer registry code is operated and maintained.		
Guidelines and perceived trustworthiness factors	**A**	**B**	**I**
Security controls and data protection methods.	•	•	•
Service Level Agreement (SLA) commitments.	•	•	•
Issuer registry planned/emergency key rotation and/or key compromise policies (including response, audit, and notification policies).	•		•
Open-source or closed-source code.	•	•	•
Existence of production and test environments and automated testing.	•		•

Tab. 3: Full governance framework developed for LER issuer registries. A, B, and I columns refer to Ability, Benevolence, and Integrity perceived trustworthiness factors from [MDS95].

5 Discussion

The highest proportions of guidelines classified under the perceived trustworthiness factor of integrity (the largest group, consisting of a total of 40 of 48 guidelines) were found in in the categories *Registry Overview and Purpose*, *Privacy Policy and Terms of Use*, *Issuer Information and Verification*, and *Security and Operations*. These categories pertain to the performance elements of the LER issuer registry. These guidelines are likely to also be evaluated over time, in line with [MDS95], which notes that integrity depends on the "extent to which the party's actions are congruent with his or her words." It is likely that integrity will be perceived as low at the launch of a new LER issuer registry, unless launched by an entity which already enjoys a high perception of integrity.

The highest proportions of guidelines classified under the perceived trustworthiness factor of ability (consisting of a total of 35 of 48 guidelines) were found in in the categories *Data Download Format*, *Technical Standards*, and *Security and Operations*. These results suggest that ability is more likely to be attributed to the portion of governance which can be verifiably checked with reference to external standards and best practices (as opposed to *Governing Parties and Oversight*, *Privacy Policy and Terms of Use*, *Registry Overview and Purpose*, and *Issuer Information and Verification*, which are based on internal characteristics). As such, there is likely some element of transitive ability which is associated with LER issuer registries: if the LER issuer registry incorporates trusted external standards and best practices, its ability may be likely to be evaluated more highly.

The highest proportions of guidelines classified under the perceived trustworthiness factor of benevolence (the least numerous group, consisting of a total of 29 of 48 guidelines) were found in the categories *Privacy Policy and Terms of Use*, *Governing Parties and Oversight*, and *Registry Overview and Purpose*. These results suggest that LER issuer registries which have weak contractual restrictions, a pure profit motive, or an unclear purpose may be less likely to be evaluated as trustworthy under the trust factor of benevolence. Furthermore, more technically-oriented guidelines may have a limited impact on promoting the belief that the LER issuer registry wants to do good to other parties.

Issuer Information and Verification and *Governing Parties and Oversight* were the largest categories in the framework, containing 14 and 8 guidelines respectively. Both categories are particularly of interest because they contain within themselves the smaller governance frameworks used between the LER issuer registry and the governing board and between the LER issuer registry and issuers.

6 Limitations

The methods used in this project were specifically targeted towards the learning and employment record (LER) space. Given that LERs span the employment, educational, and government spaces it would be useful to consider in which ways the governance framework

developed in this area is specific to the LER space and in which ways it can also apply to other issuer registry types. Additionally, the kind of trust registry targeted in this research is an issuer registry, which provides a trusted list of issuers. A verifier registry would focus on providing a trusted list of verifiers and may not share the same governance categories or guidelines.

Two potential components of integrity according to [MDS95] are "the consistency of the party's past actions [and] credible communications about the trustee from other parties." Given that the governance guidelines in this study are meant to be applicable to any LER issuer registry, including new issuer registries, we are unable to evaluate the backward-looking or inter-organizational integrity aspects in this study although they may still be relevant. The trustor's "propensity to trust" [MDS95] will likely also need to be measured to understand why or why not certain LER issuer registries are trusted in practice.

One limitation of the study was the method used to recruit advisory group members and collect data/responses: advisory group members were recruited via snowball sampling, limited to a certain group of online participants who were willing, able, and informed enough to contribute to guideline development for governance for LER issuer registries. Not all of the advisory members provided feedback as a part of the study.

7 Future Work

An area of study particularly relevant to the LER domain is the consideration of issuer registry and issuer longevity, due to the fact that educational credentials usually are designed to last for the lifespan of the credential holder (or longer) [BA24]. Archiving is a consideration in the *Issuer Information and Verification* category of this study's governance framework and is an active consideration in industry in the EU/EEA Trusted Lists [Eu25b]. A relevant future area of research is the extent to which LER issuer registries should cache or archive the information of issuers who join the issuer registry and how this information may support trust.

One notable insight in [SMD07] is that ability, benevolence, and integrity trust factors may develop over different time periods: specifically, benevolence and integrity may be more correlated over short time scales but may become more separable over longer periods. This will likely be a relevant factor in the implementation of LER issuer registries due to the longevity of the issuer registries themselves. Additionally, further research on trust repair, as mentioned in [SMD07], may be useful in the LER context due to potential bad faith actors (such as "diploma mills") who may try to unscrupulously leverage issuer registries.

One area not measured in this study was the difference in trust or importance of different LER issuer registry governance guidelines based on stakeholder role (issuers, verifiers, third parties, other issuer registries) or background (academia, state and national governments, private companies, standards bodies, and nonprofits). These measurements may be useful

for understanding how to promote trust when designing the user experience of issuer registry websites, wallet applications, or other user-facing digital elements. Relevant studies in this area could be completed using prototypes or real versions of LER issuer registries, paired with governance documentation implemented according to the framework developed in this paper.

8 Conclusion

This research work opens up new possibilities in the emerging LER issuer registry literature, contributing a governance framework which was developed among a group of 81 subject matter experts/implementers and 7 higher education consortium/non-profit facilitators. This theoretically-grounded framework can help promote effective and trusted implementations of LER issuer registries, classified under the ability, integrity and benevolence perceived trustworthiness factors from [MDS95]. By providing a governance framework for LER issuer registries, increasingly relevant LER issuer registry technology can be more easily studied and adopted.

9 Funding Disclosure

Funding for this research was provided by a 2024-2025 grant from Walmart.org.

References

[1E24] 1EdTech: Open Badges Specification 3.0, 2024, https://www.imsglobal.org/spec/ob/v3p0, accessed: 01/31/2025.

[AD11] Attewell, P.; Domina, T.: Educational imposters and fake degrees. Research in Social Stratification and Mobility, New Directions in Educational Credentialism 29 (1), pp. 57–69, 2011, https://www.sciencedirect.com/science/article/pii/S0276562410000685, accessed: 01/31/2025.

[Am25] American Association of Motor Vehicle Administrators: Verified Issuer Certificate Authority List, 2025, https://vical.dts.aamva.org/, accessed: 01/31/2025.

[BA24] Bochnia, R.; Anke, J.: Long-Lived Verifiable Credentials: Ensuring Durability Beyond the Issuer's Lifetime. In: Proceedings of the 19th International Conference on Availability, Reliability and Security. ACM, Vienna Austria, pp. 1–9, 2024, https://dl.acm.org/doi/10.1145/3664476.3669933, accessed: 01/31/2025.

[Bo24] Board of Nursing, Maryland Department of Health: Frequently Asked Questions (FAQs) About RN/LPN Licensure and APRN Certification, 2024, https://health.maryland.gov/mbon/Documents/FAQs-Licensure.pdf.

[Di21] Digital ID and Authentication Council of Canada: Pan-Canadian Trust Framework Trust Registries, 2021, https://diacc.ca/trust-framework-components/pctf-trust-registries/, accessed: 01/31/2025.

[EE12] Edelenbos, J.; Eshuis, J.: The Interplay Between Trust and Control in Governance Processes: A Conceptual and Empirical Investigation. Administration & Society 44 (6), Publisher: SAGE Publications Inc, pp. 647–674, 2012, https://doi.org/10.1177/0095399711413876, accessed: 01/29/2025.

[ET16] ETSI: ETSI TS 119 612 V2.2.1 (2016-04) Electronic Signatures and Infrastructures (ESI); Trusted Lists, 2016, https://www.etsi.org/deliver/etsi_ts/119600_119699/119612/02.02.01_60/ts_119612v020201p.pdf, accessed: 01/31/2025.

[Eu25a] European Commission: EU Digital Covid Certificates Trust List, 2025, https://health.ec.europa.eu/ehealth-digital-health-and-care/ehealth-and-covid-19/eu-dcc-trust-list_en, accessed: 01/31/2025.

[Eu25b] European Commission: EU/EEA Trusted List Browser, 2025, https://eidas.ec.europa.eu/efda/trust-services/browse/eidas/tls, accessed: 01/31/2025.

[Ga23] Garber, E. et al.: GAIN in 2023, 2023, https://openid.net/wp-content/uploads/2023/07/GAIN-in-2023.pdf.

[Go22] Goger, A. et al.: Going Digital: How Learning and Employment Records Shape Access To Quality Education and Jobs, tech. rep., Center for Universal Education at Brookings, 2022, https://www.brookings.edu/wp-content/uploads/2022/12/WEB_Report_Going-digital_FINAL122022.pdf.

[He24] Hedberg, R. et al.: OpenID Federation 1.0 - draft 41, 2024, https://openid.net/specs/openid-federation-1_0.html, accessed: 01/31/2025.

[IS21] ISO/IEC: ISO/IEC 18013-5:2021, 2021, https://www.iso.org/standard/69084.html, accessed: 01/31/2025.

[MDS95] Mayer, R. C.; Davis, J. H.; Schoorman, F David: An Integrative Model of Organizational Trust. The Academy of Management Review 20 (3), pp. 709–734, 1995, https://doi.org/10.2307/258792.

[My16] Myhre, S. L. et al.: eRegistries: governance for electronic maternal and child health registries. BMC Pregnancy and Childbirth 16 (1), p. 279, 2016, https://doi.org/10.1186/s12884-016-1063-0, accessed: 01/27/2025.

[PV09] Puranam, P.; Vanneste, B. S.: Trust and Governance: Untangling a Tangled Web. The Academy of Management Review 34 (1), Publisher: Academy of Management, pp. 11–31, 2009, https://www.jstor.org/stable/27759983, accessed: 01/29/2025.

[Sc18] Schreieck, M. et al.: The Challenge of Governing Digital Platform Ecosystems. In (Linnhoff-Popien, C.; Schneider, R.; Zaddach, M., eds.): Digital Marketplaces Unleashed. Springer, Berlin, Heidelberg, pp. 527–538, 2018, https://doi.org/10.1007/978-3-662-49275-8_47, accessed: 01/27/2025.

[SMD07] Schoorman, F. D.; Mayer, R. C.; Davis, J. H.: An Integrative Model of Organizational Trust: Past, Present, and Future. The Academy of Management Review 32 (2), Publisher: Academy of Management, pp. 344–354, 2007, https://www.jstor.org/stable/20159304, accessed: 01/29/2025.

[Sp25] Sporny, M. et al.: Verifiable Credentials Data Model v2.0, 2025, https://www.w3.org/TR/vc-data-model-2.0/, accessed: 01/31/2025.

[Tr24a] Trust Over IP: ToIP Announces the Implementers Draft of the Trust Registry Query Protocol Specification V2.0, 2024, https://www.trustoverip.org/blog/2024/04/03/toip-announces-the-implementers-draft-of-thetrust-registry-protocol-specification-v2-0/, accessed: 01/31/2025.

[Tr24b] Trusted Learner Network: Trusted Learner Network - Governing Charter 1.0, 2024, https://tln.asu.edu/documentation/trusted-learner-network-governing-charter-10, accessed: 01/31/2025.

[Va24] Vaghetti, D.: eduGAIN OpenID Federation Pilot, Rome, IT, 2024, https://www.garr.it/en/docs/7172-sessione-6-0611-04-vaghetti-0611/file.

[Ve23] Velocity Network Foundation: Governance - Overview, 2023, https://www.velocitynetwork.foundation/main2, accessed: 01/31/2025.

"Then I clicked something" – Helping Users to Report Security Incidents with Digital Identity Wallets

Daniela Pöhn [1]

Abstract: The European Digital Identity Wallet (EUDIW) introduced under eIDAS 2.0 will offer individuals a user-centric method for managing and controlling their digital identities, mirroring the principles of the self-sovereign identity model. By giving users control over their data, the digital identity wallets reduce reliance on identity providers and address privacy concerns in federated identity management systems. However, the adoption of these wallets introduces new challenges. Since users self-sovereign control their data, they could also play an important role in security incident response processes. However, current identity wallets lack user-friendly mechanisms for reporting security incidents, and research approaches do not target this issue. Hence, in this paper, we propose a step-wise conceptual design for reporting security incidents within a wallet. We first explore possible security incident response processes before defining suitable categories based on six expert interviews. Related to these findings, we present mockups that include a gamification element that was tested in a small qualitative user study. By giving the user guidance in reporting incidents, they gain more control, and that may increase trust in the wallets.

Keywords: Self-sovereign identity, digital identity, wallet, security incident response process, security incident, user interaction

1 Introduction

The introduction of the eID wallet under eIDAS 2.0 marks an important step in empowering individuals with a user-controlled way to manage their digital identity. Similar to the concept of self-sovereign identities (SSI), the wallet provides a user-centric tool that gives users control over information sharing, allowing them to reveal only the required attributes [Mü18]. Thereby, the data is stored, managed, and presented by the users and not the identity providers. This concept mitigates privacy concerns that exist in federated identity management that heavily relies on providers. However, the growing reliance on wallets also necessitates knowledge about vulnerabilities and attack vectors, as well as having robust mechanisms for detecting, reporting, and addressing security incidents.

However, there are several obstacles, for example (a) the threat landscape is rapidly evolving; (b) the usage of wallets in the context of digital identities is new, meaning also the protocols and implementations are comparably new; and (c) since the user is in control over their data, the user or their device has more knowledge about the reasons for incidents. As social engineering is the most used attack to gain access to networks, this may become challenging [Te23]. Following this, enabling users to play an active role in safeguarding their

[1] Universität der Bundeswehr München, RI CODE, Werner-Heisenberg-Straße 39, 85577 Neubiberg, Germany, daniela.poehn@unibw.de, https://orcid.org/0000-0002-6373-3637

doi:10.18420/OID2025_04

digital identity is critical. Currently, most wallets do not provide the in-app functionality to report information about possible security incidents. As users play an active role in sharing information, a user-centric approach to incident reporting can significantly enhance the overall security and create a proactive feedback loop for continuous improvement. By providing a simplified process to report suspicious activities, unauthorized access, phishing attempts, or other security or privacy incidents, users are empowered to act as the first line of defense to protect their identities. Subsequently, they could be in the loop to see the progress of the incident process.

This paper explores the concept of incorporating such a user-centric incident submission functionality within a digital identity wallet based on possible security incidents and related work. The categories used in incident reporting are the result from six expert interviews, which can be seen in mockups that are created to showcase the idea. One mockup additionally introduces a mascot that explains the reporting form. This mascot was positively evaluated in a qualitative user study with eight participants. The paper thereby contributes the following: (1) step-wise workflow to report security and privacy related incidents within a wallet with (2) categories derived from expert interviews and (3) mockups to visualize the idea including a mascot as a gamification element. Thereby, the user is put into a central position within incident handling and is empowered as first line of defense, which is in line with the idea of self-sovereign control.

The remainder of the paper is as follows: In Section 2, we provide a background and a summary of related work. This is followed by a description of possible security incidents in Section 3. In Section 4, we describe the concept of user-centric incident submission in digital identity wallets. Based on the concept, we discuss implications for future research and development in Section 5. Finally, we conclude the paper with a summary and outlook.

2 Background and Related Work

In this section, we first provide the background on SSI and digital identity wallets, like the European Digital Identity (EUDI) wallet. Then, we summarize the related work on the user perspective on wallets, especially digital identity wallets, and on security incident response processes with a focus on incident reporting. Finally, we summarize our findings and provide reasons for our approach.

2.1 Self-Sovereign Identities and the European Digital Identity Wallet

Allen [Al16] proposed the ten principles that form the design goals of self-sovereign identities (SSI). These principles put the users (*holder*) in the center of the transactions by letting them control their data through a digital *wallet*. The wallet is either on a smartphone, in the cloud (web application), or both (hybrid). Holders receive data (*verifiable credentials*) [W324]

that describe specific attributes (*claims*) from identity providers (*issuers*). Following this, the users can show the data (*proof*) to the service providers (*verifiers*) to gain access to their services. Information about the participating entities is stored in a decentralized manner, typically utilizing distributed hash tables (DLTs), such as a blockchain. Some SSI approaches also utilize smart contracts.

In contrast, the EUDI wallet is an application that allows European citizens to store and manage their credentials, for example, personal identification data (PID) and electronic attestation of attributes (EAAs), that they have received from, for example, a government agency. Using these credentials, users can authenticate themselves online and offline with different eIDAS levels of assurances (LoAs) [Eu] across specific use cases. In contrast to the concept of SSI, the use cases are restricted, and DLTs are currently not described in the architecture and reference framework (ARF) [Eu24]. Thereby, federated identity protocols with newly developed extensions, such as the OpenID for Verifiable Presentations (OID4VP) [Te24], OpenID for Verifiable Credential Issuance (OID4VCI) [LYL24], and OpenID Federation [He24], might be more likely.

2.2 User Perspective on Wallets

Although wallets have been in use for years, the application of SSI and digital identity wallets is relatively new and still evolving. Also, the security threats are not fully explored. Despite the urgent need for security, only a few approach target blockchain-based SSI [Gr23; NGJ21; PGH23; PGH24] and even less focus on EUDIW [Sh25]. Concerning the usage of wallets to manage digital identities, Kostic and Poikela [KP22] analyzed the user perception of identity wallets, whereas Krauß et al. [KKS23] presented various user scenarios. Korir et al. [KPD22] conducted a user study to analyze usability, security and perspectives on user control. The authors learned about misconceptions and users' judgments that are important to regard when implementing and designing identity wallets. Ansaroudi et al. [An23] classify trust and controlled credential sharing. Teuschel et al. [Te23] analyzed the privacy-preserving manner of sharing data and privacy designs in a qualitative user study. The authors noticed that the participants tend to share more data than necessary. In contrast, Doruntina et al. [Mu25] propose the usage of public security user interfaces for spontaneous engagement with IT security. Although public interfaces have a different purpose, they also feature the key objectives of triggering actions and providing control.

2.3 Incident Response

If security incidents happen, they have to be identified, reported, and then mitigated. Governmental agencies, the private sector, and researchers have published several norms and guidelines for incident management, such as International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC) 27035 [In16], National

Institute of Standards and Technology (NIST) Special Publication (SP) 800-61 [Ci12], and European Union Agency for Cybersecurity (ENISA) Good Practice Guide for Incident Management [EN10], that mainly concentrate on one central entity. Tondel et al. [TLJ14] show that users typically do not know when and where to report incidents. In contrast, Busetti and Scanni [BS25] do not consider the user as a reporter of an incident in their theory-based evaluation of incident reporting. Menges and Pernul [MP18] summarize different incident reporting formats, with structured threat information expression (STIX) fulfilling most requirements. However, some aspects such as indicator of compromise might be too technically for users. Winckler et al. [WBB13; WBB16] regarded the user experience when reporting incidents via smartphones. Since their publication, smartphones have changed. Schnoy [Sc14] implemented an assistant for reporting security incident-related information, however, in the form of a generic web service. Redmiles [Re19] conducted a series of qualitative interviews to examine account security incident responses. The author notices that the cultural context influences the threat models and perception of warnings. Wu et al. [WED22] structure security and privacy-related social behavior into four groups. The authors conclude that while social behaviors impact preferences and needs, existing systems are designed with little understanding of these behaviors.

In the context of federated identity management for the research and education interfederation eduGAIN, the security incident response trust framework for federated identity (Sirtfi) guideline [RE22] developed by a research and education federations (REFEDS) [RE24] working group describes the requirements and processes for security incident response (SIR) in this setting. Regarding decentralized SIR, we only found limited literature. Adebayo et al. [Ad19] propose a blockchain network based on Bitcoin for anonymous sharing of incident data and reaction strategies between affected organizations. However, the assumption that Bitcoin creates anonymity is questionable [RH13]. Similarly, [Ar20; CRF18; Ro20] propose a blockchain-based process that can be used to share incident-related information. These approaches have a deficit related to exchange processes and power consumption. Michail [Mi20] introduces a decentralized incident reporting application. BISCUIT, proposed by Putz et al. [PVP22], implies that a subset of organizations has enough stake to invest in local reporting. Furthermore, it is unclear how every user can participate based on knowledge and resources.

2.4 Need for User-Centric Security Incident Reporting

To conclude, while digital identity wallets give more control over their data to users, not all challenges are solved so far. Norms for security incident response processes focus on centralized structures, whereas decentralized incident response is less explored. Typically, the user perspective of security incident reporting is not regarded. By providing the user with the control, the user receives information about actions and, following this, security incidents. However, there are no suitable approaches that regard reporting security incidents via wallet functionalities by holders that are typically not trained to participate in SIR

processes. Therefore, we first provide an overview of possible security incidents before proposing a first concept for user-driven incident reporting with wallets.

3 Possible Security and Privacy Threats

In order to illustrate threats that the users (holders) of identity wallets may recognize, we provide selected examples based on [Sh25] and the supplementary material [Sh24]. These examples show that not all security and privacy incidents are easy to detect and that these comprise all components and entities, such as wallet and third-party services, of the wallet-based identity infrastructure.

Theft of Sensitive Information such as Credentials. Due to the vulnerabilities of insecure communication (V1), insecure data storage (V2), insecure backup or recovery (V3), insecure data exchange protocol (V21), human weaknesses (V9), and injection vulnerabilities (V14), sensitive information might get stolen from the issuer, verifier, holder, or wallet provider data storage or interactions. One specific case is the theft of credentials that may target the issuer or wallet by injection vulnerabilities (V14) or insecure exchange protocols (V21). For example, an attacker may misuse the flaws in the wallet key management system or use exfiltration methods, such as covert channels, malware, or side-channel attacks, to obtain the private key of the holder (AT5.1). If sensitive information are stored in plaintext by the wallet, an attacker can exfiltrate them (AT6.1). Regarding credential theft, an attacker can try to exploit a flaw in the presentation request. The attacker may inject a stolen or leaked credential in the communication with the verifier to access the service (AT2.4). The holder may notice the theft by, for example, receiving specially crafted phishing emails, unusual account activities, notifications from other entities, or identity verification issues.

Unavailability. Due to denial-of-service (DoS) or distributed denial-of-service attacks, services at the different entities may be become unavailable. The vulnerabilities may comprise defense and resource allocation flaws (V5) and lack of secure domain name system (DNS) services or protocol (V6). The attack techniques hence focus on DoS. For example, an attacker or malicious holder could compromise the availability of the involved entities by tampering with the connectivity, for example, by sending too many requests to the issuer or verifier, or rejecting the issuer's credential responses (A10.1). The attacker may attack the infrastructure of the wallet provider by sending too many requests. This may have the consequence of making the setup phase inaccessible and, consequently, make the wallet service unavailable (A10.3). The holder will notice that a service that they want to use is not available. This can be at different stages, for example, the wallet may not load, or the verifier is unavailable.

Malicious app. Malicious apps, like those for cryptocurrencies [Sc24], may be installed using human weaknesses, summarized as social engineering. These apps and scams may have the target to steal wallet data. The attack technique is described as follows: An attacker publishes a malicious app into the app store and uses social engineering techniques to

deceive holders into installing the app on their devices (AT16.5). Variants of this technique may involve, for example, Android packages (APKs) and malicious scripts. The holder may notice a malicious app by unusual account activities, identity verification issues, unusual smartphone behavior, leaked information, or notifications from other entities.

Data Minimization Violation. A data minimization violation may be caused by an excessive data exposure (V29) of holder data like holder private key and holder credentials. This threat may be caused by a verifier that requests extensive or unnecessary attributes. By doing that, the verifier reveals holder's information (AT29.2). As shown by Teuschel et al. [Te23], it is difficult for holders to recognize excessive data exposures. Even with awareness designs, several participants have chosen to send the attributes, since they anyway share lots of information on the Internet every day.

4 Concept of User-Centric Incident Reporting

Based on the previous section, we know possible security and privacy threats that can lead to incidents. The holders may notice these incidents in different ways, either directly (unavailability) or after some time. Since the introduction of wallets should provide the user with self-sovereign or at least more control over their data, the user should also be involved in the reporting of incidents. This also means that the user has to be kept in the loop with appropriate information. To facilitate convenient reporting, we first explain the decentralized incident response processes (see Section 4.1, before we showcase the categories (see Section 4.2. Finally, we introduce the user interface (see Section 4.3). Since users interact with the wallet, this could be a place for this functionality. Additionally, a website could provide another interface for the user in cases where the usage of the wallet is not possible or secure. However, in this case, the inclusion of log files is not possible anymore.

4.1 Decentralized Incident Response Processes

In order to identify and describe decentralized incident response processes, we analyzed related work, possible threats, the architecture, capabilities, and use cases. The detailed SIR processes are described by Ziegler et al. [Zi25].

Global TTP SIR. Similar to federated identity management, user-centric identity management with the usage of wallets may involve centralized entities, commonly named trusted third parties (TTPs). This entity then can act as a contact point for incidents that coordinates the incidents and, hence, should have an incident response team (IRT). This is similar to Sirtfi [RE22]. If no such IRT exists, then a new team should be created. The documentation, analysis, prioritization, mitigation, recovery, and lessons learned can be done similarly to federated identity management. Such a global TTP IRT could also guide the user during the process.

Network SIR. Reflecting the SSI architecture, essential assets are the single nodes and agents, key management, and registry data. Even with a user-centric approach, agents and maybe key management are typically operated. Hence, the responsible parties are the node operators and the governance entity. Depending on the type of model (without or with a TTP), either a local IRT or a global TTP IRT coordinates the incident. If the incident requires more resources than the local IRT has, then the global TTP IRT may provide help.

Issuer and Verifier SIR. The goal of this SIR is to ensure that incidents affecting specific entities and the trust relationship between them are effectively mitigated. Assets include the issuer and verifier infrastructures with authenticating data, processes of issuing data, and validation processes. Incidents affecting verifiers potentially lead to privacy concerns among holders, whereas a compromise of issuers potentially affects identity management services across the network. Similar to the network SIR, the responsibility for this SIR could be in the hands of the entity's IRT or the global TTP IRT. A possible SIR for federated identity management that can be applied for user-centric identity management is described by the Germany national research and education network DFN [DF25].

Holder SIR. The most significant difference from current identity management is the involvement of the user. Due to the holder's self-sovereignty, they are ultimately responsible for their data. In comparison to the SIRs described above, the holder has limited capabilities. No substantial knowledge for incident handling can be expected from the average holder. Since traditional SIR processes do not consider single users to be in control of their data, the holders are dependent on the wallet developers. With wallets that are currently on the market, the user has two main types of control: revocation and recovery. The first ensures that if the wallet gets stolen, it cannot be used to access existing trust relationships. The second limits the impact of inaccessibility. To provide further possibilities, the wallets with the graphical user interface have to implement such SIR functionalities.

4.2 Reporting Incidents

The categories and incident descriptions were derived by a combination of (a) brainstorming categories based on those derived by Sharif et al. [Sh24; Sh25] and (b) qualitative unstructured expert interviews with six individuals who work or worked in first-level support at different organizations. Five are knowledgeable about identity and access management, with one knowing SSI. The duration varied between 15 and 45 minutes. The interviewees did not receive any compensation. Since the procedure was in line with the guidelines of the university, the ethical board was not involved.

Although Sharif et al. [Sh24; Sh25] describe attack techniques AT1 to AT34, most terms are not known to users. One exception might be "phishing and social engineering", since phishing is targeted in several awareness campaigns. However, even "social engineering" is not known to everyone. Menges and Pernul [MP18] conclude that STIX is most suitable for reporting incidents, however, several terms are not known or too technically for users.

When reporting incidents in practice, one way is to select the concerned service and have a description field and the possibility to upload any further data. The service itself does not fit the purpose. Since we want to encourage the user to report suspicious behavior, while descriptions may not provide the necessary information, and the user should be able to report the incident on a smartphone, we use more categories, similar to the templates applied by Schnoy [Sc14].

First of all, categories of problems are derived. Several may correspond to incidents or problems each entity: "phishing or scam", "service not available", "something is not working correctly", "data breach", "suspicious activity", and "I did something that I did not want". As the interviewees emphasized, "something is not working correctly" is the most often used problem. This might also be the case with wallets. In the following, the user should specify what functionality or service is not working, add a description, and maybe also screenshots or other proofs. The interviewees emphasized that most problems are due to lack of knowledge or something the user did; only few problems are issues caused by the services themselves. Additionally, malware and phishing seems to be a problem on computers. These observations are handled by "phishing or scam","suspicious activity", and "I did something that I did not want".

Concerning wallets, two specific problems are "lost/compromised device" and "cannot log in". As the user might not be able to report these incidents within the wallet, there should also be a web interface the user can access via their browser. However, in this case, the user has to provide information about their account. Both problems have been named concerning notebooks by the interviewees and, therefore, seem to be likely to occur. Finally, one interviewee mentioned that users notice and report changes in the user interface. Those problems can be categorized as "something looks differently". Although this is most likely a user-issue, it could also indicate a malicious app.

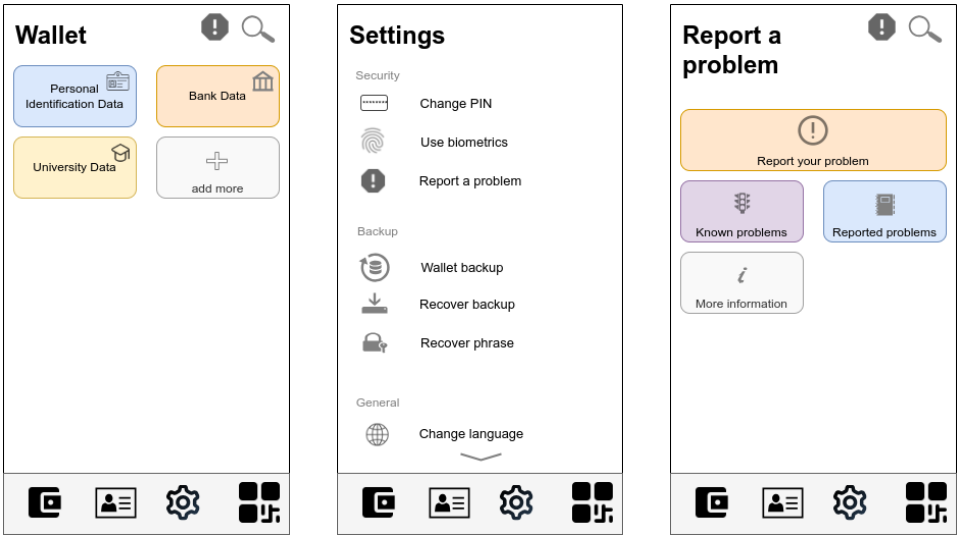
Related to the verifier, we found "service asks for much information", "clicked OK to an unwanted request", and "service is not as expected". The latter can be everything from expectations not met to a malicious verifier that offers its service to gain access to data. These problems also show that directing them to the verifier without any intermediate might be problematic. Following this, a global TTP IRT or a wallet IRT are preferred. The categories should be reevaluated when digital identity wallets are more mature. Regarding most categories, they may be related to interactions with other entities that should appear in the activity log.

After using the described categories, the concerned entity can be selected, i. e.., issuer/identity provider, verifier/service provider, and wallet. The differentiation of wallet provider, wallet, and theoretically also holder and smartphone used by Sharif et al. [Sh24; Sh25] might be too specific; therefore, the simplification of "wallet" is proposed. This category, together with the further information, helps to identify the entity that receives the reported incident. Since the technical details of user-centric identity management with wallets are hidden from the user, the correct categorization has to be evaluated by an IRT. With the decentralized character

of digital identity wallets and SSI, a global IRT for coordinating the cross-organizational SIR processes might be introduced. This IRT can consist of members of the federation or consortium the user-centric identity management is used in. Hence, either this global (TTP) IRT, the wallet IRT, or the IRT of the selected entity have to review the incident.

Even with the categories and the possibility to describe the incident, not all information might be sent to the IRT. Information that is often missing are, for example, operating system, client name, client version, and since when the problem exists. Especially for the first information, it might be easier if they are collected by the wallet, given that it has the permissions. The latter might be provided by the holder. Similarly, the exact issuer and verifier might be given by the holder. Depending on the incident, log files, telemetry data, and screenshots might be helpful. This shows that the wallet has data that might be helpful. With the user being in self-sovereign control, the user should allow the wallet to collect and send that data. Furthermore, the holder should have the possibility to upload screenshots, emails, and other files. As a limitation, the type of information that should be provided has to be improved iteratively based on practical experiences.

4.3 User Interface



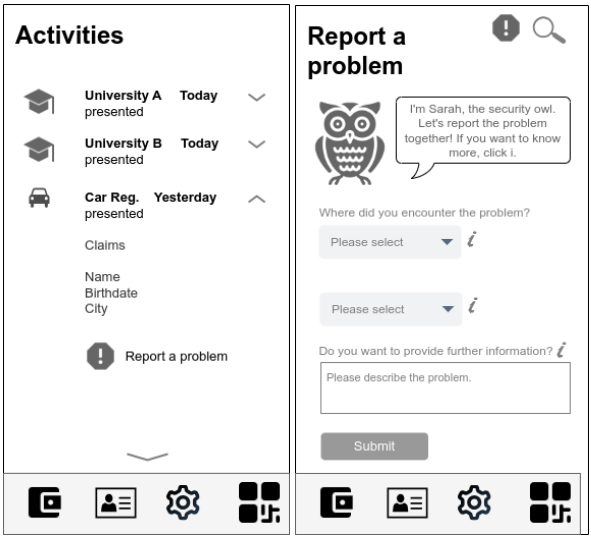
(a) Wallet home screen. (b) Settings menu. (c) Reporting a problem.

Fig. 1: Reporting a problem in a wallet 1/2.

The user interface presented in Figure 1 is a mockup, illustrating the different workflows. Based on the previous section, the user should have the possibility to report problems from their wallet, but also on a dedicated web page in case the user cannot access the wallet anymore. To provide an easy way for reporting, a reporting button should be placed on the

home screen of the wallet (see Figure 1a). In addition, an item in the settings menu (see Figure 1b) provides another possibility to access the reporting functionality.

Both paths lead the user to the overview page for problem reporting. Besides the actual reporting of problems or incidents, this page provides more functionality. Information about the status of the reported problem can be displayed and the IRT can request further information or help the user by, for example, giving them a guideline on how to proceed. Additionally, known problems and further information can be displayed.



(a) Reporting a problem with an activity.

(b) Reporting a problem with the mascot Sarah, the security owl.

Fig. 2: Reporting a problem in a wallet 2/2.

As several incidents that can occur might be based on interactions of the user with other entities and that these interactions should appear in an activity log, the activity log is the third way to report an incident. An example is given in Figure 2a. The information from the activity should be included in the incident report for convenience.

On the next page, when clicking the reporting button, the holder can report the problem they encountered, as shown in Figure 2b. The categories and required information identified in Section 4.2 are requested. They comprise the impacted service, the problem or incident, a description, device and app information, and maybe screenshots or log files. The mascot Sarah, the security owl, leads the users through the reporting page and gives hints on the questions. An owl was selected as a gamification element, since owls are associated with knowledge (see “owl of Athena” and “owl of Minerva”). The mascot can also lead the holder through the usage of wallet, for example, if they have the issue to send credentials to

a specific verifier. If the tutorial is not helping, an incident with the information available could be generated.

The usage of an owl as a mascot was positively evaluated in a small qualitative user study (n=8). Asked if the participants would prefer email or web application, they argued for the application due to the offered incentives and the guided approach. Comparing reports made per mail to a gamified version, the latter included more detailed information due to the input fields. As a limitation, real incidents were not tested due to the current status of the EUDIW.

Although ITILv4 defines a problem as "a cause or potential cause of one or more incidents", while an incident is a single unplanned event that causes a service disruption, the word "problem" is commonly the difference between the current and the desired state. Consequently, more problems than security or privacy incidents might be reported. This can serve as input for further improvements of the wallet, the ecosystem, and also the reporting functionality.

5 Implications for Future Research and Development

Based on our approach and the related work, we want to outline possible future research and development implications and topics.

Human-Centric Security Research. With the user-centric incident reporting functionality, further studies could try to understand the user behavior and the psychological factors influencing their willingness to report incidents, among others. Users should feel supported rather than blamed for potential mistakes. Also, the perceived risks might be interesting. The risks may differ between cultural groups. The results may be used as input for usability and user experience.

Usability and User Experience. Although we proposed a simplified reporting interface, and we evaluated the mascot, further qualitative and quantitative user studies are needed to test the interface and possible gamification elements. Another aspect could be the simplified terminology and the explanations given by the mascot. Additionally, the approach should be accessible to all users, regardless of their technical expertise, age, or physical abilities. This aspect could be regarded in future user studies.

Incident Categorization. Further work is required to refine and standardize the categories proposed in this paper. The categories should contain all possible incidents while using terms that are easy to understand for users. Standardized categories ensure that incidents are classified similarly across wallets and infrastructures. With an evolving threat landscape, the categories might need adaptations in the future.

Privacy Measures for Incident Reporting. Digital identities are personally identifiable information. Incident reports may contain not only digital identities, but also other personal, behavioral, or security-related data. Furthermore, screenshots and log files or telemetry data may also contain sensitive information. Data protection applies to the incident reports.

Further research could examine whether pseudonymization might be possible and if that could be automated.

Proactive Incident Management. By providing the user with a reporting functionality, the user can report an incident and should receive information about the progress of solving this incident. This requires real-time feedback with informative texts and guidelines that can help the user take the required actions. The users, hence, can be seen as human sensors. Incident prediction might be another aspect to look at. Similarly, this feedback loop can be used to educate the user, for example, on how to recognize and avoid common attack vectors.

Awareness Designs and Warnings. Based on Teuschel et al. [Te23], different actionable warnings, awareness designs, and context-aware security information could be developed and evaluated in user studies, for example, showing and explaining risks in simple language and with visual help or providing steps for mitigation.

Scalability. Given the amount of possible users with a self-sovereign control over their identities and, with this approach, the possibility to report problems, this could result in scalability issues. For example, by monitoring services, a technical impact on the availability can be measured better than with users. However, users may report it as well. Following this, these reports have to be reviewed in the designated SIR processes.

6 Conclusion

With the introduction of the EUDI wallet by eIDAS 2.0, users are empowered to control and manage their identity data. Thereby, this concept mitigates privacy concerns in current federated identity management. However, with the shift of power, the reliance on wallets is growing, which comes also with risks. To enable users to play an active role in safeguarding their digital identities, we proposed user-centric security and privacy incident reporting as a wallet functionality. We first outlined incident response processes based on the architecture and literature. Then, we conducted expert interviews and used brainstorming based on the described incidents to gather suitable categories. The incident reporting functionality was visualized in mockups that included Sarah, the security owl that provides hints and motivation. Finally, we discussed implications for future research and development. This paper highlights the potential of such functionality to also reinforce trust and resilience in the adoption of eID wallets. In future work, we want to extend the user studies and include real incidents to improve the approach.

Acknowledgement

We would like to thank Klara Rößler and Lara Rühl for their input regarding the holder SIR and the gamification element mascot.

References

[Ad19] Adebayo, A. et al.: Blockchain-enabled Information Sharing Framework for Cybersecurity. In: Blockchain for Distributed Systems Security. Wiley, New York, NY, USA, pp. 143–158, 2019.

[Al16] Allen, C.: The Path to Self-Sovereign Identity, Accessed May 5, 2025., 2016, http://www.lifewithalacrity.com/%202016/04/the-path-to-self-soverereign-identity.html.

[An23] Ansaroudi, Z. E. et al.: Control is Nothing Without Trust a First Look into Digital Identity Wallet Trends. In: Data and Applications Security and Privacy XXXVII (DBSec 2023). Lecture Notes in Computer Science, vol 13942. Springer, pp. 113–132, 2023.

[Ar20] Ardagna, C. A. et al.: A Blockchain-based Trustworthy Certification Process for Composite Services. In: Proceedings of the 17th IEEE International Conference on Services Computing (SCC), Beijing, China, 07-11 Nov 2020. IEEE, Piscataway, NJ, USA, pp. 422–429, 2020.

[BS25] Busetti, S.; Scanni, F. M.: Evaluating incident reporting in cybersecurity. From threat detection to policy learning. Government Information Quarterly 42 (1), p. 102000, 2025, https://www.sciencedirect.com/science/article/pii/S0740624X24000923.

[Ci12] Cichonski, P. et al.: NIST Special Publication 800-61 Revision 2 Computer Security Incident Handling Guide, Special Publication, Accessed May 5, 2025., National Institute of Standards and Technology, 2012.

[CRF18] Carminati, B.; Rondanini, C.; Ferrari, E.: Confidential Business Process Execution on Blockchain. In: Proceedings of the 25th IEEE International Conference on Web Services (ICWS), San Francisco, CA, USA, 02-07 July 2018. IEEE, Piscataway, NJ, USA, pp. 58–65, 2018.

[DF25] DFN: Security Incident Response in der DFN-AAI, Accessed May 5, 2025., 2025, https://doku.tid.dfn.de/de:aai:incidentresponse.

[EN10] ENISA: Good practice guide for incident management, Accessed May 5, 2025., 2010, www.enisa.europa.eu/publications/good-practice-guide-for-incident-management.

[Eu] European Commission: eID Documentation- eIDAS Levels of Assurance (LoA), Accessed May 5, 2025., https://ec.europa.eu/digital-building-blocks/sites/display/DIGITAL/eIDAS+Levels+of%20+Assurance.

[Eu24] European Commission: The European Digital Identity Wallet, Accessed May 5, 2025., 2024, https://eu-digital-identity-wallet.github.io/eudi-doc-architecture-and-reference-framework/1.4.0/.

[Gr23] Grüner, A. et al.: Analyzing and comparing the security of self-sovereign identity management systems through threat modeling. Int. J. Inf. Secur. 22 (5), pp. 1231–1248, 2023.

[He24] Hedberg, R. et al.: OpenID Federation 1.0 - draft 41, Accessed May 5, 2025., 2024, https://openid.net/specs/openid-federation-1%5C_0.html.

[In16] International Organization for Standardization: Information technology — Security techniques — Information security incident management — Part 1: Principles of incident management, Standard, Geneva, Switzerland, 2016.

[KKS23] Krauß, A.-M.; Kostic, S.; Sellung, R. A.: A more User-Friendly Digital Wallet? User Scenarios of a Future Wallet. In: Open Identity Summit 2023. Gesellschaft für Informatik e.V., Bonn, pp. 73–84, 2023.

[KP22] Kostic, S.; Poikela, M.: Do Users Want To Use Digital Identities? A Study Of A Concept Of An Identity Wallet, Accessed May 5, 2025., 2022, https://www.usenix.org/system/files/soups2022-poster95%5C_kostic%5C_abstract%5C_final.pdf.

[KPD22] Korir, M.; Parkin, S.; Dunphy, P.: An Empirical Study of a Decentralized Identity Wallet: Usability, Security, and Perspectives on User Control. In: Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022). USENIX Association, Boston, MA, pp. 195–211, 2022, https://www.usenix.org/conference/soups2022/presentation/korir.

[LYL24] Lodderstedt, T.; Yasuda, K.; Looker, T.: OpenID for Verifiable Credential Issuance - draft 15, Accessed May 5, 2025., 2024, https://openid.net/specs/openid-4-verifiable-credential-issuance-1%5C_0.html.

[Mi20] Michail, A.: Tackling the Challenges of Information Security Incident Reporting: A Decentralized Approach, PhD thesis, University of East London, 2020.

[MP18] Menges, F.; Pernul, G.: A comparative analysis of incident reporting formats. Computers & Security 73, pp. 87–101, 2018, https://www.sciencedirect.com/science/article/pii/S0167404817302250.

[Mü18] Mühle, A. et al.: A survey on essential components of a self-sovereign identity. Computer Science Review 30, pp. 80–86, 2018.

[Mu25] Murtezaj, D. et al.: Public Security User Interfaces: Supporting Spontaneous Engagement with IT Security. In: Proceedings of the New Security Paradigms Workshop. NSPW '24, Association for Computing Machinery, New York, NY, USA, pp. 56–70, 2025.

[NGJ21] Naik, N.; Grace, P.; Jenkins, P.: An Attack Tree Based Risk Analysis Method for Investigating Attacks and Facilitating Their Mitigations in Self-Sovereign Identity. In: Proceedings of the 7th IEEE Symposium Series on Computational Intelligence (SSCI). Pp. 1–8, 2021.

[PGH23] Pöhn, D.; Grabatin, M.; Hommel, W.: Modeling the Threats to Self-Sovereign Identities. In (Roßnagel, H.; Schunck, C. H.; Günther, J., eds.): Proceedings of the 10th Open Identity Summit (OID), Heilbronn, Germany, 15-16 June 2023. Gesellschaft für Informatik e.V., Bonn, Germany, pp. 85–96, 2023.

[PGH24] Pöhn, D.; Grabatin, M.; Hommel, W.: Analyzing the Threats to Blockchain-Based Self-Sovereign Identities by Conducting a Literature Survey. Applied Sciences 14 (1), 2024.

[PVP22] Putz, B.; Vielberth, M.; Pernul, G.: BISCUIT - Blockchain Security Incident Reporting based on Human Observations. In: Proceedings of the 17th International Conference on Availability, Reliability and Security. ARES '22, Association for Computing Machinery, Vienna, Austria, 2022, https://doi.org/10.1145/3538969.3538984.

[Re19] Redmiles, E. M.: "Should I Worry?" A Cross-Cultural Examination of Account Security Incident Response. In: 2019 IEEE Symposium on Security and Privacy (SP). Pp. 920–934, 2019.

[RE22] REFEDS: A Security Incident Response Trust Framework for Federated Identity (Sirtfi) Version 2, Framework, Accessed May 5, 2025., Cambridge, UK: REFEDS, 2022.

[RE24] REFEDS: Refeds - About, Accessed May 5, 2025., 2024, www.refeds.org.

[RH13] Reid, F.; Harrigan, M.: An Analysis of Anonymity in the Bitcoin System. In: Security and privacy in social networks. Springer, New York, NY, USA, pp. 197–223, 2013.

[Ro20] Rondanini, C. et al.: Blockchain-based controlled information sharing in inter-organizational workflows. In: Proceedings of the 17th IEEE International Conference on Services Computing (SCC), Beijing, China, 07-11 Nov 2020. IEEE, Piscataway, NJ, USA, pp. 378–385, 2020.

[Sc14] Schnoy, B.: Intelligent Assistent Lösung für die teilautomatisierte Erfassung von Incidents und Service Requests am Leibniz-Rechenzentrum, Bachelor Thesis, Ludwig-Maximilians-Universität München, 2014, https://www.nm.ifi.lmu.de/pub/Fopras/schn14/.

[Sc24] Scharfman, J.: Wallet Drainers, Crypto Stealers and Cryptojacking. In: The Cryptocurrency and Digital Asset Fraud Casebook, Volume II: DeFi, NFTs, DAOs, Meme Coins, and Other Digital Asset Hacks. Springer Nature Switzerland, Cham, pp. 271–306, 2024, https://doi.org/10.1007/978-3-031-60836-0%5C_10.

[Sh24] Sharif, A. et al.: Protecting Digital Identity Wallet: A Threat Model in the Age of eIDAS 2.0, Accessed May 5, 2025., 2024, https://sites.google.com/view/eu-digital-identity-wallet/home.

[Sh25] Sharif, A. et al.: Protecting Digital Identity Wallet: A Threat Model in the Age of eIDAS 2.0, Proceedings of the International Conference on Risks and Security of Internet and Systems (CRiSIS) 2024, LNCS 15456, 2025.

[Te23] Teuschel, M. et al.: 'Don't Annoy Me With Privacy Decisions!' – Designing Privacy-Preserving User Interfaces for SSI Wallets on Smartphones. IEEE Access 11, pp. 131814–131835, 2023.

[Te24] Terbu, O. et al.: OpenID for Verifiable Presentations - draft 23, Accessed May 5, 2025., 2024, https://openid.net/specs/openid-4-verifiable-presentations-1%5C_0.html.

[TLJ14] Tøndel, I. A.; Line, M. B.; Jaatun, M. G.: Information security incident management: Current practice as reported in the literature. Computers & Security 45, pp. 42–57, 2014, https://www.sciencedirect.com/science/article/pii/S0167404814000819.

[W324] W3C: Verifiable Credentials Data Model v2.0, Accessed May 5, 2025., 2024, https://www.w3.org/TR/vc-data-model-2.0/.

[WBB13] Winckler, M.; Bach, C.; Bernhaupt, R.: Identifying User Experience Dimensions for Mobile Incident Reporting in Urban Contexts. IEEE Transactions on Professional Communication 56 (2), pp. 97–119, 2013.

[WBB16] Winckler, M.; Bernhaupt, R.; Bach, C.: Identification of UX dimensions for incident reporting systems with mobile applications in urban contexts: a longitudinal study. Cognition, Technology & Work 18 (4), pp. 673–694, 2016, https://doi.org/10.1007/s10111-016-0383-1.

[WED22] Wu, Y.; Edwards, W. K.; Das, S.: SoK: Social Cybersecurity. In: 2022 IEEE Symposium on Security and Privacy (SP). Pp. 1863–1879, 2022.

[Zi25] Ziegler, L. et al.: Designing a security incident response process for self-sovereign identities. EURASIP Journal on Information Security 2025 (1), p. 12, 2025, https://doi.org/10.1186/s13635-025-00195-6.

Economic Challenges of Cybersecurity Innovations Pursuing an Ecosystems Approach

Michael Kubach [1]

Abstract: The adoption of innovative cybersecurity solutions is essential for protecting digital infrastructures. However, it is often confronted with significant challenges, especially when these innovations aim to establish collaborative ecosystems. Yet, to be effective, many security innovations like identity and trust infrastructures must build ecosystems of diverse participants. In this context, economic aspects are frequently overlooked in favor of an excessive focus on technical excellence. To address this imbalance, we examine the economic obstacles inherent in the diffusion of security innovations with an ecosystem approach. Utilizing theoretical frameworks, we analyze general challenges including organizational acceptance and information asymmetries as well as specific issues associated with ecosystem approaches, such as network effects and governance structures. Building on insights from a research project, we discuss strategies to address these challenges. Our findings underline the importance of integrating economic perspectives into the development and adoption of sustainable cybersecurity infrastructures.

Keywords: Adoption, Cybersecurity, Economics, Ecosystems, Innovation, IT Security

1 Introduction

The escalating complexity and interconnectivity of digital infrastructures have heightened the imperative for innovative cybersecurity solutions. As cyber threats evolve in sophistication and frequency, organizations are compelled to adopt advanced technologies to safeguard their assets. However, the diffusion of these innovations is often hindered by significant adoption challenges. A predominant issue lies in the excessive emphasis on technical excellence, which frequently overshadows critical economic considerations. This oversight can lead to underestimation of factors such as perceived value, expected return on investment and integration into existing organizational processes, ultimately impeding the widespread adoption of effective security measures.

Moreover, many cybersecurity innovations necessitate the formation of collaborative ecosystems to function optimally. This is particularly the case for identity and trust management infrastructures, such as the European Union's eIDAS 2.0 framework [eIDAS24] with the EU Digital Identity (EUDI) Wallet [Eu24] and self-sovereign identity (SSI) solutions. These systems aim to establish interoperable, secure, and user-centric digital identities, relying on a diverse network of participants of relying parties, users, and

[1] Fraunhofer IAO, Nobelstr. 12, 70569 Stuttgart, michael.kubach@iao.fraunhofer.de,
https://orcid.org/0000-0001-5105-1113

doi:10.18420/OID2025_05

service providers to create a functioning ecosystem. While holding significant promises for enhancing digital trust and security, their adoption faces additional complexities. The diverse network of participants with individual interests introduces additional complexities to just setting up a functioning technical infrastructure, including coordination among multiple stakeholders, governance structures, and management of network effects. These factors contribute to information asymmetries and organizational resistance, further complicating the adoption process.

Understanding these multifaceted challenges requires a comprehensive analysis grounded in established theoretical frameworks. By employing theoretical frameworks, we can systematically examine the barriers impacting the adoption of cybersecurity innovations. They allow for the exploration of general challenges, such as user and organizational acceptance, as well as specific issues inherent to ecosystem approaches.

This paper aims to shed light on the often-overlooked economic obstacles in the diffusion of cybersecurity innovations. Through an analysis of both general and ecosystem-specific challenges, we seek to underscore the critical role economic factors play in the adoption process. Drawing insights from the Mantra project – a research initiative focused on enhancing cybersecurity through collaborative efforts [Ma25] – we discuss practical strategies to address these challenges. In doing so, we contribute to the discourse on how to effectively promote the adoption of cybersecurity innovations that rely on collaborative ecosystems – which we regard as highly relevant for the development of identity and trust infrastructures. Recognizing that technical prowess alone is insufficient, we advocate for a balanced focus that equally prioritizes economic viability and organizational readiness. This approach is essential for fostering resilient digital ecosystems capable of withstanding the dynamic landscape of cyber threats.

2 Adoption Challenges of Cybersecurity Innovations

We structure our analysis of adoption challenges of cybersecurity innovations into two sections. First, we consider general challenges for cybersecurity innovations that are due to their innovative characteristics and their localization in the cybersecurity domain with respective particular challenges. Second, we focus more closely on specific challenges of cybersecurity innovations that pursue an ecosystem approach.

Our work builds and extends previous analysis by focusing specifically on cybersecurity innovations with an ecosystem focus. [HA20] develop a theoretical model for the adoption process of cybersecurity innovations in organizations. By integrating four theoretical frameworks – the Diffusion of Innovation (DOI) theory, the Technology Acceptance Model (TAM), the Theory of Planned Behaviour (TPB), and the Technology-Organization-Environment (TOE) framework – the authors propose a comprehensive model that encompasses both organizational adoption processes and user acceptance of cybersecurity innovations. The model outlines an adoption process (pre-adoption, adoption-decision, and post-adoption) and provides practical implications for effectively

implementing security measures within organizations. Hence, the focus slightly differs to our paper that focuses more on the general market adoption of a cybersecurity innovation instead of on the implementation in a specific organization.

[VY24] investigate the factors influencing the adoption of cybersecurity innovations in organizations in a literature analysis. The authors find that traditional technology adoption models like the TAM and DOI, are insufficient for explaining cybersecurity adoption due to the unique nature of cybersecurity technologies, which focus on preventing future incidents with limited direct positive gain. Therefore, they point to models with roots in preventive health like TPB and additionally institutional models that integrate other, often external, forms of pressure like regulation. A finding that is pretty much in line with the approach of [HA20].

In the following we will first derive adoption challenges from technology adoption models and models with roots in preventive health with a perspective on the general market adoption. Second, we will focus more on the specific challenges that are related to ecosystem approaches – we were not able to find any papers specifically focusing on these.

2.1 General Adoption Challenges for Cybersecurity Innovations

[Ro03] *Diffusion of Innovations (DoI)* theory provides a framework to understand how new technologies spread through social systems. It is a classical technology adoption model. The theory identifies five key attributes that influence the adoption rate of an innovation: relative advantage, compatibility, complexity, trialability, and observability.

Relative advantage refers to the degree to which an innovation is perceived as better than the idea it supersedes. In the context of cybersecurity, organizations must recognize the tangible benefits that new solutions offer over existing systems. However, this is challenging in the context of cybersecurity as the advantages may not be immediately evident to all stakeholders. Quantifying a return on investment (ROI), such as cost savings from reduced fraud, improved operational efficiencies, or increased customer trust can be challenging, as the primary value is in preventing threats that have not/will not materialize. Approaches like the Return On Security Investment (ROSI) aim to address this [SAS06].

Compatibility is the extent to which an innovation fits with the existing values, past experiences, and needs of potential adopters. For cybersecurity innovations compatibility issues arise when the new systems require changes to organizational processes, cultural shifts, or updates to legacy technologies. For example, there may be misalignments between the user-centric principles of SSI and traditional centralized organizational control over identity management with respective measures for restoring user accounts. These incompatibilities can create resistance unless the economic benefits of resolving them are clear and compelling.

Complexity refers to how difficult the innovation is to understand and use. Cybersecurity solutions often involve sophisticated technologies like certain cryptographic protocols.

Complex implementation can deter organizations lacking specialized technical expertise. Moreover, the training required for employees and users to effectively engage with new systems adds to the perceived burden. Economic factors amplify this challenge: costs associated with hiring skilled personnel, investing in employee training, and potential disruptions during the transition period.

Trialability is the degree to which an innovation can be experimented with on a limited basis. Cybersecurity innovations often have low trialability due to their scope and the risks associated with partial implementation. Testing new identity infrastructures may require significant upfront investment and cannot be easily isolated from core operational systems. A lack of trialability poses economic risks.

Observability relates to how visible the results of an innovation are to others. In cybersecurity, the protective nature of solutions means their value is often realized through the mitigation of threats that do not materialize. One could go as far as to talk about a paradox: the better the security measure function, the less apparent their benefits become. This intangible benefit makes it challenging to demonstrate immediate advantages or to showcase successful use cases. For identity management systems the benefits may be indirect or long-term, such as enhanced customer trust or compliance with future regulations. The difficulty in observing and quantifying these benefits complicates the economic justification for adoption.

The Diffusion of Innovations theory by [Ro03] highlights critical factors that influence the adoption of cybersecurity innovations. High complexity, low trialability, and limited observability create economic uncertainties that deter organizations from adopting cybersecurity innovations. Addressing these challenges requires a deliberate focus on the economic dimensions of each adoption factor. By articulating clear economic value propositions, reducing implementation costs, and making benefits more observable, the appeal of cybersecurity innovations might be enhanced.

The *Technology-Organization-Environment (TOE) framework*, as applied for example by [Wa20], explicitly introduces the organizational perspective. It provides a structured lens to analyze adoption barriers by categorizing them into three contexts: technological, organizational, and environmental. Applying this framework to cybersecurity innovations, illuminates the multifaceted challenges hindering their widespread adoption. Regarding the technological context, complexity, and compatibility with existing systems often deter adoption. This context overlaps strongly with the complexity attribute of the DoI as discussed, which is why it is not covered further here.

Within the organizational context, factors such as leadership support, organizational readiness, and culture play critical roles in the adoption of cybersecurity innovations. Resistance to change is a common barrier, as adopting new identity infrastructures may require significant alterations to established workflows and employee practices. Moreover, insufficient commitment from top management can impede necessary investments in training and infrastructure upgrades. Organizations may prioritize immediate operational needs over long-term strategic investments in cybersecurity. In a

case study, [Ha22] illustrates the importance of organizational context factors. She highlights the importance of involving top management, communicating in a non-technical language, cultural factors, and making end users feel a personal connection to cybersecurity.

Environmental factors, e.g., regulatory requirements, competitive dynamics, and industry standards, significantly influence adoption decisions – aspects that are also highlighted in institutional models [AJA09]. External pressures like the implementation of the General Data Protection Regulation (GDPR) have driven organizations to adopt advanced data protection measures. However, navigating regulatory landscapes can be complex and resource-intensive. Organizations may face uncertainties regarding legal interpretations, compliance costs, and the potential penalties for non-compliance. Competitive dynamics also play a role. Companies may hesitate to adopt innovations like SSI if industry peers are not doing the same, fearing a lack of interoperability or wasted investment.

In summary, additionally to the DoI, the TOE framework highlights particularly at the one hand the importance of factors coming from inside the organization (culture and different stakeholders: top management, employees) but also from outside the organization (regulation and industry peers).

Innovation and technology adoption models have been criticized for failing to integrate the concept of threat that is certainly relevant in the context of cybersecurity. Therefore, frameworks with roots in *preventive health theory* can be of assistance as preventive health also involves proactive measures to avert undesirable outcomes. An example is the *Fear Appeals Model (FAM)* as applied by [JW10]. It can be used to explain that individuals' perceptions of threats (e.g., the severity of cyber threats and their susceptibility), along with perceived benefits and barriers to take action, significantly influence their motivation to adopt preventive measures. Cues to action, such as security warnings or incidents, can further prompt adoption by heightening awareness of the threat.

For instance, the decision to adopt a cybersecurity innovation is contingent upon recognizing a threat and believing in their efficacy as preventive measure. In this context, social influence and fear-inducing communications can have varied effects – some may be motivated to adopt cybersecurity measures, while others may avoid the issue altogether. If individuals perceive low ability to prevent a threat, they may resort to emotion-focused coping, such as denial or resignation. Applying these models to cybersecurity underscores the importance of not only communicating risks but also enhancing individuals' confidence in their ability to effectively implement security measures.

The adoption of cybersecurity innovations is further complicated by the presence of *information asymmetries in the market*. Buyers of cybersecurity solutions often lack the expertise or means to fully evaluate the quality and efficacy of products before purchase. This phenomenon parallels the issue described in [Ak70] seminal work on the market for used cars, where quality uncertainty and asymmetric information between buyers and sellers can lead to market failure. In such a market, low-quality products ("lemons") can overshadow high-quality offerings due to a lack of transparency, ultimately diminishing

overall market quality. One illustrative example of this phenomenon could be seen in the market for virus scanners aimed at end users. Historically, there has been an absence of a functioning market for paid antivirus solutions. Apparently, there is a widespread unwillingness among end users to pay for antivirus software. Users have struggled to recognize the limitations of free solutions compared to premium products, leading to underinvestment in more effective, but costly software. The result was a market dominated by free solutions and those bundled with operating systems.

In cybersecurity markets, information asymmetry arises because vendors possess more knowledge about the strengths and weaknesses of their complex solutions than potential buyers (adopters). Organizations looking to adopt these technologies may find it challenging to assess their security robustness, compliance levels, interoperability, and long-term viability. The technical sophistication required to evaluate these aspects often exceeds the expertise available within adopting organizations. As has been shown, markets with information asymmetry can lead to adverse selection and erroneous price formation on the market. High-quality innovations may struggle to justify their premium pricing when buyers cannot discern the difference in quality, leading to a market dominated by cheaper, low-quality alternatives. As a result, organizations might opt for less effective solutions that appear cost-efficient but offer inadequate protection.

For identity and trust management solutions for example, information asymmetry can significantly impede adoption. Organizations may be reluctant to invest in new identity infrastructures due to uncertainty about their security and privacy benefits. Without clear evidence of effectiveness, the perceived risk of adoption increases. Moreover, the intangible nature of cybersecurity benefits makes it difficult to quantify the return on investment. Preventing a potential data breach is less observable than improving production efficiency, making it harder to justify investing in advanced identity solutions. Finally, the proliferation of existing free or cheap, potentially less secure identity solutions (cloud/platform identities) can saturate the market, further obscuring the advantages of innovations with a high focus on security and privacy like the EUDI Wallet.

We see that information asymmetry presents a significant economic barrier to the adoption of cybersecurity innovations. The inability of potential adopters to fully assess the quality of an innovation can lead to market and pricing inefficiencies where inferior solutions predominate. To overcome information asymmetries, independent certifications and promoting transparent performance metrics have been proposed as strategies. By reducing information gaps, these measures help ensure that qualitative cybersecurity solutions are recognized and valued appropriately, facilitating their adoption.

2.2 Specific Adoption Challenges of Ecosystem Approaches

Many cybersecurity innovations, like identity and trust management infrastructures are special cases as they rely on collaborative ecosystems. As such, they face unique challenges stemming from their inherent dependence on widespread participation and

coordination among diverse stakeholders. Such ecosystems face complexities that go beyond the adoption barriers discussed in the last subchapter. To analyze them we turned to ecosystem innovations with security relation and beyond, such as social-networks and dataspaces (Gaia-X, Catena-X). Key among the identified challenges are network effects, that are closely related to the chicken-and-egg problem and the necessity of achieving a critical mass of participants, as well as aspects related to the short/long-term business and governance model of the ecosystem. TradeLens, a blockchain-based global trading platform initiated by IBM and Maersk, can be taken as a recent example. Despite being backed by major players in the IT, shipping, and logistics industries, it struggled to establish a viable ecosystem. While TradeLens aimed to enhance transparency and efficiency among various stakeholders, it ultimately faced hurdles in achieving widespread adoption. The challenges included integrating diverse participants with varying interests, concerns over the sharing of sensitive data, and the failure to establish a viable long term business model [Ce22].

Network effects occur when the value of a product or service increases with its number of users. As shown by [KS21] for SSI ecosystems, the effectiveness and utility of solutions significantly enhances as more participants – users, relying parties, and identity infrastructure providers – adopt and integrate them. The network effect creates a positive feedback loop: as adoption grows, the benefits to each ecosystem participant multiply, further encouraging additional uptake. However, network effects can also pose barriers to initial adoption. When the user base is small, the perceived value of joining the ecosystem is limited. Early adopters do not experience its full advantages due to the lack of widespread participation. For instance, an SSI system offers maximum interoperability and convenience only when a substantial number of service providers recognize and accept decentralized credentials. Without a critical mass, the innovation may fail to deliver on its promises, discouraging potential users.

Closely related to network effects is the *chicken-and-egg problem*, a situation where the value proposition for each group of users depends on the presence of the other. An example from cybersecurity are threat intelligence sharing platforms. To facilitate the sharing of threat intelligence they rely on active participation from multiple organizations to be effective. However, organizations may be unwilling to contribute valuable security insights if they perceive that others are not reciprocating. The initial lack of shared information reduces the platform's value, discouraging new participants [JBC23]. The mutual dependency creates a stalemate, blocking the diffusion of the innovation. Hence, breaking this cycle is critical for the successful adoption of ecosystem-based cybersecurity solutions.

Achieving a critical mass of participants is essential to address the challenges posed by network effects and the chicken-and-egg problem. The term refers to the minimum number of users required for the ecosystem to become self-sustaining and for the innovation to realize its full value proposition. Once this threshold is surpassed, the adoption rate typically accelerates as the benefits become more apparent and compelling to additional users. To overcome initial adoption barriers and reach a critical mass, it is imperative to

design effective incentives for early adopters. These incentives can be both economic (e.g., discounts for early adopters) and non-economic (e.g., assistance to reduce complexity, reputation enhancement to be recognized as a leader in cybersecurity). Designing these incentives requires a deep understanding of the motivations and constraints of potential adopters. For instance, small and medium-sized enterprises (SMEs) may be more sensitive to cost incentives, while larger organizations might value strategic advantages.

Another particular challenge for ecosystem lies in establishing and maintaining *sustainable business models*. In the startup phase, *initial financial investment* represents a significant hurdle. Developing the technological infrastructure needed for a secure ecosystem requires substantial upfront costs. A requirement for (new) participants to help cover development and operational expenses raises adoption barriers. For many organizations these costs can be prohibitive without clear, immediate benefits. The high initial investment is particularly challenging in cybersecurity as the returns are often intangible or realized over a long period through avoided incidents. Moreover, the paradox in cybersecurity (prevented breaches and attacks are invisible successes), make it challenging to demonstrate the value of the investment to stakeholders. This invisibility complicates efforts to articulate a compelling business case for the ecosystem, thereby hindering the recruitment of initial participants necessary to achieve critical mass. So until a critical mass of participants is achieved, substantial financing must be secured from strategic stakeholders who believe in the ecosystem's value.

Ensuring the *long-term viability of a cybersecurity ecosystem* presents further challenges. Clearly, the initial phase of financing by strategic stakeholders has to somehow transition into a long-term business model. However, monetization in an ecosystem is particularly challenging. A significant hurdle lies in the uneven distribution of efforts, investments, and benefits among ecosystem participants. The contributions required to develop and maintain the ecosystem – and the positive outcomes or value generated – may not be evenly shared. Some participants might be able to monetize the value from the ecosystem, gaining substantial benefits from its services and opportunities, while others may invest more resources or effort than they receive in direct returns. This disparity can create tension and disincentives for continued participation, especially for those who perceive an imbalance between their contributions and benefits. Identity management infrastructures like SSI give a good example. For instance, issuers may bear significant costs to provide secure credentials, while relying parties benefit from utilizing verified identities without proportionally contributing to the underlying system leading to a free-rider problem that is prominent in open infrastructures (that SSI aims to be one). To address this, it is crucial to establish value-sharing mechanisms among the participants for an equitable distribution of costs and benefits. Such mechanisms could involve collaborative funding models, revenue-sharing agreements, or cross-subsidization strategies that align incentives across the ecosystem. However, it is difficult to balance these revenue-sharing strategies to not disincentivize potential adopters.

An example for the challenge of balancing efforts and revenues can be observed in the development and maintenance of open-source cybersecurity tools, such as the OpenSSL

project, that can be regarded as an open ecosystem. Despite its critical role, the project was historically underfunded and relied on a small team of volunteer developers, leading to the Heartbleed vulnerability in 2014. Major corporations and organizations had benefited from OpenSSL without proportionally contributing to its development and upkeep. However, the vast impact of the vulnerability led to the establishment of initiatives like the Core Infrastructure Initiative – now Open Source Security Foundation [Op25]. Here, leading organizations pool resources to provide funding and support for essential open source projects. By sharing the costs and ensuring that those who benefit also contribute accordingly, the ecosystem became more robust and sustainable.

The *governance* of collaborative ecosystems, especially those in a cybersecurity context, presents further complex challenges. Effective governance structures are essential for ensuring an ecosystem's functionality and long-term viability, with trust among its participants being a major success factor. Additionally, different governance structures applicable to cybersecurity ecosystems (centralized, hierarchical, and decentralized) have different implications regarding the power balance, single points of failure, flexibility, and the management of malicious actors within the ecosystem.

In centralized governance models, decision-making authority and control are concentrated in a single entity or a small group. This offers clear lines of accountability and the ability to implement decisions rapidly. However, it introduces significant challenges in cybersecurity ecosystems. Centralization can create single points of failure: if the central authority is compromised, the entire ecosystem becomes vulnerable, making it an attractive target for malicious actors. Moreover, concentrated power may lead to issues of trust among participants, who may be reluctant to cede control to a central body, especially when sensitive security matters are involved.

Hierarchical governance structures represent a middle ground, with authority distributed across different levels but still maintaining a degree of central oversight. While this can provide a balance between coordination and autonomy, hierarchical models may suffer from bureaucracy, slower decision-making processes, and potential inflexibility.

Decentralized governance, the vision of many open ecosystems, aims to distribute decision-making authority and control among all participants. This model aligns with the principles of SSI. Decentralization enhances resilience by eliminating single points of failure and can foster trust as participants have equal influence over governance decisions. However, decentralized governance also presents challenges. Consensus-building among diverse stakeholders can be time-consuming and complex, potentially hindering an ecosystem's ability to react quickly new developments. Coordinating security measures and responses to threats to achieve a unified and effective response can be challenging. Malicious actors may exploit fragmentation within the ecosystem to advance their objectives. Additionally, the lack of a central authority can make enforcing compliance with security standards and regulation, as well as identifying clear (legal) responsibility – which might be important for potential adopters – more difficult.

Managing malicious actors in the ecosystem is another significant governance challenge.

Malicious participants can undermine trust, compromise security, and cause systemic harm. Governance structures must include mechanisms to detect, mitigate, and remove threats posed by such actors. In centralized models, the central authority can enforce rules, monitor compliance, and take swift action against malicious actors. This can be effective but relies on the central authority's ability to detect wrongdoing and act impartially. Decentralized governance can complicate the management of malicious actors. Without a central enforcement body, participants must collectively agree on actions to address threats, which can be difficult to coordinate. Establishing clear protocols for identifying and responding to malicious behavior is essential.

Finally, the question of *power balance* is critical. In ecosystems, power imbalances may arise, leading to perceptions of unfairness, unequal treatment or concerns about strategic disadvantages. Participants may fear that an ecosystem authority could exploit its position, make unilateral decisions that favor certain parties or compromise participants' autonomy and interests. This is especially pertinent in cybersecurity ecosystems where trust is paramount, and participants are cautious about sharing control over security-related aspects. Differences in participants' resources, expertise, and influence can still lead to de facto power imbalances even in decentralized structures. Large organizations with significant stakes in the ecosystem might exert disproportionate influence. Establishing mechanisms to ensure equitable participation and prevent dominance by any single entity is essential for maintaining trust and collaboration.

General Challenges		
Relative Advantage	Trialability	Environmental Factors
Compatibility	Observability	Information Asymmetry
Complexity	Organizational Factors	Threat Perception
Ecosystem-specific Challenges		
Network Effects	Initial Financial Investment	Managing Malicious Actors
Chicken-and-Egg Problem	Long-term viability (Revenue sharing)	Power Balance
Critical Mass Requirement	Governance Challenges	

Tab. 1: Adoption Challenges of Cybersecurity Innovations

Addressing the multifaceted challenges is essential for the successful adoption and sustained operation of cybersecurity ecosystems. By carefully designing strategies that promote equitable participation, transparent governance, and balanced value distribution, collaboration and trust within the ecosystem can be fostered. Table 1 finally gives an overview of general and ecosystem-specific challenges for cybersecurity innovations.

3 Learnings from a Cybersecurity Ecosystem Project

The theoretical work described before is now complemented by insights gained from a research project. The increasing complexity and frequency of cyberattacks on critical infrastructures underscore the necessity for innovative approaches to IT security information exchange. The Mantra research project aims to address this imperative by developing a secure framework for the real-time exchange of cyberattack patterns, creating a "cybersecurity mesh" through federated learning within a peer-to-peer ecosystem [Ma25]. As such, the project faces the challenges mentioned before and aims to overcome the neglection of non-technical factors. In the following, we will briefly describe how we addressed them in the first phase of the unfinished project and the insights we gained.

3.1 Addressing General Adoption Challenges

Recognizing the importance of understanding the specific situations and motivations within stakeholder organizations, the project prioritized early-stage engagement with potential users and contributors. We conducted interviews and workshops to gather insights from various organizational roles, including top management, IT security professionals, and end users. By building personas and identifying individual needs and drivers, we tailored the development process to address the differing motivations and concerns within organizations. This approach acknowledged that drivers and obstacles vary between organizational leadership and the users applying the solution. Leadership might prioritize strategic objectives such as enhancing the organization's security posture and compliance, while end users might focus on usability and integration with existing workflows. By considering these perspectives, we developed features and communication strategies that resonated with each group, thereby facilitating acceptance and adoption.

Moreover, interviews and workshops underscored the significance of human and organizational factors in the successful adoption of the innovation. Understanding and addressing softer aspects such as organizational culture, interpersonal relationships, and individual attitudes were seen as crucial. The human factor influences trust levels, willingness to share information, and acceptance of new processes. To address these aspects, we incorporated change management strategies and communication plans that highlighted the benefits of participation and aligned with organizational values. Training programs and support resources were provided to ease the transition and build confidence among users. By fostering positive relationships and addressing concerns empathetically, we enhanced engagement and encouraged proactive involvement in the ecosystem.

3.2 Addressing Specific Adoption Challenges of Ecosystem Approaches

Understanding the challenges associated with network effects and the chicken-and-egg problem, Mantra aimed to implement strategies to accelerate the achievement of a critical

mass of participants. One key method was integrating already existing security information and threat intelligence feeds before the critical mass was reached. By incorporating existing data sources into the ecosystem, we enhanced its immediate value proposition. This provided tangible benefits to early adopters, even with a limited number of participants. To further reduce adoption barriers, especially for SMEs, we aimed to eliminate costs associated with joining the ecosystem, making participation more accessible for organizations with limited resources. This inclusivity not only aimed to expand the ecosystem's user base but also to enrich the diversity of contributions, strengthening the overall security mesh.

Project work revealed that implementing decentralized governance is highly challenging in practice. While decentralization, as shown above, brings many advantages, its complexities necessitated a more pragmatic approach. We also observed that many initiatives that had started with decentralized intentions eventually adopted hierarchical models to manage operational demands effectively. As a result, we chose a hybrid governance model, combining elements of centralized oversight with decentralized participation.

4 Conclusion

In this paper, we explored the economic obstacles inherent in the diffusion of cybersecurity innovations that pursue an ecosystem approach. This is highly relevant for identity and trust management infrastructures such as eIDAS 2.0 with the EUDI Wallet that is currently being developed. Utilizing theoretical frameworks, we shed light on the multifaceted barriers that impede the effective adoption of security innovations within collaborative ecosystems.

Our findings show how successful innovation depends not merely on visionary technical invention but on effective adoption by users. The challenges identified – including network effects, establishing sustainable business models, and navigating complex governance structures – are also acute in the context of identity and trust infrastructures. The experiences from the Mantra project underscore the necessity of integrating economic considerations, stakeholder engagement, and human-centric approaches. By addressing both technical and non-technical factors, we can enhance the development and adoption of sustainable cybersecurity infrastructures, ultimately contributing to more secure and trustworthy digital ecosystems.

Our paper is subject to a number of limitations. We primarily focus on identifying and analyzing the challenges associated with the adoption of cybersecurity innovations, rather than providing comprehensive strategies to address them. Further research is needed to develop and validate solutions to these challenges. Additionally, the Mantra project is ongoing, and the learnings discussed are provisional. As the project progresses, more concrete insights and practical approaches are expected to emerge, which will further inform the development and adoption of sustainable cybersecurity infrastructures.

Bibliography

[Ak70] Akerlof, G. A.: The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. In (Diamond, P.; Rotschild, M., ed.): The Quarterly Journal of Economics, vol. 84, pp. 488, 1970.

[AJA09] Appari, A.; Johnson, M.; Anthony, D.: HIPAA Compliance: An Institutional Theory Perspective. AMCIS 2009 Proceedings, pp. 252, 2009

[Ce22] Cecere, L.: Tradelens Discontinues Operations. Why You Should Care. Forbes. 2022. https://www.forbes.com/sites/loracecere/2022/12/05/tradelens-discontinues-operations-why-you-should-care/.

[eIDAS24] Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework (2024). http://data.europa.eu/eli/reg/2024/1183/oj/eng.

[Eu24] European Commission: EU Digital Identity Wallet Home - EU Digital Identity Wallet, 2024. https://ec.europa.eu/digitalbuildingblocks/sites/display/EUDIGITALIDENTITYWALLET/.

[HA20] Hameed, M. A.; Arachchilage, N. A. G.: A Conceptual Model for the Organizational Adoption of Information System Security Innovations. In (Ramesh, C. J.; Gupta, B. B., ed.) Advances in Information Security, Privacy and Ethics, IGI Global, pp. 317-339, 2020.

[Ha22] Hansen, J.: Zero Trust Adoption: Qualitative Research on Factors Affecting the Adoption of Zero Trust, 2022. https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-486838.

[JBC23] Jesus, V.; Bains, B.; Chang: V.: Sharing Is Caring: Hurdles and Prospects of Open, Crowd-Sourced Cyber Threat Intelligence. IEEE Transactions on Engineering Management, vol. 71, pp. 6854 – 6873, 2024.

[JW10] Johnston, A. C.; Warkentin M.: Fear Appeals and Information Security Behaviors: An Empirical Study. MIS Quarterly 34/3, pp. 549–66, 2010.

[KS21] Kubach, M.; Sellung, R.: On the Market for Self-Sovereign Identity: Structure and Stakeholders. In (Rossnagel, H.; Schnuck, C. H.; Mödersheim, S., ed.): Open Identity Summit 2021, Bonn: Gesellschaft für Informatik e.V., pp. 143–54, 2021.

[Ma25] Mantra Project Website, 2025. https://asvin.io/mantra-projekt/

[Op25] OpenSSF., About – Open Source Security Foundation, 2025. https://openssf.org/about/.

[VY23] Vestad, A.; Yang, B.: Adoption of Cybersecurity Innovations—A Systematic Literature Review. In (Onwubiko, C.; Rosati, P.; Rege, A.; Erola, A.; Bellekens, X.; Hindy, H.;

Jaatun, M. G., ed.): Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media, Singapore: Springer Nature, pp. 285–304, 2024.

[Ro03] Rogers, Everett M. Diffusion of Innovations. 5. Ed. New York: Free Press, 2003.

[SAS06] Sonnenreich, W.; Albanese, J.; Stout, B.: Return on security investment (ROSI) - A practical quantitative model. Journal of Research and practice in IT 38/1, pp 45-56, 2006.

[Wa20] Wallace, S.; Green, K.; Johnson, C.; Cooper, J.; Gilstrap. C.: An Extended TOE Framework for Cybersecurity-adoption Decisions. Communications of the Association for Information System 47/1, pp. 51, 2020.

Tracking the Shift to Decentralised Identity: A Case Study in International Air Travel

Erran Carmel[1], Edgar Whitley[2] , Emma Grubb[3]

Abstract: This study explores how the air travel industry, provoked by the pandemic, shifted from a centralised paradigm towards a more decentralised one, in the form of verifiable health credentials. The study chronicles this shift, drawing on interviews with key decision–makers, participant observation and analysis of key documents. At the centre of the transition is IATA, representing the airlines, and ICAO, which sets standards for passports and digital travel credentials. The foundations for the shift began before the pandemic with the convergence of technologies (blockchain, biometrics and wallets) with new visions (frictionless travel experience, minimal handling of passenger data). The pandemic saw a travel crisis, leading to intense debate. Almost as soon as the vaccinations began, health status proof–of–concept applications were put into place using a variety of decentralised implementations. As the pandemic dissipated, the effort focused on health credentials dissipated as well, but what remained was the paradigm shift towards decentralised identities for international air travel.

Keywords: Self-sovereign identity, pandemic, digital health passport

1 Introduction

While the key benefits of decentralised identity credentials are widely understood and accepted, there is limited understanding of what is needed to support their implementation at scale. Nevertheless, one sector, the air travel industry, has begun actively deploying decentralised solutions for travel credentials. The industry began exploring the potential of decentralisation around 2015 and the process was catalysed by the need to incorporate health credentials (proof of vaccination / recovery) during the COVID–19 pandemic.

However, adoption is slow because of the range of potential technologies underpinning decentralised identities and the differing decentralisation visions [Schn19]. Progress toward decentralisation must offer a solution that is at least equivalent to existing technological architectures. To date, this has proven challenging, with many commentators noting that wide–scale adoption of decentralised identity credentials is still

[1] American University, Kogod School of Business, Washington, D.C. USA, carmel@american.edu https://orcid.org/0000-0002-3643-6805

[2] London School of Economics, London, United Kingdom, E.A.Whitley@lse.ac.uk https://orcid.org/0000-0003-1779-0814

[3] https://orcid.org/0009-0000-7453-7044

doi:10.18420/OID2025_06

some way off [Gi23].

This paper studies a shifting paradigm in the way that digital identities and attributes are used for international air travel. It seeks to answer the following research question: How did international air travel shift from a centralised credentials paradigm towards decentralised credentials?

It does this by providing a longitudinal analysis of the development of a particular form of decentralised travel credentials: for identity and health attributes. It shows how decentralisation visions brought together industry stakeholders. These activities unfold over distinct historical periods: before, during and after the pandemic.

Through interviews with key industry players and detailed archival document analysis it shows how the International Air Travel Association (IATA) explored decentralisation visions that diverged and converged. Over the same periods, the International Civil Aviation Organisation (ICAO) sought to promulgate its existing, well–established approach to international travel documents (e.g. passports). This was—and still is—based on a centralised technical architecture with public key cryptography.

Given the range of empirical materials, the paper considers the shift towards novel technological futures and draws on theories that "engage with the future as an object of critique in its own right" [BrMi03, p.5] and specifically the insight that visions of the future have three characteristics that can aid in their analysis: objectives, orders, and technologies [Berk06, p.302].

The rest of the paper is structured as follows: The setting and background are described in Section1, followed by a brief methods description in Section 2. The chronology of Section 3 is the core of the paper, chronicling, through four time periods, how the pandemic catalysed IATA's initiative to deploy decentralised identity credentials.

2 The Setting: Air travel and digitalisation

In 2018, international air travel was booming and forecasted to double to 8.2 billion air travellers by 2037 [Ia18a]. The industry was eager to use digitalisation to address pain points, including making traveller identification a "frictionless" experience [Ia18b, In17, We17]. These approaches were frequently built on biometric identification techniques.

Two key organisations, IATA and ICAO, play significant leadership roles [Ei22]. Both have headquarters in Montreal's "international district" but have different institutional arrangements, remits and methods.

IATA (International Air Transport Association) is a trade association [Ia24]. IATA built the Timatic database which provides real–time information on travel document requirements [Ia23]. This allows airlines to check that a traveller has the correct documentation before they begin their journey. IATA also initiated the One ID

programme, first proposed in 2018, that would result in a unified process allowing the passenger to arrive at the airport flagged as 'ready to fly' without needing to "check–in" [Ia18c].

ICAO (International Civil Aviation Organisation) has been a specialised agency of the United Nations since 1947 [Ic23]. ICAO developed 12,000 international standards and recommended practices—all by member consensus (currently 193 nations). It has soft enforcement power. ICAO's Traveller Identification Programme (TRIP) supports security processes including evidence of identity, as well as document issuance and control [Ic24a]. A key aspect of this is Doc 9303, ICAO's evolving specifications and guidance for passports, which ICAO terms "machine readable travel documents" [Ic21a].

In the context of air travel, identification is normally based on state–issued travel documents (passports and visas) where questions about how the identity is established have already been addressed [e.g. Ia18b]. Manual authentication is now making way to biometric authentication [KlPl20]. Finally, authorisation (supported by systems such as Timatic) allows authenticated passengers to board planes and pass through border control points.

In parallel, there has been a shift in thinking about identity and credentials: that individuals control their own identity and attributes rather than relying on state–based intermediaries or monolithic identity providers. That vision is to enhance individual privacy against the emerging risks of state or corporate surveillance. These citizen–centric solutions offer a vision of decentralised identity known as self–sovereign identities [Al16], many of which are based on blockchain technologies.

The paper presents a longitudinal study of digital identity for international air travel, centred around the period of the COVID–19 pandemic. This period involved a major shock to the air travel industry as travel restrictions led to collapse in passenger travel. At the same time, many solutions for reopening societies and international travel were sought. In parallel, technological developments and standards for self–sovereign identities continued to mature.

Whilst the pandemic played a major role in the transformation of digital identities for international air travel, this is not just another story of the role of the pandemic on digitalisation. Instead, the pandemic accelerated visions of a future for digital identity that had been under development before the pandemic struck.

This paper follows Berkhout's [Be06] suggestion that visions of the future have three characteristic features: objectives; orders; and technologies (p. 302). Applying these features makes it easier to draw insights from the study. They provide a narrative for the various stakeholders, bind together communities of interest and of practice as well as focusing resources. They structure "the technical, institutional and behavioural problems that need to be resolved in order for a particular vision to be realised" [Be06, p.305]. The three features are realized as follows: Objectives (the qualitative or quantitative expression of novel future outcomes) are presented as visions for the future in think tank reports, blueprints and press releases; Orders (the social and institutional relationships in which

these objectives can be met) are presented as institutional and governance relationships that exist to coordinate the various stakeholders that are trying to deliver the stated objectives; and Technologies (The means for achieving objectives) are presented as specific technological artefacts, as well as documented standards that determine their operation and use.

3 Methods

The data collection for the study combined participant observation, interviews and document analysis as described below.

Participant observation. The analysis draws on the authors' extensive experience in the policy debates and standards setting for digital identity systems as well as specific digital identity responses to the pandemic [references removed for peer review]. Significantly, the participant observation provided the authors with unique contextual knowledge [Da21] providing a hermeneutic reading of the selected documents.

Interviews. For this study, the authors conducted fourteen interviews with prominent players in the digital identity space. All interviewees were selected because of the role they played in the shift towards decentralised, self–sovereign digital identities in the period 2017–2023. Most are recognized leaders in the identity field and were central global players during the period of interest.

Interviews took place in the years 2021, 2022 and mostly 2023. Interviews were semi–structured, recorded and transcribed and lasted about one hour. Some of the interviewees participated in follow–up email exchanges with the authors and / or shared additional documents. The authors guaranteed the interviewees anonymity to themselves as individuals, and as members of their organisations, hence pseudonyms are used throughout. Participants P01 to P04 are IATA / ICAO executives; participants P05 to P14 are consultants and vendors. The role descriptions are therefore deliberately vague, as many of the interviewees have been in prominent positions in the air travel ecosystem and their occasionally judgmental quotations must be unattributable.

Timeline. These three above sources of data (participant observation, interviews and documents) were combined to produce the timeline below. The timeline has three main components: key documents and organisations (marked as squares), events (marked as ovals) and influence arrows. The authors independently went through each of these components to determine their significance for inclusion into the timeline and their influence in relation to one another. Each author's ratings were then compared and differences resolved by dialogue.

4 Chronicle of the Shift

The timeline is presented in Figure 1 is divided into four eras chronicled in this section: Pre–Pandemic, Pandemic Stage1, Pandemic Stage2 and Post–Pandemic.

4.1 Pre–Pandemic (prior to 2020)

During the pre–pandemic era, many of the objectives for future outcomes for the international air travel industry were focused on incremental improvements to existing processes. For example, ICAO's Doc 9303 (2021a) played a central role in standardising the format of the passport as a globally interoperable, machine–readable document with both a visual inspection zone and a machine–readable zone. Since 2005 ICAO also introduced detailed guidance about checking the integrity of the travel document, including checking the digital signatures used to sign the chip on ePassports.

This verification relies on a centralised architecture based around the ICAO Public Key Directory (PKD) which is a central database serving as the repository of Document Signer Certificates and Certificate Revocation Lists issued by participating countries, together with a system for the distribution of the certificates and lists worldwide [Ic24b]. ICAO maintains the PKD on behalf of participating countries to facilitate validation in ePassports.

Whilst Doc 9303 defines the protocols that are used, interoperability was centralised because, according to P07, "the fastest way to get interoperability [across nation states] is to have a dictator".

Many of the ideas behind the use of the PKD to validate ePassport data were also being explored in relation to Visible Digital Seals (VDS) for non–electronic documents such as visas "to ensure the authenticity and integrity of non–electronic documents in a comparatively cheap, but highly secure manner using asymmetric cryptography" [Ic18, S.4].

This centralised approach largely worked for the majority of the billions of travelers, but many understood its gaps-- notably that scores of smaller nations did not use the PKD to validate the data on ePassports [Ne18], as was confirmed by P06:

And, by the way, half the countries don't actually have keys in the directory and that kind of doesn't work. But they don't say that part out loud. You only find that out when you work with the systems people. So, the idea that the ICAO system works is a myth.

In contrast to the broad success in rolling out the PKD, there was no significant progress at digitising international health credentials until the late 2010s and, even then, progress was quite limited. CAPSCA, an ICAO–led collaboration for health crises had little impact and often was focused on issues like disinfection [Ic23].

From an industry perspective there was an emerging vision to change the international traveller experience and make it seamless, "frictionless", secure and scalable; elevating traveller convenience and reducing business costs for airlines. For example, the World Economic Forum (WEF) in concert with technology consulting firms released reports advocating cross–border data sharing [We16, We17].

In response, IATA began to formulate its vision, known as One ID [Ia18c]. Its underlying architecture included an Identity Management Platform that allowed stakeholders (airlines, airports etc.) to store, share, reuse and amend passenger information in a collaborative and trusted manner" [Ia18c]. This was followed by a think tank report [Ia18b] that explicitly called for "having a central repository to get safely and securely (sic) information … using a unique, centralised and trusted by the industry authority" [Ia18b, S.10].

Meanwhile, a number of elements began to coalesce around 2017. These included technological developments that could be used to support digital identities including distributed trust via blockchains, enhanced biometrics, digital wallets and improved cryptographic techniques. At the same time, fundamental assumptions about digital identity systems were being reevaluated. For example, P07 described being "sucked in" to digital identity because of a personal interest in blockchain technologies and was soon working on blockchain–based digital identity applications supporting refugees. Many of these initiatives included major industry players including software developers and consulting companies.

In January 2018, two years before the pandemic, the WEF unveiled its vision of a "Known Traveller Digital Identity" (KTDI). A key feature of this initiative was to respond to the "risks related to the existence of centralised databases" [We18, p.13] by building on blockchains. The WEF organised an event at Davos that year which demonstrated the potential of this approach, including an endorsement by Marriott CEO Sorensen. According to P13, this level of support was important for the global travel sector given Marriott's status as one of the largest hotel chains in the world.

At this time there were a number of competing visions for the future of international air travel, utilising different technologies and standards. The social and institutional relationships between the visions were, however, limited and whilst both IATA and ICAO were aware of more decentralised approaches, neither felt the need to actively consider them and continued to advocate for the paradigm of centralised approaches.

By the time WEF presented another report on digital identity at Davos in January 2020 [We20] there was a sense that "everyone was talking about decentralised credentials" (P07). They were not, however, talking about health credentials and when the pandemic hit two months later, KTDI was mothballed. Nonetheless, the decentralised vision behind KTDI alongside like–minded initiatives would continue to gain traction, evolving into various forms under the pressure of the global pandemic.

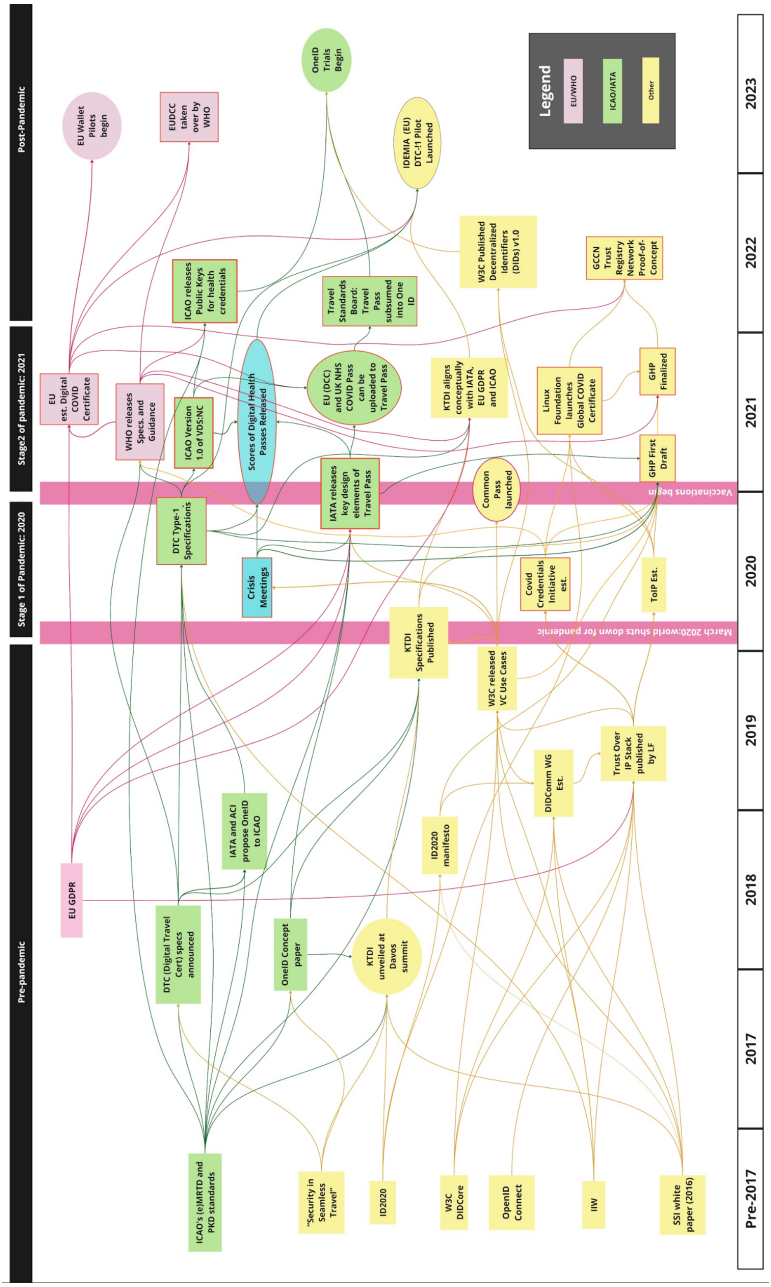


Figure 1. Timeline of events and influence in the decentralisation shift, centred on the pandemic years.

4.2 Pandemic Phase 1: before vaccination roll–out (2020)

Whilst the problem was understood, the choice of solution was less clear and in the urgency of a global crisis and through countless online meetings, debates unfolded within the community as to whether it was best to build on ICAO's well–established, centralised PKD or to explore more decentralised alternative approaches that were currently in vogue. For example, P02 noted that inside ICAO there were traditionalists opposed to anything decentralised, arguing to leverage existing solutions rather than consider a novel approach that was untested and "unproven".

P01 recalled:

We all went out and did our own unique things. There was definitely a lack of consensus. … ICAO got pushed aside in the battle with WHO … and then IATA were saying: we gotta get this industry back on its feet … We gotta get passengers back in the air.

Meanwhile, in April 2020, one interviewee, P03, was impacted by a personal flight travel ordeal at the height of the pandemic lockdown. This provided an additional, personal, motivation to create a solution quickly. By June 2020 P03 felt there was a "convergence" in the debate: the tipping point may have been the realisation that there was a compromise between the familiar centralised view and a radical decentralised vision:

It was when I discovered the notion of public permissioned blockchain: anyone can read, but only authorised entities can write. That was a real breakthrough for me, because it gave a level of assurance of trust in issuers, while still allowing anyone, anywhere to verify the token.

Whilst trust in issuers was relatively straightforward for travel documents it was more complicated in the case of health credentials. For example, in the process of drafting the Good Health Pass and the work of the COVID Credentials Initiative, P06 noted the need to create a secondary infrastructure to give "coherence to a decentralised system". Of particular importance was the role of the issuers of health credentials which, at that time, would have been based on COVID test results. For the credentials to operate successfully, there needs to be confidence in both the issuer of the credentials and their ability to bind test results to individuals.

At the time, the core entity offering validity to credentials was ICAO's Doc 9303 standards and PKD. Alternative trust registry standards simply did not (yet) exist. Thus, progress with the more decentralised visions required a focus on infrastructure and governance.

ICAO was still working to establish foundational identity elements that it had been developing for some time and so it launched the Digital Travel Credential (DTC) Type 1 guidelines [Ic20] in late 2020 [Idem21, S.5], thus establishing another standard. This would later lead to a number of vendor implementations. DTC Type 1 is the first step in migrating the passport onto the mobile phone and opening interoperability to health credentials.

By the end of 2020, several vendors, recognizing an opportunity, were beginning to develop wallet–based apps for digital health credentials, including [Ib22] and CommonPass [Busb20] based on a range of technological approaches.

However, the shift in thinking that decentralised approaches could work was solidifying. In November 2020 IATA announced Timatic as "the base for the global registry and verification of health requirements" and the One ID concept approved in 2019 became the "base for the IATA Contactless Travel App for identity verification that will also manage the test and vaccination certificates"[Ia20].

4.3 Pandemic Phase 2: The first year of vaccination roll–out (2021)

When the very first vaccinations were administered in December 2020, the vision for reopening travel switched from immunity passports to what were then called "Digital Health Passports". During the first half of 2021 there was a frenetic pace as many organisations around the world worked on elements of national and global health credential technology systems and their associated governance mechanisms.

The intensity of effort to develop new standards and technological artefacts in this period was indicated by the interviewees. For example, P06, one of the key individuals of the ten drafting groups in the Interoperability Working Group for Good Health Pass, noted that they worked "intensively for six weeks". During this same period, P03 reported working "7 days a week for four months straight".

P10 described "many prototypes" which came out at this time including domestic "green passes" and equivalents, with limited geographical scope but the same underlying technical requirements: securely binding individuals to their health (vaccination) attributes. There were trials of the IATA Travel Pass in conjunction with Singapore Airlines in March 2021 [Si21], British Airways on flights from London to Switzerland by June 2021 [Br21] and Finnair flights to Helsinki in July 2021 [Ro21].

During this period ICAO continued to work on "technical solutions aimed at meeting [the pandemic] challenge" [Ic21b, S.2] building on their work of Visible Digital Seals to allow for travel related public health proofs [Ic21c]. ICAO acknowledged that other approaches existed but remained unconvinced as to whether the verifiable credentials standard was suitable for implementation globally.

A proposal on digital travel credentials presented to the ICAO facilitation panel argued that it was "of critical importance" that digital travel credentials "can be validated using the travel document issuing authority's public key infrastructure" that is currently deployed for ePassports [Ic21d, S.2].

While leveraging existing standards appeared a natural choice for ICAO, the approach was anathema to many advocates of more decentralised digital identities because it violated many of the basic privacy principles. For example, the visible digital seals could not offer

selective disclosure of health attributes [Hard21, YaYo21] and one interviewee (P06) simply stated that they were "horrified" at the proposals.

P04 made a similar point about IATA's motivations: "It's ironic that every business wants to collect consumer data, but we want to get away from it". IATA argued that "the verification of health credentials … is a responsibility for governments and should result in the issuance of an authorization confirming that the passenger meets the government requirements including health requirements and is 'OK to travel'" [Ia21a, S.14]. Doing this protects passenger data privacy and ensures that airlines would "not be required to access, validate or store travellers' medical data" [Ia21a, S.14]. This was in marked contrast to earlier IATA thinking that envisaged the creation of a Digital Centralized Authority to handle the assurance of passenger data [Ia18b, S.11].

Many stakeholders were actively seeking more detailed guidelines to develop more decentralised approaches and several of the interviewees pointed to the Good Health Pass (GHP) blueprint [Good21], published February, as a major milestone in this regard.

The governance challenges faced by the issuers of COVID credentials were transformed by the need to record and share authoritative vaccine information. Vaccines were normally administered by recognised health providers—which required a governance framework for determining which claims about someone being vaccinated could be relied upon.

Not all the discussions were about the dichotomy of centralised versus decentralised. Personality issues in all camps. For example, ID2020 was reluctant at first to work on GHP. There were many sceptical stakeholders of the more radical self–sovereign identity vision, with consultant P09 noting "that a system like that doesn't work. The trust triangle doesn't work" adding that there was still a need to reduce fraud because of the "flawed" design.

In March 2021 a number of influential non–airline COVID certificates were released as well including the EU Digital COVID Certificate (EU DCC) [Eu21].

By August 2021, it was announced that the EU DCC, as well as the UK NHS COVID Pass, could be uploaded onto the IATA Travel Pass [Ia21b]. The cross–organizational collaborations, as well as interoperability between European credentials and IATA Travel Pass architecture, gave the IATA credential credibility and weight. In October, IATA continued to announce numerous airlines agreeing to implement the Travel Pass.

ICAO, at this same time, continued to push the Visible Digital Seal as the recommended format. However, ICAO also noted "existing solutions should be considered and could incorporate a Visible Digital Seal or other interoperable formats from regional or global intergovernmental bodies" [Ic21e, S.i]. It appeared that while ICAO remained on a centralised mindset, it was recognizing other solutions, such as IATA Travel Pass. Indeed, P01 suggested that ICAO "threw up its hands" and gave in to IATA.

As 2021 ended, major airlines defaulted to their homegrown 'OK to travel' health verification solutions and travel remained complicated for passengers and stakeholders. There were many reasons for this and much finger-pointing. For example, P09, raised implementation concerns: "the system implementation was not successful because IATA licensed a vendor … to build the system. IATA shouldn't be in the systems building business".

Despite criticisms of the Travel Pass, it paved the way for greater gains for IATA with its return to One ID. In December of 2021, as P03 notes, IATA's Travel Standards Board "requested that the Travel Pass technical specifications become the other [specification] for One ID" and asserted that Travel Pass "really took us years ahead". Thus, post hoc, it became an effective proof of concept.

4.4 3.4 Post–Pandemic: 2022 and beyond

By the beginning of 2022, interest in digital health credentials waned rapidly for several reasons, perhaps most importantly, because of vaccine efficacy. As a result, most nations had removed vaccine travel requirements by mid–2022 and airlines discontinued use of health credential apps [Pa22]. Perhaps too late for this pandemic, in late 2021 ICAO released its Health Master List and policies with a global directory of public keys required for authentication of health credentials [Ic21f]. The KTDI initiative resumed, but as secondary to other developments. Several important initiatives closed the following year. The COVID Credentials Initiative folded in July 2023. The WHO absorbed the EU Digital COVID Credential as EU funding came to an end [Eu23]. By the end of 2023, despite the global health catastrophe, there was still no international digital health trust architecture to underpin health credentials. Nevertheless, pandemic–era projects for digital travel credentials appear to have served as major catalysts for the next steps in the transition towards more decentralised approaches in travel. P08 also noted that the EU DCC paved the way for more recent advancements like digital wallet credentials. P08 suggested that, post–pandemic, three major international digital identity models continued to progress. The first two are foundational identities that rely on centralised authority: the well–established ICAO Digital Travel Credential model for passports and the less–mature ISO mobile Drivers License (mDL) model (which enables selective disclosure). The third model is the decentralised verifiable credentials approach.

1 Concluding Discussion

The paper chronicled the shift towards a vision of decentralised digital identity in international air travel in terms of competing visions of the future

Early visions of the future of international air travel were based on attempts at digitalisation of existing processes and this objective was broadly shared across the whole sector. Its likely implementation was grounded in an understanding of the capabilities of

emerging technologies, particularly biometrics and cryptographically assured documents. The sector increasingly needed to rely on passenger data in a secure and trusted manner, ideally with minimal human intervention. At the same time there were growing concerns about how best to handle passenger data in a privacy–preserving manner. The COVID–19 pandemic accelerated this shift. For air travel, there was a need to shift from checking if a passenger was 'ready to fly' to determining if they were 'OK to travel'. That is, moving beyond visa details to sensitive health information (e.g., details of vaccination) -- with the related risks of holding such data.

ICAO, with its expertise in high–assurance travel documents, explicitly sought to leverage existing solutions and presented a solution "which combines practicability, security and ease of verification reusing the existing ePassport trust model to address this challenge" [Ic21b, p.2]. In particular, ICAO extended its work around Visible Digital Seals [Ic18] to include "travel–related public health proofs" [Ic21b]. ICAO acknowledged other initiatives proposing more decentralised approaches but noted that they are "not yet used in the global border management and travel document continuum" [Ic21b, p.3], highlighting, instead, the valuable role of global standards setting bodies such as the WHO and ICAO in "establishing global baselines for common approaches" [Ic21b, p.3].

IATA drew on very different institutional and social orders to embrace a more decentralised approach to deal with sensitive health information. Its organisational structure allowed it to liaise with independent organisations seeking innovative solutions to verifiable credentials. In particular, it worked closely with Evernym [Andr20], Linux Foundation Public Health [Linu23], Good Health Pass initiative [Go21] and the COVID Credentials Initiative [Co22] amongst others. To address the question of how best to handle passenger data including health data, IATA worked closely and urgently with these groups to develop new technologies and standards in the verifiable credentials space.

IATA's more open approach enabled it to trial its Travel Pass far more rapidly than ICAO's alternative. As a result, IATA learned valuable lessons about implementing verifiable credentials at scale, giving it the confidence in this approach that it continues to develop for non–pandemic processes. While both IATA and ICAO sought to achieve the vision of determining whether an individual was 'OK to travel', the pandemic highlighted the complexity by adding health attributes. For example, whilst a passport might be valid for ten years, a COVID test result might only be useful for three days.

Digital identities often rely on single authoritative sources (civil registration databases, trusted government-issued credentials such as passports, driving licences). In contrast, the attributes for determining whether someone is 'OK to travel' shifted from immunity following infection—to details of vaccination or a recent negative test. Moreover, the authoritativeness of these attributes varied by location with some countries recording vaccinations in a national health registry, while others provided vaccinations in a more decentralised manner (e.g. local pharmacies) with potentially less rigorous binding of the attribute to the person who was vaccinated.

Bibliography

[Al16] Allen, C. (2016). *The Path to Self-Sovereign Identity. Life with alacrity.* https://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html

[Andr20] Andrade-Walz, Alex: *COVID-19 and the lasting shift to contactless ID verification* : BiometricUpdate, 2020

[Be06] Berkhout, Frans: Normative expectations in systems innovation. In: *Technology Analysis & Strategic Management* Bd. 18 (2006), Nr. 3/4, S. 299–311

[Br21] British Airways: *British Airways to Trial IATA Travel Pass* : https://mediacentre.britishairways.com/pressrelease/details/86/0/13143, IATA, 2021

[BrMi03] Brown, Nik ; Michael, Mike: A sociology of expectations: Retrospecting prospects and prospecting retrospects. In: *Technology Analysis & Strategic Management* Bd. 15 (2003), Nr. 1, S. 3–18

[Bu20] Busby, Mattha: *Digital „health passport" trials under way to aid reopening of borders* : The Guardian, https://www.theguardian.com/world/2020/oct/17/digital-health-passport-trials-commonpass-travel-covid-19, 2020

[Co22] COVID Credentials Initiative: *(Public Release) DHCS Implementation Guide V1.pdf* : COVID Credentials Initiative, 2022

[Da21] Davison, Robert M.: From ignorance to familiarity: Contextual knowledge and the field researcher. In: *Information Systems Journal* Bd. 31 (2021), Nr. 1, S. 1–6

[Ei22] Eilstrup-Sangiovanni, Mette: Ordering global governance complexes: The evolution of the governance complex for international civil aviation. In: *The Review of International Organizations* Bd. 17 (2022), Nr. 2, S. 293–322

[Eu21] European Commission: *A framework for the issuance, verification and acceptance of interoperable certificates on vaccination, testing and recovery to facilitate free movement during the COVID-19 pandemic (Digital Green Certificate* : European Commission, https://ec.europa.eu/info/sites/info/files/en_green_certif_just_reg130_final.pdf , 2021

[Eu23] European Commission: *EU Digital COVID Certificate 2023* : European Commission, https://commission.europa.eu/strategy-and-policy/coronavirus-response/safe-covid-19-vaccines-europeans/eu-digital-covid-certificate_en , 2023

[Gi23] Giannopoulou, Alexandra: Digital Identity Infrastructures: a Critical Approach of Self-Sovereign Identity. In: *Digital Society* Bd. 2 (2023), Nr. 2, S. 18

[Go21] Good Health Pass: *Good Health Pass A Safe Path to Global Reopening* : Good Health Pass, 2021

[Ha21] Hardman, Daniel: *A Better, More Secure, and More Private Approach to COVID Credentials* : Evernym, 2021

[Ia18a] IATA: *IATA Forecast Predicts 8.2 billion Air Travelers in 2037* : IATA, https://www.iata.org/en/pressroom/pressroom-archive/2018-press-releases/2018-10-24-02/ , 2018

[Ia18b] IATA: *Airline Industry Retailing (AIR) Think Tank 2018* : IATA, https://www.iata.org/contentassets/a46387f9bc6b42368c0a72664f6f930f/air-tt-white-paper.pdf , 2018

[Ia18c] IATA: *One ID: Concept paper Version 1* : IATA, 2018

[Ia20] IATA: *IATA Travel Pass Key to Reopening Borders Safely* : IATA, https://www.iata.org/en/pressroom/pressroom-archive/2020-press-releases/2020-11-23-01/ , 2020

[Ia21a] IATA: *From Restart to Recovery: A blueprint for simplifying air travel* : IATA, 2021

[Ia21b] IATA: *EU and UK Digital COVID Certificates Recognized by IATA Travel Pass* : IATA, 2021

[Ia23] IATA: *Timatic* : IATA, https://www.iata.org/en/publications/timatic/ , 2023

[Ia24] IATA: *Vision and Mission* : IATA, 2024

[Ib22] IBM: *IBM Digital Health Pass* : IBM, https://www.ibm.com/downloads/cas/YBNOVQP3 , 2022

[Ic18] ICAO: *Visible Digital Seals for Non-Electronic Documents (Version 1.7)* : ICAO, 2018

[Ic20] ICAO: *Guiding core principles for the development of a Digital Travel Credential (DTC) Version 4.4* : ICAO, https://www.icao.int/Security/FAL/TRIP/PublishingImages/Pages/Publications/Guiding%20core%20principles%20for%20the%20development%20of%20a%20Digital%20Travel%20Credential%20%20%28DTC%29.PDF , 2020

[Ic21a] ICAO: *Doc 9303 Machine Readable Travel Documents* : ICAO, https://www.icao.int/publications/pages/publication.aspx?docnum=9303 , 2021

[Ic21b] ICAO: *Guidance for Visible Digital Seals (VDS-NC) for Travel-Related Public Health Proofs Release 1.0* : ICAO, 2021

[Ic21c] ICAO: *Guidelines - VDS for Travel-Related Public Health Proofs* : ICAO, 2021

[Ic21d] ICAO: *The digital travel credential (DTC) FALP 12* : ICAO, 2021

[Ic21e] ICAO: *Testing and cross-border risk management measures manual* (Nr. 10152) : ICAO, 2021

[Ic21f] ICAO: *International Civil Aviation Authority (ICAO) Health Master List Policy and Procedure* : ICAO, 2021

[Ic23a] ICAO: *About ICAO* : ICAO, https://www.icao.int/about-icao/Pages/default.aspx 2025

[Ic23b] ICAO: *Collaborative Arrangement for the Prevention and Management of Public Health Events in Civil Aviation - CAPSCA* : ICAO, https://www.icao.int/safety/CAPSCA/Pages/default.aspx , 2023

[Ic24a] ICAO: *ICAO Traveller Identification Programme* : ICAO, 2024

[Ic24b] ICAO: *ICAO Public Key Directory* : ICAO, 2024

[Id21] IDEMIA: *Digital Travel Credential Position Paper* : IDEMIA, 2021

[In17] Innovate Identity: *Transforming the Airline Passenger Journey* : OIXUK, https://openidentityexchange.org/networks/87/item.html?id=180 , 2017

[KlPl20] Kloppenburg, Sanneke ; van der Ploeg, Irma: Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences. In: *Science as Culture* Bd. 29 (2020), Nr. 1, S. 57–76

[Li23] Linux Foundation Public Health: *Linux Foundation Public Health - Open Source Software for Digital Health* : Linux Foundation Public Health, 2023

[Ne18] Newman, Lily Hay: *US Border Patrol Hasn't Validated E-Passport Data For Years* : Wired, 2018

[Pa22] Park, Kyunghee: *Vaccine Passports Likely to Become Unnecessary, Airline Alliance CEO Says.* : Bloomberg.com, 2022

[Ro21] Robinson, James: *Finnair to trial IATA „Travel Pass" app on flights between Heathrow and Helsinki* : Daily Mail, 2021

[Sc19] Schneider, Nathan: Decentralization: an incomplete ambition. In: *Journal of Cultural Economy* Bd. 12 (2019), Nr. 4, S. 265–285

[Si21] Singapore Airlines: *Singapore Airlines First In World To Pilot IATA's Travel Pass App* : Singapore Airlines, https://www.singaporeair.com/en_UK/dk/media-centre/press-release/article/?q=en_UK/2021/January-March/ne0921-210308 , 2021

[We16] WEF: *Security in Travel Promoting Seamless and Secure Travel through Cross-Border Data Sharing and Collaboration* : World Economic Forum, https://www3.weforum.org/docs/IP/2016/MO/WEF_AT_SecurityinTravel.pdf , 2016

[We17] WEF: *Digital Borders Enabling a secure, seamless and personalized journey* : World Economic Forum, https://www3.weforum.org/docs/IP/2017/MO/WEF_ATT_DigitalBorders_WhitePaper.pdf , 2017

[We18] WEF: *The Known Traveller Unlocking the potential of digital identity for secure and seamless travel* : World Economic Forum, https://www3.weforum.org/docs/WEF_The_Known_Traveller_Digital_Identity_Concept.pdf , 2018

[We20] WEF: *Reimagining Digital Identity: A Strategic Imperative* : World Economic Forum, https://www.weforum.org/whitepapers/reimagining-digital-identity-a-strategic-imperative/ , 2020

[YaYo21] Yang, Lucy ; Young, Kaliya: *The Global COVID Certificate Landscape* : Linux Foundation Public Health, 2021

Phishing and Identity Manipulation trough Audiovisual Channels

Analysis, Classification, and Protective Measures

Lara Eberl, Lisa-Marie Engländer, Caroline Löhle, Dennis Jahnecke, Aylin Baris, Donjeta Seljaci, Schabnam Shamsi, Jochen Günther[1]

Abstract: Phishing and identity manipulation via audiovisual channels, particularly in videoconferencing, have become significant cybersecurity threats. Advances in artificial intelligence (AI) enable attackers to convincingly mimic voices and appearances, exploiting human vulnerabilities. This paper examines the mechanisms behind such attacks, focusing on deepfake technologies used for financial fraud, political disinformation, and social engineering. Our experimental results reveal that over 70% of participants were unable to detect manipulated content, highlighting the limitations of human perception. The study identifies real-time detection tools, awareness campaigns, and organizational measures as key protective strategies. By combining technical and legal approaches, this research aims to develop effective solutions to counter these threats and maintain trust in digital communication.

Keywords: Audiovisual phishing (vishing), Video manipulation, Deepfake, AI-driven impersonation, Cybersecurity and human factors

1 Introduction

Phishing via audiovisual channels, particularly in videoconferencing environments, has emerged as a critical cybersecurity threat. The convergence of advanced artificial intelligence (AI) technologies and the increasing reliance on virtual communication platforms has enabled attackers to exploit vulnerabilities using convincing impersonation tactics. These attacks frequently leverage deepfake audio and video technologies to mimic the voices and appearances of trusted individuals, undermining the authenticity of real-time interactions. For example, AI-generated voice cloning has been used to defraud organizations by emulating executives' voices, with one notable case involving the theft of €220,000 through a fake audio directive [GN25; KJW23]. Similarly, studies have demonstrated that deepfakes are intentionally used to manipulate public opinion, discredit individuals, and commit financial fraud [APA24].

[1] Heilbronn University of Applied Sciences, Faculty of Business, Max-Planck-Straße 39, Heilbronn, DE-74081, jochen.guenther@hs-heilbronn.de, https://orcid.org/0000-0002-2509-7422

doi:10.18420/OID2025_07

Deepfake technology, based on advanced AI methods such as Generative Adversarial Networks (GANs) and Autoencoders, enables the manipulation of audio, images, and videos to create highly realistic yet false content [Ah22]. While the technology has found positive applications in various fields, such as the film industry and medical research, it also poses significant risks. Particularly concerning is the spread of misinformation and identity theft, as deepfakes can undermine trust in digital media [Co19; APA24]. In an increasingly digital world, deepfakes pose a threat to political processes, corporate security, and personal integrity [La23; SAK24].The sophistication of these AI-driven techniques presents significant challenges for detection and prevention. Deepfake audio, generated through GANs or text-to-speech models, can produce speech that is nearly indistinguishable from genuine human voices [KJW23]. Likewise, video deepfakes can convincingly replicate facial expressions and movements, further complicating the identification of fraudulent activities during videoconferences [MDW25]. The ability to produce fake audio recordings and manipulate facial expressions in real-time exacerbates the risk of misuse [SAK24]. As the generation of deepfakes becomes more advanced, it is increasingly difficult to detect such deception. Consequently, the development of deepfake detection methods must keep pace with the rapid advancements in generation techniques [JS21].These developments have not only heightened risks for financial fraud but have also raised concerns about the erosion of trust in digital communications and evidence. A particularly serious issue is the use of deepfakes in videoconferencing, where manipulated identities can be exploited to impersonate superiors or colleagues and gain access to confidential data [SAK24].

This paper presents a comprehensive survey that makes the following contributions. We provide an overview focusing on the effectiveness and perception of deepfake voice and video content. This survey also includes a detailed summary of up-to-date publicly accessible online tools to generate deepfake audio and video content, highlighting their potential misuse in phishing attacks and the associated security risks. By addressing individual perception, tools and implications, this research seeks to contribute to a comprehensive understanding of how to safeguard against these emerging threats in an increasingly digital world.

The remainder of this paper is organized as follows. Section 2 outlines the methodology, which includes both a literature review and a survey to analyze the effectiveness and perception of deepfake-based phishing attacks. Section 3 presents the findings from the literature review and the results of our survey, providing insights into existing research and empirical data on the risks and challenges associated with AI-driven impersonation. Sections 4 provides a summary of the identified fields of action and an outlook on further research.

2 Methodology

2.1 Research Design

This research follows a Design Science Research (DSR) framework as outlined by Hevner, which seeks to build and evaluate artifacts by integrating rigorous academic foundations with practical relevance. This approach is particularly well-suited for addressing complex and evolving issues such as audiovisual phishing and deepfake-based impersonation. The DSR process enables the combination of theoretical insights and real-world observations, leading to the generation of actionable, practice-oriented outcomes [HC10]. Within this framework, existing knowledge was systematically reviewed with the help of a structured literature analysis, and empirical data were collected through user surveys to design evidence-based recommendations.

2.2 Literature Review Methodology

The literature review was conducted using the systematic, concept-centric methodology by Webster and Watson [WW02], extended by Fettke [Fe06]. This method organizes literature around key themes and concepts rather than individual sources, allowing for a deeper understanding of research streams and identifying conceptual gaps.

The search was performed across multiple reputable academic databases, including Google Scholar, Springer Link, AIS eLibrary.

To ensure the relevance and quality of the literature, only peer-reviewed publications and academic works published from 2017 onward were included. This timeframe reflects the point at which deepfake technologies began receiving broader attention and practical deployment. Searches were refined using specific keywords in the title field, such as:

- “Voice phishing”
- “Deepfake”
- “Video manipulation”
- “Cybersecurity and human factors”

The initial search yielded a substantial number of results across academic databases. Due to the volume, a complete review of all entries was not feasible. Instead, a multi-stage filtering process was implemented:

1. Elimination of duplicates and low-quality or non-academic content.
2. Pre-screening of titles and abstracts to assess thematic fit.
3. In-depth review of high-relevance publications, focusing on conceptual contribution and methodological rigor.

Ultimately, this process resulted in the identification of a core set of 14 highly relevant academic sources, which serve as the conceptual foundation of this paper. These works were selected based on their methodological rigor, conceptual depth, and direct relevance to topics such as audiovisual phishing, AI-generated impersonation, and human perception. By grounding the research in this carefully curated body of literature, the study offers a well-rounded synthesis of technical, psychological, and practical perspectives.

2.3 Deepfake Technology and Mechanism

Deepfakes are AI-generated or manipulated audio, image, and video content created using deep learning techniques such as Generative Adversarial Networks (GANs) and Autoencoders [Ah22].

The technology enables the replication of faces, voices, and even full-body movements, making it appear as though individuals have said or done things they never did [APA24]. While deepfake technology has beneficial applications in fields such as the film industry and medical research, it is also used for harmful purposes, including the spread of misinformation and identity theft [La23].

Literature highlights the role of perception, awareness, and detection in audiovisual phishing and deepfake technologies.

- *Perception*: Deepfakes exploit human cognitive and psychological vulnerabilities, making it hard to distinguish real from fake content [Ah22]. Humans tend to have a natural tendency to accept visual information as genuine unless there are clear indications of fabrication (“authenticity bias”) [GEF22]. Furthermore humans often overestimate their ability to identify manipulated media. This overconfidence can lead to a lack of critical scrutiny, increasing susceptibility to deepfake deceptions (“overconfidence bias in detection”). Individuals are more inclined to believe information that aligns with their pre-existing beliefs. Deepfakes that confirm these beliefs are less likely to be critically evaluated and more readily accepted as true [MH20].
- *Awareness*: Low public awareness and low and digital literacy heightens the success of phishing attacks [APA24]. Many individuals lack the necessary knowledge to critically assess digital content, leading to increased susceptibility to fake news, impersonation scams, and manipulated videos
- *Detection*: Detection tools lag behind deepfake advancements, lacking real-time capabilities as models require intensive computational resources and fail to provide instant verification, limiting their practical application in everyday scenarios [JS21]. Furthermore many existing solutions rely on outdated forensics techniques that are ineffective against the latest AI-generated manipulations.

2.4 The Impact of Deepfakes in the Digital Age

The increasing prevalence of deepfake technologies is reshaping how digital media is consumed and evaluated. The ability to produce highly realistic forgeries undermines trust in visual and audio evidence [Co19].

A key concern is the erosion of credibility in digital recordings. Individuals could deny past statements or actions by falsely claiming they were deepfakes, a phenomenon known as the "Liar's Dividend" [La23]. Additionally, deepfakes pose a growing threat to democratic processes, as documented cases show that manipulated political speeches have been used to spread misinformation or discredit candidates [APA24].

Cybercriminals also exploit deepfake technology in CEO fraud schemes, using fake voices to deceive employees [SAK24]. This presents new challenges for forensic investigations and corporate security measures [Co19].

2.5 Survey Methodology generally

In response to the literature review findings, which revealed critical research gaps regarding the perception, awareness, and detection of deepfake technologies in academic settings, two empirical studies were designed within the framework of Design Science Research (DSR) [HC10]. These studies aimed to examine how manipulated audiovisual content is perceived and assessed by students in realistic academic scenarios.

To explore these dimensions, two independent but complementary surveys were conducted—one based on a fabricated audio message, and the other involving a deepfake video. The empirical data collected was intended to both validate the conceptual gaps identified in the literature and generate actionable insights into human vulnerabilities in educational environments.

2.6 Voice-Based Survey

The first empirical study examined how students respond to manipulated audio message in an academic setting. A fabricated voice message was created to simulate a realistic university-related scenario. To capture both emotional and critical responses, a questionnaire was administered immediately after the audio playback.

The survey was conducted in two different academic contexts: one in a third-semester *Software Engineering* course, selected due to its technical focus and high cognitive demands, and another with fourth-semester students from a less technical course, where academic pressure was assumed to be lower. The Software Engineering group was chosen for their advanced technical background, which was expected to influence their sensitivity to digital manipulation, while also allowing for analysis of how cognitive load affects perception and critical judgment.

Participants were informed beforehand that the message was artificially generated. This ensured ethical transparency while still enabling authentic responses to a plausible, yet fabricated, audio scenario.

2.7 Deepfake Video Survey

The second empirical study explored how students respond to manipulated video content in an academic context. A short video message was created using the real-time deepfake tool *DeepLiveCam*, convincingly impersonating a university professor. The aim was to simulate a realistic communication scenario involving audiovisual deception.

The video was shown to first-semester Business Informatics students at Heilbronn University. This group was chosen for their early stage in academic training and diverse backgrounds, allowing insights into initial user sensitivity toward deepfake technologies. To ensure spontaneous reactions, participants were not informed in advance that the video was manipulated.

After viewing, an anonymous questionnaire captured students' responses across three key dimensions: perceived realism, detectability of manipulation, and overall authenticity. This approach enabled a focused assessment of how convincingly low-threshold deepfake tools can imitate identity and how users perceive such content without prior warning.

2.8 Integrated Analysis

By combining these two distinct but complementary methods, a voice-based survey and a deepfake video experiment, the study was able to capture a broad spectrum of student behaviors, perceptions, and detection capabilities. The audio-based component explored reactions to manipulated content within a known context, integrating prior knowledge and psychological pressure factors, while the video-based experiment provided insight into spontaneous detection capabilities in a realistic, unsuspecting scenario.

Together, the studies offer a nuanced understanding of how individuals in academic contexts perceive and respond to digitally manipulated content. They directly address the literature-identified challenges of perception, awareness, and detection, and highlight the need for improved educational strategies and technological countermeasures in the face of increasingly accessible and convincing deepfake technologies.

3 Analysis & Results

3.1 AI Voice Generation: Comparative Analysis

As part of this study, several publicly available AI voice generation tools were evaluated. The selection of tools was based on the fact that they were publicly available at the time of the survey. The aim is to assess the capabilities of these applications for impersonating individuals.

Evaluation of the tools was based on key criteria, including voice quality, language support, customization options, usability, and distinctive features. The comparative analysis revealed notable differences in performance and user experience across platforms.

ElevenLabs stood out with excellent voice quality and support for 29 languages. Its advanced features, such as instant and professional-grade voice cloning, make it particularly suitable for high-quality, professional use cases. The platform also offers an intuitive user interface, ensuring ease of adoption even for non-experts. Based on these strengths, ElevenLabs was selected for the implementation in this study. Its combination of high-quality output, broad language support, and ease of use provided the most balanced and efficient solution for the given requirements.

Murf AI provided very good voice quality and supports over 20 languages. It allows for a high degree of customization, including control over speaking style and expression, making it well-suited for dynamic content creation in diverse contexts.

LOVO AI offers similarly high voice quality with support for more than 100 languages. Its wide range of voice styles and real-time generation capabilities enhance its flexibility, while its user-friendly design contributes to broad accessibility.

Play.ht supports over 130 languages and delivers good, lifelike voice output. While it offers extensive application potential due to its linguistic diversity, its usability is moderate, requiring users to invest some time in familiarization.

Speechify is optimized for social media and content creation, offering natural-sounding voices and a high level of user-friendliness. Although its voice quality is rated as good rather than excellent, it remains a strong option for scenarios that prioritize speed and simplicity.

Overall, the evaluation demonstrates that AI voice generation technologies have become increasingly sophisticated and accessible. These tools enable the creation of realistic voice content without requiring extensive technical expertise. However, their growing capabilities also raise important ethical and security concerns, particularly regarding potential misuse for impersonation and manipulation, which must be considered in future research and regulation.

3.2 Deepfake Video Generation: Comparative Analysis

In the context of this study, various real-time deepfake video tools freely available online. Just as with the voice applications, the selection of tools here was also based on the fact that they were publicly available at the time of the survey. The aim is to assess the capabilities of these applications for impersonating individuals.

The applicationswere systematically evaluated with regard to their usability, realism, system requirements, and unique features. The comparative analysis revealed notable differences in performance, technical prerequisites, and potential use cases.

DeepLiveCam an open-source face-swapping tool, stood out for its lightweight design and minimal system requirements. It produces convincing results using only a single reference image and simulates a live video stream in real-time. Based on these advantages, DeepLiveCam was selected for use in this study. Its ease of use, combined with realistic output and low hardware demands, made it ideal for demonstrating deepfake capabilities in an educational and controlled research setting without requiring extensive resources or technical knowledge.

DeepFaceLive enabled smooth real-time face-swapping with high visual fidelity. The setup process was simple, making the tool easily accessible. Its convincing results, especially in live video scenarios, highlight its suitability for realistic identity manipulation.

FaceFusion, a GAN-based deepfake application, delivered high-quality facial transformations. However, it required significantly more computing power and fine-tuning efforts to achieve optimal results, making it more appropriate for advanced users seeking precise control and output quality.

Xpression Camera allowed for quick avatar-based manipulations in video conferencing settings. While it provided fast results and ease of use, the generated videos appeared less realistic compared to other tools, limiting its effectiveness in scenarios requiring high visual authenticity.

Overall, the evaluation showed that creating highly realistic deepfake videos is surprisingly easy with freely available tools, even without advanced knowledge. This growing accessibility underlines the need to address the ethical and security implications of such technologies, especially with regard to potential misuse for deception or manipulation in academic and public discourse.

3.3 Experimental Study: Voice phishing

As AI-driven voice technologies have become increasingly advanced, the ability to evaluate the authenticity of spoken messages plays a crucial role in protecting against manipulation and voice-based phishing attacks (vishing). This experiment investigated participants' ability to assess the credibility of audio messages and identify signs of

manipulation, highlighting vulnerabilities in digital communication and the importance of human awareness.

Demographic Overview: the audio experiment was conducted with 56 Business Informatics students from Heilbronn University aged between 18–24 years. Some participants had diverse professional backgrounds, including IT consulting, retail, and industrial work, offering a broad perspective on digital audio communication and trust evaluation.

Experiment Setup: the audio message featured a person delivering information and tasks to the students while using AI-based voice conversion technology to imitate the voice of a fellow student. This "voice conversion" scenario was designed to evaluate how effectively participants could detect manipulated audio content and assess its credibility.

Findings

1. **Credibility of the Audio Message**

    - *High Credibility Ratings*: 91% (51 of 56 participants) found the audio message credible, while only 9% (5 of 56) perceived it as untrustworthy.
    - *Factors Supporting Credibility*: Participants based their trust primarily on the logical structure of the message (55%, 31 of 56), convincing speech style (50%, 28), familiarity of the voice (43%, 24), lack of manipulation indicators (41%, 23), a genuine tone (32%, 18), and high audio quality (11%, 6).

2. **Indicators of Suspicion**

    - *Reasons for Distrust*: A minority of participants expressed doubts due to artificial-sounding voice elements (3 mentions), suspicious content (1), manipulation cues (2), and a staged tone (1). These subtle indicators were enough to trigger skepticism in some cases.

3. **Awareness of AI-Generated Content**

    - Some participants noted unusual characteristics such as “artificial-sounding voices” or signs of manipulation, yet most admitted they would not have suspected fraud without prior knowledge of potential manipulation.

4. **Suggestions for Protection**

- To counter the risk of voice phishing, participants recommended tools for verifying message authenticity (mentioned by over 30), public awareness campaigns (20 mentions), practical detection guides (15), and educational content distributed via social media platforms. Interestingly, background noise was perceived as increasing the authenticity of the audio.

3.4 Experimental Study: User Perception of Deepfakes

This experiment analyzed participants' ability to assess the credibility of video messages and detect manipulation, focusing on vulnerabilities in digital communication and the impact of visual deepfake technologies on trust.

Demographic Overview: the video experiment was conducted with 34 first-semester Business Informatics students (aged between 18–25 years) from Heilbronn University. Some participants had diverse professional backgrounds, including IT consulting, retail, and industrial work, offering a broad perspective on digital video communication and credibility assessment.

Experiment Setup: the video featured a person delivering information and tasks to the students while using AI-generated face-swapping technology to mimic the appearance of their professor. This "face-swapping" scenario aimed to evaluate how effectively participants could recognize manipulated visual content and assess its credibility. The video was created using *DeepLiveCam*, a publicly available face-swapping tool.

Findings

- *Detection Rates*: 74% (25 out of 34 participants) did not recognize that the video was manipulated before being informed.
- *Perceived Realism*: On average, participants rated the video as highly convincing, with a mean score of 4.3 out of 5 for authenticity. (n = 34)
- *Noted Irregularities*: The few who suspected manipulation mentioned cues such as unnatural eye movements, slight lip-sync delays, and low video quality.
- *Open Feedback*: Suggestions for improvement included better resolution, more natural facial expressions, and improved synchronization between audio and facial movement.

These results indicate that even simple, publicly accessible deepfake tools can produce highly convincing results, challenging participants' ability to detect manipulation.
Notably, even participants familiar with the impersonated individual largely failed to notice the deception. This suggests that visual and behavioral cues such as prosody, gestures, and facial expressions, though highly individual, played only a minor role in the recognition process. The findings highlight the strong influence of cognitive biases

like overconfidence and confirmation bias and emphasize the importance of the human factor in the detection process. Therefore, reliance solely on technical detection measures may be insufficient; awareness and media literacy must also play a key role in addressing the threats posed by deepfake technologies.

4 Conclusion

This study explored audiovisual phishing, focusing on AI-driven deepfakes in videoconferencing. These technologies exploit trust in audiovisual communication by realistically imitating voices, facial expressions, and behaviors, posing significant threats such as fraud, disinformation, and social engineering.

Experimental results showed that over 70% of participants failed to detect manipulated content, even when familiar with the impersonated individuals. This is particularly concerning in trusted settings such as corporate or educational environments.

Effective real-time detection tools are essential but still face considerable challenges, especially in multilingual and live scenarios. These tools must be scalable, user-friendly, and easily integrable into common communication platforms. Current research highlights persistent gaps in robust, cross-platform detection capabilities [ML21] and emphasizes the need for interdisciplinary strategies to counter the growing threat of deepfakes [CC19].

Beyond technical solutions, raising public awareness and offering training are crucial. Understanding why people fail to recognize deepfakes can help design more effective educational programs. Organizations should implement clear verification protocols, while cross-sector collaboration is necessary to build comprehensive defense strategies.

Finally, legal frameworks must evolve to ensure accountability, ethical AI use, and privacy protection. Interdisciplinary cooperation, combining technical, legal, and educational perspectives, is essential to address the complex risks posed by deepfakes, while also exploring their potential positive applications in fields such as education, healthcare, and digital communication.

Bibliography

[Ah22] Ahmed, S. R.; Sonuç, E.; Duru, A. D.; Ahmed, M. R.: Analysis Survey on Deepfake Detection and Recognition with Convolutional Neural Networks. 2022.

[APA24] Akhtar, Z.; Pendyala, T.L.; Athmakuri, V.S.: Video and Audio Deepfake Datasets and Open Issues in Deepfake Technology. 2024.

[CC19] Chesney, R.; Citron, D.: Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107(6), pp. 1753–1819, 2019.

[Co19] Collins, A.: Forged Authenticity: Governing Deepfake Risks. 2019.

[Fe06] Fettke, P.: State-of-the-Art des State-of-the-Art. In: Wirtschaftsinformatik, vol. 48, pp. 257–266, 2006.

[GN25] Galyashina, E. I.; Nikishin, V. D.: AI Generated Fake Audio as a New Threat to Information Security: Legal and Forensic Aspects. Department of Forensic Expertise, Kutafin Moscow State Law University, 2025.

[GEF22] Groh, M.; Epstein, Z.; Firestone, C.; Picard, R.: Deepfake Detection by Human Crowds, Machines, and Machine-informed Crowds. Proceedings of the National Academy of Sciences Jan 2022, 119 (1). https://doi.org/10.1073/pnas.2110013119

[HC10] Hevner, A.; Chatterjee, S.: Design Science Research in Information Systems. In: Design Research in Information Systems: Theory and Practice, pp. 9–22, 2010.

[JS21] Jeong, H.; Sumi, K.: Frontiers of Computer Vision. 2021.

[KJW23] Khanjani, Z.; Watson, G.; Janeja, V. P.: Audio Deepfakes: A Survey. Frontiers in Big Data 5:1001063, 2023.

[La23] Laas, O.: Deepfakes and Trust in Technology. 2023.

[MDW25] Müller, N. M.; Dieckmann, F.; Williams, J.: Attacker Attribution of Audio Deepfakes. In (Ferrari, E., ed.): Proc. Int. Conf. on Cybersecurity and Privacy, 2025. Fraunhofer AISEC, Technical University of Munich, University of Southampton.

[ML21] Mirsky, Y.; Lee, W.: The Creation and Detection of Deepfakes: A Survey. ACM Computing Surveys, 54(1), 1–41, 2021.

[HM20] Menczer, F.; Hills, T.: The Attention Economy. Scientific American, 323(6):54–61, Dec, 2020.

[SAK24] Sharma, S.K.; AlEnizi, A.; Kumar, M.; Alfarraj, O.; Alowaidi, M.: Detection of Real-Time Deepfakes and Face Forgery in Video Conferencing Employing Generative Adversarial Networks. 2024.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: MIS Quarterly, vol. 26, no. 2, pp. xiii–xxiii, 2002.

[Yi23] Yi, J.; Wang, C.; Tao, J.; Zhang, X.; Zhang, C. Y.; Zhao, Y.: Audio Deepfake Detection: A Survey. Journal of LaTeX Class Files, Vol. 14, No. 8, August 2023.

Supporting trust decisions of holders and verifiers: a unified model of trust management in SSI

Eric Biele [1], Daniel Richter [1], Ricardo Bochnia [1], and Jürgen Anke [1]

Abstract: Credential exchange between holders and verifiers using Self-Sovereign Identity (SSI) requires mutual trust. We investigate which acceptance criteria are necessary for both parties to ensure trust when requesting proofs and verifying presentations, highlighting how machine-readable governance can support both in their trust decisions. Our analysis of trust management approaches reveals a fragmented landscape with limited interoperability. We present a conceptual trust management model with four key segments: organizational trust, governance, authoritative data, and technical components. This model provides a unified perspective on establishing trust through interoperable, machine-readable processes, offering a shared understanding and a common language for novel trust management approaches. Finally, we discuss how bridging mechanisms like the Trust Registry Query Protocol (TRQP) could enable interoperability across different trust management systems.

Keywords: Self-sovereign identity, Trust management, Machine-readable governance, Organizational trust, Trust graph, Trust registry

1 Introduction

The paradigm of self-sovereign identity (SSI) is based on the exchange of signed data, typically issued as verifiable credentials (VC). Holders of VCs can create proofs for third parties (verifiers) in the form of verifiable presentations (VP). In the VC trust model, there is no direct connection between the issuer and the verifier of a VC [SLC22]. However, verifiers must decide whether they can accept claims made by an issuer about the holder of a VC. Similarly, holders must ensure that only trustworthy and authorized verifiers process their data.

Trust in SSI has two dimensions: technical trust and organizational trust. *Technical trust* refers to IT security goals in data exchange, such as integrity, authenticity, and confidentiality. Improving technological components, including the development of credential formats and communication protocols, contributes to the goal of establishing technical trust [JCK22]. But solely relying on technical trust is insufficient for the trust decision stated above [RJV21].

An illustrative example is a university that issues a diploma to a graduate in the form of a VC. In the context of an application process, a company must verify the authenticity and

[1] HTW Dresden, Digital Service Systems Group, Friedrich-List-Platz 1, 01069 Dresden, Germany,
eric.biele@htw-dresden.de, https://orcid.org/0009-0009-8819-0614;
daniel.richter@htw-dresden.de, https://orcid.org/0000-0003-1549-5467;
ricardo.bochnia@htw-dresden.de, https://orcid.org/0009-0007-4317-1810;
juergen.anke@htw-dresden.de, https://orcid.org/0000-0002-9324-9387

doi:10.18420/OID2025_08

legitimacy of the VP received and decide whether to accept it. However, the information contained in a VP is insufficient to make this decision. Therefore, trust frameworks are necessary to establish organizational trust [Wi23, p. 67].

Organizational trust refers to the authority and trustworthiness of the actors involved in an SSI-based ecosystem. Trust frameworks augment the VC trust model by defining ecosystem governance through binding rules, guidelines, and specifications. Trust frameworks provide the information on which holders and verifiers base their trust [Re21]. To enable informed and timely decisions, at least some rules defined by trust frameworks must be available in a machine-readable format and accessible via well-defined interfaces and protocols.

This paper focuses on mechanisms that support the establishment of trust during the proof request and the verification of VPs between the holder and the verifier. We assume a typical setup where the holder is a natural person and the verifier is an organization. Other configurations are possible, but not in the scope of this paper. Furthermore, we assume that verifiers aim to verify VPs automatically. When an organization acts as a verifier, it typically utilizes software systems for workflow automation. In contrast, it is up to a human holder to decide whether or not to share the relevant data with a verifier. However, it is desirable to support natural persons in this decision. This results in the necessity to provide certain information of a trust framework in machine-readable form to enable well-founded (i.e., supported by additional information) or even automated trust decisions as described above.

As a result, various approaches have been proposed to enable organizational trust decisions during proof requests and verification of VPs. However, there is currently no unified overview or connection between these as they use different technologies and terminologies, making it difficult to compare and evaluate them. We aim to provide a more unified conceptualization of SSI trust management by gathering common concepts and relationships. Therefore, we pose the following research questions:

- RQ1: How do practical approaches to trust management for SSI-based proof requests and verifications of VPs contribute to trust decisions for holders and verifiers?
- RQ2: What are common concepts and relationships that describe existing trust management approaches during the proof request and verification of VPs between the holder and verifier?

The article is structured as follows: First, we introduce the context of requesting proofs and verifying VPs and identify the trust decisions and acceptance criteria for both holders and verifiers. We then review relevant trust management approaches, illustrating the fragmented landscape and interoperability challenges. Next, we propose a conceptual model that comprises four key segments: organizational trust, governance, authoritative data, and technical components. Following that, we discuss the applicability and limitations of our model, highlighting how bridging mechanisms like the Trust Registry Query Protocol can enhance interoperability. Finally, we present our conclusions and identify promising directions for future research.

2 Methodology

Trust decisions relevant to the holder and the verifier were first identified and underpinned with acceptance criteria derived from the existing literature on trust and SSI. Afterward, we reviewed the literature, specifications, and practice-oriented approaches to identify relevant sources for trust establishment approaches. On this basis, we selected theoretical and practical approaches that deal with the provision for well-founded or automated trust decisions in machine-readable form. While this selection is comprehensive, it is not exhaustive.

The selected approaches were then analyzed for key features and recurring concepts. The findings formed the basis for developing a conceptual trust management model using a heuristic search and synthesis approach [GM14]. A continuous alternation between problem structuring and model development characterized it. First, the problem was defined, the most important requirements for the model were determined, and the findings from the analysis of the approaches were consolidated into a preliminary model structure. Through several iterative cycles, insights from model evaluations and approach analysis were continuously synthesized, enabling refinement of the problem structure and model adaptations. This iterative process was continued until a stable model was achieved.

3 Trust decisions of verifier and holder

To determine the necessary information for establishing trust we analyze the trust decisions made during the proof request and the verification of VPs between the interacting parties. Tab. 1 provides an overview of trust decisions and associated criteria.

#	**Holder trust decision: Acceptance of the proof request**
CH1	Authorization of verifier for proof request
CH2	Identification of verifier
CH3	Legitimacy of request purpose in given context
#	**Verifier trust decision: Acceptance of the VP**
CV1	Authorization of issuer for issuing VC
CV2	Identification of issuer
CV3	Validation of VP
CV4	Approval of VC schema

Tab. 1: Trust decisions and related acceptance criteria

3.1 Holder trust decision

Typically, the verifier initiates the interaction by sending a proof request to the holder. This leads to the first trust decision on the holder's side: accepting or rejecting the request. This decision depends on various criteria.

CH1: Is the verification of the VCs used in the VP restricted to a predefined set of actors, and if so, can the requesting verifier present a valid authorization? Such restrictions are common when requesting proofs that contain sensitive personally identifiable information [Re21]. For example, German eID data can only be requested by actors with a valid authorization certificate (German: "Berechtigungszertifikat"). In relevant cases, criterion CH1 can be checked automatically by the holder's wallet as authorizations are typically machine-readable.

CH2: Beyond authorization, what is the identity of the verifier? Whether the requesting verifier needs authorization or not, the holder can base their trust decision on the existence of the verifier's clear identity [Sz03]. This criterion can also be checked automatically [BRA24], and its acceptance is based, e.g., on the holder's experience with the verifier or other information about the relationship.

CH3: Is the processing purpose of the verifier's request legitimate in the given context? Finally, the holder can base their decision to accept the verifier's request on an assessment of the *context* the requested information exchange is embedded into [Wi23, p. 68]. For example, the sharing of personal identification data is appropriate for the contracting of a telecommunications service. However, it is probably not appropriate for the use of the service's customer service, as a customer number, for example, is also sufficient. Currently, there is no standardized mapping of process context to appropriate data requests, leaving such decisions entirely to holders.

The relevance of this trust decision is dependent on the ecosystem governance. For valuable VCs, such as personal identification data, the presented acceptance criteria are of greater significance than for lower-risk VCs. Regardless of whether all criteria can be fully checked, the holder will make the final decision. The criteria serve less as strict exclusion factors, but rather as a support to make an informed decision. As this is a non-automated trust decision, the relevant information should be provided to the holder in a human-readable, easily understandable form [EKA23; Te23].

3.2 Verifier trust decision

With the acceptance of the proof request, the holder presents the requested data as a VP. In this case, the verifier must also make a trust decision on whether to accept the presented VP. This decision can be checked based on the following criteria:

CV1: Is the issuer authorized to issue VCs of the requested type? Issuers must be authorized to issue specific credential types to prevent unauthorized actors from issuing VCs [Re21; CB21]. This prevents arbitrary claims from being made about subjects.

CV2: Beyond authorization, what is the identity of the issuer? There are cases in which it is necessary to check who the issuer of a VC is to establish trust [Sz03], e.g., it may be necessary to identify the issuer of an employment reference to find out which employer issued it and thus make a well-informed evaluation of the information.

CV3: Is the VP valid, i.e., are the embedded VCs neither revoked nor expired? The validity of the VCs used in VPs is essential for their usage [CB21]. If a VC is revoked or expired, e.g., due to errors or time limits, it loses its validity and becomes unusable.

CV4: Is the VC schema approved in the respective context at the time of issuance? There must be a source of information indicating that a VC schema belongs to an ecosystem to put the data in context [BRA24]. This is necessary to prevent duplicate schemas from being created fraudulently.

If the verifier wants to make an automated decision, it must be possible to answer at least the acceptance criteria CV1 or CV2, as well as CV3 and CV4 in a machine-readable form. The application of CV1 and CV2 depends on the requirements of the verification policies of the Credential Governance Framework [RPA23].

4 Approaches to trust establishment when using verifiable credentials

Various initiatives have developed approaches to providing the required information in machine-readable form to support the trust decisions stated above. This section briefly presents the following five approaches: Credential Trust Establishment, Trust Management Infrastructure, European Blockchain Services Infrastructure VC Framework, OpenID Federation Wallet Architecture, and the eIDAS Architecture Reference Framework.

Credential Trust Establishment (CTE) is a document format proposed by the Decentralized Identity Foundation (DIF) that is based on the principles of the specification of Trust Establishment [CS23]. It defines the VC schemas valid in the ecosystem and specifies roles that include the rights to issue and verify specific VC types according to these schemas, as well as the entity that has granted these rights. Additionally, these roles are assigned to the participants listed in the document. Therefore, the CTE is designed to manage the above information with additional metadata, such as the version number, in a single downloadable document. However, it is not designed to manage revoked VCs with a CTE Document. The specification only states that the document must be made available via a URI and signed by the CTE publisher, ensuring that integrity and authenticity are maintained. How the document is replicated, updated, or saved is not defined. It should also be mentioned that this approach supports trusted governance: the CTE enables the mutual recognition of other CTE documents and thus creates the basis for trust across ecosystem boundaries [Na24].

The **Trust Management Infrastructure** (TRAIN) offers a trust infrastructure based on the Domain Name System (DNS). It builds on the results of the EU-funded projects LIGHTest, NGI ESSIF-Lab TRAIN, and NGI Atlantic "Next Generation SSI Standards" and addresses the challenges of providing information about trusted entities in an ecosystem, including associated policies and automated decision-making [KR21] or decision support [Ch23]. This works by anchoring pointers in the DNS by the owners of corresponding DNS entries that refer to trust lists that comply with the ETSI TS 119 612 standard [JCK22]. Thus,

owners of DNS records can act as governance authorities, which leads to a high degree of decentralization. They can reference the trust lists they manage and publish, which contain issuers, references to their associated VC schemas, and verifiers within the respective trust framework. A policy server, a public repository for verifiers to publish their requests, regulates the requested data [Ch23]. The URLs of these are also stored in trust lists. In addition, TRAIN provides open-source components that enable automated resolution of these references and retrieve the relevant trust information [Ch23]. Due to the high flexibility of TRAIN, other trust frameworks can be referenced via additional pointers in DNS entries and therefore allow trust across ecosystem boundaries as well, similar to CTE. However, TRAIN's approach does not include the revocation of VCs.

The **European Blockchain Services Infrastructure (EBSI) VC Framework** provides an infrastructure for the trustworthy handling of VCs. It is based on a trust chain which in general is a sequence of certificates that can be validated up to a trust anchor that is considered trustworthy [Sm21]. At EBSI, it is used to accredit trusted entities such as issuers. At this point, it should be noted that the EBSI VC Framework distinguishes between accreditations and attestations [Eu24d]. Accreditations qualify entities for issuing accreditations or attestations, while attestations are domain-specific VCs. Trusted issuers and associated schemas are stored in registries on the EBSI ledger [Eu24c]. For the authorization of verifiers to request a specific VC, the EBSI VC Framework provides that trusted issuers can issue a verifier authorization in the form of a VC, which in turn can be checked by verifying the chain of trust [Eu24e]. These authorizations can be used in presentation policies defined by the issuers to limit the requests of the issued VC. In addition, there may be request policies in which verifiers specify which data they request and for what purpose [Eu24e]. To check whether the VC is valid, the status can be determined in various ways based on business requirements, such as the issuer hosting the status information and the verifier retrieving it as needed [Eu24b].

The **OpenID Federation Wallet Architecture** offers an approach for a federated trust infrastructure and is published by the OpenID Foundation. It also uses a trust chain to establish trust and has similarities with the Public Key Infrastructure [Wi23, pp. 120–123]. Within this chain, issuers and verifiers that are considered trustworthy are named Leaf Entities (LEs) and are authorized by superordinate entities (such as Intermediates or Trust Anchors). They form the end of the trust chain. Using metadata and metadata policies in subordinate statements makes it possible to apply policies to immediately subordinate entities or to all entities within the trust chain. These mechanisms enable the authorization of issuers and verifiers concerning the issuance or request of VCs [De24]. However, the authentication and authorization of a verifier can also be carried out via an attestation [Te25]. Furthermore, the OpenID for Verifiable Credential Issuance specification states that VC issuers manage revocation. However, no concrete specification is given [LYL24]. It is also not specified how VC schemas are managed.

The **Architecture Reference Framework** (ARF) proposes the architecture of the European Digital Identity Wallet (EUDI wallet) ecosystem, which is managed by the European Digital

Identity Cooperation Group (EDICG). A central component of this framework is the trust model, which defines the mechanisms for building trust within the ecosystem. It differentiates the handling of VCs depending on the different categories existing in the eIDAS context. For Personal Identification Data (PID), Qualified Electronic Attestations of Attributes (QEAAs), and VCs from the public sector (Pub-EAAs), i.e., highly regulated VCs, the trust model provides for issuers to be included in trusted lists. These lists are managed and provided by Trusted List Providers (TLPs). Non-qualified electronic attestations of attributes (EAAs) do not fall within this scope as they may adopt alternative trust models [Eu25]. To authenticate issuers and verifiers, the ARF proposes the use of access certificates issued by Certification Authorities (CAs) anchored in trusted lists [Eu25]. To restrict verifiers, registration certificates are issued by a registrar, which specify which attributes they intend to request [Eu25]. Attestation schemas are defined by Attestation Schema Providers, such as the European Commission or other organizations, and are published in Attestation Rulebooks, which also contain mechanisms for authentication and authorization for specific VCs if required [Eu25]. To verify the status of a VC, issuers maintain a status list that allows for continuous validation [Eu25].

The approaches share the same goal of establishing trust between the interacting parties, but differ significantly in the use of the underlying technologies and the scope they address, as well as in the stage of development. The ARF, e.g., is currently still under development and is intended to outline the architectural design of the EUDI wallet ecosystem and comply with the eIDAS Regulation [Eu25], while EBSI's scope is the European public sector [Eu24a].

5 Conceptual model of trust management in SSI

The analyzed approaches show that providing the information required for trust decisions when requesting proofs and verifying VPs within an ecosystem is possible in machine-readable form and thus enabling intra-ecosystem trust. However, different technologies are used, some of which focus on different aspects, such as the discoverability of this information. As neither a specific approach has achieved a dominant position nor uniform standards have been established, it can be assumed that multiple approaches will co-exist in the future. This assumption is further supported by the fact that SSI ecosystems have to be established in different regulatory environments with specific requirements. Accordingly, interoperability between these approaches must be ensured.

To accomplish this, the Trust Registry Query Protocol (TRQP) was proposed by the Trust over IP (ToIP) Foundation [OD24]. This protocol allows read-only access to the required data but does not support operations for creating, updating, or deleting information, and does not make any decisions. The main advantage is that interoperability can be established between the analyzed approaches, thus enabling trust across ecosystem boundaries, i.e., inter-ecosystem trust. Another advantage of this approach is the possibility of using existing systems via a bridge so that existing trust management infrastructures are retained.

Using the trust decisions and associated acceptance criteria, the results of the analysis, and the aforementioned proposal for bringing these approaches together, we present a conceptual model (Fig. 1). The purpose is to understand the concepts and relationships that describe existing trust management approaches. It provides a common terminological foundation and a structured representation of a trust infrastructure within the scenario under consideration.

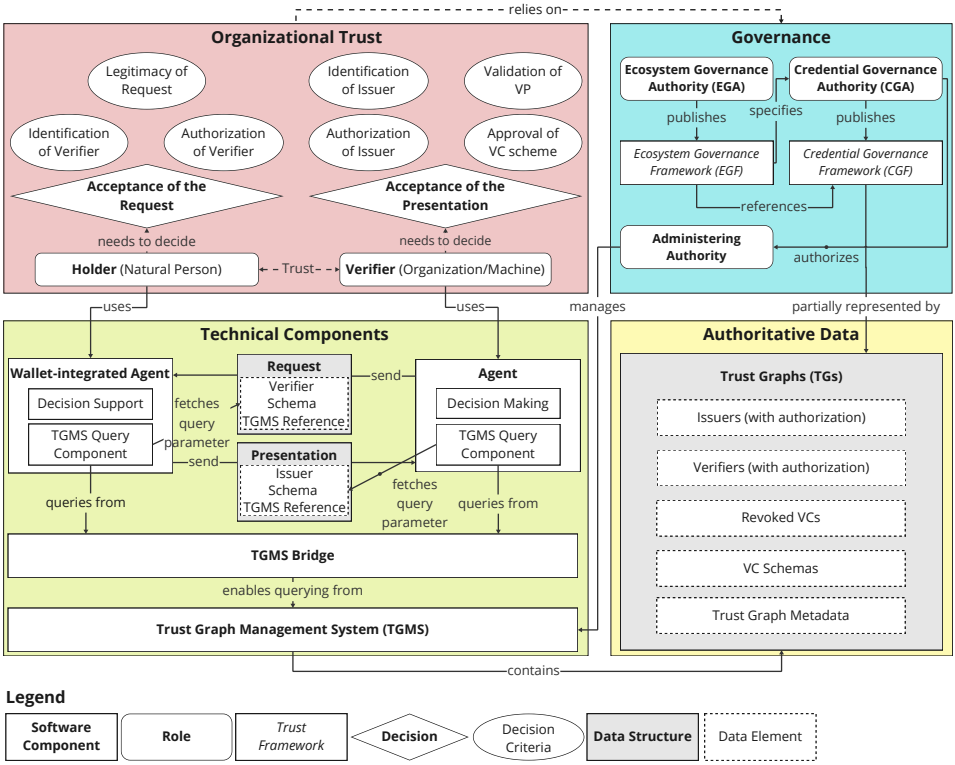


Fig. 1: Conceptual model of trust management in SSI

Several assumptions were made for the proposed model. First, it is assumed that the VC was issued to the holder by an issuer in a previous, independent process. Second, the security of the technical components, such as the wallet, is considered a given. Third, it is assumed that the information provided can be considered trustworthy. This implies that the model provides for the operational use of existing data and that aspects such as the onboarding of participating entities are out of scope.

With this in mind, the conceptual model was developed, comprising four segments: governance, organizational trust, authoritative data, and technical components. These segments and their specific contents are explained in detail below.

Organizational Trust: The organizational trust segment shows the two roles, holder and

verifier, which aim to establish trust when requesting proof and verifying a VP. This results in the trust decisions to be made and their associated acceptance criteria, as defined in Sect. 3. This segment thus represents not only the scenario under consideration but also the problems and requirements that arise with the effort to establish organizational trust.

Governance: The governance segment illustrates governing actors that are necessary for establishing trust between the holder and the verifier. These include the role of the Ecosystem Governing Authority (EGA), which is responsible for publishing the Ecosystem Governance Framework (EGF). The EGF serves as a governance framework for application ecosystems, wherein SSI technologies are being used and includes business, legal, and technological rules regarding this use. EGFs include specific provisions on the structure and use of VCs of a specific type, which are called Credential Governance Frameworks (CGFs) [RPA23]. They are published by Credential Governing Authorities (CGAs), whose jurisdiction is defined in the EGF. The CGAs determine the conditions under which a VC and its schema may be used and authorize an Administering Authority (AA), which is responsible for the operational management of the technical representation of the authoritative data related to the CGF. The approaches analyzed use different terminology for these roles and documents, such as TRAIN, which uses the term Trust Scheme Operator for the EGA. In this segment, we have used the terms of the ToIP Foundation as they provide precise definitions.

Authoritative Data: We define authoritative data as data that either contains information about entities or rules and objects of a governance framework and is provided by governance authorities or authorized entities. This includes information that is contained in governance frameworks (e.g., the rules and regulations that apply) but also information that arises as a result of the application of the rules (e.g., trustworthy entities). To fulfill the acceptance criteria for the proof request and verification of VPs and to support or automate the aforementioned trust decisions, some of this data must be available in machine-readable form. This includes information on verified issuers and verifiers, as well as on revoked VCs and approved VC schemas of the respective domain. Furthermore, authoritative data can support the fulfillment of the criterion of context legitimacy by, for example, providing information about the intended use of the data by a verifier. As mentioned in Sect. 4, there are significant differences in the approaches used to provide authoritative data. TRAIN, for example, uses trust lists to anchor trustworthy entities. EBSI, on the other hand, uses a mixed approach, consisting of a trust chain in the form of chained VCs for the authorization of entities while using registries for approved schemas or revoked VCs.

The model summarizes how the machine-readable authoritative data is provided as a trust graph (TG). According to the definition of the ToIP Foundation, TGs are data structures that map trust relationships between entities and can take on various forms, such as a trust chain or trust list [OD25]. Metadata, such as a version number, is also stored within it. A TG thus represents a subset of the information of a CGF, and it is possible to represent several CGFs within a single TG. This can be seen, for example, in a CTE document in which several schemas and associated issuers and verifiers are grouped together in a single file.

Technical Components: In the segment about the technical components of the model, the Trust Graph Management System (TGMS), which manages the machine-readable authoritative data in the form of TGs, is introduced first. This system is used both to store this data and to carry out operations by AAs, as well as to provide agents with access to it.

A holder uses a wallet-integrated agent to answer proof requests from a verifier and transmit VPs. To check a request according to the defined acceptance criteria, three essential parameters must be extracted from the proof request: the verifier, the VC schemas of the requested proof, and the TGMS reference. The wallet-integrated agent can retrieve the required data from the TGMS using a TGMS query component and these parameters. A decision support component within the wallet-integrated agent processes this data. It provides the holder with information within his wallet application that informs him whether the required acceptance criteria have been met. This is the case with TRAIN, for example, which warns the user in their wallet if the verifier is not trustworthy [Ch23].

In contrast, the verifier only needs an agent, as wallet functionality is not mandatory for verification. This agent also uses a TGMS query component that works the same way as the holder component but extracts the VC issuer from the holder's presentation instead of the verifier. As soon as all the relevant information from the TGMS is available, the agent can use a decision-making component to make an automated trust decision based on the acceptance criteria applied. This distinguishes it from a wallet-integrated agent, where the final decision is left to the holder.

Furthermore, the analysis of the approaches presented in Sect. 4 shows a heterogeneous technological landscape. Therefore, a TGMS Bridge is required to enable the unified querying of the relevant information from the respective TG and the associated TGMS based on the specific TGMS reference as outlined by the TRQP specification. This results in a standardization of the TGMS query component, which is independent of the underlying TGMS. Providing the bridge and thus complying with the required format by the TRQP specification is the responsibility of the implementers of the specific ecosystem.

6 Discussion

Our analysis of trust management approaches revealed a fragmented landscape characterized by different terminologies and limited interoperability. We addressed these challenges by proposing a conceptual model. Researchers and practitioners can use this conceptual framework and terminology to develop, compare, and evaluate solutions for decision support or automated trust decisions during the proof request and the verification of VPs.

6.1 Implications

While our model shows that existing approaches share a common ground, they still differ considerably in their implementations. Even if a single mechanism is used within an

identity ecosystem, accepting credentials from other ecosystems will require interoperability. Therefore, the agents must be able to obtain the authoritative data from various trust management approaches. The TRQP is a potential solution that addresses this problem by defining a unified query language and thus serves as a specification for the development of TGMS bridges. Accordingly, it is possible to extract the trust decision-relevant information from the fragmented information sources mentioned above. Consolidation is also conceivable, with only a few approaches gaining wide adoption.

Furthermore, our model underscores the need for the evolution of digital wallets and agents. Agents require decision-support or decision-making features, such as automated checks of the verifiers' authorization, to assist holders. Wallet providers will have to aim for solutions that do not require cybersecurity knowledge and can be understood by the average user.

6.2 Limitations

Our modeling approach was based on analyzing existing approaches to establish trust when using VCs. Thus, our model offers an abstract view of these approaches. However, it does not provide any additional components that are not found in any of these approaches. Since most of the mentioned approaches are not deployed in large-scale production, additional elements that are not captured by our model may be necessary.

Additionally, we only focused on the proof request and the verification of VPs and excluded the issuer. The interaction between issuer and holder faces similar trust decisions and, therefore, our model should (with minor modifications) apply to the issuance and storage of VCs. A second model, consisting of issuer and holder, would likely have significant overlap, as organizational trust depends on the same governance and technical components.

While focusing on the interaction between natural persons as holders and verifiers as organizations, the holder may be an organization, too. Depending on the scenario, the holder might want a higher degree of automation for their trust decision than a natural person. Future research could explore whether this has an impact on, e.g., the acceptance criteria (since checking for context legitimacy automatically is difficult) or the technical components. Still, a higher degree of automation could also be desirable for natural persons, in that recurring contacts are regarded as trustworthy.

6.3 Future Work

To establish organizational trust in an ecosystem, it is necessary to decide upon a trust management system. As pointed out before, the requirements of different ecosystems, e.g., health, public administration, or banking, have different conditions and requirements concerning governance and its technical support. It is, therefore, useful to analyze the ecosystems' requirements to determine which trust management is best suited to establish trust. This

will have to be accompanied by an investigation of the current trust management approach concerning characteristics, such as the topology, scalability, latency, and expressiveness of rules that might have to precede this.

Future work could extend our model by examining how enrolling actors into specific roles within an ecosystem, e.g., as trusted issuers or verifiers, could be incorporated. Some initiatives have already begun developing onboarding processes for issuers and verifiers, such as TRAIN's Trust Registries Enrollment Tool (TRET) [Jo24].

Finally, introducing process templates could be a potential approach to addressing the criterion of the legitimacy of request purpose in a given context. These templates could define a process, including its workflows and the associated data exchange. Such process templates might be subject to a dedicated lifecycle, including approval or assessment by third parties, which increases the level of automation for trust decisions.

7 Conclusion

In this paper, we investigated technical approaches for establishing organizational trust. A variety of trust management systems are currently in development to provide the information to support or even automate trust decisions for holders and verifiers. We systematize criteria that holders and verifiers can use to decide whether to accept a request or present data.

Trust management systems aim to provide the required information for such decisions by maintaining a trust graph as a machine-readable representation of relevant ecosystem governance aspects. Our findings show that existing approaches vary widely in their conceptual and technical approach to trust establishment (RQ1). They all use different technologies to achieve their goals. For instance, CTE uses special JSONs hosted on a web server or cached locally, while TRAIN relies on DNS and a trust list, whereas EBSI uses a blockchain and credential chaining.

To facilitate research and development of trust management systems, we derive a conceptual trust management model that unifies the main elements and their relationships of the analyzed trust management systems (RQ2). The model consists of four major segments: organizational trust, governance, authoritative data, and technical components. Each part is vital to the trust establishment between the holder and verifier. By offering a unified understanding and a common language for novel trust management approaches, our model benefits SSI research in discussing organizational trust establishment independent of specific technical implementation proposals. Similarly, by providing a common terminology our model supports standardization efforts and technology choice in organizational practice.

References

[BRA24] Bochnia, R.; Richter, D.; Anke, J.: Self-Sovereign Identity for Organizations: Requirements for Enterprise Software. IEEE Access 12, pp. 7637–7660, 2024, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2023.3349095.

[CB21] Chadwick, D. W.; Burnett, D.: Verifiable credentials. In (Preukschat, A.; Reed, D., eds.): Self-Sovereign Identity. Manning Publications, Shelter Island, NY, pp. 126–156, 2021, ISBN: 9781617296598.

[Ch23] Chadwick, D. W.; Kubach, M.; Sette, I.; Johnson Jeyakumar, I. H.: Establishing Trust in SSI Verifiers. In: Open Identity Summit 2023. Gesellschaft für Informatik e.V., Bonn, pp. 15–26, 2023, ISBN: 978-3-88579-729-6, DOI: 10.18420/OID2023_01.

[CS23] Caballero, J.; Schreiber, J.: Trust Establishment 1.0.x, ed. by Cohen, G.; Buchner, D.; Ebert, M.; Curren, S., 2023, URL: https://identity.foundation/trust-establishment/, visited on: 01/23/2025.

[De24] De Marco, G.; Hedberg, R.; Jones, M.; Bradley, J.: OpenID Federation Wallet Architectures 1.0, 2024, URL: https://openid.net/specs/openid-federation-wallet-1_0.html, visited on: 01/23/2025.

[EKA23] Ebert, S.; Krauß, A.-M.; Anke, J.: Towards informed choices: A decision model for adaptive warnings in self-sovereign identity. In: Mensch und Computer 2023 - Workshopband. 2023, DOI: 10.18420/muc2023-mci-ws11-322.

[Eu24a] European Blockchain Services Infrastructure: About Us, ed. by European Blockchain Services Infrastructure, 2024, URL: https://ec.europa.eu/digital-building-blocks/sites/display/EBSI/About+us, visited on: 02/21/2025.

[Eu24b] European Blockchain Services Infrastructure: Choosing a Credential Status Strategy, ed. by European Blockchain Services Infrastructure, 2024, URL: https://hub.ebsi.eu/vc-framework/credential-status-framework/strategies, visited on: 02/06/2025.

[Eu24c] European Blockchain Services Infrastructure: EBSI W3C VCs and VPs, ed. by European Blockchain Services Infrastructure, 2024, URL: https://hub.ebsi.eu/vc-framework/ebsi-w3c-vc-vp, visited on: 02/06/2025.

[Eu24d] European Blockchain Services Infrastructure: Issuer Trust Model, ed. by European Blockchain Services Infrastructure, 2024, URL: https://hub.ebsi.eu/vc-framework/trust-model/issuer-trust-model-v3, visited on: 01/23/2025.

[Eu24e] European Blockchain Services Infrastructure: Policies for Verifiable Credentials Presentation and Request, ed. by European Blockchain Services Infrastructure, 2024, URL: https://hub.ebsi.eu/vc-framework/trust-model/policies, visited on: 02/06/2025.

[Eu25] European Commission: Architecture and reference framework - EUDI Wallet, 2025, URL: https://eu-digital-identity-wallet.github.io/eudi-doc-architecture-and-reference-framework/latest/architecture-and-reference-framework-main/, visited on: 02/19/2025.

[GM14] Gregory, R. W.; Muntermann, J.: Research Note: Heuristic Theorizing: Proactively Generating Design Theories. Information Systems Research 25 (3), pp. 639–653, 2014, ISSN: 10477047, 15265536, URL: http://www.jstor.org/stable/24700315, visited on: 02/20/2025.

[JCK22] Jeyakumar, I. H. J.; Chadwick, D. W.; Kubach, M.: A novel approach to establish trust in verifiable credential issuers in Self-sovereign identity ecosystems using TRAIN. In (Roßnagel, H.; Schunck, C. H.; Mödersheim, S., eds.): Open Identity Summit 2022. Gesellschaft für Informatik e.V, Bonn, pp. 27–38, 2022, DOI: 10.18420/OID2022_02, URL: https://dl.gi.de/handle/20.500.12116/38702.

[Jo24] Johnson Jeyakumar, I. H.; Kubach, M.; Vargas, J.; Walker, J.: A Trust Registries Enrollment Tool Supporting Decentralized Ecosystem Governance: Use Case Healthcare. In: Open Identity Summit 2024. Gesellschaft für Informatik e.V., Bonn, pp. 103–114, 2024, ISBN: 978-3-88579-744-9, DOI: 10.18420/OID2024_09.

[KR21] Kubach, M.; Roßnagel, A.: A lightweight trust management infrastructure for self-sovereign identity. In (Roßnagel, H.; Schunck, C. H.; Mödersheim, S., eds.): Open Identity Summit 2021, Fachtagung vom 01.-02. Juni 2021 Copenhagen, Denmark. 312, GI-Edition. Proceedings 312, Gesellschaft für Informatik, Bonn, pp. 155–165, 2021, ISBN: 9783885797067.

[LYL24] Lodderstedt, T.; Yasuda, K.; Looker, T.: OpenID for Verifiable Credential Issuance, 2024, URL: https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0.html, visited on: 02/07/2025.

[Na24] Nazarenko, S.: Credential Trust Establishment 1.0, ed. by Curren, S.; Ebert, M., 2024, URL: https://identity.foundation/credential-trust-establishment/, visited on: 01/23/2025.

[OD24] O'Donnell, D.; Antti; Andor; Latour, J.; Martin, C.: ToIP Trust Registry Query Protocol v1.0 Specification, ed. by Trust Over IP Foundation, 2024, URL: https://trustoverip.github.io/tswg-trust-registry-protocol/, visited on: 02/11/2025.

[OD25] O'Donnell, D.; Griffin, K.; Dwarshuis, K.; Thomson, N.; Hickman, N.; Joosten, R.: ToIP Glossary, ed. by Reed, D.; van Cann, H., 2025, URL: https://glossary.trustoverip.org/, visited on: 02/20/2025.

[Re21] Reed, D.: SSI governance frameworks. In (Preukschat, A.; Reed, D., eds.): Self-Sovereign Identity. Manning Publications, Shelter Island, NY, pp. 248–273, 2021, ISBN: 9781617296598.

[RJV21] Reed, D.; Joosten, R.; Van Deventer, O.: The basic building blocks of SSI. In (Preukschat, A.; Reed, D., eds.): Self-Sovereign Identity. Manning Publications, Shelter Island, NY, pp. 22–38, 2021, ISBN: 9781617296598.

[RPA23] Richter, D.; Praas, C. R.; Anke, J.: Beyond Paper and Plastic: A Meta-Model for Credential Use and Governance. In: ECIS 2023 Research Papers. 2023, URL: https://aisel.aisnet.org/ecis2023_rp/371/.

[SLC22] Sporny, M.; Longley, D.; Chadwick, D. W.: Verifiable Credentials Data Model v1.1: W3C Recommendation, ed. by Sporny, M.; Noble, G.; Longley, D.; Burnett, D.; Zundel, B.; den Hartog, K., 2022, URL: https://www.w3.org/TR/vc-data-model/, visited on: 02/14/2025.

[Sm21] Smith, S.: Decrentralized key management. In (Preukschat, A.; Reed, D., eds.): Self-Sovereign Identity. Manning Publications, Shelter Island, NY, pp. 220–247, 2021, ISBN: 9781617296598.

[Sz03] Sztompka, P.: Foundations of trust. In (Sztompka, P., ed.): Trust: A Sociological Theory. Cambridge University Press (CUP), New York, pp. 69–102, 2003, ISBN: 0521598508.

[Te23] Teuschel, M.; Pöhn, D.; Grabatin, M.; Dietz, F.; Hommel, W.; Alt, F.: 'Don't Annoy Me With Privacy Decisions!' – Designing Privacy-Preserving User Interfaces for SSI Wallets on Smartphones. IEEE Access 11, pp. 131814–131835, 2023, DOI: 10.1109/ACCESS.2023.3334908.

[Te25] Terbu, O.; Lodderstedt, T.; Yasuda, K.; Looker, T.: OpenID for Verifiable Presentations, 2025, URL: https://openid.net/specs/openid-4-verifiable-presentations-1_0.html, visited on: 02/07/2025.

[Wi23] Windley, P.: Learning Digital Identity: Design, Deploy, and Manage Identity Architectures. O'Reilly Media, 2023, ISBN: 9781098117658.

Towards the European Business Wallet

Tina Hühnlein [1], Detlef Hühnlein [1], Steffen Schwalm[2], and Carsten Stöcker[3]

Abstract: Against the background of the recent announcement of the European Commission to create a "European Business Wallet" (EUBW), which aims at facilitating trustworthy electronic business and government processes, the present paper recalls the relevant regulatory and technical background for the EUBW, contributes to the necessary standardisation process for legal person identities and outlines how the EUBW might in the end look like from a technical perspective.

Keywords: EUDI, EUDIW, EUBW, eIDAS, Wallet, Verifiable Credential (VC)

1 Introduction

On January 29th 2025 Ursula von der Leyen presented the European Commission's plan to boost Europe's competitiveness and turn it into a place where future technologies, services, and clean products are invented, manufactured, and put on the market. The presented "Competitiveness Compass" [EC25a; EC25b] serves as roadmap and sets out the course the Commission will follow in the next five years. Among the announced "Flagship Actions enablers" for the year 2025 is the "European Business Wallet" (EUBW), which will build upon the amended eIDAS-Regulation (EU) No. 910/2014 and is envisioned to be "the cornerstone of doing business simply and digitally in the EU, providing a seamless environment for companies to interact with all public administrations."

While the closely related "European Digital Identity Wallet" (EUDIW) is defined in Art. 3 (42) of (EU) No. 910/2014 to be "an electronic identification means which allows the user to securely store, manage and validate person identification data and electronic attestations of attributes for the purpose of providing them to relying parties and other users of European Digital Identity Wallets, and to sign by means of qualified electronic signatures or to seal by means of qualified electronic seals", it is not entirely clear yet, what the announced EUBW exactly will be from a technical perspective.

But it is already very clear today, that EUDIW-like solutions for server-based environments and related trust services for the management of machine-readable legal person identities will play a crucial role for the envisioned trustworthy automation of electronic business and government processes across Europe and hence the present paper aims at carrying together

[1] ecsec GmbH, Sudetenstr. 16, 96247 Michelau, Germany,
tina.huehnlein@ecsec.de, https://orcid.org/0009-0003-0634-6563;
detlef.huehnlein@ecsec.de, https://orcid.org/0000-0002-0316-7936
[2] msg group, Robert-Bürkle-Str. 1, 85737 Ismaning, Germany, steffen.schwalm@msg.group
[3] Spherity GmbH, Emil-Figge-Straße 80, 44227 Dortmund, Germany, carsten.stoecker@spherity.com

doi:10.18420/OID2025_09

aspects, which seem to be relevant for paving the way for the upcoming design phase of the EUBW and related standardisation activities.

Section 4 briefly outlines how the EUBW may look from a technical perspective and how it may be integrated into the existing and emerging system landscape and Section 5 finally summarises the present paper and provides a brief outlook towards possible future developments.

Note: This document is an *extended abstract* only to meet the given page limits. Please reach out to the lead author to request the full version of the paper and do not hesitate to contribute to the "EU Business Wallet Initiative" initiated by the non-profit go.eIDAS Association.

2 Background for European Business Wallets

2.1 Relevant Regulatory Framework

2.1.1 eIDAS-Regulation (EU) No. 910/2014

Among the most relevant legal acts for the EUBW is certainly the so called "eIDAS-Regulation" (EU) No. 910/2014, which introduces the closely related EUDIW in its Art. 5a. While the eIDAS-Regulation formally addresses both natural and legal persons and defines the relevant terms in Art. 3 in a way, which covers both natural and legal persons, looking at the details of the first batch of endorsed implementing acts[4] and the currently available "Architecture and Reference Framework" [ARF], which is part of the common "Union Toolbox" recommended in (EU) 2021/946 for a coordinated approach towards the "European Digital Identity Framework", it becomes clear that the EUDIW is – at least until now – mainly geared to be used in processes for natural persons using mobile phones.

This claim is supported by the fact that the current "Wallet Reference Implementation" is only available for Android and iOS as well as by the simple observation, that the technical details for the "Person Identification Data" (PID) for legal persons, called "Legal-person PID" or simply "LPID" in the following, are not yet defined in a precise technical manner, which would give rise for interoperable implementations.

While the *legal* definition of PID in Art. 3 (3) of (EU) No. 910/2014 is "a set of data that is issued in accordance with Union or national law and that enables the establishment of the identity of a natural or legal person, or of a natural person representing another natural person or a legal person", and therefore covers *both* natural and legal persons, the *technical* specification in the Annex of (EU) 2024/2977 only contains "data identifiers", or "claim names", for natural persons yet. As the Tables 3 and 4 of the Annex of (EU) 2024/2977 for

[4] See (EU) 2024/2977, (EU) 2024/2979, (EU) 2024/2980, (EU) 2024/2981 and (EU) 2024/2982.

LPID unfortunately do not yet contain the necessary "data identifiers" for legal persons, it is currently not possible to implement systems issuing, handling and presenting LPID in an interoperable manner.

In a similar manner, the so called "PID Rulebook" in Annex 3.1 in the [ARF] does – at least currently – not address legal persons at all and "Topic 28 (Wallet Unit for legal persons)" in Annex A.2.3.28 of the [ARF] only states that

- "Legal-person PIDs" and "Wallet Units for legal persons" will be discussed with the Member States in the future,
- a "Legal-person PID" is issued under an eID scheme[5] and that
- the Commission shall develop a "Legal-person PID Rulebook" to specify the attribute scheme and other technical details applicable for "Legal-person PIDs.

The current list of notified eID schemes is available in the Official Journal of the European Union [OJ24] and the related notification documents are available in the Wiki of the eIDAS Collaboration Network. As only the eID schemes in Austria, Bulgaria, Croatia and the Netherlands seem to involve legal person identities today, it will be interesting to see whether the EUBW-related legislation will introduce changes here in order to facilitate the issuance and use of LPID across Europe.

We will come back to this topic in Section 3, when we outline, how an LPID may look like from a technical point of view.

2.1.2 Regulations related to legal entity identifiers

As the LPID will contain one or more legal entity identifiers, it is worthwhile to consider the existing regulations, which define or use such legal entity identifiers, as listed in Table 4 within the Annex of (EU) 2024/2977 and in a similar manner also in the draft of the implementing act according to Article 5b (11) of (EU) No. 910/2014:

- Value Added Tax Identification Number (VATIN) according to the Council Directive 2006/112/EC
- Taxpayer Identification Number (TIN)
- European Unique Identifier (EUID) according to (EU) 2017/1132, (EU) 2020/2244 and (EU) 2021/1042], which is based on ISO/IEC 6523-1 [ISO23]

[5] Note, that this statement is perfectly in line with the eIDAS-Regulation, as Art. 5a (11) of (EU) No 910/2014 states that "European Digital Identity Wallets shall be provided under an electronic identification scheme with assurance level high." This stipulation is somewhat unfortunate for the EUBW because there is no differentiation with respect to identities for natural and legal persons here and most use cases for EUBW only require a substantial level of security, which can be achieved without special purpose cryptographic hardware.

- Legal Entity Identifier (LEI) according to (EU) No 2022/1860, which is based on ISO 17442-1 [ISO20a]
- Economic Operator Registration and Identification Number (EORI-No) according to (EU) No 1352/2013
- Excise Number provided in Article 2(12) of Council Regulation (EU) No 389/2012

Furthermore the recently enacted directive[6] (EU) 2025/25, which amends Directive (EU) 2017/1132, should be taken into account in Section 3 and for the standardisation of a future legal entity identity to be used together with the EUBW.

The novel Directive (EU) 2025/25 outlines in Art. 16b the "EU Company Certificate" for limited liability companies (Nr. 2) and partnerships (Nr. 3). According to (Nr. 6), the issued "EU Company Certificate" shall be authenticated by means of trust services according to (EU) No 910/2014 in a way, which is compatible with the EUDIW.

In a similar manner Art. 16c of this Directive outlines the "Digital EU power of attorney", which authorises a person to represent the company and which shall also be authenticated by means of trust services according to (EU) No 910/2014 while the granting, amending or revocation of the power of attorney shall be compatible with the EUDIW.

2.1.3 Other regulations with relevance for the EU Business Wallet

In addition to the eIDAS-Regulation (see Section 2.1.1) and the regulations related to legal entity identifiers (see Section 2.1.2) there are numerous additional regulations which are relevant for the EUBW, because they introduce *application specific requirements for identification and authentication* of business partners, including consumers as well as other enterprises or government authorities, or data objects with varying security and policy requirements.

We only provide a few first examples here and refer to the full version of the paper for a more systematic treatment of the existing EU regulations in the 32 policy areas of [EC25c]:

- *Second Payment Services Directive (PSD2)* (EU) 2015/2366 requires in Art. 34 of its implementing act (EU) 2018/389 to use qualified certificates for website authentication and/or qualified certificates for electronic seals for the authentication of payment service providers.
- *Anti Money Laundering Regulation (AMLR)* (EU) 2024/1624 requires in its Art. 22 Nr. 6. that obliged entities shall identify their customers, or any person acting on their behalf, by either using a conventional identity document or passport or by electronic

[6] According to Art. 4 of (EU) 2025/25 the EU Member States shall adopt and publish, by 31 July 2027, the laws, regulations and administrative provisions necessary to comply with this Directive and apply the measures from 31 July 2028.

identification means according to the eIDAS-Regulation (EU) No 910/2014 with assurance levels 'substantial' or 'high' or relevant qualified trust services according to this regulation to prevent the use of the financial system for the purposes of money laundering or terrorist financing.

- *Energy Labelling Regulation* (EU) 2017/1369 requires in Art. 4 of its implementing act (EU) 2024/994 to use qualified electronic seals for adding data to the "European Product Registry for Energy Labelling" (EPREL) data base.
- *Ecodesign for Sustainable Products Regulation* (EU) 2024/1781 requires in Art. 11 that the "Digital Product Passport" ensures authentication, reliability and integrity with a high level of security and privacy.
- *Single Digital Gateway (SDG) Regulation* (EU) 2018/1724 requires in Art. 13 that cross-border users are able to identify and authenticate itself and sign or seal documents electronically based on means of the eIDAS-Regulation.
- *European Health Data Space (EHDS)* (EU) 2025/327 requires that recognised eID-means according to Art. 6 of (EU) No. 910/2014 shall be used by Health Professionals (Art. 12) and Citizen (Art. 16) for health data access.
- *Data Governance Act* (EU) 2022/868 introduces in Art. 2 (20) a so called "Secure Processing Environment", which requires confidentiality, integrity and accessibility of data.
- *Electronic Freight Transport Information (eFTI)* (EU) 2020/1056 requires in Art. 4 that information in a machine-readable format shall be made available via an authenticated and secure connection.
- *électronique Convention relative au contrat de transport international de Marchandises par Route (eCMR)* is the related electronic consignment note specified by the "Open Logistics Foundation", which uses advanced electronic seals according to Art. 36 of (EU) No. 910/2014 to enable trustworthy electronic bills of lading.

2.2 Most Relevant Standards and Specifications

This section lists some of the seemingly most relevant standards and specifications for the EU Business Wallet. While a complete survey is beyond the scope of the present paper, it is nevertheless hoped that this section might provide a first starting point for a more systematic survey as basis for the upcoming design and standardisation phase.

2.3 Standards for Machine Readable Data Formats

As the EUBW is envisioned to automate business and government processes and exchange data in a machine readable fashion, it is obvious that standards like JSON [Br17], XML [XML], CBOR [BH20] and ASN.1 [ASN.1] are relevant for the EUBW.

2.4 Standards for Application Programming Interfaces and Business Data Exchange

Based on the machine readable data formats above, there are standards to create "Application Programming Interfaces" (API) and web services, which include *OpenAPI*, *AsyncAPI*, *SOAP*, *ebXML* and protocols like the *Dataspace Protocol*.

2.5 Standards for Credential Management and Authorisation

For the identification, authentication, authorisation and related security services, the participants in a communication need appropriate credentials. The most important credential standards available today seem to be as follows:

- [X.509] certificates represent the most classical form of credential, which is issued by a "Certification Authority" in order to bind the public key of an entity to its identity.
- Within the XML-based era [SAML]-based credentials played a major role, whereas an "Identity Provider" issued so called "Assertions" about a subject, which typically includes a set of attributes. For cross-border eID-use there is the [eID] profile, which seems to be worth to consider for the specification of identities for legal entities and for the specification of LPID in Section 3.
- Today [SAML] is about to be replaced by the more lightweight protocol stack based on OAuth [Ha12] and "OpenID Connect" (OIDC) [SB+23], in which an "OpenID Provider" issues JSON-based [Br17] "ID Tokens", which are in fact "JSON Web Tokens" (JWT) [JWT].
- The classical "Federated Identity Management" architecture is about to be replaced by architectures based on the "Self-Sovereign Identity" (SSI) paradigm. Here an "Issuer" issues "Verifiable Credentials" [W3C22; W3C25b] based on JSON [Br17], JSON-LD [SL+20] or CBOR [BH20] within the "mdoc" format for the mobile drivers licenses (mDL) according to ISO/IEC 18013-5 [ISO21], for example. There is also the option that the credentials support "Selective Disclosure" (SD-JWT [FYC25] and SD-JWT VC [TFC24]). The most important protocol for the issuance of credentials to the wallet today is "OpenID for Verifiable Credential Issuance" (OID4VCI) [LY+24] and for the presentation of claims to the "Verifier" there are protocols like "OpenID for Verifiable Presentations" (OID4VP) [TL25] and DIDComm [CLT].
  Currently there are at least the following different credentials formats:
  - JSON-based and signed as JWT [JWT] according to the W3C Verifiable Credential Data Model (VCDM) [W3C22; W3C25b],
  - JSON-based and prepared for supporting selective disclosure and signed as JWT [JWT] (SD-JWT [FYC25] and SD-JWT VC [TFC24]),

- CBOR-based and protected with corresponding "CBOR Object Signing and Encryption" (COSE) [Sh22] mechanisms, as used in ISO/IEC 18013-5 [ISO21] for example,
- JSON-LD-based and signed as JWT [JWT] according to the W3C VCDM [W3C22; W3C25b] or
- protected by some alternative "Data Integrity Proof" [W3C25c] using "Controlled Identifiers" (CIDs) [LS+25] or "Decentralized Identifiers" (DIDs) [SL+22] with a suitable "DID-Method" [W3C25a]. This approach is used in some European communities with "DID:EBSI" with first "Business Wallets" and related use cases within the "European Blockchain Services Infrastructure" [EB25a; EB25b; EB25c].

An alternative approach, which is not yet fully standardised as basic mechanism, but nevertheless has been used for the creation of the "Verifiable Legal Entity Identifier" (vLEI) according to [ISO24] is to use the "Key Event Receipt Infrastructure" [KERI] and "Authentic Chained Data Containers" [SF] for the creation of credentials.

3 Towards Standards for Legal Person Identities

In order to create the missing standard(s) for Legal Person Identities, we propose the following approach, which consists of three simple steps, which are explained in more detail in the following sub-sections:

1. Create abstract data model for legal person identities

2. Specify the mapping of the data model to standardised credential formats

3. Define the details of the trust framework for the management of the legal person identities

3.1 Data Model for Legal Person Identities

In a first step we create an abstract data model for legal entities and legal persons based on the existing regulations related to identifiers as outlined in Section 2.1.2 considering previous work [SEM24; Sp24] and in particular the recent work of the European Commission [EC25d] for the notification of wallet-related providers according to the amended eIDAS-regulation (EU) No. 910/2014 and the specific implementing act (EU) 2024/2980.

In addition to the core elements already present in [EC25d], a `LegalPerson` may have additional attributes as part of its envisioned "EU Company Certificate" according to (EU) 2025/25, as explained in the full paper.

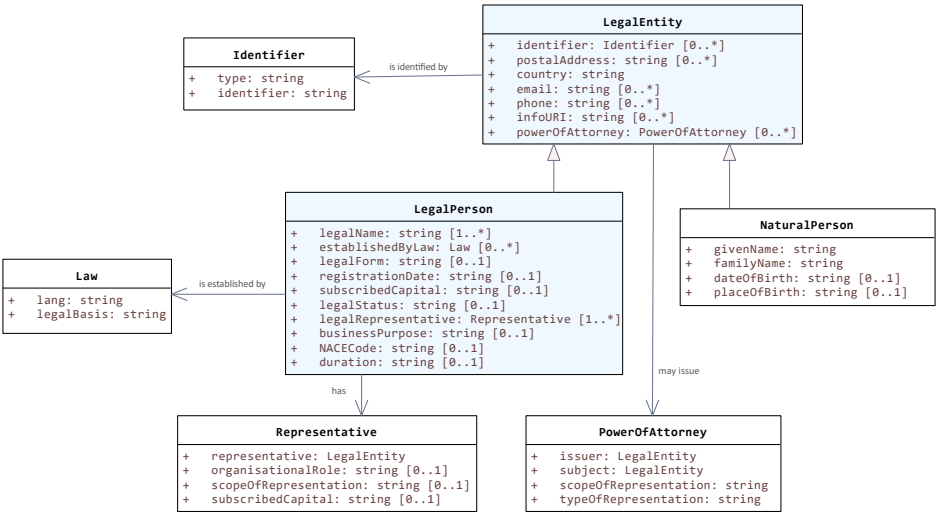


Fig. 1: Data Model for Legal Person Identities

3.2 Mapping of Data Model to Standardised Credential Formats

The abstract data model outlined above may, of course, be serialised to any machine readable data format mentioned in Section 2.3. Furthermore, it may also be mapped to any of the standardised credential formats listed in Section 2.5 in a more or less straightforward manner.

Only the mapping to [X.509] and related standards produced by ISO, ITU, IETF and last but not least the standards produced by ETSI ESI seems to be worth to be explicitly explained here in more detail, as these certificates are already widely used within the electronic identification schemes and the trust services according to the eIDAS-Regulation (EU) No. 910/2014 and also in the envisioned trust framework around the EUDIW according to (EU) 2024/2980.

Note, that the issuance of an LPID in form of a suitable [X.509] certificate, which can be used for the creation of at least advanced electronic seals, allows that the *subject* of the certificate can, more or less immediately, act as *issuer* of electronic attestations of attributes according to Art. 3 (EU) No. 910/2014, provided that such an “Enterprise EUBW” supports the functionality for issuing credentials using suitable protocols, such as OID4VCI [LY+24] for example.

Note, that [ISO20b] defines a specific `roleExtension`, which can be used for the encoding of the `organisationalRole` attribute, but some other attributes from the data model depicted in Figure 1 do not seem to be specified in international standards yet.

LegalPerson attribute	[X.509] field	Reference(s)
legalName	organizationName commonName	[ET25d, § 4.2.1] [ET25d, § 4.2.1]
establishedByLaw	QcPSB (esi4-qcStatement10)	[ET25e] with CR5
identifier	organizationIdentifier	[ET25d, § 4.2.1] and [ET25c, § 5.1.4] with CR7, see also Section 2.1.2
postalAddress	postalAddress	[X.520, § 6.6.1]
country	countryName	[ET25d, § 4.2.1]
email	rfc822Name	[X.509, § 9.3.2.1]
phone	telephoneNumber	[X.520, § 6.7.1]
infoURI	uniformResourceIdentifier	[X.509, § 9.3.2.1]

Tab. 1: Mapping of LegalPerson attributes to [X.509]

3.3 Trust Framework for Legal Person Identities

The "canonical trust framework" for Legal Person Identities used by the EUBW is defined by the eIDAS-Regulation (EU) No. 910/2014, which both regulates *electronic identification (eID)* (see Section 3.3.1) for natural and legal persons as well as *trust services* (Art. 3 (16)) (see Section 3.3.2).

Furthermore, we will briefly outline the main currently available options for EU Member States for the issuance of LPID in Section 3.3.3.

3.3.1 Electronic Identification

The legal framework for eID is more or less unchanged since 2014 and contains in particular the concept of *notification* (Art. 9), which involves a *peer-review* (Art. 12 (5)) of the electronic identification scheme (Art. 3 (4)) by the other EU Member States and finally leads to the *mutual recognition* (Art. 9) of the involved electronic identification means (Art. 3 (2)).

In this domain the EU Member States are more or less free to use their favorite technologies, the trust is induced by the scrutiny of the rather sophisticated and laborious peer-review process and the interoperability is induced, at least in theory, by the [SAML]-based interoperability framework (Art. 12 and the implementing act (EU) 2015/1501).

3.3.2 Trust Services

The initial eIDAS-Regulation from 2014 has replaced the infamous "Electronic Siganture Directive" 1999/93/EG and in particular already contained the concept of qualified and

non-qualified trust services for the issuance of [X.509] certificates to natural persons for advanced electronic signatures (Art. 26) and to legal persons for advanced electronic seals (Art. 36).

The amendment in 2024 has introduced the EUDIW (Art. 5a) and additional trust services, which are relevant for the EUBW, such as the services for the electronic attestations of attributes (Art. 45b-45h), for example.

Here, one may distinguish between

- simple *electronic attestations of attributes (EEA)* (Art. 3 (44)),
- well-regulated *qualified electronic attestations of attributes (QEEA)* (Art. 3 (45)) and
- *electronic attestations of attributes issued by or on behalf of a public sector body responsible for an authentic source* (PubEEA) (Art. 3 (46)), which aim at a similar level of trust as QEEA, but directly involve the authentic sources (Art. 3 (47)), which are usually provided by public sector bodies.

The trust services for electronic attestation of attributes may be combined with other trust services including the newly introduced trust service for electronic ledger (Art. 45l), as provided by the "European Blockchain Services Infrastructure" [EB25a], for example. This approach has been followed in projects such as EBSI-VECTOR and EBSI-Trace4EU for the issuance and usage of LPID and the management of Powers of Attorney by responsible natural persons.

3.3.3 Options for the issuance of LPID and their use in the EUBW

The available options for the issuance of LPID may be summarised as follows:

- EU Member States, which have not done this yet, may become active in order to integrate aspects related to legal person identities into their existing eID schemes with a high level of assurance (see Art. 5a (11)) and amend the corresponding notifications (see [OJ24]). As soon as this has happened, they can designate one or more LPID issuers, which may integrate suitable authentic sources, as foreseen in Art. 3 of (EU) 2024/2977, Annex II Nr. 3 of (EU) 2024/2980 and [EC25d].
- On the other hand, an EU Member State may leave the issuance of LPID more or less to the market of (qualified) trust service providers, which may in turn may collaborate with business registries, so that the management of LPID would essentially become a (qualified) trust service for (Q)EAA.
- A third and somewhat intermediate strategy would use PubEEA-providers for the issuance of LPID based on direct access to the relevant registers and authentic sources (Art. 3 (47)).

It is not surprising that the different options for establishing a dedicated LPID issuer, which may be a (Q)TSP or PubEAA issuer, are currently subject to a variety of large scale pilot projects across Europe. It will be interesting to see, which strategies are in the end preferred by the different EU Member States, what the timeline for the individual actions will be, what influence any changes of the legal framework due to the announced EUBW may have and last but not least, which impact the forthcoming standard ETSI TS 119 479 [ET25b] for the "*Specification of interfaces related to Authentic Sources*" may have on these exciting developments.

4 Solution Outline for a European Business Wallet

This section sketches how the EUBW might look like in the end from a technical perspective.

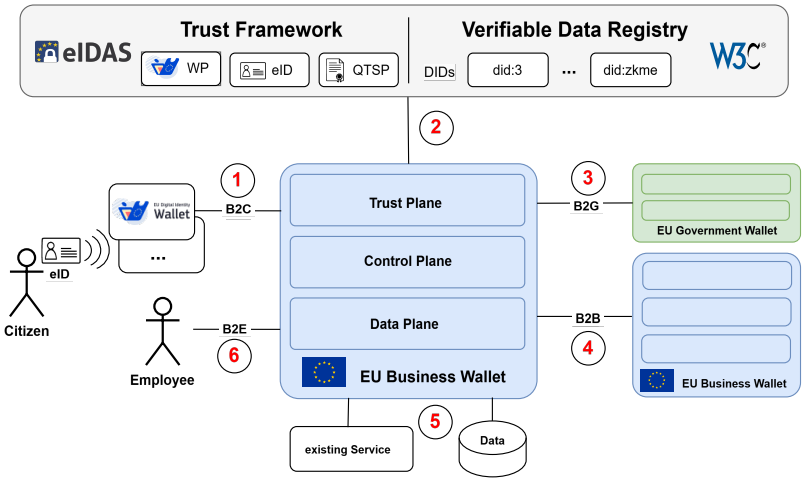


Fig. 2: Outline of System Architecture with EU Business Wallet

As outlined in the Figure 2, the EUBW will **(1)** support B2C processes together with the EUDIW and hence the EUBW shall be able to act as "Wallet-Relying Party" and use "Access Certificates" in protocols like OID4VP [TL25] and DIDComm [CLT] for example.

The EUBW will **(2)** request the necessary "Access Certificates" as well as further potentially required certificates and attestations (e.g. LPID) and interact with other suitable trust service providers according to the eIDAS-Regulation (EU) No. 910/2014.

The EUBW will **(3)** hopefully also be usable for electronic government processes, as soon as there are "EU Government Wallets" around, which support normalised interfaces and standards according to the state of the art.

An EUBW-instance will, of course, be able to conduct trustworthy electronic business processes with other EUBW-instances **(4)** and will need to smoothly integrated with other business IT systems, as well as existing services **(5)** and finally support B2E processes, which involve employees **(6)**.

5 Conclusion and Outlook

The present document aims at contributing to the standardisation of the recently announced EUBW by providing a compact survey of the probably most relevant European regulations (see Section 2.1) and seemingly relevant international technical standards (see Section 2.2), before we address Legal Person Identities in Section 3, which are crucial for the successful use of the EUBW. The paper concludes in Section 4 by outlining a technical solution for the envisioned EUBW.

The EUBW is geared towards supporting the most important European initiatives and regulations mentioned above, as well as a large variety of hardly regulated business processes, such as the "Digital Product Passport" or "Data Spaces", which are emerging for the secure and trustworthy sharing of sensitive data in different industries. For this purpose it is of utmost importance to define a modular and coherent European standardisation framework to achieve these goals and contribute to the design, development and operation of a trustworthy and sovereign "EuroStack".

The present paper has already been provided as input to the "EU Business Wallet Initiative", which aims at further paving the way for the EUBW by incubating standardisation activities at the European and later international level.

While the exciting development around the EUBW is in a rather early phase, we already thankfully acknowledge, that the present paper has greatly benefited from very fruitful discussions with numerous experts in pertinent national and international projects and standardisation bodies, which are currently working on the standardisation of a trustworthy EUDIW and LPID.

References

[ARF] European Commission: The European Digital Identity Wallet Architecture and Reference Framework, URL: https://github.com/eu-digital-identity-wallet/eudi-doc-architecture-and-reference-framework/blob/main/docs/architecture-and-reference-framework-main.md.

[ASN.1] ITU-T: Information technology – Abstract Syntax Notation One (ASN.1): Specification of basic notation, Recommendation X.680, 2021, URL: https://www.itu.int/rec/T-REC-X.680.

[BH20] Bormann, C.; Hoffman, P.: Concise Binary Object Representation (CBOR), IETF RFC 8949, URL: https://datatracker.ietf.org/doc/html/rfc8949.

[Br17] Bray, T.: The JavaScript Object Notation (JSON) Data Interchange Format, IETF RFC 8259, URL: https://datatracker.ietf.org/doc/html/rfc8259.

[CLT] DIDComm Messaging v2.x, Editor's Draft, DIF Ratified Specification, URL: https://identity.foundation/didcomm-messaging/spec/.

[EB25a] European Commission: European Blockchain Service Infrastructure, 2025, URL: https://ec.europa.eu/digital-building-blocks/sites/display/EBSI/Home.

[EB25b] EBSI: EBSI Verifiable Credential Framework, 2025, URL: https://hub.ebsi.eu/vc-framework.

[EB25c] EBSI: EBSI-VECTOR Project, 2025, URL: https://www.ebsi-vector.eu/en.

[EC25a] European Commission presents its compass to boost Europe's competitiveness in the next five years, 2025, URL: https://ec.europa.eu/commission/presscorner/detail/en/ac_25_385.

[EC25b] Commission, E.: Competitiveness Compass for the EU, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 2025, URL: https://commission.europa.eu/document/download/10017eb1-4722-4333-add2-e0ed18105a34_en.

[EC25c] Commission, E.: Summaries of EU legislation, 2025, URL: https://eur-lex.europa.eu/browse/summaries.html.

[EC25d] Specification of systems enabling the notification and subsequent publication of Provider information, 2025, URL: https://tinyurl.com/STS-TS2.

[eID] eIDAS Cooperation Network: eID Services – eID-Profile, URL: https://ec.europa.eu/digital-building-blocks/sites/display/DIGITAL/eID+Profile.

[ET25b] ETSI TS 119 478: Electronic Signatures and Infrastructures (ESI); Specification of interfaces related to Authentic Sources, Draft, 2025, URL: https://tinyurl.com/TS-119478.

[ET25c] ETSI EN 319 412-1: Electronic Signatures and Infrastructures (ESI); Certificate Profiles; Part 1: Overview and common data structures, 2025, URL: https://www.etsi.org/deliver/etsi_en/319400_319499/31941201/01.06.00_20/en_31941201v010600a.pdf.

[ET25d] ETSI EN 319 412-1: Electronic Signatures and Infrastructures (ESI); Certificate Profiles; Part 3: Certificate profile for certificates issued to legal persons, 2023, URL: https://www.etsi.org/deliver/etsi_en/319400_319499/31941203/01.03.01_60/en_31941203v010301p.pdf.

[ET25e] ETSI EN 319 412-5: Electronic Signatures and Infrastructures (ESI); Certificate Profiles; Part 5: QCStatements, 2025, URL: https://www.etsi.org/deliver/etsi_en/319400_319499/31941205/02.05.00_20/en_31941205v020500a.pdf.

[FYC25] Fett, D.; Yasuda, K.; Campbell, B.: Selective Disclosure for JWTs (SD-JWT), 2025, URL: https://datatracker.ietf.org/doc/draft-ietf-oauth-selective-disclosure-jwt/.

[Ha12] Hardt, D.: The OAuth 2.0 Authorization Framework, IETF RFC 6749, URL: https://datatracker.ietf.org/doc/html/rfc6749.

[ISO20a] ISO: ISO 17442-1: Financial services — Legal entity identifier (LEI) – Part 1: Assignment, 2020.

[ISO20b] ISO: ISO 17442-2: Financial services — Legal entity identifier (LEI) – Part 2: Application in digital certificates, 2020.

[ISO21] ISO/IEC: ISO/IEC 18013-5: Personal identification — ISO-compliant driving licence – Part 5: Mobile driving licence (mDL) application, 2021.

[ISO23] ISO/IEC: ISO 6523-1: Information technology — Structure for the identification of organizations and organization parts — Part 1: Identification of organization identification schemes, 2023.

[ISO24] ISO: ISO 17442-3: Financial services — Legal entity identifier (LEI) – Part 3: Verifiable LEIs (vLEIs), 2024.

[JWT] Jones, M.; Bradley, J.; Sakimura, N.: JSON Web Token (JWT), IETF RFC 7519, URL: https://datatracker.ietf.org/doc/html/rfc7519.

[KERI] Smith, S. M.: KERI Resources - Key Event Receipt Infrastructure, URL: https://keri.one/keri-resources/.

[LS+25] Longley, D.; Sporny, M.; Sabadello, M.; Reed, D.; Steel, O.; Allen, C.: Controlled Identifiers v1.0, 2025, URL: https://www.w3.org/TR/cid-1.0/.

[LY+24] Lodderstedt, T.; Yasuda, K.; Looker, T.: OpenID for Verifiable Credential Issuance, 2024, URL: https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0.html.

[OJ24] Electronic identification schemes notified pursuant to Article 9(1) of Regulation (EU) No 910/2014 of the European Parliament and of the Council on electronic identification and trust services for electronic transactions in the internal market, Official Journal of the European Union, 20.12.2024, 2024, URL: http://data.europa.eu/eli/C/2024/7543/oj.

[SAML] OASIS: Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0, OASIS Standard, 2005, URL: https://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf.

[SB+23] Sakimura, N.; Bradley, J.; Jones, M.; de Medeiros, B.; Mortimore, C.: OpenID Connect Core 1.0 incorporating errata set 2, 2023, URL: https://openid.net/specs/openid-connect-core-1_0.html.

[SEM24] SEMIC-Community: EU Company Vocabulary, 2024, URL: https://semiceu.github.io/Core-Business-Vocabulary/releases/2.2.0/.

[SF] Smith, S.; Feairheller, P.: Authentic Chained Data Containers (ACDC), v0.9 Draft, URL: https://trustoverip.github.io/tswg-acdc-specification/.

[Sh22] Schaad, J.: CBOR Object Signing and Encryption (COSE): Structures and Process, IETF RFC 9052, URL: https://datatracker.ietf.org/doc/html/rfc9052.

[SL+20] Sporny, M.; Longley, D.; Kellogg, G.; Lanthaler, M.; Champin, P.-A.; Lindström, N.: JSON-LD v1.1, W3C Recommendation, 2020, URL: https://www.w3.org/TR/json-ld/.

[SL+22] Sporny, M.; Longley, D.; Sabadello, M.; Reed, D.; Steel, O.; Allen, C.: Decentralized Identifiers (DIDs) v1.0, 2022, URL: https://www.w3.org/TR/did-core/.

[Sp24] Spherity: EU Company Vocabulary, 2024, URL: https://oid.spherity.com/eucc.

[TFC24] Terbu, O.; Fett, D.; Campbell, B.: SD-JWT-based Verifiable Credentials (SD-JWT VC), 2024, URL: https://datatracker.ietf.org/doc/draft-ietf-oauth-sd-jwt-vc/.

[TL25] Terbu, O.; Lodderstedt, T.; Yasuda, K.; Looker, T.: OpenID for Verifiable Presentations, 2025, URL: https://openid.net/specs/openid-4-verifiable-presentations-1_0.html.

[W3C22] Sporny, M.; Longley, D.; Chadwick, D.: Verifiable Credentials Data Model v1.1, 2022, URL: https://www.w3.org/TR/vc-data-model/.

[W3C25a] W3C: DID Methods, W3C Group Note, 2025, URL: https://www.w3.org/TR/did-extensions-methods/.

[W3C25b] Sporny, M.; Longley, D.; Chadwick, D.; Herman, I.: Verifiable Credentials Data Model v2.0, W3C Candidate Recommendation Draft 27, 2025, URL: https://www.w3.org/TR/vc-data-model-2.0/.

[W3C25c] Longley, D.; Sporny, M.; Herman, I.: Verifiable Credential Data Integrity 1.0, W3C Candidate Recommendation Draft, 2025, URL: https://www.w3.org/TR/vc-data-integrity/.

[X.509] ITU-T: Information technology - Open Systems Interconnection - The Directory: Public-key and attribute certificate frameworks, Recommendation X.509, 2019, URL: https://www.itu.int/rec/T-REC-X.509-201910-I/en.

[X.520] ITU-T: Information technology - Open Systems Interconnection - The Directory: Selected attribute types, Recommendation X.520, 2019, URL: https://www.itu.int/rec/T-REC-X.520-201910-I.

[XML] Bray, T.; Paoli, J.; Sperberg-McQueen, C. M.; Maler, E.; Yergeau, F.; Cowan, J.: Extensible Markup Language (XML) 1.1 (Second Edition), W3C Recommendation, 2006, URL: https://www.w3.org/TR/xml11/.

Further Conference Contributions

Securing Engagement: Key Takeaways from Onboarding Users into Data Protection Apps

Lennart Kiss [1], Rachelle Sellung [2]

Abstract: This paper examines persistent challenges users/citizens face in exercising their data protection rights under the General Data Protection Regulation (GDPR), particularly in everyday digital contexts. It presents findings from the BMBF-funded Tester project, which developed a privacy assistant to support transparency, intervenability, and data protection literacy in the context of self-tracking and health-monitoring devices. Focusing on the onboarding experience, this study explores how users engage with a GDPR-aligned privacy assistant through an iterative, human-centered design process. Based on mixed-method user evaluations, we validate key usability strategies—such as personalization, grouped customization, and adaptive content—that can promote awareness and improve informed user decision making. The results offer lessons for the design of onboarding experiences in privacy-focused applications and contribute to broader efforts to make digital privacy rights more accessible, transparent, and usable.

Keywords: Privacy Assistant, Usability, User Experience, GDPR, Data Protection, Privacy

1 Introduction

The General Data Protection Regulation (GDPR), introduced in 2018, grants individuals the right to control their personal data. However, practical implementation of these rights remains challenging for many users [GML20, ZMM24]. Transparency around data processing and effective mechanisms for intervenability are still limited, particularly in consumer-facing applications that handle sensitive personal data. While the legal framework is well defined, its usability in everyday contexts often falls short [JF20]. This paper addresses those challenges through the lens of onboarding in data protection applications. For simplicity, the onboarding process is generally defined as the "sequence of steps that introduces users to a new product or service" [Ca17]. In our case specifically in the context of a privacy assistant for users of self-tracking and health-monitoring devices developed within the Tester Project funded by the German federal ministry for education and research (BMBF). These third-party self-measurement devices, often used for fitness or chronic disease management, collect large volumes of personal and sensitive data. Yet users frequently remain unaware of what data is

[1] University of Stuttgart, Institute of Human Factors and Technology Management (IAT), Identitymanagement, Allmandring 35, Stuttgart, 70569, lennart.kiss@iat.uni-stuttgart.de https://orcid.org/0009-0000-8839-5470
[2] Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering IAO, Identitymanagement, Nobelstrasse 12, Stuttgart, 70569, rachelle.sellung@iao.fraunhofer.de https://orcid.org/0000-0003-1235-030X

doi:10.18420/OID2025_10

collected, how it is processed, and by whom. Traditional onboarding flows in consumer applications focus on basic feature introductions or account setup to convert first time users into long term users [Ca22, SF23]. In contrast, onboarding in privacy-focused tools must also serve as a critical educational and trust-building function. The onboarding process studied in this work aimed to introduce a GDPR-aligned privacy assistant for improving transparency and user control over self-tracking data. The assistant includes features such as a dashboard to show data flows, a help desk for exercising GDPR rights, and an information center with educational materials tailored to the user's data protection literacy. Therefore, the onboarding was designed to build a user's understanding of relevant data flows and to prepare them to make informed decisions about their data processing events later in the application experience. The contribution of this paper is twofold. First, it presents an empirical evaluation of how users engage with privacy-related onboarding in a real use case based on qualitative study data. Second, it identifies design strategies that support user empowerment through transparency, especially in data ecosystems where users have low visibility and high dependency on third-party device ecosystems. The rest of the paper is structured as follows: Section two reviews related work on user experience in privacy assistants and onboarding. Section three outlines the design and evaluation methodology. Section four presents key findings from user studies. Section five discusses the implications for privacy tool design. Section six reflects limitations and future research, and Section seven concludes with a summary.

2 Related Literature

This chapter outlines related research on three relevant topics to this work. First, this section will explore the literature on the user experience perspective of Data Protection, in particular related to GDPR. Next, it will summarize related research on privacy assistants. Lastly, there is a brief overview of the topic of user experience in onboarding.

Despite its legal clarity, the GDPR offers limited guidance on how to implement user rights in a usable way. [JF20] highlight the lack of concrete implementation guidance for usability in data protection systems. Additionally, [AM21] report that UX designers often rely on informal practices when addressing GDPR compliance with no long-term sustainability, indicating a need for structured methodologies. Privacy preferences are often highly subjective and context-dependent, varying by data type and retention period [AJL13, Em21]. This complexity underscores the challenge of creating understandable interfaces for privacy control. Studies also show that users struggle with the length and legal complexity of privacy policies [Eb21, Ac16, WZH17], with estimates suggesting it would take hundreds of hours annually to read them thoroughly [Gl16]. There have been attempts to improve privacy policies in how they are visualized for users [ST20, CMU25, CGA06]. Recent research into privacy assistants explores how digital tools can support users in managing personal data, particularly in contexts like the Internet of Things (IoT) [AZ19, Da18] and mobile apps [Li16]. However, many of these tools are developed without sufficient user involvement or attention to human-centered interface

design [St23]. Furthermore, [St23] emphasizes the need for privacy assistants to support key functions such as privacy configuration, awareness building, and motivation, and found that users expect both automation within settings configuration as well as alerts about potential privacy threats. Onboarding has been studied extensively in general UX literature [Ca03, SGV18], but much less in the context of privacy and data protection. [Ga19] outlines three key components of effective onboarding—UI patterns, contextual educational content, and contextual communication. The Minimalist Instruction framework by [Ca14] advocates for action-oriented, error-resilient onboarding flows. [SF23] presents a larger summary of his work. However, this identifies similar themes of engagement and learning presented in [Ga19]. Additionally, privacy onboarding differs from commercial onboarding in that its goal is not just immediate feature engagement, but long-term user empowerment, trust, and informed control over personal data.

3 Methodology

This study applied a design science approach in accordance with [He04] to develop and evaluate a privacy assistant focused on onboarding users in the context of self-tracking and health-monitoring devices. The objective was to design an onboarding process that supports transparency and user empowerment under the GDPR, particularly in complex, third-party device ecosystems. The methodology followed an iterative cycle of design, evaluation, and refinement. The initial prototype was based on prior user research within the project on user needs, self-tracking behaviors, and expectations regarding transparency and intervenability [Fl22, Fl24]. The resulting onboarding process was then refined through multiple rounds of expert reviews and user testing.

The onboarding flow was structured around five sequential steps, each designed to reduce cognitive burden and progressively increase user awareness of personal data flows. It begins with a brief, non-technical introduction that communicates the assistant's purpose and benefits, setting the stage for what users can expect. This is followed by an account setup phase or an optional guest mode for anonymous exploration. Subsequently, users are guided through a device selection process, where they connect their self-tracking devices. This step provides the assistant with the necessary context to inform users about what data is being collected and by which devices. Next, users configure their privacy preferences by selecting one of three pre-configured privacy profiles—ranging from restrictive to permissive—with the option to manually adjust settings if desired. The final step asks users to self-assess their familiarity with data protection concepts. Based on this input, the assistant tailors the tone and depth of educational content offered within the app. Importantly, all steps in the onboarding flow are optional and skippable, allowing users to engage at their own pace.

The first development stage of the privacy assistant was assessed by means of expert reviews. Subsequently in the following stages user testing was conducted in two iterations, involving a total of 26 unique participants. Due to logistical constraints, the

same sample could not be retained across both sessions. To ensure consistency in participant profiles, each participant was asked to identify with one of the predefined personas established earlier in the project [Fl24]. Quantitative evaluation employed the System Usability Scale (SUS) to measure perceived usability, the User Experience Questionnaire (UEQ) to assess both hedonic and pragmatic aspects of the interaction, and A/B testing to compare variations in transparency and intervenability features. Qualitative data were gathered through post-task and post-test questionnaires. Key coding dimensions included usability factors such as user-friendliness, interface clarity, terminology, and error comprehension; user expectations; consistency and aesthetics of design; and accessibility. Additionally, the analysis addressed data protection attitudes. Error analysis covered both the source and severity of errors. Reflections captured exclusively in the post-test phase included overall satisfaction, engagement patterns, perceived usage barriers, comparisons to similar tools, and perspectives on the assistant's impact on transparency and intervenability.

4 Findings and Lessons Learned

This section summarizes the outcomes of user testing conducted across three development iterations of the privacy assistant's onboarding process. The qualitative findings reflect recurring themes in usability, user engagement, and comprehension. They are grouped by strengths, identified challenges, and design lessons that emerged during the iterative process.

Overall, users responded positively to the onboarding experience. The introductory screen effectively communicated the assistant's purpose without overwhelming users. Participants appreciated the clean interface, visual aids, and flexible design, including the ability to skip steps or enter via a guest mode. The account setup process was perceived as intuitive, with clear guidance for actions. Feedback mechanisms—such as progress indicators and confirmation screens—were frequently cited as helpful in structuring the process and motivating users to continue. These design elements contributed to a sense of clarity and control during onboarding and reflect the effectiveness of the design principles discussed in the related literature. Despite the generally positive reception, several challenges were observed in later onboarding steps. While connecting self-measurement devices was straightforward, the purpose and implications of privacy preference configuration were often unclear. Users struggled to understand how their selections would affect the app's future behavior. This gap suggests a need for more explicit communication of how these settings influence the assistant's reports and recommendations. Participants also reported difficulty differentiating between the pre-configured privacy profiles. Some chose at random or based on assumed "safety," without a clear understanding of the trade-offs. Although manual configuration was available, it was rarely used due to perceived complexity. The adaptive educational content, which adjusted based on users' self-assessed privacy knowledge, was well received. However, some users expected this adaptability to extend

beyond the Information Center into other parts of the app, resulting in mismatched expectations. The following insights summarize design strategies that improved onboarding effectiveness and highlight areas for future enhancement:

- **Personalization must be tied to Tangible Outcomes:** Users valued adaptive features but needed clear, real-time feedback connecting their choices to meaningful results. Visual previews or usage scenarios could improve comprehension.
- **Iterative Refinement is Non-Negotiable:** Repeated testing revealed subtle but significant usability barriers that were not evident in initial expert reviews. Frequent feedback loops ensured alignment with user expectations and interaction patterns.
- **Adaptive Content should anticipate User Expectations:** Users should be informed about which parts of the system are personalized. Without this clarity, adaptive features risk generating confusion rather than empowerment.
- **Feedback Mechanisms are integral to Engagement:** Features like progress bars and confirmation screens are not cosmetic—users consistently cited them as motivational and informative. They should be considered essential components in onboarding design.
- **Simplified Complexity through Grouped Customization:** To reduce cognitive load, future iterations should organize privacy preferences by themes or data types. This would help users engage more confidently without requiring expert-level understanding.

5 Discussion

This study contributes to ongoing efforts to improve data protection by examining how onboarding can support user empowerment in privacy-focused applications. Through empirical evaluation of a privacy assistant designed for users of self-tracking devices, this work identifies actionable design strategies that address key usability barriers while maintaining relevance to GDPR principles of transparency and intervenability. In the context of data protection, onboarding becomes a strategic point of engagement: it sets the foundation for user understanding of complex data flows and initiates a mental model for managing privacy preferences. This framing distinguishes privacy onboarding from traditional feature onboarding and underscores its domain specificity. Our work demonstrates that onboarding can be structured to increase data protection literacy without requiring immediate, high-stakes decisions. Pre-configured privacy settings and adaptive educational content, when properly explained and scoped, reduce decision fatigue while still providing control. These design elements support privacy as a process, rather than a one-time opt-in. This paper contributes to a growing but still limited body of research on human-centered privacy design. While many privacy tools aim to enforce compliance or automate protection, few address how users are introduced to these tools and the role of first-time interactions in building trust and enabling autonomy. By

focusing on onboarding, we highlight a phase in the privacy interaction lifecycle that has tangible effects on user behavior and retention. Our study also validates several design principles in a privacy-specific context. These include the importance of progressive disclosure, grouped customization, and feedback-rich interactions, which together create a more navigable and engaging onboarding process. Finally, the iterative nature of our approach demonstrates how empirical UX research can shape the development of effective privacy tools. Our mixed-methods evaluation revealed not only usability strengths and weaknesses, but also user expectations, misinterpretations, and the unintended consequences of unclear feature boundaries. These insights are critical for translating GDPR rights into usable, real-world applications.

6 Limitations and Future Work

While this study provides important insights into the onboarding of privacy-focused applications, its limitations should be acknowledged. First, the privacy assistant evaluated in this work was tested in a simulated environment, with certain features (e.g., account creation, authentication) mimicked for the purpose of usability testing. No real user data was processed or stored, and the assistant was not deployed in a live system. This limitation restricts our ability to assess how users would respond to real-world consequences of their interactions, particularly in terms of long-term trust and data protection behavior. Second, the study focused primarily on short-term usability and user experience metrics. While the findings illustrate how onboarding supports comprehension and engagement at the initial interaction stage, they do not account for retention, behavior change, or sustained use over time. Longitudinal studies would be necessary to evaluate whether the onboarding experience leads to more informed or proactive privacy management in the long run. Additionally, expanding the adaptive content model to cover more interaction layers of the application—while managing user expectations—could further enhance usability without compromising simplicity.

7 Conclusion

This paper examined how onboarding can be leveraged to support user engagement and comprehension in privacy-focused applications, specifically through the development and evaluation of a GDPR-aligned privacy assistant for self-tracking scenarios. By applying a design science approach and involving users throughout the development cycle, we identified onboarding strategies that reduce complexity, build trust, and improve early understanding of personal data flows. Our findings demonstrate that onboarding is not merely a functional prelude but a critical phase in the privacy interaction lifecycle. Design elements such as grouped customization, adaptive content, and feedback mechanisms were linked to enhanced usability while directly supporting privacy goals, including transparency and user empowerment. The scientific contribution

of this work lies in bridging the gap between legal rights and practical usability. We show how empirical, user-centered methods can produce onboarding experiences that not only lower the barrier to entry but also prepare users to act meaningfully within the constraints of today's digital environments. By focusing on securing engagement at the earliest stage of interaction, this research contributes to the design of more effective, human-centered privacy tools that support not just compliance, but comprehension, agency, transparency and intervenability.

Bibliography

[Ac16] Acquisti, A. et al.: The economics of privacy. Journal of Economic Literature, 54, pp. 442–492, 2016.

[AJL13] Acquisti, A.; John, L. K.; Loewenstein, G.: What is privacy worth? In: The Journal of Legal Studies, Bd. 42, Nr. 2, S. 249-274, 2013.

[AM21] Almeida, F.; Monteiro, J. A.: Exploring the Effects of GDPR on the User Experience. In: Journal of Information Systems Engineering and Management, Bd. 6, Nr. 3, em0140, 2021.

[AZ19] Andrade, L.; Zorzo, S.: Privacy Everywhere: A Mechanism for Decision Making and Privacy Assurance in IoT Environments. 2019.

[Ca03] Carroll, J. M.: Introduction: Toward a Multidisciplinary Science of Human-Computer Interaction. In: HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science, S. 1–9, Elsevier Inc., 2003.

[Ca14] Carroll, J.: Creating Minimalist Instruction. In: International Journal of Design for Learning, Bd. 5, Nr. 2, 2014.

[Ca17] Cascaes Cardoso, M.: The Onboarding Effect: Leveraging User Engagement and Retention in Crowdsourcing Platforms. In (ed.): Proc. 2017 CHI Conf. Ext. Abstr. on Human Factors in Comput. Syst., 2017.

[Ca22] Cascaes Cardoso, M.: User Onboarding Design in Citizen Science: A Path to Grow Engagement and Participation. PhD thesis, 2022.

[CGA06] Cranor, L. F.; Guduru, P.; Arjula, M.: User Interfaces for Privacy Agents. In: ACM Transactions on Computer-Human Interaction, Bd. 13, S. 135–178, 2006.

[CMU25] Carnegie Mellon University, www.usableprivacy.org, accessed: 30/01/2025.

[Da18] Das, A. et al.: Personalized Privacy Assistants for the Internet of Things: Providing Users with Notice and Choice. IEEE Pervasive Comput. 17, pp. 35–46, 2018.

[Eb21] Ebbers, F. et al.: User Preferences for Privacy Features in Digital Assistants. Electron. Markets 31, pp. 411–426, 2021.

[Em21] Emami-Naeini, P. et al.: Privacy Expectations and Preferences in an IoT World. In (ed.): Proc. SOUPS 2017. USENIX Association, Berkeley, CA, pp. 399–412, 2017.

[Fl22] Floris, A.: Who is the self-tracker? Establishing an initial instrument to holistically

assess self-tracking behaviours. Master's thesis, University of Twente, 2022.

[Fl24] Floris, A.; Astfalk, S.; Sellung, R.; Roßnagel, H.: Tracking or being tracked: How much do self-trackers care about their data's privacy? In (ed.): Sicherheit 2024. Gesellschaft für Informatik e.V., pp. 121-136, 2024.

[Ga19] Galavan, R.: C.A.R.E. – A Simple Customer Onboarding Framework. In: Intercom Onboarding, S. 20–28, 2019.

[Gl16] Gluck, J. et al.: How Short is Too Short? Implications of Length and Framing on the Effectiveness of Privacy Notices. In (ed.): Proc. 12th USENIX Conf. on Usable Privacy and Security (SOUPS '16). USENIX Association, Berkeley, CA, pp. 321–340, 2016.

[He04] Hevner, A. R. et al.: Design Science in Information Systems Research. In: MIS Quarterly, Bd. 28, Nr. 1, S. 75–105, 2004.

[JF20] Johansen, J.; Fischer-Hübner, S.: Making GDPR Usable: A Model to Support Usability Evaluations of Privacy. In (ed.): Privacy and Identity Management. Data for Better Living: AI and Privacy. Privacy and Identity 2019. IFIP Advances in Information and Communication Technology, Volume 576, Springer, Cham, 2020.

[Li16] Liu, B. et al.: Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions. In (ed.): Proceedings of the 12th USENIX Conference on Usable Privacy and Security (SOUPS '16). USENIX Association, Berkeley, CA, pp. 27–41, 2016.

[SF23] Svanström, N.; Fredriksson, A.: User Onboarding Experience – Increasing Users' Perceived Motivation of Engaging More with a Product. Master's thesis, Chalmers University of Technology, 2023.

[SGV18] Strahm, B.; Gray, C. M.; Vorvoreanu, M.: Generating Mobile Application Onboarding Insights Through Minimalist Instruction. In: Proceedings of the 2018 Designing Interactive Systems Conference, S. 361–372, 2018.

[ST20] Soumelidou, A.; Tsohou, A.: Effects of Privacy Policy Visualization on Users' Information Privacy Awareness Level. Information Technology & People 33, pp. 502–534, 2020.

[St23] Stöver, A. et al.: Investigating How Users Imagine Their Personal Privacy Assistant. Proceedings on Privacy Enhancing Technologies Symposium 2023/2, pp. 384–402, 2023.

[WZH17] Mihale-Wilson, C.; Zibuschka, J.; Hinz, O.: About User Preferences and Willingness to Pay for a Secure and Privacy-Protective Ubiquitous Personal Assistant. In (ed.): Proc. 25th European Conf. on Information Systems (ECIS), Guimarães, Portugal, pp. 32–47, 2017.

[ZMM24] Zaguir, N. A.; de Magalhães, G. H.; de Mesquita Spinola, M.: Challenges and Enablers for GDPR Compliance: Systematic Literature Review and Future Research Directions. IEEE Access 12, pp. 81608–81630, 2024.

How can Design Thinking benefit Cybersecurity?: Insights from Don Norman's The Design of Everyday Things

Mrudula Arunkumar [1], Christian Schunck [1], Salma Ben Mamia[1], Heiko Roßnagel [1]

Abstract: Human error and insufficient security awareness remain the largest cyber-risk factors for organizations. Despite the prevalence of security training, employees often fail to translate knowledge into secure behavior leading to a gap between security awareness and secure behaviour. Hence, the integration of human factors beyond awareness in cybersecurity is crucial wherein the focus lies on steering the actions executed by people rather than the technical protection offered by the security systems. Donald A. Norman's The Design of Everyday Things is one of the pioneering books that introduces how intended actions can be achieved through a user-centric product design. Consequently, it provides a lens to rethink the various security policy designs that are developed to enforce cybersecurity. This short paper therefore proposes a new framework involving design thinking principles to help design better security policies with a human factor focus.

Keywords: Human Factors, Design thinking, Cybersecurity, Human Behaviour, Security Policy

1 Introduction

The increasing prevalence of cyberattacks necessitates improved cybersecurity measures, and most companies of late have focussed on strengthening the human defense against cyber attack [ZR19] through security awareness training. However this still does not translate to employees behaving securely due to the persistent gap that exists between having awareness and behavioural actions. Therefore, there needs to be additional methods to capitalize on the human factors beyond awareness to promote secure behaviour in the work environment. Having in depth knowledge on the general

[1] Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering IAO, Team Identity Management, Nobelstraße 12, Stuttgart, 70569, firstname.lastname@iao.fraunhofer.de,
https://orcid.org/0000-0002-6441-9623
https://orcid.org/0000-0002-7917-8180
https://orcid.org/0000-0001-7057-8404

doi:10.18420/OID2025_11

cognitive factors that influence the behavioural actions and user interaction with technology first paves a way into assessing the key issues that determine the security compliance. From a security perspective, this could entail (re)designing security policies. Making security measures more intuitive and usable, termed as *usable security*, has been a topic of research since early 2000s [PE08, SF05] focussing on the design vulnerabilities that are present especially concerning specific aspects of security like authentication and encryption methods. However, based on a recent literature review [DTO23] it is clear that the research field lacks a concrete and systematic approach to implement better design for security applications. The current approach to address the human factor in security often fails [BSN19] because it focuses solely on knowledge delivery and compliance enforcement, neglecting the empathetic and intuitive understanding of employees' daily workflows. Therefore, applying the concepts and insights from design thinking on cybersecurity policy engineering that deeply affect everyday workflows for employees should provide more tangible outputs .

Donald A.Norman's *The Design of Everyday Things* [No13] is one of the pioneering book on product design from a human factors perspective, making it a valuable resource for creating security policies in cybersecurity. Given that cyberattacks stem from human error and intentional policy violations, it is crucial to optimize solutions that encourage secure behaviors. Norman emphasizes that design must accommodate human tendencies and has infact noted the lack of usability in the cybersecurity field namely the issues of authentication such as password design [No09]. However his insights in *The Design of Everyday Things* can be used as a template to apply design thinking at a much more systemic level to cyber security especially in the context of security policy design to promote secure behavioural actions. Thus, this short paper will overview design thinking principles relevant to empowering the human factor in cybersecurity and provide guidelines to expand their application. An improved security design should promote compliance, thus serving as a useful tool for security officers responsible for policy design and compliance.

2 Gaining insights from The Design of Everyday Things

One of the key insights from the book *The Design of Everyday Things* is that human behaviour cannot be effectively altered through coercion or unrealistic expectations. This highlights the necessity of designing systems that are aligned with human cognition and typical behavioural patterns. This principle is particularly relevant in cybersecurity, where employees often face competing demands for attention, leading to security fatigue and reliance on automatic actions that attackers exploit. In cybersecurity, these "everyday things" refer to the security policies and systems that employees interact with. Therefore, the principle of designing systems based on an understanding of human behaviour—rather than relying on coercion or intensive training to modify human actions—should be pursued to promote positive change towards secure behaviour.

2.1 *To err is human* – non-compliant employee or bad security design?

Human error is frequently cited as the cause of many cyberattacks. However, it is often prematurely concluded that the issue ends with human error. A crucial question to consider, as Norman emphasizes, is *why* these human errors occur. To effectively address human error, it is imperative to understand its root causes that can enlighten about the type of the error. This is especially evident in cybersecurity incidents, such as phishing attacks, which are often attributed solely to user negligence. Such assessments halt root-cause analysis at the point of human error.

As Norman describes, while ultimately executing an action, an error can occur either as an *action-slip* or as a *mistake*. In cybersecurity, an action slip occurs when habitual actions override intended ones, such as opening a phishing email attachment or mistakenly sending sensitive data to the wrong recipient due to autofill. Notably, *description similarity slips*, are particularly relevant; like clicking on a phishing link may resemble the routine action of clicking a trusted link. Conversely, a *mistake* involves sharing credentials to attackers due to misjudgement of trustworthiness indicating a false intention of doing the typical action. To address human error in cybersecurity, one must delve into the "why" behind actions to determine whether they stem from automatic tendencies or poor choices. Both types of errors can arise from poorly designed systems.

Users may deliberately violate security policies due to poorly designed systems that hinder compliance. Norman notes that ineffective designs can lead users to bypass security measures to meet their needs, often prioritizing work over policy adherence, resulting in *deliberate violations*. Awareness is not necessarily the issue as employees could intentionally choose to override policies due to work priorities and these *deliberate violations* are often dismissed as recklessness. Security measures that neglect user experience push individuals toward insecure behaviors like password reuse or disabling security features. Consequently, employees may resort to intuitive actions that support their workflows, ignoring security protocols despite being trained on them [KFR18]. Therefore, re-training the employees with a blame-and-train approach will be ineffective when deliberate violations stem from complex security policies.

Norman also warns of hindsight bias that skews perceptions of security non-compliance, leading to the mislabeling of such slips and violations that resulted from overly restrictive policies or routine behaviour patterns as negligence following an incident. This mindset is prevalent in the cybersecurity field, that focusses on retrospective evaluations rather than proactive planning. Organizations should recognize these errors as indicators of design failure rather than personal failings and adjust policies to support secure behavior without imposing unnecessary friction.

2.2 Information overload burdens the user

The aspect of cognitive load is a crucial factor that is especially relevant in the realm of

cybersecurity given how fluctuating this can be in the day-to-day work environment. Norman argues that high cognitive load can significantly contribute to errors, leading to security fatigue. This fatigue restricts employees' ability to act securely, as high cognitive load creates a narrow focus where individuals rely on automatic actions that may diverge from intended security policies [No22, Pa16]. Organizations often respond with intensive training that increases the burden through a blame-and-punish approach, telling employees to "be more careful," which fails to scale and does not prevent non-compliant behavior. Another prevalent design flaw in cybersecurity is the overuse of security warnings. While Norman emphasizes the importance of affordances and signifiers in guiding user behavior, excessive warning messages can overwhelm employees. Constant system warnings and policy reminders may lead to a "click-through" habit, where individuals disregard essential security steps, facilitating non-compliant behavior. Additionally, constant exposure to warnings can desensitize users, leading to risky behaviors that could have been prevented if warnings were applied more judiciously. Consequently, users may perceive warnings as excessive or unnecessary, undermining their intended role as affordances or signifiers for secure actions.

3 Applying Design Thinking in Cybersecurity

The above-mentioned insights from *The Design of Everyday Things*, can be directly applied within the cybersecurity context especially in the domain of exploring the human error and designing security policies that enable intuitive secure actions.

3.1 Error-tolerant systems

First and foremost, it is important to acknowledge that human error is inevitable and that many errors – slips or mistakes can be prevented through design. Adaptive security policies should guide users toward recovery rather than penalizing failure, an approach Norman describes as "designing for error." By introducing restrictions or limitations in a system already through its design, what Norman refers to as *constraints*, it can help guide users towards the intended secure action. Errors can be prevented by halting action at early stages of interaction with the policy or a security system, allowing for reflection and correction. Effective security policies embed constraints that subtly steer users from unsafe behaviour. Norman discusses the significance of natural mappings in error prevention, suggesting that security design should provide intuitive corrections that align with natural workflows and integrate seamlessly with existing behavioural actions. To address specific types of errors, such as action slips, design adaptations can force a change in routine procedures when encountering malicious threats, reducing the likelihood of habitual responses. Rather than assuming users can avoid errors through training alone, cybersecurity should prioritize better feedback, intuitive design, and safeguards that minimize the impact of human fallibility. Cognitive load management is also crucial; security policies and protocols should reduce mental strain, ensuring

compliance. Intuitive interfaces and minimal decision fatigue enhance security adherence without imposing excessive cognitive burdens. This principle applies to warning announcements. A well-designed warning signifier (e.g., an icon or color) can help users quickly grasp the importance of the warning and the necessary actions. Norman aptly states that, “Designers must strive to find a balance between providing necessary warnings and overwhelming users with unnecessary information," to prevent warning desensitization and negligence. By leveraging affordances, reducing cognitive load, and embedding security into workflows, organizations can create policies that are both effective and user-friendly.

3.2 Enhancing security culture

Another critical factor in fostering secure and compliant behavior is the environmental influences that shape motivation and emotional engagement with intended secure actions. Norman posits that design should encompass not only functionality and usability but also emotional responses and experiences. Security training and policies must be crafted to resonate with emotional intelligence, making security personally meaningful. This approach enhances satisfaction, loyalty, and overall user experience. By emphasizing positive reinforcement over punitive measures, organizations can cultivate intrinsic motivation for secure behavior [VD15]. Positive reinforcement encourages the repetition of desirable actions rather than focusing on mistakes through blame-based approaches. Feedback loops can support these positive reinforcements, helping individuals understand correct actions and increasing motivation. A robust security culture should be established as a collective norm, where secure behaviour is perceived as an inherent organizational value rather than an externally imposed obligation. This can be achieved by empowering users to take responsibility for their interactions with security systems through clear and transparent policies. Providing visibility into security operations, particularly regarding the design process, is essential so that the people feel personally involved and thus, responsible. Norman’s emphasis on emotional engagement in learning reinforces the importance of narratives, scenarios, and experiential learning in embedding security within one’s own identity.

4 Guidelines to implement Human-Centered Security Policies

Implementing human-centred security policies thus requires a structured approach that aligns with organizational workflows, minimizes friction, and fosters a culture of security. The following steps provide a pragmatic guide to designing and deploying security policies that align with Norman’s principles and human behavioral tendencies:

Conducting User Research: Understanding user behavior is crucial for designing effective security policies. Organizations should conduct stakeholder analysis, surveys, and user experience testing to identify common security pain points, habitual

workarounds, in security protocols. Observing users in real-world work environments can provide insights into natural behavior patterns that should inform policy design. **Prototyping and Testing Policies:** Security policies should be treated as dynamic prototypes rather than rigid mandates. By developing low-friction security solutions and testing them with small user groups, organizations can evaluate user experience, compliance rates, and unintended consequences before full-scale implementation. Iterative testing ensures that security policies evolve to meet both security and user experience needs. **Training Through Experiential Learning:** Traditional security training is often ineffective because it lacks engagement and real-world relevance. Security policies should be reinforced through hands-on experiential learning, simulations, and scenario-based training. Employees should experience simulated security threats in a controlled environment to better internalize secure practices. **Establishing Feedback Loops:** Security policies should be continuously refined based on real-world application. Organizations should create feedback mechanisms where employees can report security challenges, inefficiencies, or workarounds without fear of reprisal. Regular user feedback sessions help identify gaps and ensure that security measures remain practical and effective. **Measuring and Adjusting Policies:** Key performance indicators (KPIs) should be used to evaluate the effectiveness of security policies. Metrics such as policy adherence rates, frequency of security incidents, and user-reported friction points should be analyzed regularly. Security policies should remain flexible, adapting to emerging threats, technological advancements, and evolving user behavior.

5 Discussion

Integrating design thinking and user-centric principles offers a fresh perspective for enhancing cybersecurity. This paper primarily emphasizes on the need for integrating design thinking principles into improving cybersecurity. With Don Norman's book *The Design of Everyday things* emphasizing on the idea on how systems need to be designed for the human and not human redesigning themselves for a system, it provides resources to have a guideline to integrating design thinking principles into cybersecurity. This fosters and builds the idea that security policies should facilitate, rather than hinder, user activities thus naturally encouraging secure behaviour without introducing unnecessary complexity.

5.1 Beyond usable security

The approach outlined here extends well beyond traditional studies in *usable security*, which primarily focus on improving user interfaces, authentication mechanisms, and reducing friction in security interactions [DTO23]. While the field of usable security has made significant strides in enhancing the user experience of security systems—by refining authentication methods, aiding developers in usability improvements, and

influencing user security behaviour through design strategies — these efforts often operate within the constraints of existing security paradigms rather than rethinking the fundamental structure of security policies themselves. In contrast, the proposed framework leverages Norman's cognitive and behavioural design principles to reimagine security policies at a systemic level. Rather than merely improving compliance through better user experiences, this approach integrates human error tolerance, workflow alignment, and intrinsic motivation to create policies that are not just *usable* but *intuitively adoptable* by users. By embedding security measures seamlessly into existing work processes and leveraging natural affordances, this framework moves beyond surface-level usability enhancements and addresses the root causes of policy circumvention, to foster sustainable, long-term security behaviour. This shift from static usability testing to an adaptive, *human-centred security policy design* represents a significant departure from conventional approaches, positioning this framework as a next-generation model for cybersecurity policy design.

5.2 The Security vs. User Experience Trade-off: A False Dichotomy

The conventional argument that security and user experience must always be traded off against each other is flawed because it assumes that one is simply reducing as the other increases. However, a security policy that is highly secure in theory but so ill-conceived that users routinely circumvent is not secure at all. Security is not measured by theoretical robustness but by its real-world implementation. A policy that accounts for actual user behaviour, cognitive load, and work processes—thereby ensuring compliance—is objectively more secure than a stricter policy that is routinely ignored, bypassed, or undermined through workarounds. A policy that balances security with human-centred design does not trade security for user experience — it ensures security by making compliance intuitive and natural. This approach recognizes that user experience is not a competing priority but a fundamental enabler of security.

6 Conclusion

Expanding the horizons into understanding the various cognitive factors in play during work can throw more light on how to enable employees to behave securely through security policy design. Traditional security policies often fail due to their misalignment with human cognitive behaviour, leading to non-compliance and circumvention. Insights from *The Design of Everyday Things* reveal fundamental principles on what influences behavioural actions. Instead of treating violations as human errors, organizations should view them as signs of design failure and explore root causes in depth, adjusting policies to support secure behavior without unnecessary friction. This paper presents a holistic framework built on the design thinking principles from *The Design of Everyday Things* and provides a roadmap for designing security policies that are intuitive, user-friendly, and effective. By focusing on human-centred design, error tolerance, and behavioural

integration, cybersecurity policies can promote secure behaviour without overwhelming users, thus fostering a secure environment.

Bibliography

[BSN19] Bada, M.; Sasse, A, M. ; Nurse, J, R, C: Cyber Security Awareness Campaigns: Why do they fail to change behaviour?, arXiv:1901.02672 [cs], 2019.

[DTO23] Di Nocera, F.; Tempestini, G.; Orsini, M.: Usable Security: A Systematic Literature Review. In: *Information* Bd. 14, Nr. 12, S. 641, 2023.

[KFR18] Kurowski, S.; Fähnrich, N.; Roßnagel, H.: On the possible impact of security technology design on policy adherent user behavior-Results from a controlled empirical experiment. In: *SICHERHEIT 2018*, 2018.

[No22] Nobles, C.: Stress, Burnout, and Security Fatigue in Cybersecurity: A Human Factors Problem. In: *HOLISTICA – Journal of Business and Public Administration* Bd. 13, Nr. 1, S. 49–72, 2022.

[No09] Norman, D, A.: The Way I See It, When security gets in the way. In: *Interactions* Bd. 16, Nr. 6, S. 60–63, 2009

[No13] Norman, D, A.: The design of everyday things. Rev. and expanded edition. Cambridge (Mass.) : MIT press, ISBN 978-0-465-05065-9, 2013.

[PE08] Payne, B, D.; Edwards, W. K.: A Brief Introduction to Usable Security. In: *IEEE Internet Computing* Bd. 12, Nr. 3, S. 13–21, 2008

[Pa16] Parkin, S; Krol, K.; Becker, I.; Sasse, M. A.: Applying Cognitive Control Modes to Identify Security Fatigue Hotspots. In: *(Proceedings) Workshop on Security Fatigue, [part of] SOUPS 2016: Twelfth Symposium on Usable Privacy and Security*, 2016.

[SF05] Sasse, M,A.; Flechais, I.: Why Do We Need It? How Do We Get It? In: Cranor, LF and Garfinkel, S, Security and Usability: Designing secure systems that people can use, S.13-30, 2005.

[VD15] Vlaev, I.; Dolan, P.: Action Change Theory: A Reinforcement Learning Perspective on Behavior Change. In: *Review of General Psychology* Bd. 19, Nr. 1, S. 69–95, 2015.

[ZR19] Zimmermann, V.; Renaud, K.:Moving from a ‘human-as-problem” to a ‘human-as-solution” cybersecurity mindset. In: *International Journal of Human-Computer Studies* Bd. 131, S. 169–187, 2019

Making Security Matter: Engaging Non-Security Users Effectively through Empathy

Anne Elisabeth Krueger[1]

Abstract: This position paper highlights the importance of empathetic communication in information security systems, so that cyberattacks can be effectively defended against at the vulnerable points of the security organization, the non-security users. By systematically integrating empathy, trust between non-security users and security users can be enhanced, resulting in increased mutual engagement and commitment, which in turn encourages safer behaviors from non-security users. To this end, initial principles are introduced to guide the systematic integration of empathy into the interactions between security and non-security users. With the overarching goal that organizations can potentially become more resilient and less vulnerable to cyber threats over time, this position paper aims to spark discussions and provide insights for designers of information security systems.

Keywords: Information Security System, Trust, Empathy, Experience Design

1 Introduction

The significant number of security incidents in organizations worldwide underscores the necessity for non-security users within an organization to engage with the topic of information security. Non-security users in departments such as finance, HR, and marketing are often the first line of defense against cyber threats [KCF21]. Many security breaches occur due to human error, such as clicking on phishing links, using weak passwords, or mishandling sensitive data. Therefore, it is essential to equip these employees with the knowledge to recognize and prevent such risks. However, Information Security is often perceived by non-security users as an additional task that interferes with their core productive activities, as potential attacks are viewed as distant and outside their immediate responsibilities.

A promising approach to address these issues seems to promote a trustful relationship-building between non-security users and security users [KCF21]. Establishing trust is essential, as it ensures both parties that they act in each other's best interest. This mutual trust then forms the foundation for upholding agreements and, for instance, non-security users adhering to security policies [KCF21]. In addition to factors such as transparent communication and consistent behavior, empathy seems crucial to building trust [Ra16; WJW22] (see Figure 1). Defined as *the ability to imagine others' emotional experiences while maintaining a clear distinction from one's own [CG11], empathy fosters understanding and creates the foundation for recognizing others' needs and perspectives*. Furthermore, the perceived empathy of the respective other simultaneously makes one feel understood and

[1] Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering IAO, Nobelstr. 12, 70569 Stuttgart, Germany, anne-elisabeth.krueger@iao.fraunhofer.de, https://orcid.org/0000-0002-3472-1186

doi:10.18420/OID2025_12

acknowledged, fostering a deeper, more respectful connection and, in turn, strengthening trust [PK12; WJW22].



Fig. 1: Empathy builds trust, leading to stronger engagement.

Thus, if security users were supported and actively sought to empathize with the perspective and emotional experience of non-security users, and this was recognized as such, an atmosphere of trust could be established. This, in turn, could lead to safer organizations, as non-security users feel obligated to remain vigilant against potential attacks due to their close relationship with the security department.

Based on these considerations, this paper pursues the following objectives:

- To highlight that empathetic communication in information security fosters trust and promotes safer behavior,'making security matter.'
- To outline key principles for designing security systems that incorporate empathy, strengthening mutual commitment, and long-term cyber resilience.

2 Food for Thought: Design Principles for Information Security Systems to Foster Empathy

Three fundamental principles to systematically introduce empathy are proposed to initiate discussion and provide food for thought for designers and other relevant stakeholders of information security systems. The aim is to support empathy and mutual understanding of their emotional experiences between non-security and security users, fostering a close relationship that encourages commitment and, in the long term, enhances organizational security. Thereby, it has to be noted, that the presented principles do not claim to be exhaustive.

2.1 Principle 1: Establish an Atmosphere of Emotional Safety

Creating an emotionally safe environment [Ed20; Ed99] is key to fostering empathy between security and non-security users. Both stakeholder-groups must feel comfortable expressing their opinions, concerns and emotional experiences without fear of judgment or retaliation. For instance, if a non-security user receives a phishing email and is uncertain about its legitimacy, they should feel comfortable asking security users without embarrassment. Likewise, they should be able to express doubts about the necessity of a security measure without hesitation. Only when prejudices are minimized and active listening is practiced such an open communication is possible, where individuals are willing and able to freely share and connect to their true (psychological) needs to build mutual understanding of the others' experience and thus, to develop empathy. In turn, an emotionally safe environment promotes trust and collaboration, ultimately strengthening the security culture and organization.

Concrete approaches, methods and inspiration on how to create emotional security in organizations can be found in the literature [Ed20; Ed99].

2.2 Principle 2: Enable both Imaginative Empathy Processes

For a security user to genuinely understand the impact of established policies on a non-security user, mere observation and external analysis are not sufficient (cf. [Er16]). Instead, they should actively engage and immerse themselves in the non-security user's experience, linking it to their own emotional perspective to develop a deeper understanding. Thus, to systematically integrate empathy into information security systems, both imaginative processes from empathy research should be considered [Da04]: In the *Imagine-Other* process, people are encouraged to consider what someone else might feel or think in a given situation. In contrast, the *Imagine-Self* process involves imagining how they themselves would feel if they were in the other person's place. While the *Imagine-Self* perspective has a stronger emotional impact on the person adopting it, making it easier to empathize with others' emotional experiences [Er16], the *Imagine-Other* perspective provides an essential level of abstraction to maintain the necessary distinction from one's own experiences (cf. [CG11]).

To support the *Imagine-Other* process, there exist methods and tools in experience design, such as Experience Interviews [Ze16], Experience Probes [Kr20] or the Needs Empathy Map [Fe15; Kr22; OP11]. For the *Imagine-Self* process, the Needs Perspectives, a newer form of the method UX Concept Inspiration [KFP15], can be applied.

2.3 Principle 3: Support the Development of Emotional Sensitivity

For security users to systematically build empathy with non-security users, it is crucial that they are attuned to their own emotional experiences (cf. [YDP13]). To understand another's

emotional experience and differentiate it from one's own, it is essential to connect with one's psychological needs and emotions, thereby developing emotional granularity and sensitivity to one's personal lived experience [CG11; Kr17; YDP13]. Thereby, it is assumed that an individual's emotional experience is based on the fulfillment or non-fulfillment of their psychological needs [Ha03; Ha13], such as for Competence, Stimulation or Connection. For example, the experience of privacy, a fundamental component of information security systems, is subject to varying perceptions based on the underlying psychological needs and associated individual cognitive frameworks [KB24]. Here, the psychological need for *Comfort* might require easy access to privacy settings, while the need for *Connectedness* necessitates the sharing confidential information in a secure environment with the aim of building relationships.

In experience design research, various tools have already been created, which can possibly facilitate this development of emotional sensitivity (e.g. [Bu23; DF20a; DF20b; DF21; DF22; Kr17; YPD15]) and thus can be used and adapted for information security systems.

3 Outlook

The overarching goal of the position paper is to *making security matter* to non-security users in organizations. For this purpose, three initial principles for the systematic integration of empathy into information security systems are introduced. This is done under the assumption that it will, in turn, lead to more trust and ultimately result in safer behavior by non-security users - making organizations less vulnerable to cyberattacks.

Additional research is planned to validate these principles, refine them if necessary, and develop tailored methods to support empathic behaviors in information security systems. For this purpose, the above-mentioned methods and materials from experience design research will be used and adapted accordingly for the context of information security systems.

Furthermore, perceived empathetic behavior can create a positive emotional experience for relevant stakeholders (cf. [WJW22]). Fostering positive emotions promotes motivation, creativity and resilience [Fr04]. Consequently, among non-security users perceived empathetic behavior of security users might lead to an increased motivation to engage in protective behaviors, thereby enhancing overall security compliance. In turn, stimulating negative emotional experiences might have the contrary effect. Thus, consideration should be given to how positive experiences can be incorporated into information security systems.

The objective is not merely to adopt a task-based, technocratic perspective on the relevant stakeholders in the information security system but to shift the focus on their felt-life, communication and relationships. This focus seems essential for effective and equitable collaboration of security and non-security users (cf. [KFR18]). Only by comprehending the entire information security system, its interdependencies and emotional levels organizations can create sustainably secure environments. Thereby, it seems important to understand

that, although the human factor can pose risks for potential security breaches, it should be regarded as an integral and essential part of a functioning information security process.

Acknowledgments. The position paper is part of the activities in the "Mittelstand-Digital Zentrum Fokus Mensch" (MDZ-FM). The MDZ-FM is part of the Mittelstand-Digital initiative funded by the German Federal Ministry for Economic Affairs and Climate Action, funding number: 01MF23003E.

References

[Bu23] Burmester, M. et al.: Bedürfnismaterialien, 2023, https://www.digitalzentrum-fokus-mensch.de/wissen/beduerfnismaterialien, Stand: 01/05/2024.

[CG11] Coplan, A.; Goldie, P.: Empathy. Oxford University Press Oxford, 2011.

[Da04] Davis, M. H. et al.: Cognitions associated with attempts to empathize: How do we imagine the perspective of another? Personality and Social Psychology Bulletin 30 (12), pp. 1625–1635, 2004.

[DF20a] Desmet, P.; Fokkinga, S.: Beyond Maslow's Pyramid: Introducing a Typology of Thirteen Fundamental Needs for Human-Centered Design. Multimodal Technologies and Interaction 4 (3), Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 38, 2020.

[DF20b] Desmet, P.; Fokkinga, S.: Thirteen chairs, thirteen fundamental needs. 2020.

[DF21] Desmet, P.; Fokkinga, S.: The House of Happiness. Inspiration Poster, 2021, https://diopd.org/wp-content/uploads/2021/05/House-of-happiness-screen-scaled.jpg, Stand: 10/17/2022.

[DF22] Desmet, P.; Fokkinga, S.: Thirteen fundamental needs, 2022, https://needtypology.com/.

[Ed20] Edmondson, A. C.: Die angstfreie Organisation: wie Sie psychologische Sicherheit am Arbeitsplatz für mehr Entwicklung, Lernen und Innovation schaffen. Vahlen, 2020.

[Ed99] Edmondson, A.: Psychological safety and learning behavior in work teams. Administrative science quarterly 44 (2), pp. 350–383, 1999.

[Er16] Erle, T. M.: A grounded approach to psychological perspective-taking, PhD thesis, Universität Würzburg, 2016.

[Fe15] Ferreira, B. et al.: Designing Personas with Empathy Map. In: SEKE. Vol. 152, 2015.

[Fr04] Fredrickson, B. L.: The broaden–and–build theory of positive emotions. Philosophical transactions of the royal society of London. Series B: Biological Sciences 359 (1449), pp. 1367–1377, 2004.

[Ha03] Hassenzahl, M.: The thing and I: understanding the relationship between user and product. In: Funology. Springer, pp. 31–42, 2003.

[Ha13] Hassenzahl, M. et al.: Designing moments of meaning and pleasure. Experience design and happiness. International journal of design 7 (3), 2013.

[KB24] Krueger, A. E.; Brandenburg, S.: Learnings from a Guided Method for Experience Design: Psychological Needs in the Context of the Privacy Value. In: Open Identity Summit 2024. Gesellschaft für Informatik eV, pp. 115–126, 2024.

[KCF21] Kurowski, S.; Cetin, F.; Fischer, R.: Why should they care? Conceptualizing the challenges of information security training. In: Open Identity Summit 2021. Gesellschaft für Informatik eV, pp. 199–204, 2021.

[KFP15] Krüger, A. E.; Fronemann, N.; Peissner, M.: Das kreative Potential der Ingenieure–menschzentrierte Ingenieurskunst. In: Stuttgarter Symposium für Produktentwicklung, SSP 2015 Entwicklung smarter Produkte für die Zukunft. Pp. 1–10, 2015.

[KFR18] Kurowski, S.; Fähnrich, N.; Roßnagel, H.: On the possible impact of security technology design on policy adherent user behavior-Results from a controlled empirical experiment. In: SICHERHEIT 2018. Gesellschaft für Informatik eV, pp. 145–158, 2018.

[Kr17] Krueger, A. E. et al.: Needs profile: sensitising approach for user experience research. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction - OZCHI '17. ACM Press, Brisbane, Queensland, Australia, pp. 41–48, 2017, http://dl.acm.org/citation.cfm?doid=3152771.3152776, Stand: 11/30/2019.

[Kr20] Krueger, A. E. et al.: Guided user research methods for experience design—a new approach to focus groups and cultural probes. Multimodal technologies and interaction 4 (3), p. 43, 2020.

[Kr22] Krueger, A. E.: Two methods for Experience Design based on the Needs Empathy Map: Persona with Needs and Needs Persona. In (Marky, K.; Grünefeld, U.; Kosch, T., eds.): Mensch und Computer 2022 - Workshopband. Gesellschaft für Informatik e.V., Bonn, 2022.

[OP11] Osterwalder, A.; Pigneur, Y.: Business Model Generation: Ein Handbuch für Visionäre, Spielveränderer und Herausforderer. Campus Verlag, 2011.

[PK12] Pavlovich, K.; Krahnke, K.: Empathy, connectedness and organisation. Journal of business ethics 105, pp. 131–137, 2012.

[Ra16] Rahman, W.: Empathy and trust: Into a better workplace environment. Journal of Business and Economics 7 (12), pp. 2025–2034, 2016.

[WJW22] Wu, Q.; Jin, Z.; Wang, P.: The relationship between the physician-patient relationship, physician empathy, and patient trust. Journal of general internal medicine 37 (6), pp. 1388–1393, 2022.

[YDP13] Yoon, J.; Desmet, P. M.; Pohlmeyer, A. E.: Embodied typology of positive emotions: The development of a tool to facilitate emotional granularity in design. In: IASDR 2013; 5th international congress of international association of societies of design research, Tokyo, Japan. Pp. 1195–1206, 2013.

[YPD15] Yoon, J.; Pohlmeyer, A.; Desmet, P.: Positive emotional granularity cards. 2015.

[Ze16] Zeiner, K. M. et al.: Das Erlebnisinterview–Methode zum Verständnis positiver Erlebnisse: Praktische Einführung und Diskussion von Implikationen zur UX-Gestaltungspraxis. In: UP 2016. Gesellschaft für Informatik eV und die German UPA eV, pp. 10–18420, 2016.

Security awareness versus secure behaviour: A bibliometric study of the state of research on human factors in IT Security

Salma Ben Mamia [1], Christian Heinrich Schunck[2], Mrudula Arunkumar[3] and Heiko Roßnagel[4]

Abstract: This bibliometric analysis examines the evolving role of human behavior and IT practices in shaping cybersecurity research and strategies. By leveraging VOSviewer, we analyse a comprehensive dataset of Scopus articles to explore key themes, keyword co-occurrence, and the intersection between human vulnerabilities and technological solutions. Our findings reveal a persistent gap in addressing the human factor, with keywords such as 'human error,' 'social engineering,' and 'awareness' frequently emerging as critical vulnerabilities. The analysis also highlights the inadequate integration of human-centric solutions by IT professionals, as evidenced by weak connections between 'employee awareness,' 'cyber hygiene,' and 'cybersecurity culture.' Despite advancements in technology, the study underscores that cybersecurity's effectiveness is often compromised by human behavior, suggesting a need for deeper engagement with user training, organizational culture, and behavioral change. This bibliometric analysis contributes to understanding the critical blind spots in cybersecurity research, offering valuable insights into the overlooked relationship between human behavior and IT practices in safeguarding digital infrastructures.

Keywords: Human Factor, Design thinking, Cybersecurity, Human Behaviour, Scopus, VOS viewer.

1 Introduction

This short paper presents a bibliometric analysis of the intersection between human behavior and IT practices in cybersecurity research. By analysing keyword co-occurrence patterns and the evolution of cybersecurity literature, we aim to uncover critical gaps in current research and identify areas where human-centric solutions and IT practices can be better integrated. Through this analysis, we contribute to the ongoing conversation around improving cybersecurity defences by considering both technological and behavioral perspectives. Thus, in this paper the focus will be on

[1] Fraunhofer IAO, Nobelstr. 12 70569 Stuttgart, Lernlabor Cybersicherheit, salma.ben.mamia@iao.fraunhofer.de, https://orcid.org/ 0000-0002-7213-9007
[2] Fraunhofer IAO, Nobelstr. 12 70569 Stuttgart, Lernlabor Cybersicherheit, christian.schunck@iao.fraunhofer.de
[3] Fraunhofer IAO, Nobelstr. 12 70569 Stuttgart, Lernlabor Cybersicherheit, mrudula.arunkumar@iao.fraunhofer.de
[4] Fraunhofer IAO, Nobelstr. 12 70569 Stuttgart, Lernlabor Cybersicherheit, heiko.rossnagel@iao.fraunhofer.de

doi:10.18420/OID2025_13

having a systematic overview through the bibliometric analysis of human factors and awareness in cybersecurity and explore the research themes and the gaps present concerning the human factors. Things, scope for future research can be examined with intended direction of capitalising on the research and practical application gaps.

2 Analysing the human factor research in cybersecurity

The integration of human factors in cybersecurity takes place in different forms ranging from awareness campaigns to training. These activities often do not result in behavioural change. Performing a bibliometric analysis, shall provide an overview that is needed to understand if there lies a gap between what the research reveals and what the principles of human centred design suggest.

An analysis of the most researched area of cyber security of published documents, then cyber security and human factor, then cyber security and awareness from 2015 to 2024, is presented. The first search term was "cybersecurity," the second "cybersecurity and human factor," and the third "cybersecurity and awareness." An analysis of published documents from 2015 to 2024 indicates that cybersecurity is the most researched area, followed by cybersecurity and the human factor, and then cybersecurity and awareness.

In doing so, the study uses the Scopus Core Collection database, using visualization of similarities (VOS) viewer software which is [Ke63] a freely available software that collects the bibliometric data for analysis and visualization by using several bibliometric indicators including bibliometric coupling, citation [Me18], co-citation analysis [Br87], co-authorship [SUP15] and co-occurrence of keywords [La17]. The VOS viewer is particularly useful for displaying large bibliometric maps in easily interpreted ways. The VOS viewer is particularly useful for displaying large bibliometric maps in easily interpretable ways.

2.1 A bibliometric analysis using cyber security as Keyword

Fig.1 is a VOSviewer map centred on "cyber security", we found 220000 related publications on Scopus, 20000 are analysed. It is expected to see technical themes (blockchain, encryption, digital technology) form a robust central cluster due to high publication volume and cross-citation. Conversely, "awareness," "human factor," and "phishing" tend to be more weekly represented in these networks, likely reflecting both lower publication volume in core computer science venues and the interdisciplinary dispersion of behavioural studies. This does not highlight a stronger academic focus on technological solutions in cybersecurity research than on human factors. Overall, the weaker presence of human factor topics in bibliometric analyses reveals a need for more integrative research and emphasizes the opportunity for interdisciplinary collaboration that addresses both technical rigor and user-centric vulnerabilities in the cybersecurity landscape.

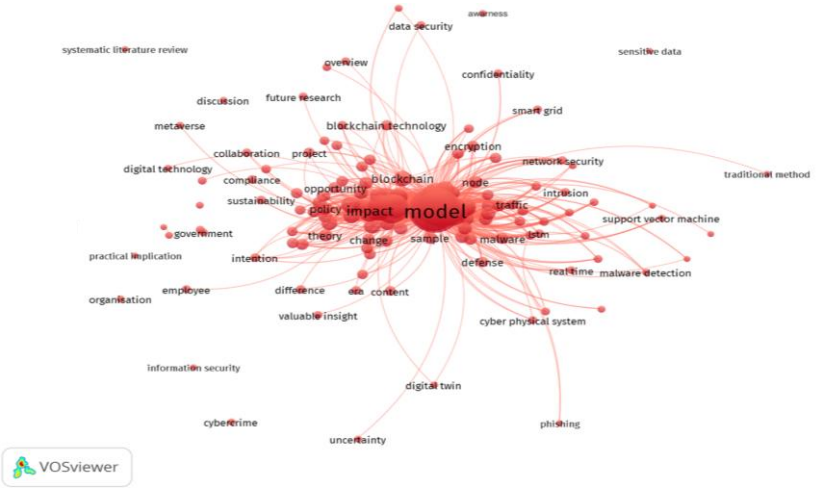


Fig. 1: VOSviewer Co-occurrence map for cybersecurity as keyword for a dataset of 20000 publications

2.2 A bibliometric analysis using cyber security and human factor as Keywords

In a refined VOSviewer analysis that specifically filters for “cyber security” and “human factor,” the overall publication count drastically decreases from 220,000 to around 16,000 articles, underlining a substantial gap in scholarship that directly addresses the human element of cyber threats. Even within this smaller corpus, the map in Fig.2 reveals that technical or innovation-driven keywords—such as “game,” “advancement,” “experiment,” “prediction,” “agent,” “defence,” “account,” “sustainability,” “ChatGPT” “evolution,” and “autonomous behavior”—form the dominant cluster alongside “cyber security,” whereas crucial user-centric terms like “human behavior,” “social engineering,” and “phishing emails” lie on the periphery with weaker co-occurrence links. This pattern emphasizes that although technical and emerging AI-driven research domains remain a focal point of cybersecurity studies, the human factor—despite its importance in real-world threat mitigation—still occupies a comparatively smaller share of the literature and is less integrated into the broader research discourse.

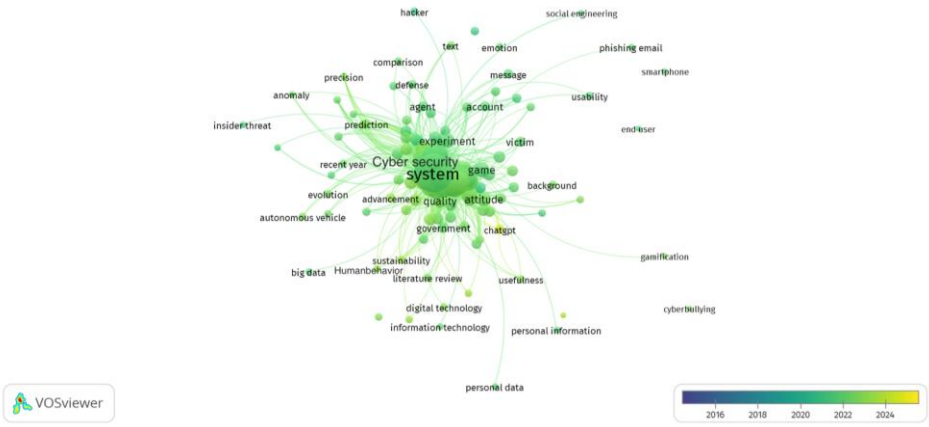


Fig. 2: VOSviewer Co-occurrence map for cyber security and human factor as keyword for a dataset of 16000 publications

2.3 A bibliometric analysis using cyber security and awareness as Keywords

In this more focused VOSviewer search combining “cyber security” and “awareness,” the publication pool contracts dramatically to 2900 articles, signalling that awareness-oriented cybersecurity research is relatively underexplored. Within this dataset, terms such as “cybersecurity,” “IoT,” and “internet” cluster more closely with “technology” and “industry,” suggesting a stronger focus on technical countermeasures, corporate policies, and tactical defences against human-targeted threats. Meanwhile, “cybersecurity awareness” forms a separate cluster linked to terms like “phishing attack” and “user,” indicating a possible disconnect between behavioral insights and practical security strategies. This fragmentation underscores the need for more integrative research that bridges behavioral science with organizational policies and training frameworks, ensuring that cybersecurity awareness is strengthened both at the individual level and within broader systems.

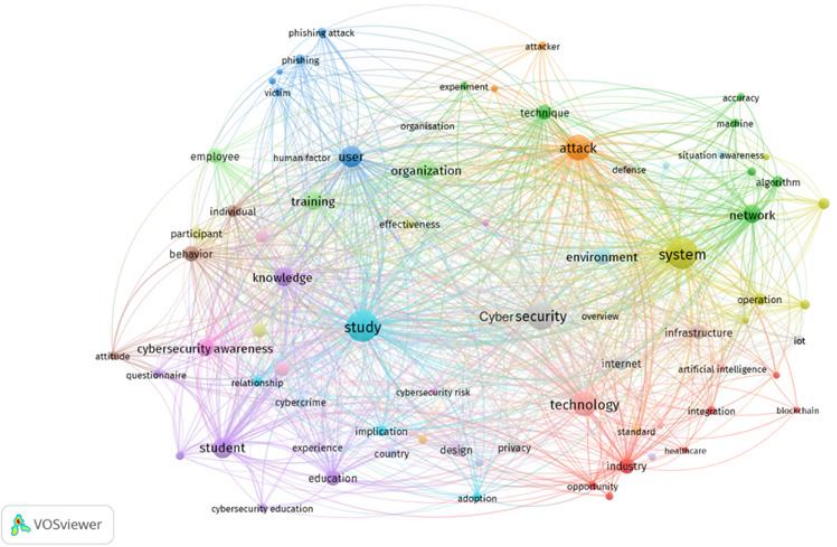


Fig. 3: VOSviewer Co-occurrence map for cybersecurity and awareness as keywords for a data set of 2900 publications

3 Social structure using cyber security and awareness as Keywords, generated by Biblioshiny (R-Studio)

In Fig.4, the collaboration network generated by Biblioshiny (R Studio) [AC17] which was based on the 3040 articles found on Scopus using “awareness” and “cybersecurity” as keywords, the USA leads cybersecurity and awareness research, forming the largest cluster with a strong collaboration with Canada, reflecting shared institutional efforts and policy development. The UK follows as the second dominant country, playing a key role in global cybersecurity initiatives. Meanwhile, Germany and Portugal exhibit a strong bilateral partnership, likely driven by EU-funded projects and regional cybersecurity strategies. Despite these efforts, there is a need for greater global collaboration and investment in cybersecurity awareness, as emerging threats continue to evolve, requiring stronger policies, education, and research to enhance digital resilience worldwide.

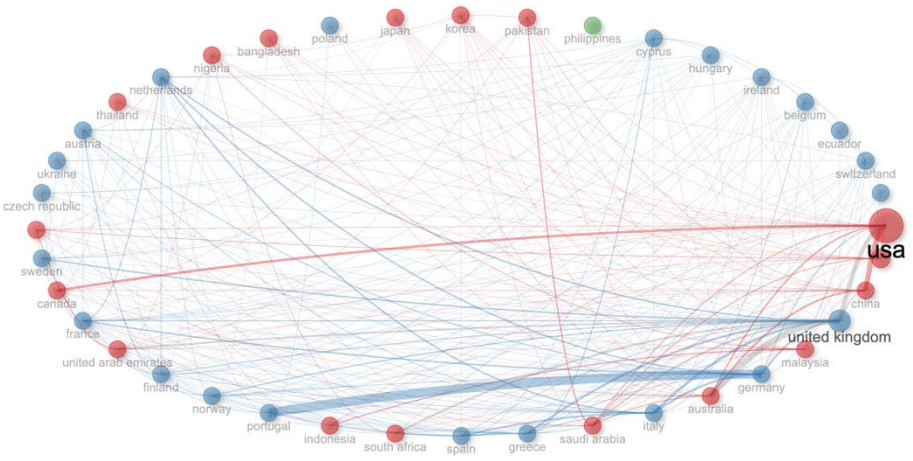


Fig. 4:Countries collaboration for a data set of 2900 publications

4 Conclusion

Overall, these analyses indicate that research on cybersecurity awareness is both relatively limited in volume and geographically uneven, highlighting the need for stronger engagement and collaboration. Germany could strengthen its role by promoting collaborative projects that integrate behavioral research, technical innovation, and organizational policy. Future work might involve developing targeted training programs for individuals and organizations, conducting longitudinal studies on awareness campaign effectiveness, and encouraging cross-disciplinary partnerships to bridge technology, psychology, and sociology. By fostering a coordinated national strategy and supporting diverse research teams, Germany can significantly advance both its own cybersecurity posture and the global understanding of awareness-driven defence measures. Building on this framework, we at the Cybersecurity Training lab keep the human factor in focus by integrating these principles and applying the strategies to training modules that focus on redesign security policies. This further pave the way to assess the user behaviour and collect behavioural data that can be used to measure the effectiveness of security policies.

Bibliography

[AC17] Aria, Massimo; Cuccurullo, Corrado: bibliometrix: An R-tool for comprehensive science mapping analysis. In: Journal of Informetrics Bd. 11 (2017), Nr. 4, pp. 959–975

[Br87] Broadus, R. N.: Toward a definition of “bibliometrics”. In: Scientometrics Bd. 12 (1987), Nr. 5–6, pp. 373–379

[Ke63] Kessler, M. M.: Bibliographic coupling between scientific papers. In: American Documentation Bd. 14 (1963), Nr. 1, pp. 10–25

[La17] Laengle, Sigifredo; Merigó, José M.; Miranda, Jaime; Słowiński, Roman; Bomze, Immanuel; Borgonovo, Emanuele; Dyson, Robert G.; Oliveira, José Fernando; u. a.: Forty years of the European Journal of Operational Research: A bibliometric overview. In: European Journal of Operational Research Bd. 262 (2017), Nr. 3, pp. 803–816

[Me18] Merigó, José M.; Pedrycz, Witold; Weber, Richard; De la Sotta, Catalina: Fifty years of Information Sciences: A bibliometric overview. In: Information Sciences Bd. 432 (2018), S. 245–268

[SUP15] Singh, Vivek Kumar; Uddin, Ashraf; Pinto, David: Computer science research: the top 100 institutions in India and in the world. In: Scientometrics Bd. 104 (2015), Nr. 2, pp. 529–553

FidNET: Flexible Infrastructure for Decentralized Trust Establishment in Industrial Supply Chain Networks

Isaac Henderson Johnson Jeyakumar [1], Michael Kubach [2]

Abstract: As industrial supply chains become increasingly complex, establishing digital trust and transparency is crucial for efficient operations. This paper introduces FidNET, a flexible infrastructure leveraging blockchain technology to enhance trustworthiness and transparency within supply chain networks. By utilizing blockchain's immutable and decentralized ledger, FidNET provides a secure platform for verifying transactions and ensuring the integrity of data shared among stakeholders. A concept and architecture are discussed, showing how blockchain technology can be effectively implemented to build digital trust and improve transparency in industrial supply chains.

Keywords: blockchain, DLT, digital identity, digital trust, industrial supply chain, sovereignty

1 Introduction

The continuing digitalisation and increasing complexity of industrial supply chains confronts organisations with trust and transparency challenges while handling their operations. Moreover, firms in more and more industries are held accountable for sustainability impacts from their supply chain through regulations such as the UK Modern Slavery Act, the EU Supply Chain Act, the German Supply Chain Due Diligence Law – but also through media and NGO attention and customer demand [NKS23]. One approach to managing these sustainability requirements are digital product passports (DPPs). A DPP is a digital tool that registers, processes, and shares product-related information among supply chain entities and other stakeholders. In the EU DPPs will even be required for some products such as industrial and electric vehicle batteries from January 2026 [ZS24]. To be put into practice, such a tool requires a digital identity and trust infrastructure. Hence, in particular for globally interconnected value networks, the establishment of a digital trust infrastructure is pivotal for legally compliant, efficient and secure business processes. However, existing approaches often reach their limits when it comes to ensuring data integrity, scalability, traceability and interoperability.

We address the mentioned challenge with an approach that builds on our work on trust infrastructures for self-sovereign identities (SSI). The TRAIN concept combines decentralized identifiers (DIDs) and Verifiable Credentials (VCs) with the established DNS(SEC) infrastructure [JCK22]. With *FidNET*, we develop this approach further and

[1] Fraunhofer IAO, Nobelstraße 12, 70569 Stuttgart, isaac.johnson@iao.fraunhofer.de, https://orcid.org/0000-0001-9397-1321
[2] Fraunhofer IAO, Nobelstraße 12, 70569 Stuttgart, michael.kubach@iao.fraunhofer.de, https://orcid.org/0000-0001-5105-1113

doi:10.18420/OID2025_14

combine it with blockchain/distributed ledger-based trust anchors. This allows us to present a flexible infrastructure for the decentralized establishment of trust in industrial supply chain networks. Leveraging the European Blockchain Services Infrastructure (EBSI), FidNET aims to offer a solution to strengthen trust and transparency through immutable and decentralized ledgers. By integrating VCs and DIDs as emerging standards, our infrastructure enables secure and efficient management of identities and authorizations within the supply chain.

This paper will first give some background on the state of trust and transparency in industrial supply chains as well as existing approaches and their limitations in section 2. We then present the concept and architecture of FidNET and discuss how it integrates with EBSI as blockchain technology in section 3. Within the limits of this article, we can only present our basic concept with architecture for discussion. A further evaluation with implementation details is subject to further research in an ongoing project.

2 Background and Related Work

The efficient management of globalized industrial supply chain networks and the fulfilment of regulatory obligations increasingly requires the exchange of data on a large scale. Data subjects include machines and sensors, products, organizations and the people associated with them (employees). The secure exchange of information in this context, the creation of trustable certificates and the establishment of the basic identities behind them requires trust management as a basic infrastructure. It must be flexible enough so that it can include diverse requirements, bridge gaps between domains of trust and allow for efficient cooperation in supply chain networks. But how can such a flexible, scalable and interoperable trust infrastructure to manage information about trusted organizations, federations of entities, persons or even individual machines and sensors (in the Internet of Things context) be realized? A central trust anchor with corresponding governance as single gatekeeper seems inappropriate to cater these diverse requirements – a certain degree of decentralization is required. Individual organizations or groups of entities (federations) must be able to define and manage their trust anchors, as well as the rules to establish trust in their context (the trust framework) in a sovereign manner.

In the next section we will have a look at existing approaches to establish digital trust in supply chain networks, before we then analyze their challenges and limitations.

2.1 Approaches to Establish Digital Trust in Supply Chain Networks

In this section, different German and European initiatives are analyzed regarding how they address of trust and transparency in industrial supply chains.

Gaia-X is a European initiative aimed at creating a federated data infrastructure based on European values to ensure data sovereignty in digital value creation ecosystems. In

Germany, Gaia-X has been applied across various domains, including logistics and manufacturing. Gaia-X ensures that data is handled in compliance with European laws, providing a secure framework for data exchange. In manufacturing supply chains, Gaia-X enables the creation of new business models through data sharing

Catena-X [Ca25] is a sovereign data space initiative primarily focused on the automotive industry. It was founded in 2021 by the German Federal Ministry for Economic Affairs and Energy (BMWK) and the German Association of the Automotive Industry (VDA), with support from 28 founding companies. Catena-X aims to improve the flow of parts and materials in automotive networks by creating a transparent and interconnected network across the entire value chain. Key use-cases include demand and capacity management, traceability of production parts, and enhancing sustainability by reducing CO2 emissions.

Manufacturing-X [In24] is an initiative that seeks to digitalize industrial supply chains across various sectors beyond the automotive sector. It aims to create an interoperable data ecosystem where companies can share data while maintaining control over it. It focuses on establishing standards that allow companies to exchange data securely and confidently without losing control over their data. Factory-X [Fa24] is a lighthouse project under Manufacturing-X, specifically targeting the industrial manufacturing sector. Started in 2024 it involves around 50 partners from industry, research, and associations. It emphasizes vertical integration by connecting shop floor operations with supply chain management to enhance efficiency and innovation.

Energy Data-X [Te23] is a consortium funded by the BMWK to build a data ecosystem for the energy industry based on Gaia-X principles. It aims to digitally network sectors like electricity, transport, and heat in a sovereign manner to achieve energy and climate protection goals. Energy Data-X supports the integration of renewable energy plants and enhances flexibility aggregation, particularly in mobility and energy sector.

These initiatives aim to contribute to building more resilient, sustainable, and digitally integrated supply chains in Germany and also enabling a blueprint for secured data exchange around the world.

2.2 Challenges and Limitations of Existing Systems

Initiatives like the ones listed in section 2.1 leverage technologies like VCs and DIDs, aiming to establish secure, decentralized means of establishing trust across stakeholders in the ecosystem. However, they are often less flexible and decentralized as claimed. As an example: in Gaia-X and Catena-X trust is established by centralized authorities like digital clearing houses [DT25] which act as a gatekeeper for the entire ecosystem and trust is established using centralized identifiers like DID:WEB. Hence, there is little room for flexibility for the stakeholders to integrate their own blockchain based identifiers for services in the ecosystem. This illustrates the requirement to develop decentralized approaches for establishing trust in such initiatives for industrial supply chain networks.

3 FidNET Framework

With TRAIN, we have developed an open source concept for trust management based on trust lists. It has been demonstrated in a number of pilots, for example with the Linux Foundation Public Health, RegiTRUST (for the UN Development Programme), and implemented by the Gaia-X Federation Services [Gx25, JK25] that have made the components publicly available [Ec24]. While TRAIN addresses many of the above-mentioned challenges it is currently based on the DNS infrastructure, which is only decentralized to a limited extent. Blockchain and distributed ledger technology (DLT) promise an option to decentralize TRAIN further, make it independent of the DNS and implement new governance structures for trust anchors.

With FidNET we aim to integrate blockchain/DLT-based trust anchors into the TRAIN concept. In a first step we analyzed which blockchain infrastructures could provide a value through integration with TRAIN. We took a closer look at the European Blockchain Services Infrastructure (EBSI) [Eb25], the Ethereum Name Service (ENS) [En25], and Handshake (HNS) [Hn25]. In our analysis of these projects, we evaluated their maturity, publicly available documentation, adoption rates, and overall suitability for integration with the TRAIN concept. After careful consideration, we determined to prioritize first EBSI and then ENS due to their level of maturity and the availability of support resources. By working with EBSI, we aim to integrate a European trust anchor with close alignment to European regulatory frameworks. ENS on the basis of Ethereum on the other hand is a global project, a decentralized autonomous organization (DAO) with focus on building an alternative to DNS. Therefore, it is a good contrast to the euro-centric, semi-centralized EBSI as access-restricted blockchain with a focus on public administration.

ENS and EBSI are global or pan-European infrastructures that are highly available and secured against unauthorized access by their respective mechanisms. Changes are also transparently traceable. In addition, their specific governance means that they are accepted as an anchor of trust by their respective stakeholders. Compared with Bitcoin both blockchains are highly energy efficient as they do not use proof-of-work consensus mechanisms so that sustainability criteria do not stand against their use. Therefore, they seem well suited for integration in FidNet. In the following, we will describe the conceptual design and architecture for the integration with EBSI. The integration with ENS will be subject to a follow-up project.

3.1 Conceptual Design and Architecture

The FidNET framework enables the management of the VC life cycle within an industrial supply chain context. Key functionalities of FidNET (see also Figures 1 and 2) include:

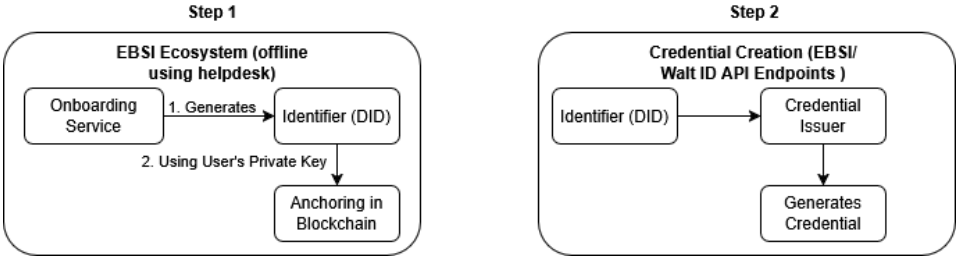


Fig. 1: Process for Creation of EBSI DID and Verifiable Credential

Step 1 - Creation of Identifier: As described in Figure 1, identifiers are crucial for issuing a credential. In the EBSI ecosystem, DID-based identifiers are required at different levels (e.g., Trusted Issuers, Trusted Accredited Organizations (TAOs) and to End-User credentials). This initial process of creating identifiers is offline in the EBSI ecosystem and happens via the EBSI Helpdesk. During this issuance process the DID Document is anchored on the EBSI Blockchain and can be resolved using EBSI DID Resolver [Eb25].

Step 2 - Issuance of Verifiable Credentials: Verifiable Credentials (VCs) are issued to stakeholders in the supply chain ecosystem using for example the Walt ID SDKs. They support EBSI Identifiers and OID4VCI [Lo24] issuance protocols. The credential issuer creates a credential offer. On successful acceptance of the offer at the stakeholder end the credential can be stored in the configured digital wallet.

Step 3 - Validation Process: During the validation process the service provider can request the credentials using the OID4VP [Te24] protocol to validate the credential. The Service Provider also has its own TRAIN instance where it can configure trusted issuer identifiers in the trust list via the TTFM component of TRAIN as described in Figure 2. In addition to that, the trustworthiness of trusted issuers can be verified using TTV component of TRAIN.

This approach establishes a secure, decentralized, transparent, and trustworthy mechanism for credential issuance and validation in the industrial supply chain, improving operational efficiency and enhancing stakeholder confidence.

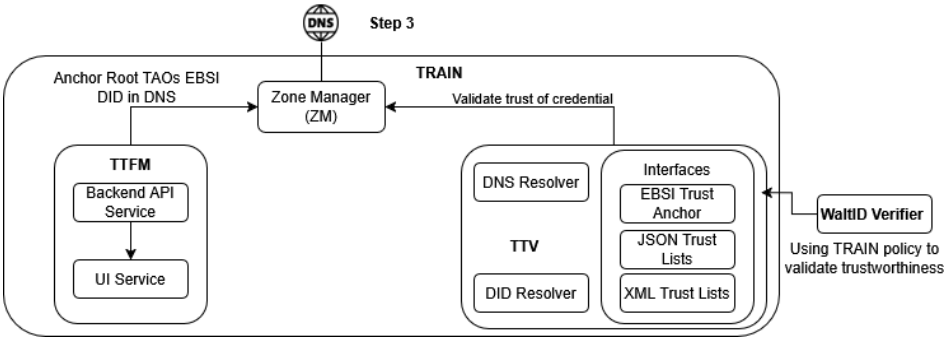


Fig. 2: Validation process of verifiable credentials using TRAIN

3.2 Role of the EBSI Blockchain in FidNET

Utilizing EBSI's decentralized ledger, FidNET benefits from a secure and decentralized platform for recording transactions and credentials.

Types of Data	Stored on the Blockchain	Stored off-Blockchain
Verifiable Credentials Metadata	Information about credentials (issuer details, issuance time, relevant standards)	Personal Identifiable Information (PII)
Transaction Records	Logs of credential issuance and validation events	Detailed Product Specifications
Trust Anchors	References to trusted identities and their associated DIDs	Specific attributes and sensitive business information
Schema	Definitions of data structures used for credentials	-
Trusted Issuer Registry	List of verified issuers and their credentials	-
Credentials of Trust Anchors	Credentials issued to Trust Anchors for verification	-
Identifier Details	Public keys and DID documents associated with identities	-

Tab. 1: Comparison of Data Stored on- vs. off-Blockchain

By ensuring data integrity through immutable records, EBSI fosters confidence among supply chain participants, enabling them to rely on the accuracy and authenticity of the information shared. The transparency offered by EBSI allows stakeholders, including

suppliers, manufacturers, and consumers, to verify product compliance and traceability, thereby enhancing accountability throughout the supply chain. The EBSI Trust Issuers Registry provides a public point which can be used to fetch the credentials and validate the trust chain of trusted issuers, TAOs and Root TAO. With this, EBSI is a powerful trust anchor that can be leveraged with FidNET, strengthening trust and efficiency of supply chain operations. Table 1 above provides an overview of the types of data stored and not stored on the EBSI blockchain, highlighting its commitment to privacy and security.

4 Conclusion

FidNET aims to advance the development of flexible, decentralized digital trust and transparency mechanisms within industrial supply chains. FidNET leverages EBSI to enhance the reliability of information and compliance in European supply chains. Its potential to integrate with initiatives such as Gaia-X and Catena-X underscores its potential to transform supply chain operations by providing a robust framework for secure and verifiable data sharing. FidNET could enable organizations to define flexible trust anchors and carry out transparent transactions.

A learning from our work with EBSI was that the onboarding process is quite complex. One reason for this is certainly the stringent verification procedures. Factors in this context are compliance requirements, like from the General Data Protection Regulation (GDPR). However, for some use cases these hurdles might be too high. Moreover, EBSI is focused only on Europe which limits its application potential further. Our future work therefore will explore the technical feasibility of integrating other blockchain technologies, such as CHEQD (https://cheqd.io) or Ethereum / ENS and see how this could bring additional value.

The project "FidNet" is financed by the Baden-Württemberg Stiftung gGmbH within the research programm "Blockchain Ideas Competition".

Bibliography

[Ca25] Catena-X, Catena-X Your Automotive Network, https://catena-x.net/en/1, accessed: 17/02/2025.

[DT25] Deutsche Telekom, Telekom Data Intelligence Hub. https://dih.telekom.com/en/glossary/gxdch-gaia-x-digital-clearing-house, accessed: 19.02.2025.

[Eb25] EBSI, DID Resolver | EBSI, https://hub. hub.ebsi.eu/tools/did-resolver. accessed: 18/02/2025.

[Ec24] Eclipse XFSC, XFSC: TRAIN - GitLab, https://gitlab.eclipse.org/eclipse/xfsc/train, accessed: 12/02/2025.

[En25] ENS, ENS Website, https://ens.domains, accessed: 12/02/2025.

[Fa24] Factory-X, Factory-X: The digital ecosystem, https://factory-x.org, accessed: 17.02.2025.

[Gx25] GXFS, Gaia-X Federation Services, https://www.gxfs.eu, accessed: 08.02.2025.

[Hn25] HNS, Handshake Website. https://handshake.org, accessed: 12.02.2025.

[In24] Industrie 4.0, Das Förderprogramm Manufacturing X, https://www.plattform-i40.de/IP/Redaktion/DE/Standardartikel/manufacturing-X-foerderprojekte.html, accessed: 17.02.2025.

[JCK22] Johnson Jeyakumar, I. H.; Chadwick, D.; Kubach, M.: A novel approach to establish trust in verifiable credential issuers in Self-sovereign identity ecosystems using TRAIN. In (Roßnagel, H.; Schunck, C.; Mödersheim, S. ed.): Open Identity Summit, Lecture Notes in Informatics (LNI), 2022. Köllen Druck + Verlag GmbH, Bonn, pp. 27-38, 2022.

[JK25] Johnson Jeyakumar, I. H.; Kubach, M.: A trust implementation model for cross-domain decentralized identity ecosystems: architecture, use case, and implementation. Procedia Computer Science, International Conference on Digital Sovereignty (ICDS) 254, 10-19, 2025.

[Lo24] Lodderstedt, T. et. al.: OpenID for Verifiable Credential Issuance. https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0-ID1.html, accessed: 18.02.2025.

[NKS23] Naffin, J.; Klewitz, J.; Schaltegger, S.: Sustainable development of supplier performance. An empirical analysis of relationship characteristics in the automotive sector. Corporate Social Responsibility and Environmental Management 30/4, pp. 1753-1769, 2023.

[Te23] Tenne, Data Ecosystem Energy Data-X is Funded by the German Government, https://www.energydata-x.eu/en/2023/10/04/datenoekosystem-energy-data-x-wird-durch-bundesregierung-gefoerdert/, accessed: 17/02/2025.

[Te24] Terbu, O. et. al.: OpenID Connect for Verifiable Presentations, https://openid.net/specs/openid-4-verifiable-presentations-1_0.html, accessed: 11.02.2025.

[ZS24] Zhang, A.; Seuring, S.: Digital product passport for sustainable and circular supply chain management: a structured review of use cases. International Journal of Logistics Research and Applications 27/12, pp. 2513-2540, 2024.

eIDAS 2.0: Evaluating the Issuance of Digital University Diplomas

Awid Vaziry [1,3], Luca Vetter[2,3], and Axel Küpper [1]

Abstract: This study examines the implementation challenges and strategic approaches for issuing eIDAS-compliant digital university diplomas. Through a mixed-methods approach that combines regulatory analysis with twelve expert interviews involving Qualified Trust Service Provider (QTSP) executives, conformity assessment specialists, and higher education stakeholders, we evaluate two primary issuance methods: qualified electronic sealed documents with qualified electronic timestamps or Qualified Electronic Attestations of Attributes (QEAA). Our findings reveal that while QEAA offers superior interoperability and holder binding, it faces regulatory uncertainty and implementation complexities. The research indicates that becoming a QTSP presents significant challenges for individual institutions, primarily due to security requirements, high operational costs, and complex conformity assessment processes. We identify three viable alternatives: outsourcing to existing QTSPs, forming university alliances to distribute costs, or implementing a 'bring-your-own-QTSP' model. This paper provides practical guidance for higher education institutions, emphasizing the importance of standardization, cross-institutional collaboration, and sustainable implementation strategies to enhance the value and portability of digital academic credentials across Europe.

Keywords: eIDAS, QTSP, QEAA, qualified trust service provider, conformity assessment, verifiable credentials

1 Introduction

The 2024 amendment of the Regulation (EU) No 910/2014 on electronic identification, authentication, and trust services (eIDAS) mandates that specific digital attributes, including educational qualifications, must be verifiable electronically by 2026 [Th24b]. However, higher education institutions lack consensus on how to issue eIDAS-compliant digital diplomas. This article serves as a guide for navigating the regulatory, technical, and operational challenges of issuing eIDAS-compliant digital diplomas.

Universities face at least two key decisions when issuing eIDAS-compliant diplomas: whether to issue electronically sealed documents (such as PDFs with qualified electronic seals) or use Qualified Electronic Attestations of Attributes (QEAA). They must also determine whether to become a Qualified Trust Service Provider (QTSP) themselves or to outsource this function. Each path involves regulatory, technical, and operational complexities, necessitating careful

[1] Technische Universität Berlin / T-Labs, Service-centric Networking, Straße des 17. Juni 135, 10623 Berlin, Germany, vaziry@tu-berlin.de, https://orcid.org/0009-0007-2192-5968;
axel.kuepper@tu-berlin.de, https://orcid.org/0000-0002-4356-5613

[2] Technische Universität Berlin , Straße des 17. Juni 135, 10623 Berlin, Germany, luca.vetter@alumni.tu-berlin.de

[3] The authors contributed equally to this work

doi:10.18420/OID2025_15

evaluation of compliance, financial implications, and available resources. This study employs a qualitative methodology, including twelve expert interviews with executives from QTSPs and conformity assessment bodies, eIDAS consultants, and other stakeholders, as well as an analysis of the eIDAS Regulation and related standards.

2 Background and related work

This section provides an overview of the eIDAS Regulation, trust services, and their relevance to digital university diplomas. The conformity assessment process and the introduction of QEAA under eIDAS 2.0 are discussed. Existing research on the issuance of digital diplomas is reviewed.

2.1 The eIDAS Regulation

The eIDAS Regulation (Regulation (EU) No 910/2014) [Th14] was established to standardize electronic identification and trust services across the European Economic Area (EEA). It ensures the legal recognition of digital transactions and supports the creation of a secure digital single market. Operational since 2016, the Regulation was amended in 2024 (eIDAS 2.0) [Th24a] to address challenges in digital identity management. Key updates include the European Digital Identity Wallet, which enables the secure storage of verifiable credentials, and QEAA, providing a standardized method for issuing and verifying certified attributes, such as diplomas.

2.2 Qualified Trust Service Providers and Trust Services

A QTSP within the eIDAS framework is an entity of higher trust and authority that has undergone a rigorous conformity assessment. A QTSP can offer qualified trust services that have legal recognition across all EU member states, such as issuing certificates for qualified electronic signatures and seals[4]. To be notified as a QTSP, the organization must be assessed by a conformity assessment body [Th24b]. The conformity assessment can be seen as a chain of digital trust that reaches from the European Commission through national accreditation bodies and conformity assessment bodies to the QTSPs. According to Article 22 of eIDAS Regulation, the member states of the EEA are obliged to provide national trusted lists to the European Commission, containing all trust service providers and their offered trust services [Th24b]. This shall be implemented in an electronically signed form that can be processed automatically. The European Commission aggregates those national lists and publishes a *List of Trusted Lists*[5].

[4] In eIDAS terminology, the qualified electronic seal of a legal entity is the counterpart of a qualified electronic signature of a natural person.

[5] https://eidas.ec.europa.eu/efda/trust-services/browse/eidas/tls

The trust service QEAA enables QTSPs to issue digitally signed electronic credentials that verify specific attributes of a subject. These attributes can be presented in a human-readable format, with the issuing trust service provider having validated their authenticity according to their established verification procedures. The attributes must meet the requirements defined in Annex V, including the possibility of revocation, issuing trust service provider information, and details on how to verify the validity of an attribute. Public bodies can act under Art. 45f as public electronic attribute authorities, allowing them to issue electronic attestations of attributes directly from authentic government sources while meeting the same reliability and trustworthiness standards as qualified trust service providers [Th24a].

2.3 Related work

The European Digital Credentials for Learning initiative within the Europass framework enables institutions to issue eIDAS-compliant credentials using qualified electronic seals (QSeal) [Th25]. However, challenges such as obtaining qualified seals, system integration, and stakeholder acceptance hinder widespread adoption. Similarly, several Blockchain-based solutions under the European Blockchain Services Infrastructure (EBSI) have been explored. For instance, Tan et al. [Ta23] piloted cross-border diploma issuance, while Zaklasnik et al. [ZFS23] focused on secure cross-border verification. Despite demonstrating technical feasibility, these blockchain-centric efforts face governance issues, standardization gaps, and integration challenges. Herbke and Yildiz [HY22] introduced a system that converts digital credentials into a Self-Sovereign Identity format. Their approach supports limited credential types and requires significant technical adjustments for institutional adoption. These initiatives often remain pilot-based or face hurdles in standardization and integration, with limited consideration given to strategic implementation. Our work fills this gap through an analysis that evaluates two issuance approaches and assesses the feasibility of universities becoming QTSPs versus outsourcing from the university's perspective.

3 Method

The overall methodology is summarized in Fig. 1, which provides an overview of the research process and the professional backgrounds of the consulted experts. To achieve the study's objectives, a mixed-methods approach was adopted, combining qualitative and quantitative methods. This included an in-depth analysis of regulatory frameworks (such as eIDAS, ETSI, and NIS2) and operational documents (such as certification practice statements), as well as structured expert interviews. Participants were carefully selected to ensure a diverse range of perspectives, including representatives from conformity assessment bodies, QTSPs, consultants, universities, and the financial sector. The interviews were conducted using a semi-structured framework comprising 53 questions, which included both open-ended and closed-ended elements. These questions were systematically organized into four thematic sections: regulatory, technical, economic, and organizational requirements. This structured yet adaptable approach allowed for a comprehensive exploration of the research objectives.

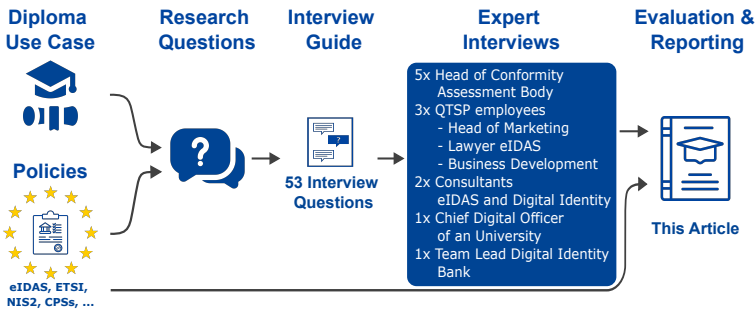


Fig. 1: Overview of research methodology and distribution of expert interview participants for the eIDAS diploma use case assessment.

4 Conformity Assessment Scheme

The conformity assessment scheme ensures the trustworthiness and compliance of QTSPs through a structured five-stage process. First, the project kick-off involves assigning assessors and aligning the milestone plan. This is followed by the document delivery stage, where QTSPs submit essential materials, such as the Certification Practice Statement and the Trust Service Provider Practice Statement. The third stage, document review, includes a remote audit to identify compliance deviations. The fourth stage involves an on-site audit to verify processes and services. Finally, a conformity assessment report is generated and sent to the QTSP, who must submit it to the supervisory body within three working days for notification. The European regulatory framework is not fully harmonized, leading to varying audit requirements across member states due to differences in national implementing laws, as seen with the NIS2 Directive and other EU regulations. The eIDAS 2.0 Regulation mandates the creation of implementing acts by May 2025 to standardize accreditation, audit requirements, and assessment schemes [Th24a].

Fig. 2 illustrates the hierarchy of responsibilities, the relevant standards, and the conformity assessment process using the example of a German entity being assessed by a German conformity assessment body. At the top of the hierarchy is the European co-operation for Accreditation (EA), which oversees the National Accreditation Body. The responsible German organizations and relevant standards are indicated, and relevant excerpts from the eIDAS Regulation are cited [Th14].

5 Use case evaluation: universities issuing diplomas under eIDAS

Recital 7 of the eIDAS 2.0 Regulation states that citizens should have access to electronic attestations of academic qualifications through the European Digital Identity Wallet [Th24a]. This chapter summarizes discussions and expert opinions on how universities could issue digital diplomas within the eIDAS framework.

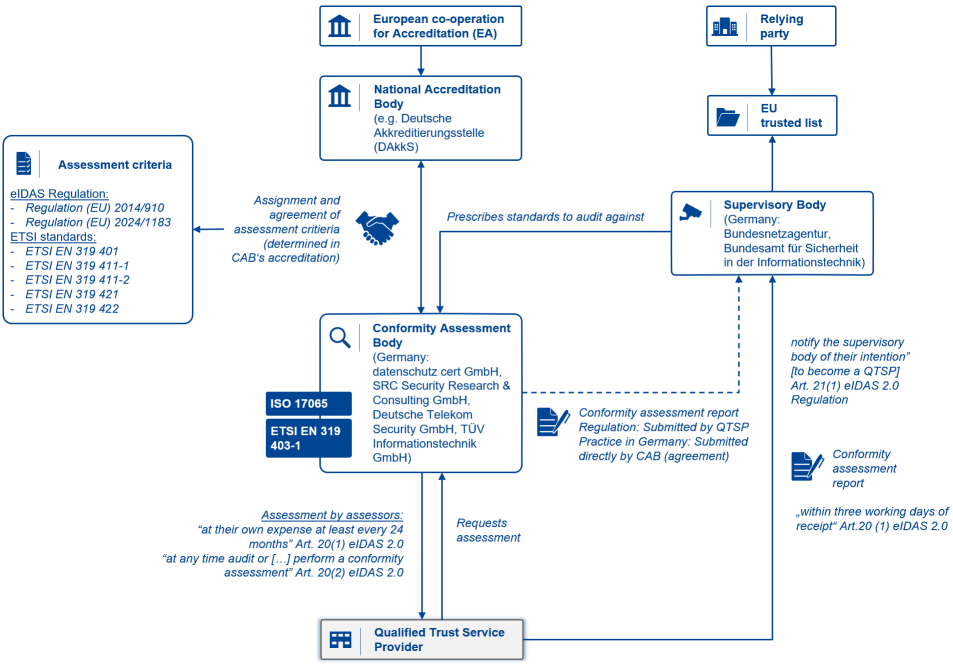


Fig. 2: eIDAS 2.0 conformity assessment scheme indicating the Regulation and assessment criteria as well as organizational authorities and dependencies.

5.1 Reviewing the requirements for universities becoming a QTSP

For becoming a notified QTSP, Universities must comply with the eIDAS Regulation, undergo assessments by accredited conformity bodies, and adhere to the NIS 2 and CER Directives, along with national implementation laws. These obligations include extensive event logging, liability for services (addressed through liability insurance or financial reserves), and significant operating costs for personnel, infrastructure, and conformity assessments, which must be renewed every 24 months [Eu24; Th14]. QTSPs must designate trusted roles, including a security officer, system administrator, system operator, and system auditor. Additional roles, such as registration and revocation officers, are needed for certificate issuance. Segregation of duties and redundancy are mandatory, and guidance from EN ISO/IEC 27002:2022 recommends further separation of roles, such as isolating the system auditor from the security officer and the system administrator from other trusted roles. SMEs may negotiate tailored role exclusion matrices with conformity assessment bodies to address specific needs. QTSPs must operate in secure, access-controlled facilities that house critical components, such as hardware security modules and qualified signature creation devices, with all physical access being logged [Eu24].

5.2 Approach 1: digital document with a qualified electronic seal and a qualified electronic timestamp

The first approach involves issuing a university diploma in a digital format (e.g., PDF, JSON, XML) and electronically sealing it using a qualified certificate for electronic seals. It also includes a qualified electronic timestamp (QTimestamp) to legally verify the issuance date, as the machine time of the qualified seal alone may not be legally binding [Eu16]. This method enables universities to implement digital diploma issuance efficiently, as all necessary components are defined in eIDAS 1.0. This method allows individuals with limited technical skills to easily view and verify digital university certificates using PDF viewers, without the need for specialized software installation. However, this approach has the drawback of not linking the presenter of the document to its rightful owner (holder binding). Another drawback relates to the identification of the document's issuer. If the university is not notified as a QTSP itself, but uses external qualified trust services instead, the external provider will be listed as the issuer of the QSeal and QTimestamp, potentially diminishing the trustworthiness compared to having the university as the issuer.

5.3 Approach 2: Qualified Electronic Attestations of Attributes

This approach issues diplomas as digital credentials called QEAA, which are securely stored and easily verifiable through digital identity wallets. Unlike Approach 1, QEAA provides enhanced interoperability and can be seamlessly integrated with various digital platforms. To issue QEAA, the student must be identified at a high assurance level by the issuer. This can be done by presenting a national ID card, such as using the eID function of the German identity card, or utilizing a credential in the EUDI Wallet. Analogous to the first approach, universities, being the QTSPs themselves, could benefit from a higher level of trust, as they would appear as issuers of the QEAA. When outsourcing the QEAA to a QTSP, the university remains the authentic source of the diploma information, but the QTSP appears as the technical issuer. This means that when a subject or relying party verifies the QEAA, they will see the QTSP's digital signature as the technical guarantor of authenticity, while the university is referenced as the attribute source within the QEAA content. For this arrangement to work securely, a trusted channel must be established between the university and the QTSP. The university will maintain a database of its students and forward their data, including name, type of degree, courses, and grades, to the QTSP. Based on those, the QTSP would create the QEAA. Currently, there is legal uncertainty surrounding the use of QEAA for digital university diplomas. They are designed for short- to medium-term validity, requiring re-signing at specific intervals and regular validity checks to ensure ongoing accuracy. Furthermore, the question of how the operation of signing the QEAA is procedurally embedded within an outsourcing scenario remains open and will be addressed through further legislation, including upcoming implementing acts and respective ETSI standards.

6 Discussion

After reviewing the eIDAS Regulation, along with supporting documentation and expert consultations, we have identified two viable approaches for issuing diplomas under eIDAS. First, issuing a digital document (e.g., PDF) with a QSeal and a QTimestamp, and second, issuing a QEAA. The question of which approach is preferable cannot be answered definitively due to some regulatory and technical uncertainty. However, in the long run, a machine consumable credential with holder binding seems desirable (e.g. QEAA).

A key consideration is whether to utilize external qualified trust services or to establish themselves as a notified QTSP. This raises the question of whether the potential benefits justify the effort required. The primary efforts involved include the costs of auditing, infrastructure, personnel, and adapting current IT systems and processes. The primary benefits of becoming a QTSP include independence from external parties and the perception of a higher level of trust in directly issued diplomas.

The issuance of qualified certificates for electronic seals and qualified electronic timestamps by QTSPs has become largely commoditized, with prices anticipated to reach the cent range in the coming years. As a result, it is unlikely that universities will choose to become QTSPs for this purpose. However, universities may choose to form a university alliance, acting as public electronic attribute authorities that handle eIDAS-compliant diploma issuance and lifecycle management on behalf of dozens or hundreds of universities, thereby significantly reducing the operating costs per entity. Another option, referred to by the authors as 'bring-your-own-QTSP', enables students to order and pay for their digital credentials directly. They can use a service such as OAuth to authorize the QTSP to access their personal data. The university's role would be limited to providing the interface and verifying the QTSP's legitimacy through the EU List of Trusted Lists. This approach could fully outsource the costs and responsibilities of the issuance.

7 Conclusion

This study has examined the eIDAS-based issuance of digital education credentials and addressed the question of whether a university should be accredited as a QTSP itself or outsource those operations. The analysis identified two primary approaches for universities to issue digital diplomas compliant with eIDAS 2.0. First, the issuance of QSealed documents with a QTimestamp, and second, the use of QEAA. Each approach has its own implications for interoperability, verification, and user experience. In both approaches, the question arises of whether the university should become a QTSP or not. Significant barriers exist to becoming QTSPs. The high financial costs, strict security requirements, and operational complexities involved in achieving QTSP status make it an impractical option for most institutions. Alternatives include partnering with existing QTSPs or forming university alliances to manage eIDAS-compliant diploma issuance collectively. Furthermore, this research highlights the significance of standardization and interoperability in issuing

digital diplomas. The transition to eIDAS-compliant credentials provides an opportunity to improve the standardization, trustworthiness, portability, and digital verifiability of academic qualifications throughout the European Economic Area. However, it also raises concerns about holder-binding and the balance between security and accessibility.

In conclusion, transitioning to eIDAS-compliant digital diplomas poses challenges but also enhances the value of digital academic credentials in the digital age. Universities need to evaluate the costs and benefits of various implementation strategies, taking into account technical feasibility, user experience, and sustainability. Ongoing collaboration among educational institutions, technology providers, and policymakers is essential for developing effective solutions. Future research should focus on standardizing data formats, forming university alliances, verifying and revoking diplomas, and optimizing EUDI Wallet processes. Additionally, it is essential to define which document will be considered the original source of truth versus a certified copy.

References

[Eu16] European Telecommunications Standards Institue: Etsi en 319 421 - v1.1.1 - electronic signatures and infrastructures (esi); policy and security requirements for trust service providers issuing time-stamps, 2016.

[Eu24] European Telecommunications Standards Institue: Etsi en 319 401 - v3.1.1 - electronic signatures and trust infrastructures (esi); general policy requirements for trust service providers, 2024.

[HY22] Herbke, P.; Yildiz, H.: ELMO2EDS: Transforming Educational Credentials into Self-Sovereign Identity Paradigm. In: 2022 20th International Conference on Information Technology Based Higher Education and Training. IEEE, Antalya, Turkey, pp. 1–7, 2022.

[Ta23] Tan, E. et al.: Verification of Education Credentials on European Blockchain Services Infrastructure (EBSI): Action Research in a Cross-Border Use Case between Belgium and Italy. en, Big Data and Cognitive Computing 7 (2), p. 79, 2023.

[Th14] The European Parliament and the Council of the European Union: REGULATION (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, 2014.

[Th24a] The European Parliament and the Council of the European Union: REGULATION (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital IdentityFramework, 2024.

[Th24b] The European Parliament and the Council of the European Union: REGULATION (EU) No 910/2014 amended by directive (EU) 2022/2555 and regulation (EU) 2024/1183. 2024.

[Th25] The European Commission: European Digital Credentials, 2025, https://europass.europa.eu/en/stakeholders/european-digital-credentials, accessed: 04/11/2025.

[ZFS23] Základsník, M.; Farana, R.; Surma, S.: Digitization of University Diplomas in the European Union. In (Silhavy, R.; Silhavy, P., eds.): Software Engineering Research in System Science. Vol. 722, Series Title: Lecture Notes in Networks and Systems, Springer International Publishing, Cham, pp. 382–390, 2023.