CEN

WORKSHOP

AGREEMENT

CWA 17525

March 2020

ICS 35.030; 35.240.01

English version

Elements of fair and functioning data economy: identity, consent and logging

This CEN Workshop Agreement has been drafted and approved by a Workshop of representatives of interested parties, the constitution of which is indicated in the foreword of this Workshop Agreement.

The formal process followed by the Workshop in the development of this Workshop Agreement has been endorsed by the National Members of CEN but neither the National Members of CEN nor the CEN-CENELEC Management Centre can be held accountable for the technical content of this CEN Workshop Agreement or possible conflicts with standards or legislation.

This CEN Workshop Agreement can in no way be held as being an official standard developed by CEN and its Members.

This CEN Workshop Agreement is publicly available as a reference document from the CEN Members National Standard Bodies.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

CEN-CENELEC Management Centre: Rue de la Science 23, B-1040 Brussels

Ref. No.:CWA 17525:2020 E

Contents

Foreword

This CEN Workshop Agreement has been developed in accordance with the CEN-CENELEC Guide 29 “CEN/CENELEC Workshop Agreements – The way to rapid consensus” and with the relevant provisions of CEN/CENELEC Internal Regulations - Part 2. It was approved by a Workshop of representatives of interested parties on 2020-01-17, the constitution of which was supported by CEN following the public call for participation made on 2018-10-12. However, this CEN Workshop Agreement does not necessarily include all relevant stakeholders.

The final text of this CEN Workshop Agreement was provided to CEN for publication on 2020-02-14.

The following organizations and individuals developed and approved this CEN Workshop Agreement:

- Markus Kalliola, Sitra
- Katri Korhonen, Sitra
- Juhani Luoma-Kyyny, Sitra
- Črt Ahlin, DataFund
- Pirkka Frosti, DigitalLiving
- Mika Huhtamäki, Vastuu Group
- Antti Kettunen, TietoEvry
- Paul Knowles, Dativa
- Teemu Kääriäinen, Nixu
- Ville Lavonius, Vastuu Group
- Robert Mitwicki, Lab10Coop
- Perttu Prusi, Fujitsu
- Mikael Rinnetmäki, SensoTrend
- Henna Suomi, DigitalLiving
- Gregor Žavcer, DataFund
- Annika Wolff, LUT University

Attention is drawn to the possibility that some elements of this document may be subject to patent rights. CEN-CENELEC policy on patent rights is described in CEN-CENELEC Guide 8 “Guidelines for Implementation of the Common IPR Policy on Patent”. CEN shall not be held responsible for identifying any or all such patent rights.

Although the Workshop parties have made every effort to ensure the reliability and accuracy of technical and non-technical descriptions, the Workshop is not able to guarantee, explicitly or implicitly, the correctness of this document. Anyone who applies this CEN Workshop Agreement shall be aware that neither the Workshop, nor CEN, can be held liable for damages or losses of any kind whatsoever. The use of this CEN Workshop Agreement does not relieve users of their responsibility for their own actions, and they apply this document at their own risk.

CEN Workshop Agreement

This CEN Workshop Agreement, CWA, contains the requirement specifications for three important building blocks of any internet service that uses personal data and in which the agency for that data is within the individual whose data is being utilized. The building blocks are: identity, consent and logging.

Before you take a deep dive into the requirements specifications, we want to give you the background of why this work was done and why agency for personal data matters. The Finnish Innovation Fund Sitra, which initiated this workshop, started a project called IHAN in 2018 to build the foundation for a fair data economy. In a fair data economy, people are in control of how their data is used and shared, while businesses need to earn the trust of people to get access to diverse sources of data. This change of data agency enables the widest circulation of data because data can be exchanged also between ecosystems which are currently data silos and only serve the businesses within the ecosystem.

The IHAN project covers a wide range of topics, including citizen engagement, business models, technical requirements and governance. Many of the outputs are tested in real life in technical and business pilots in Finland and elsewhere in Europe.

This CWA is a part of the technical work package of the IHAN project and builds on the work that was previously published under IHAN Technical Blueprint document. While the CWA takes a deep dive into three of the most important parts of the Blueprint, we continue to work also on the other fronts of technical requirements and acknowledge other European initiatives working on the same topics. Therefore, this is not the end of the technical requirements work, but hopefully a good start to something that finally will make a meaningful impact towards the free flow of data in Europe. We hope that many other European projects will be inspired by this work and that it can be utilized in research and development by research institutions, universities and private enterprises. It is also our goal that this CWA will be further developed by CEN either as a technical committee document or as a European Standard (EN).

This CWA was approved by consensus with the experts listed in the document. The work was done during 2019 in three work streams which were open and free of participation costs. The pre-release version was in public consultation during November 2019.

We would like to express our appreciation to all stream leads as well as other contributors of the final CWA. Many thanks to the secretariat SFS Finland for the practical arrangements with CEN, and final thanks to CEN for allowing the workshop to take place.

We hope you enjoy reading this!

Markus Kalliola	Juhani Luoma-Kyyny	Katri Korhonen
Senior Lead	Senior Lead	Specialist
Sitra	Sitra	Sitra

0 Introduction

0.1 General background

The data economy is about creating services by using new or re-using existing information, and especially by combining the information in previously untested ways. The basic principle behind a fair data economy is value creation according to "data principles":

- human centric (from organization- or technology-centric to human centric)
- thriving (unlocking the use of data to scale services)
- balanced (data sharing benefits all)

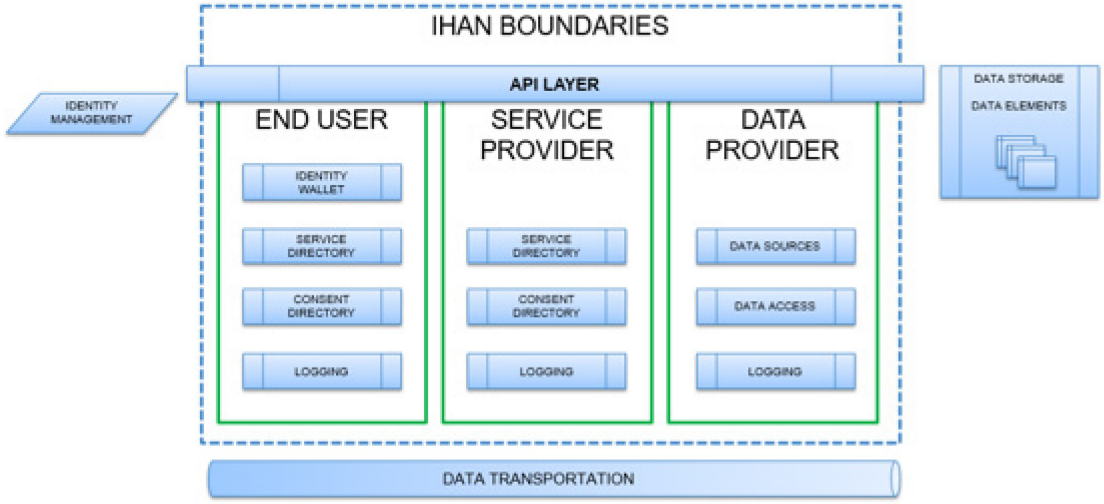
("A roadmap for a fair data economy" https://www.sitra.fi/en/publications/roadmap-fair-data-economy/)

A fair data economy needs a well-functioning architecture and infrastructure. The basic components for that are:

- identity
- consent
- logging

0.2 IHAN Logical view

The diagram below describes the logical components of a possible service implementation according to IHAN requirements. It also draws the boundaries of IHAN – data transportation technologies, data storages and identity management are outside IHAN boundaries. For example, identity is an essential element of a personal data-based service, but IHAN requirements do not define how and where identity should be managed. Also, data transportation and storing are required to implement services, but they are decisions made by companies providing services.



In the diagram, logging and consent are clear layers across the logical "roles" (end user, service provider and data provider). Identity is the "starting point" of the structure.

In this CEN Workshop Agreement CWA, three of the logical components in IHAN are described in detail.

1 Identity

1.1 About identity

A human-centric data economy runs around us individuals. Human biology, needs, ambitions, limitations, capabilities and the understanding of our existence defines us. The way we behave and interact with others and the physical world either limit or empower our everyday life. The global economy and societies are built on the system of connected individuals.

The wellbeing of the world relies on open, transparent and sustainable societies where every citizen can be provided fair and safe access to society. Identity is a basic human right. Our future wellbeing and existence rely on machines and automation. The better the machines can understand us as humans, the better the world we can build with the efficiency automation brings.

Identity is the foundation on which all the other data economy capabilities are built. The way we understand the digitalization of a natural person today is narrow and driven by limited industries, technologies of the past and the interests of a few. In this document, we aim to gather a new perspective on how to approach the digitalization of a natural person and thus create the first broader definition of a true global digital identity.

Identity within the digital world is hard due to the nature of data, which can be in multiple places at the same time, can be easily duplicated, transferred or compromised. As a result, we need to have mechanisms which allow us to create the digital representation of our identity in a way that serves us in the digital world without compromising our privacy or losing control over it.

Based on the nature of the identifiers, we could distinguish between identifiers which are strongly coupled with human being-like biometrics and those which are purely digital identifiers which have no connection to the analogue world.

1.2 In scope

This document:

- defines a digital identity for a natural person for the contextual processing by information systems and machines;
- sets the background for all the other components needed to use and utilize the digital identities within a decentralized data economy, such as consent, logging, data transport, services, etc;
- focuses on providing a solid and focused background to deliver a practical approach for future development and still covers the digital identity definitions from a wide enough perspective to not limit its use in today's needs, technologies or industrial use cases;
- produces a neutral, objective and generic definition for all humans that can then be scaled up based on the industry, use case and technology it is applied to based on this core definition;
- covers also the basic mechanisms for use in the digital services (contextual use), trust and identity management that are within the scope of the digital identity itself;

- defines a well-considered overview on the individual's digital properties, their usage and needed core processes for further consideration on standardization; and

- describes the need for truly decentralized identity for every human being in the digital age.

1.3 Current identity markets and the use of digital identities

Digital identities have existed since the first software that enabled the use of user accounts and registers turned digital with the electronic revolution after the Second World War. There are multiple digital representations of us in various systems, registers and even offline hardware-based solutions, such as driver settings in cars or user profiles in washing machines. It does not matter if it is a rather limited pseudonym user or a governmental identity in the population register, it is still a digital identity representing a natural person.

Market sizes

Digital identity

Digital identities will continue to play an even greater part as the foundation for the digital economy and the next generation of the Internet. *"Extending full digital ID coverage could unlock economic value equivalent to 3 to 13 percent of GDP in 2030" [1]* The global market for Personal ID credentials was valued at $8.7 billion in 2016 and is forecast to reach $9.7 billion by 2021 [2]. The identity & access management market will grow from $8.09 billion in 2016 to $14.82 billion by 2021, at a CAGR of 12.9% from 2016 to 2021 [3].

Blockchain (with Digital Identity)

The global blockchain market is expected to reach USD 16.3 billion by 2024. Extensive use across varied business applications including logistics & transportation, BFSI, retail, healthcare, media & entertainment, manufacturing, legal and government is anticipated to fuel market growth. Additionally, the high efficiency of this technology for intra-organizational execution, asset management, transaction processing, and identity management has significantly resulted in mounting adoption by enterprises to better handle resources and produce superior results [4].

The state of the current digital identity markets

" Nearly one billion people globally lack a legally recognized form of identification, according to the World Bank ID4D database. The remaining six billion *people have some form of identification, but over half cannot use it effectively in today's digital ecosystems."* [1]

The incumbent markets, such as in the Northern Europe, have achieved their late maturity and need overhaul. The regulated, strongly authenticated digital identity market has been dominated by the banks and governmental solutions. The majority of the global population does not have a digital identity capable of being used in strong authentication. The market standard is set by fragmented digital user accounts and the GAFA provided digital identities directly linked to these companies' customership. The decentralised solutions mainly based on distributed ledger technologies (DLT) are in their early stages. There is a significant need to develop new digital identity architectures, standards and global solutions in favour of a sustainable, open and profitable data economy.

Some of the biggest challenges facing the digital identity markets today are:

- Cost of authentication resulting in slow digitalization
  Digital services and innovation cannot be developed due to costly strong authentication and a serious lack of interoperability between different digital services.

- Non-inclusive services

Most of the global population does not have trusted and fair access to the digital society. Underage persons and dependents cannot be fully included in the digital society without a digital identity.

- Lack of digital trust and usability

  The use of passwords and physical copies of documents as proof creates increased risks for identity theft and hinders the personalization of online services.

- Missing the AI opportunity

  Personal data cannot be released to the use of AI without reliable user authentication and linking digital identities the machines can understand.

- Significant economic competitive advantage is lost

  Societal and industrial services are severely hindered by a lack of automated trust, resulting in the poor automation of meaningful services for citizens and consumers.

From a technical perspective

IDPro Body of Knowledge [5] categorizes identity management solutions roughly into two categories:

- Workforce IAM / Internal IAM
- Consumer / Citizen IAM

Workforce and Internal IAM focuses mainly on the IAM processes of:

- Managing the lifecycle of identities through the joiner-mover-leaver process
- Ownership of HR of the identity information
- Provisioning (on-boarding and off-boarding) of identity information
- Role Management and access governance
- Re-certification of access

Consumer / Citizen IAM focuses more on consumer journeys through processes, such as registration of consumers and authentication assurance with different levels of assurance.

1.4 Need for standardization

Digital identity is an application area with a multitude of standards covering the key functional areas of identity management, federation, authentication, access management and consent management. Contributors to standardization include, for example, NIST, ISO/IEC, OpenID Foundation, W3C, OASIS and Hyperledger. In the NIST and ISO/IEC standards, the focus has especially been on centralized and federated identity mechanisms. This does not take into account the decentralized identity solutions. Due to this, there is a need for standards that bring these two approaches together to form a coherent basis for identity standardization.

Current centralized, federated and user-centric identity solutions provide a well-standardized basis for building identity management solutions. These are, however, defined purely from the viewpoint of centralized methodologies without taking into account the recent advancements in the areas of decentralized identity technologies. The aim of this standardization work is to bridge the gap between centralized and decentralized identity methodologies to guarantee that the developed identity solutions are compliant with the existing standards built for centralized solutions while taking into account the novel decentralized ways of working.

The path from centralized to decentralized identity solutions is covered in detail in Christopher Allen's "The Path to Self-Sovereign Identity" [1]. The evolution has happened through following phases:

- Phase 1: Centralized Identity
    - Created in the early days of the Internet, the centralized authorities became the issuers and authenticators of digital identity.
- Phase 2: Federated Identity
    - The next major advancement for digital identity occurred at the turn of the century when a variety of commercial organizations moved beyond hierarchy to debalkanize online identity in a new manner.
- Phase 3: User-Centric Identity
    - User-centric methodologies tend to focus on two elements: user consent and interoperability. By adopting them, a user can decide to share an identity from one service to another and thus debalkanize his digital self.
- Phase 4: Self-Sovereign Identity
    - Rather than just advocating that users are at the centre of the identity process, self-sovereign identity requires that users are the rulers of their own identity.

This evolutionary model and the need to have standards that support identity solutions built for different phases of the evolution is complemented by Kim Cameron's "Laws of Identity" [2] that define the essential laws for managing digital identities. These laws need to provide the foundations for building the standardization of digital identity:

- 1: User Control and Consent
    - Technical identity systems must only reveal information identifying a user with the user's consent.
- 2: Minimal Disclosure for a Constrained Use
    - The solution that discloses the least amount of identifying information and best limits its use is the most stable long-term solution.
- 3: Justifiable Parties
    - Digital identity systems must be designed so the disclosure of identifying information is limited to parties with a necessary and justifiable place in a given identity relationship.
- 4: Directed Identity
    - A universal identity system must support both "omni-directional" identifiers for use by public entities and "unidirectional" identifiers for use by private entities, thus facilitating discovery while preventing unnecessary release of correlation handles.
- 5: Pluralism of Operators and Technologies

    - A universal identity system must channel and enable the inter-working of multiple identity technologies run by multiple identity providers.

- 6: Human Integration

    - The universal identity metasystem must define the human user as a component of the distributed system integrated through unambiguous human-machine communication mechanisms offering protection against identity attacks.

- 7: Consistent Experience Across Contexts

    - The unifying identity metasystem must guarantee for its users a simple, consistent experience while enabling the separation of contexts through multiple operators and technologies.

The following universal capabilities should be considered for the next generation of digital identities:

- persistent digital unique identifier for a person
- capability to link various digital identities online and in various systems to the universal identity
- support for multiple levels of trust from regulated strong authentication to various lower levels of trust
- option to support multiple levels of enrolment and trust for one digital identity
- possibility to represent other digital identities, e.g for children, people with special needs or under legal custody
- detachment from the hardware devices and user interfaces, e.g. capability to support biometric, augmented, ubiquitous and ambient technologies and solutions where no human-computer-interfaces are needed
- possibility to support ethical and user-centric models for fair data use
- support for both centralized and decentralized identity and data architectures
- capability to enable the use of digital rights, consents, permits, semantics and the linking of data based on the authenticated digital identities
- independency from the issuers' own customership and solutions
- semantic classification of identities and their reference data needed to promote the use of AI and automation of trust for the machines
- cross industry and sector use of digital identities from the citizens’ and global consumers' perspective
- capability to give digital identities also for machines as actors in the digital identities
- support for authentication also in the physical services, e.g. logistics, over the counter services, etc.

1.5 Stakeholders

The generic roles and stakeholders for digital identities can be simplified to:

Primary tier:

- Identity holder - the natural person the identity represents
- Identity issuer - the party issuing and attesting the digital representation of the person, customer or a citizen, e.g. government, company or any such party
- Relying party - a service provider, company or other party providing data and digital services for the identity holders

Secondary tier:

- Identity providers (IdPs) - an entity that creates, maintains, and manages identity information for the first-tier identity holders while providing authentication services to relying applications within a federation or distributed network for the issuers and relying parties
- Identity broker - party that links, authenticates or federates different identities for the primary tier users
- Regulator - government officials, law makers, international standardization bodies and such parties capable of controlling and monitoring the identities and their use

The natural person identity is created most often in the following processes and lifecycles:

- citizenship - a governmental process for issuing new digital identities to citizens
- customership - a digital user account or customer identity is provided to support the service provisioning for the customers

The shift in existing roles

The new processes and issuers for digital identities should also be recognized:

- self-sovereign identity - a self-sovereign identity can be created and maintained by an individual themselves but also verified by decentralized issuers for validity in different contexts and use cases
- identity through interaction - interaction between people, physical spaces and services - a personification of any data even on a pseudonymous and anonymous level can still constitute a very low level of authenticated identity. Also, the metaphysical and temporarily identity created, for example, by standing fifth on line in a room at a specific time can loosely be considered as an identity. Identity is also needed for any interaction between people in physical and digital interaction without any proper legal status or connection to any organization or determined party
- machines as users - natural persons are increasingly represented through a relation by various machines and robots we own and mandate to operate on our behalf. They should be considered as having a similar identity to natural persons as they act in the similar capabilities as future customers, tenants and operators in the physical and virtual world.

The shift in existing relationships

The introduction of the decentralized and linked identities poses new opportunities and challenges for the existing systems and legislation. The transition from relying on a centralized and regulated single party to the use of, for example, self-sovereign identities and a network of decoupled issuers and relying parties should be taken into consideration when creating new standards and legislation for future digital identities.

1.6 Identifiers

A natural person's identity can be defined through the perspective of different identifiers, a set of the attributes which uniquely identify a person. Those identifiers can be static, such as date of birth, which never change, but also dynamic and changing over time, such as a passport number. We could also distinguish between the self-issued, such as DID [3] (decentralized identifiers) and those which are automatically given to us, such as a national security number issued by a government.

Collecting all identifiers in one place is very difficult, probably even impossible. The reason for this is those attributes change all the time. Some of them could be treated as an identifier in one location or culture, whereas in another they could be treated as a regular attribute.

To deal with this problem, we have to define what we mean by identifiers and how we can distinguish them within the vast number of other attributes. This is very important as incorrectly handling those identifiers can lead to privacy violations and serious consequences for individuals.

We defined identifier as follows:

An identifier is an attribute that allows, alone or together with another attribute, unique identification of a person.

Those identifiers are commonly known as PII (Personal Identifiable Information).

This definition is generic enough to not limit what attributes can be used but specific enough to serve the purpose of protecting a person.

To cope with the vast amount of PII and the complexity of the digital world, the Kantara Initiative [1] published a comprehensive list of attributes called Blinding Identity Taxonomy (BIT) [2], which allows identifying sensitive information in any data sets which are collected or used by parties in a secure and privacy concerned way. It is a living list as over time we could develop new identifiers or, due to the progress of technology, some attributes which had so far not been treated as PII could be treated as such.

1.6.1 Biometrics

Digital representation of human properties: retina, fingerprint, voice, face, heartbeat, etc.

Any identifier that is tied to the human body or physical world provides strong assurance that a person is who they claim to be. A big disadvantage of such identifiers is that, once compromised, you cannot change. Due to the progress with technology, it is increasingly easier to steal such identifiers and use them against a user. They are important components of a set of digital identifiers, but they cannot be the only ones.

- Characteristics:
    - hard to lose
    - once compromised they are hard or impossible to change
    - cannot revoke without losing access
    - uniquely bound to the human being

- Examples [1]:
    - DNA Matching
    - Ear
    - Fingerprint recognition
    - heartbeat
    - facial recognition
    - Eyes - iris recognition
    - Eyes - retina recognition
    - fingerprint geometry recognition
    - gait
    - hand geometry recognition
    - odour
    - typing recognition
    - vein recognition
    - voice - speaker identification/verification/authentication
    - signature recognition

1.6.2 Non-biometric

Cryptographic material - keys, password, tokens, etc.

Identifiers based on any type of the data not tied to the physical world can be easily changed and randomly generated, the user can use different identifiers against different systems, that is, the user can have different personas which can be their digital avatars without the need to reveal too much information.

it is easier to maintain and create such identifiers and the user can easily create new ones if they get lost.

- Characteristics:
    - easy to create new
    - possibility to revoke
    - can have more than one
    - huge variation of types
- Example:
    - DID [2]

- Public/Private key
- password
- token

1.7 Classification of properties

What makes us really us is a set of the data points around us. Data points can be represented by simple attributes, more complex ones and events. To better understand the types of data which compose our identity, we can classify them based on their properties. The most generic classification would be static vs dynamic. A static identifier would be a property that is assigned once to a human being and never changes. For example, date of birth or fingerprint. Those attributes are especially sensitive as once leaked they cannot ever be changed. The opposite of static identifiers is dynamic identifiers. These are attributes which change over time, such as a passport number, height, weight, address.

Further on, we can distinguish between identifiers that are self-issued vs issued by third parties. The difference here is a level of trust within the given attribute. All attributes which are self-issued require a verification process of third parties to make sure that what you are claiming is true. Those attributes can be verified by an authority and converting them into issued attributes adds to them a higher level of trust or taken as it is and builds trust over time by the network.

Another way of looking at attributes is the nature of control of the attribute. Over some of the identifiers you have full control and you can decide to change them at any time. However, some are outside of your jurisdiction and it is up to the network or authority that assigned them to you. This is especially true of some events which generated a specific footprint on your identity, such as committing a crime, which in some cases would not be up to you to reveal or not.

One of the most important classifications from the personal point of view is the classification based on the level of sensitivity. Revealing some attributes could have serious consequences depending on the context, such as revealing sexual orientation within countries where that orientation is not tolerated. The level of sensitivity also includes information on how many properties we need to correlate in order to identify uniquely one person. Taking into consideration DNA, which can be used globally as a unique identifier, it has a higher sensitivity than, for example, date of birth, which could be shared by hundreds of people. To identify such sensitive information, the Kantara Initiative published a Blinding Identity Taxonomy [2] which lists all the most common very sensitive properties attached to a human being.

Another important data category is the Personally Identifiable Information (PII). This is defined by ISO/IEC 29100:2011 [1] as any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal. To determine whether a PII principal is identifiable, account should be taken of all the means which can reasonably be used by the privacy stakeholder holding the data, or by any other party, to identify that natural person.

Generic categorization of identity properties

- static vs dynamic - sensitivity and immutability
- self-issued vs third party - level of trust
- controlled vs uncontrolled

- level of sensitivity
- human properties
- data categories from Kantara - sensitivity level of the bucket
- data privacy vocabulary [3]

1.8 Identity management

As defined in ISO/IEC 24760-1:2019, identity management generally refers to a mechanism comprising policies, procedures, technology and other resources for maintaining identity information including associated metadata. An identity management system is typically used for identification or authentication of entities. It can be deployed to support other automated decisions based on identity information for an entity recognized in the domain for the identity management system. [1]

Identity information associated to metadata specifies, for instance, its origin, scope of use, and period of validity. Identity information metadata can itself be identity information and can be included in the identity it relates to. [1]

Identity information and its associated metadata can be changed. As defined in ISO/IEC 24760-1:2019, the procedures and conditions for changing, updating, and creating identity information can include, for example, the following activities [1]:

- Requesting and receiving information from external sources
- Verifying and validating the identity information
- Qualifying and categorizing
- Recording
- Provisioning
- Archiving
- Deleting

Identity management covers the lifecycle of identity information from initial enrolment to archiving or deletion. It includes the governance, policies, processes, data, technology, and standards, which can include:

- An Identity Register
- Authentication of the identity
- Establishing provenance of the identity information
- Establishing the link between identity information and an entity
- Maintaining the identity information
- Ensuring the integrity of the identity information
- Providing credentials and services to facilitate authentication

- Mitigating the risk of identity information theft

1.9 Trust

Trust in the context of digital identities is built around three central concepts: identity proofing, digital authentication and federation.

For centralized and federated identity solutions these core concepts are mainly defined by NIST SP 800-63-3. It defines identity proofing as the process of establishing that a subject is who they claim to be. Digital authentication establishes that a subject attempting to access a digital service is in control of one or more valid authenticators associated to the subject's digital identity. Successful authentication provides reasonable risk-based assurance that the subject accessing the service is the same as the one who has accessed the service previously. Federation is the process of conveying the results of authentication and relevant identity information to Relying Parties. [1]

The different actors and their interactions as defined by NIST SP 800-63-3 are described in the diagram below.

This model is compliant with the way in which trust is established in decentralized identity-based models. For instance, the W3C Verifiable Credentials Data Model [3] defines the interactions using the following kind of model:

When comparing traditional centralized / federated identity models to the decentralized model, it can be stated that:

- Role of the Credential Service Provider (CSP) in the centralized / federated model is equivalent to the role of Issuer in the decentralized model. Similarly, as CSP issues an authenticator to the end-user after performing enrollment and identity proofing, the Issuer issues a Credential (that is equivalent to an authenticator) to the end-user after performing the same steps. The difference between the models is that in the centralized / federated model the end-user must enroll the identity to the CSP, whereas in the decentralized model the end-user can perform the enrollment directly to the Verifiable Data Registry without any centralized identity provider.

- Roles of the Verifier and Relying Party (RP) in the centralized / federated model are equivalent to the role of Verifier in the decentralized model. The acts of authentication, validation of authenticator / credential binding, passing of attribute and assertion information and creation of authenticated session are equivalent to the process of End-User presenting the credential to the Verifier that can then validate it against the Verifiable Data Registry without the need for any centralized identity provider.

1.10 Identity proofing

When dealing with centralized / federated identity models, the following roles defined in NIST SP 800-63-3 can be identified in the context of digital identity enrollment and identity proofing [1]:

- Applicant: End-user undergoing process of enrollment and identity proofing.

- Credential Service Provider (CSP): A trusted entity that issues or registers subscriber authenticators and issues electronic credentials to subscribers.

- Subscriber: End-user who has received a credential or authenticator from a CSP.

In the context of decentralized identity models, e.g. based on W3C Verifiable Credentials Data Model [3], the corresponding roles are the following:

- Applicant is equivalent to Subject. Subject is an entity about which claims are made.
- CSP is equivalent to the Issuer. Issuer is an entity that can assert claims about a Subject.
- Subscriber is equivalent to Holder. Holder is an entity that possesses one or more verifiable credentials and can generate presentations from them.

NIST SP 800-63-3 defines the basic process of enrollment and identity proofing as follows [1]:

1. An applicant applies to a CSP through an enrollment process.
2. The CSP proofs the identity of the applicant. Upon successful proofing, the applicant becomes a subscriber.
3. Authenticator(s) and corresponding credential(s) are established between the CSP and the subscriber.
4. The CSP maintains the credential, its status, and the enrollment data collected for the lifetime of the credential. The subscriber maintains his or her authenticator(s).

The equivalent of the process for decentralized identities is the following:

1. Subject enrolls a digital identity for him/herself and potentially stores the identifier information to the Verifiable Data Registry.
2. Subject sends a credential request to the Issuer.
3. Issuer issues a new verifiable credential to the Subject which uses the Holder component to store the verifiable credential.
4. Issuer maintains the lifecycle of the credential and updates the revocation registry in the Verifiable Data Registry when the credential is revoked. Holder is responsible for maintaining the credential on behalf of the Subject.

This process for decentralized identities is compliant with the model described in “Aries RFC 0036: Issue Credential Protocol 1.0” [4].

1.11 Authentication

When dealing with centralized / federated identity solutions, in the context of digital authentication, NIST SP 800-63-3 defines the following roles [1]:

- Claimant: A subject whose identity is to be verified using one or more authentication protocols.
- Verifier: An entity that verifies the claimant's identity by verifying the claimant's possession and control of one or two authenticators using an authentication protocol. To do this, the verifier may also need to validate the credentials that link the authenticator(s) to the subscriber's identifier and check their status.

In the context of decentralized identity models, e.g. based on W3C Verifiable Credentials Data Model [3], the corresponding roles are the following:

- Claimant is equivalent to Subject whose identity is to be verified using the credentials in the possession of the Subject through a Holder.

- Verifier in centralized / federated identity solution is equivalent to a Verifier in decentralized identity model. In decentralized identity solutions, the Verifier verifies the Verifiable Presentations generated by the Holder.

NIST SP 800-63-3 defines the basic process of digital authentication as follows [1]:

1. The claimant proves possession and control of the authenticator(s) to the verifier through an authentication protocol.

2. The verifier interacts with the CSP to validate the credential that binds the subscriber's identity to their authenticator.

In some cases, there is no need for the verifier to interact directly with the CSP to complete authentication. This may be, for example, the case when using digital certificates as the credential.

NIST SP 800-63-3 defines the process for CSP credential lifecycle management [1]. Namely, the CSP maintains the status information of the credentials that it issues. In general, CSP assign credentials that have a finite lifetime. When the status of the credential changes or are close to expiration, credentials may be renewed or reissued. Alternatively, the credential may be revoked or destroyed. In a typical use case, the subscriber authenticates to the CSP with their existing authenticator and credential to request a new authenticator and credential. If the subscriber fails to perform this operation prior to expiration or revocation, they may need to perform the enrollment and identity proofing operation again to obtain a new authenticator and credential.

1.12 Federation

In centralized / federated identity schemes NIST SP 800-63C [2] defines federation as a three-party relationship between subscribers, identity providers (IDP) and relying parties (RP). Similarly, federation on the decentralized side can be seen as a three-party relationship between Issuers, Holders and Verifiers. Depending on the protocol, different information passes between different participants.

In a traditional federation setup, the subscriber usually communicates through a browser with the IDP and RP, whereas in decentralized setups the interaction between Holders and Issuers or Verifiers can happen through any user agent as defined in the W3C Verifiable Credentials Data Model [3].

In a centralized / federated setup, RP and IDP may communicate either with front channel calls or with backchannel calls. In a decentralized setup, there is no need for the Verifier (equivalent to RP) to communicate directly with the Issuer (equivalent to IDP) since the Holder is responsible of presenting the necessary Verifiable Credentials information (VC) as a Verifiable Presentation directly to the Verifier. The Verifier is then responsible of using the Verifiable Data Registry to determine that the VC was issued by a trusted party and through the revocation registry check that the VC has not been revoked.

As defined in NIST SP 800-63C [2], the subscriber authenticates to the IDP and result of the authentication is passed as an assertion to the RP. In this transaction, IDP acts as the verifier of the credential. The IDP can also make attribute statements about the subscriber as part of the process. These attributes and authentication information are passed to the RP in the assertion. The assertion is conceptually similar to the Verifiable Credential that the Issuer provides to the Holder in case of decentralized identity schemes. It contains information about the Subject, the Issuer and about the claims that the Issuer has made about the Subject. Similarly, as assertion is passed to the RP in centralized / federated schemes, the Holder can use the Verifiable Credential to make Verifiable Presentations to be passed on to the Verifier.

1.13 Creating identities

1.13.1 Enrollment

ISO/IEC 24760-1:2019 defines enrollment as the process that results in the creation of one or more identities for the enrolled party. Created identity information is registered as the enrolled entity's identity in a domain. The values of the unique attributes of the created identity can be chosen by an entity or can be assigned by the identity management system. The enrollment process can also include capturing of biometric data as identity information for the enrolled entity [1].

Enrollment in the context of digital identities consists of three phases:

- Identity proofing
- Enrollment
- Authenticator Issuance and Maintenance

1.13.1.1 Identity proofing

NIST SP 800-63-3 defines identity proofing as the process of verifying the subject's association with their real-world identity. The party to be proofed is called an applicant. When the applicant successfully completes the proofing process, they are referred to as a subscriber. [1] As defined in ISO/IEC 24760-1:2019, the identity proofing involves a verification of the provided identity information and can include uniqueness checks, possibly based on biometric techniques. [2]

Identity proofing is a policy-based operation that is based on enrollment policy which should include specification of the verification criteria of the identity evidence to be provided by the applicant. [2]

As defined in ISO/IEC 24760-1:2019, the identity proofing process is based on identity evidence provided by the applicant to achieve a specific level of assurance. Identity evidence in this context refers to the information that can support validating identity information. It is presented and gathered information related to an applicant that provides the attributes needed for a successful identification at a specific level of assurance. [2]

1.13.1.2 Enrollment

ISO/IEC 24760-1:2019 defines enrollment as the process of making the subscriber known within a particular domain. The process includes the collection and validation of identity information from the identity proofing phase and the collection of identity information required for identity registration, followed by the identity registration itself. [2]

1.13.1.3 Authenticator Issuance and Maintenance

NIST SP 800-63-3 outlines registration of subscriber's credentials and tracking the authenticators issued to the subscriber as duties of the CSP in the context of identity proofing and enrollment. [1]

NIST SP 800-63-3 defines multiple ways in which CSP may bind the authenticator to the subscriber [1]:

- The authenticator may be given to the subscriber at the time of enrollment,
- The CSP may bind authenticators the subscriber already has, or
- The authenticators may be generated later as needed.

Subscribers have the duty to maintain control of their authenticators and comply with CSP policies in order to maintain active authenticators. The CSP maintains enrollment records for each subscriber to allow recovery of authenticators, such as when they are lost or stolen. [1]

1.13.2 Registration

Registration is the process that follows identity enrollment. ISO/IEC 24760-1:2019 defines it as the process of recording an entity's identity information in an identity register [1]. After registration, an entity has become known in the domain and the lifecycle of its identity has started.

Registration can be for a specific or indefinite duration. National legislation can impose restrictions on the actual duration of indefinite registration, including when and how indefinite registration can end.

1.14 Changing identities

ISO/IEC 24760-1:2019 defines the process of maintaining identity information as the process performed by the identity management system to perform maintenance of identity information it has registered by changing one or more of the attribute values in an identity [1].

An identity management system shall specify mechanisms for maintaining the integrity and accuracy of attributes it stores. It shall maintain the identity information stored in the register as an accurate representation of the identity.

1.15 Deleting identities

ISO/IEC 24760-1:2019 defines the lifecycle model for identity information [1]. Initially, the entity is unknown and there is no identity information present. Identity information is established through identity creation. Finally, after deleting all identity information for an entity, the entity becomes unknown again. In addition to being in Unknown, Established and Active states, the identity can be:

- Suspended: Identity information is present in the identity management system specifically to indicate that the entity cannot utilize the resources of the domain. This is analogous to identity revocation.

- Archived: Identity information for an entity is still present in the identity register, even though the entity no longer exists in the domain. Archived information is not available for recognizing the entity except possibly during re-enrollment. When the entity re-enrolls, the archived information can be used to establish a new identity for the entity, which can include some of the archived information (restore).

Suspension is the process of marking some of the identity information stored in the identity register for an entity as being temporarily unavailable for use. Suspension can be achieved by removing access rights expressed in the stored identity information.

Archival is the process of moving the identity to an Archived state. It is the partial removal of identity information from the identity register for an entity.

Deletion is the complete removal of the identity information in an identity register. However, there may be legal requirements that prevent complete deletion of the identity information and may require the identity register to retain some identity information for auditing purposes.

1.16 Managing multiple identities

ISO/IEC 24760-1:2019 defines a reference identifier as an identifier in a domain that is intended to remain the same for the duration an entity is known in the domain and is not associated with another entity for a period specified in a policy after the entity ceases to be known in that domain [1]. A reference identifier for an entity can change during the lifetime of an entity, at which point the old reference identifier is no longer applicable for that entity. The identity register that is used to store the identity information is assumed to be indexed by a reference identifier.

The identity information stored in an identity register can include multiple reference identifiers. A reference identifier can be used to indicate a particular partial identity for the entity in a domain. This makes it possible to manage multiple identities in the identity management system.

1.17 Merging digital identities

As defined in ISO/IEC 24760-3:2016, merging of digital identities from different authorities in two distinct domains is a very typical requirement when two organizations are merging in a federation [1]. In these use cases, there is a need to define procedures to resolve collisions and inconsistencies, such as those within the resulting domain:

- Reference identifiers are unique.
- Pseudonyms are used where applicable.
- It is not possible to associate an identity with the wrong entity.

ISO/IEC 24760-3:2016 calls for the need for arbitration between authorities of identity providers that contain conflicting identity information.

1.18 Linking identities

As defined in ISO/IEC 24760-1:2019, the linking of identities should be based on pseudonymous identifiers that contain minimal identity information sufficient to allow establishing a link between known identities [1]. Pseudonyms are usually used to reduce the risk associated with privacy impact related to the ability to link identities across domains.

In addition to pseudonymous identifiers, ISO/IEC 24760-3:2016 defines the following other types of identifiers that can be used when linking identities:

- Verinymous Identifier
    - A verinymous identifier is a persistent identifier in its domain that can be used in identity linking across domains and allows Relying Parties to obtain further identity information about the user.
    - Common verinymous identifiers include, for example, email addresses, mobile phone numbers, passport numbers and social security numbers.
    - Verinymous identifiers may allow identity information in different domains to be correlated. While correlation is fine if it is allowed by the user, unexpected correlation such as profiling may have a negative privacy impact.
    - In case verinymous identifier information is leaked, an attacker may perform correlation and materialize threats from the information.

- Ephemeral Identifier
    - Only used for a short period of time and only within a single domain. It may change for multiple uses to the same service or resource.
    - When used correctly, using ephemeral identifiers makes it difficult for two visits of an entity to be correlated.
    - An ephemeral identifier is often used in the context of attribute-based access control where access to a resource is granted if the entity has a particular attribute.

1.19 Storing identities

An identity register refers to a persistent repository used to store identities. A typical identity register is indexed by a reference identifier. The identity information authority in a particular domain typically uses its own identity register. However, an identity register can be shared between related domains, such as within the same commercial entity. The reliability of the identity information in an identity register is determined by the identity proofing policies used during enrollment. [1]

ISO/IEC 24760-1:2019 defines an identity register as being either centralized or distributed [1]:

- centralized — A fully centralized system has a single identity register and a single point of control over enrollment and access to the stored identity information.
- distributed — An identity management system can have multiple identity registers and multiple points of control over enrollment and access to registered identity information.

A more centralized system typically displays less complexity but is more rigid in structure.

1.20 Delegation/Guardianship

For instance, in identity enrollment cases, there may be occasions where an individual cannot meet the identity evidence requirements to enroll a digital identity. For these kinds of cases, there is a need for guardianship and NIST SP 800-63A [1] defines a process where a trusted referee may assist in the identity proofing of the applicant.

These guardians or trusted referees can include, for example, notaries, legal guardians, medical professionals or persons with a power of attorney. Their role is to vouch for or act on behalf of the applicant in accordance with applicable laws, regulations or other policies. The trusted referee may act either remotely or in an in-person process.

NIST SP 800-63A mandates that there is a need to have in place a written policy and procedures on how a trusted referee is determined and the lifecycle by which the trusted referee retains their status as a valid referee. The policy must also include any restrictions as well as revocation and suspension procedures.

Re-proofing of the subscriber should be performed at regular intervals.

Special attention needs to be put on cases where the information of minors is being handled:

- Relevant regulation (e.g. COPPA in the United States) requirements need to be followed.
- Parent or legal adult guardian should be involved as a trusted referee for an applicant who is minor.

1.21 Interoperability

In his article “The Path to Self-Sovereign Identity”, Christopher Allen outlines interoperability as one of the key characteristics of a successful digital identity scheme. The core message is that: “Identities should be as widely usable as possible.” [1] This means that the digital identity information should be made widely available, crossing international boundaries without losing user control.

As defined in ISO/IEC 24760-3:2016 [2], one of the ways in which the benefits of interoperability can materialize is through standardized federation protocols where the objective is to offer interoperable exchange of identity information and leverage the benefits derived from it, such as expanded consumer productivity and efficiency that in turn enables growing and prosperous digital economy.

The World Bank’s ID4D Practitioner’s Guide [3] takes the definition of interoperability a step further and defines a three-level model about the ways in which interoperability should be guaranteed in digital identity systems. In their definition, interoperability should be present:

- Between ID subsystems (components/devices)
- With other domestic systems
- With ID systems in other jurisdictions

Through this definition it is possible to achieve a range of benefits, such as:

- Technology and vendor neutrality
- Ensuring integrity of identity information
- Efficient administration
- Reducing fraud and improved targeting
- Improved user experience
- Innovation and new use cases

2 Consent

2.1 In scope

2.1.1 Consent lifecycle management

Consent follows a lifecycle that has to be managed - from requesting Consent, granting it, modifying it and revoking it. At this stage, requesting, granting and revoking should be covered by solutions. If Consent is not revoked, it should in any case expire after a certain agreed upon date. Several IHAN Blueprint components take part in the Consent lifecycle:

- **Personal consent directory (PCD)** stores all End User's Consents given to the Service providers. Service providers will use this information to access data from Data providers (IHAN Blueprint [12]).
  Note: An End user need not be an individual. In that case, a "personal" consent directory might be a "corporate" consent directory.

- **Service provider consent directory (SPCD)** contains records of all received Consents from all End user Service provider's Services (IHAN Blueprint [12]). There are two identical versions of Consent - End users and Service providers. Both parties have the possibility to prove the contents of the consent.

- **Data Access Control (DAC)** is a component that orchestrates the process of receiving data requests, identifying individuals and associated data, accessing the data and delivering it to Service Provider(s) (IHAN Blueprint [12]). It stores (part) of the Consent for its own reference - the authorization part and the required data part.

2.1.2 Structure of Consent

Parts that constitute the Consent and enable the legal, technical and business views of the Consent shall be described at a conceptual level of requirements.

2.2 Out of scope

- Defining the schemas for data / metadata (naming of data fields)

- Defining the APIs

- Modifying of consent in the Consent lifecycle

2.3 The facets of Consent

The term "consent" can be understood in the general sense of agreement to something was given. In the legal sense, defined by the General Data Protection Regulation (GDPR), the act of receiving consent means that certain kinds of data processing activities can be done on Personally Identifiable Information (PII). The legal basis for data processing in that case is based on consent (other legal bases for processing exist). In the IHAN Blueprint, Consent is viewed in a more technical aspect - with consent you not only get the agreement but also technical means and authorization to get access to the data in question. Also, Consent in IHAN should not be viewed as exclusively needed in the legal sense - you might use a different legal basis for data processing according to GDPR, not necessarily consent, but the technical part is still needed. We can look at consent in different ways: from the legal, technical and business point of view.

In the text, when we use uppercase "Consent", we mean it in the form as is needed by IHAN. When we use lowercase "consent", we mean it in a general way of the act of agreeing.

2.4 Legal view of Consent

As IHAN is based in the European Union environment, the General Data Protection Regulation [4] should be our legal framework for establishing the Fair data economy. Several Articles of GDPR touch upon consent and have to be taken into account, listed below for an overview:

- Article 4 (11) of GDPR defines *consent* as:

> 'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;

- Article 6 (1a) of GDPR defines having consent of the data subject for processing of his personal data (for one or more purposes) as one of the possible six legal bases for lawful processing of personal data.
- Article 7 of GDPR defines the conditions for consent:
    - It is up to the controller to demonstrate that the data subject has consented to processing of his personal data.
    - If in written form, the request for consent must be clearly distinguishable from other matters and clearly understandable.
    - The data subject has the right to withdraw consent at any time.
    - Consent must be freely given.
- When processing personal data of children under a certain age, consent must be given by the holder of parental responsibility (GDPR Article 8).
- When processing special categories of personal data (more sensitive in their nature - e.g. health data), explicit consent has to be given unless other exceptions apply (GDPR Article 9).
- Article 13 of GDPR defines the information that has to be provided to the data subject when personal data are collected. This applies also to cases where given consent is the legal basis for processing of personal data, but not exclusively to them. The data subject has to be informed of any further processing outside of the initially communicated purposes.
- In the case that personal data are not obtained directly from the data subject, Article 14 defines similar information about terms of usage of his personal data has to be provided to the data subject.
- Article 41 of GDPR describes monitoring of compliance with codes of conduct by supervisory authorities. Consents should therefore be recorded in such a way to facilitate the monitoring process including tasks described in Article 57 of GDPR (e.g. (a, f)).

2.5 Technical view of Consent

IHAN Blueprint requires that with the act of giving consent, the technical means to access the data to be processed are also given to the Data controller. This means that with consent,

authorization details to get to the data residing with the Data provider have to be given. The authorization details have to be properly protected to be only accessible by the required parties. The metadata about the naming / structure of the data have to be passed in the consent, for semantic interoperability. The Consent receipt therefore has to have additional data that enables this.

Structure of Consent:

- Human readable part describing the relationship between End user and Service provider (what kind of data is used, purposes, etc. - allowing for GDPR compliance). Kantara specification [2] should be used to generate a Consent receipt. Publicly available vocabularies can be used for describing the fields (e.g. W3C Data Protection Vocabulary [3]).

- Data specification part describing how to get to the data, the APIs, metadata about the data. This should include descriptions of which fields might contain personal identifiers, sensitive data, etc., and implying how the data should be processed.

- Authorization part with authorization details to access the data at a Data provider.

2.6 Business view of Consent

With properly recorded consents, the level of regulatory compliance for companies should increase. Not only that, but the trust the individuals assign to those companies should increase as well - as now individuals have greater control and overview of how their personal data is used.

To make interoperability as seamless as possible, the structure of the Consent receipt in the IHAN ecosystem must be common. This makes it possible for companies to more easily join the IHAN Fair data economy and get to the data they need. It also makes it easier for individuals to have greater control over the data they shared with a Service provider or are keeping at a Data source. With a standardized Consent receipt, an individual could be offered services on top of his “shoebox” of consent receipts, to see which companies have which kind of data about him and exercise different actions on them (updating, revocation, GDPR portability right, etc.).

If for every kind of data usage by a Data controller (company) one would issue a Consent receipt to the individual, the individual saving his Consent receipts in his “shoebox” would in time get a total overview of what kind of data companies keep about him, where his data resides and what it is being used for. (Note that not every usage of data is based on consent, so a more appropriate term for such a receipt might be a Data receipt.) Assuming also that the information in the Consent receipt is semantically equipped, this would open the possibility for him to leverage services that are based on that data, even if they are not under his direct control (e.g. are stored somewhere else). This, for example, opens opportunities for various analytics services to offer him insights based on that data, as well as enabling more complex scenarios.

2.7 Summarizing Consent

Consent in the IHAN sense could be viewed from the three facets described - legal, technical and business. It enables data exchange between Data provider and Service provider, while giving the individual End user control over how their data is used. It can and should be used even if no data exchange is required - e.g. if the Service provider is also the Data provider. The Consent must be recorded in digital form and available to both the individual (End user) and the Service provider, possibly in part also to the Data provider(s). The individual must also have the option to view (relevant parts) of the contents of the Consent in human

readable form. A digital receipt recording the details about the act is called a Consent receipt.

2.8 Structure of Consent

The Consent is technically comprised of three parts: human readable part, data specification part and the authorization part.

2.8.1 Human readable part of Consent

From the existing standardization standpoint, the consent is defined in detail in ISO/IEC 29100:2011[5]. It defines consent as the PII principal's freely given, specific and informed agreement to the processing of their PII. This definition is compliant with the way in which consent is defined in the GDPR. ISO/IEC 29100:2011 extends the definition of consent to list the items through which consent principles may be followed. These include the following examples:

- Presenting the choice whether to allow PII processing.
- Obtaining the opt-in consent for collecting or processing sensitive PII.
- Informing about the rights of the PII principal.
- Providing information indicated by the openness, transparency and notice principle.
- Explaining the implications of granting or withholding consent.

For a PII controller, adhering to the consent principles means the following:

- Providing clear, prominent, easily understandable, accessible and affordable mechanisms to exercise choice and give consent.
- Implementing the PII principal's preferences as expressed in the consent.

An appropriate form to record this information in digital form is called a *consent receipt*. It should contain the required information and it can contain additional information, extending its possible uses. Kantara's consent receipt specification is a de facto standard for the structure and contents of the Consent receipt. It has also been adopted by some IHAN Pilot projects. The specification is part of the ISO New Work Item Proposal on how to become part of an ISO standard (NWIP TS Privacy technologies – Consent record information structure; ISO/IEC TS 2043 [6]). A human readable rendering of the consent receipt must be presented to the individual for his approval and for historic reference. It should be rendered from the digital Consent receipt. Different viewers/renderings can be available to the individual, depending on which information might be relevant to him. Both the individual and the Data controller save their own copy of the Consent receipt to facilitate the GDPR compliance process towards the regulatory authorities or any future negotiations between themselves.

The Kantara consent receipt specification defines the elements that should be present in the human readable consent proposal and receipt, available to the individual (and issued by data controller). A standardized Consent receipts allows for easier record keeping and building services on top of the data available in the Consent receipts (such as GDPR functions).

To allow for flexibility, the basic Kantara Consent receipt schema might be upgraded with Overlays Data Capture Architecture [7]. This would enable more fine-grained control over what is relevant to be shown as a human readable representation: hiding less relevant

fields, localizing the language/labels of fields shown and possibly including additional formatting information.

To describe the possible terms to categorize personal data handling in accordance with GDPR, the Data Privacy Vocabulary [3] was constructed under the umbrella of a W3C Data Privacy Vocabularies and Controls Community working group (DPVCG). It includes classes and properties for different GDPR categories of actors and actions. It also includes comprehensive vocabularies of possible personal data categories, purposes, processing categories, technical and organizational measures and others, as well as ontologies connecting these vocabularies. Although not a standard, it is a quite comprehensive set of vocabularies and ontologies that can be used in the context of the human readable consent receipt, bringing the data inside it into the machine-readable realm and establishing linked data concepts.

2.8.2 Data specification part of Consent

For specifying the data format of the data to be exchanged, an open and vendor neutral data format should be used to facilitate interoperability and future proofing. Vocabularies enable a common understanding of data by different parties.

For behavioral data collection and processing, OASIS Classification of Everyday Living (COEL) [8] could be used. An example in healthcare is the OpenEHR [10], part of which also defines the semantics of the Electronic Health Record (EHR). In the absence of a standard, community developed schemas for data could be used, benefiting from being semantically equipped and being maintained with interoperability in mind. schema.org [11] maintains a large collection of linked data schemas that might be considered in the absence of a standard.

It should be noted that data might be exchanged across national borders and different languages could be used. Therefore, among other things, specification of exchanged data should consider that data might be specified in a language foreign to the data provider. The ODCA architecture [7] could be used to solve the problem of internationalization and interoperability in describing the data.

The data that could immediately identify a person are identifiers of different kinds. These must be treated especially carefully as they can “unblind” and otherwise anonymous data set. A list of these is compiled under the name of Blinding Identity Taxonomy (BIT) [2] and resides with the Kantara initiative.

2.8.3 Authorization part of Consent

ISO/IEC 29146:2016 [9] defines the standard process of authorization in the context of access management as the decision performed by a Policy Decision Point (PDP) to deny or allow access to the resource based on a given policy and an access token issued to convey the result of the decision. The PDP implements the access control policy or policy set for the resource. Based on a defined set of policies, the PDP decides whether the subject may access the resource. Security Token Service (STS) is the service used to issue the access tokens.

The PDP is supported by the policy information point (PIP). This component acts as a source of attribute values (e.g. resource, subject, environment condition) that are used by the PDP to make the authorization decision.

Controlled access to resources happens through the following steps:

1. An authenticated user discovers the location of the PDP and STS.

2. User requests access authorization to certain resources from PDP. Based on the policy or policy set provided by the PIP, the PDP determines whether to grant authorization.

3. If access is granted, an access token is issued and passed to the user.

4. The user can use the access token to access the resources.

2.9 Requirements

The following are general IHAN and more specific Consent-related functional and non-functional requirements that must be fulfilled.

- Several solutions might be developed for the same component, but the solutions should be technically compatible to avoid creating software silos that prevent open data exchange (IHAN Blueprint).

- Functionalities and software interfaces between components should be standardized to a point where interoperability is straightforward to implement (IHAN Blueprint).

- There should be a standard way for metadata exchange, consent management and usage and actual data exchange between the End users, Service providers and Data providers (IHAN Blueprint).

- Consent contains criteria for data usage purposes and may contain rules for what must happen to the data at Service provider after service provision (IHAN Blueprint).

- Consent must include credentials to access the data at the Data provider's Data Access Control component.

- The ecosystem must allow for fair value exchange (IHAN Blueprint).

- The ecosystem must remain distributed and contains no design decisions that create centralized solutions (IHAN Blueprint).

- The ecosystem must be secure as it is handling personal data that is governed by GDPR and other regulations (IHAN Blueprint).

- Each Consent defines all Data access records that will be used to request data from Data providers (IHAN Blueprint).

- Consents must have parts to enable (modified from IHAN Blueprint):

    - Part 1 must be readable only to the Service Provider and must contain information about the Data Providers (interfaces for metadata, and data request).

    - Part 2 must enable secure access by the Service provider to the data about the End user held by the Data provider. There could be multiple Part 2 –type elements, one for each Data Provider.

    - There should be an additional, Part 3, of the Consent where the relationship between the End user and Service provider is described, as needed per GDPR.

- The Consent structure should be known and follow an established standard.

- All communication between distributed components should be using secured encrypted connections.

- Identifiers should be used in such a way that the provenance of a dataset could be possible to establish when warranted (in the spirit of the IHAN Identifier, IHAN Blueprint).

2.10 Functionality achieved for actors

Functionalities achieved for actors in the IHAN Fair data economy will be available through different implementations, considering the constraints of the IHAN Blueprint requirements.

As interoperability is one of the key requirements, it should be possible to "mix" different implementations of components when using them, as much as possible. This requires agreed upon APIs and data structures, which are not yet defined. The flow should also be unified for this to be possible and should follow the conceptual flow described in the IHAN Blueprint.

For example, "centralized" and "decentralized" variants of the consent lifecycle are possible, which differ due to technical reasons. Yet, we should attempt to unify them through abstracting out the unneeded steps and unifying the usage interface. For instance, storing a Consent receipt in a wallet should be a similar API call, no matter what kind of storage underlies it (e.g. object storage or decentralized storage).

2.11 Reference implementation (centralized approach): Tilaajavastuu MyData Platform Consent Management

Sharing personal information following MyData principles can be achieved in two ways:

- With a centralized approach using OAuth 2.0 type of access tokens and consents managed
  centrally in a data operator.

- With a decentralized approach using Verifiable Credentials and consents managed in a user's
  personal wallet.

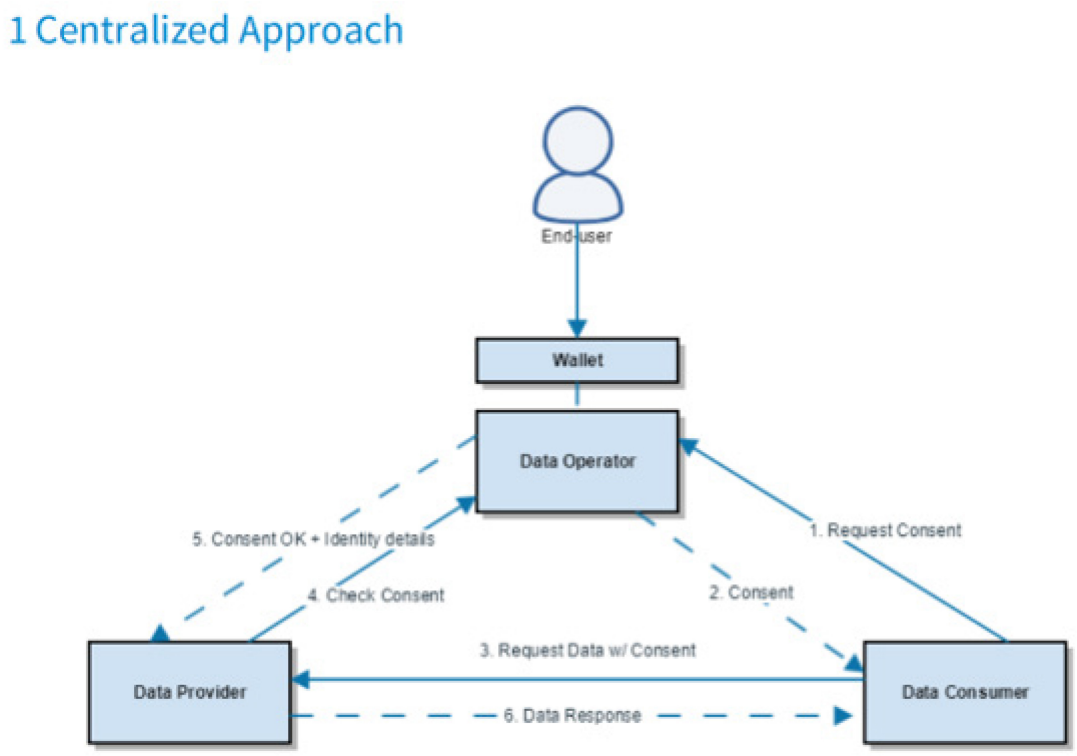
These models are complementary and can co-exist in parallel so that part of the integrations are carried out with a centralized approach and other part with the decentralized approach.

In the centralized MyData setup, there are three counterparts:

- Data Providers that offer personal data resources.

- Data Consumers that want to utilize the personal data resources offered by Data Providers.

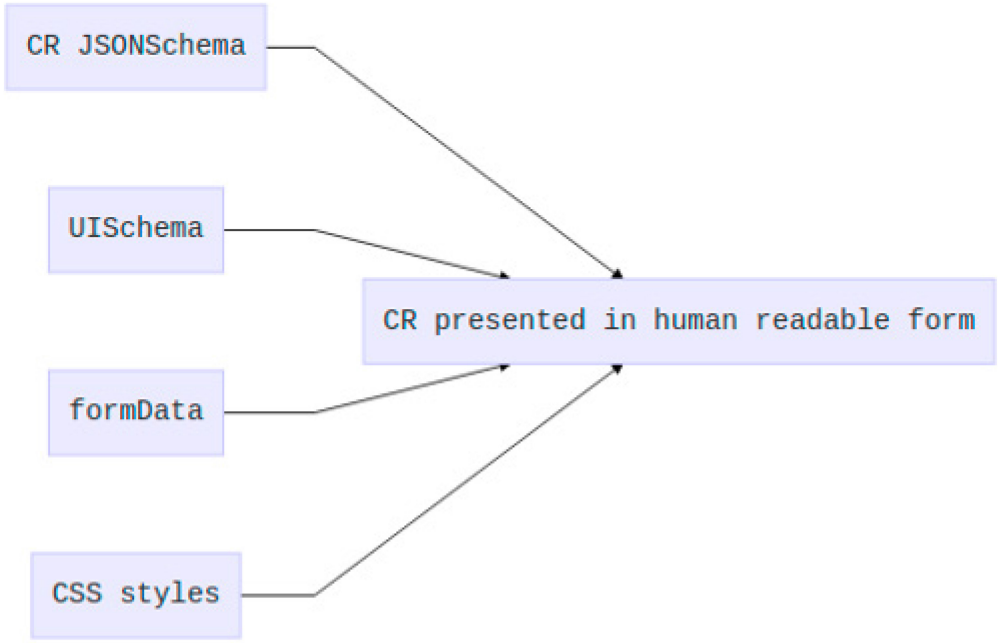
- End user who has data assets in the Data Providers and wants to be in control of how his/her personal data is shared among Data Providers and Data Consumers.



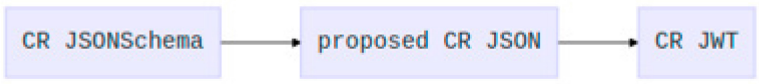
In the centralized approach, the user controls data access through a Data Operator whose functionalities are offered to the end user through a Wallet. Through the Wallet the end user can explore his/her data in different Data Providers and manage consents given to individual Data Consumers for accessing data in Data Providers.

2.12 Reference implementation (decentralized approach): Datafund Kantara compliant Consent Receipt Suite

Building upon the React JSONSchema framework, the presentation of CR on screen in either the human readable and read only form or the more flexible editable form used by the PII controller is achieved by a combination of CR JSONSchema (defining the structure and allowed values), UISchema (defining the controls / menus displayed on screen), formData (defining the default / displayed values) and CSS styles (defining the look of the UI). All the parts are saved in the form of a JSON project file, which can be used as a template.



The "CR JSONSchema" defines the structure of a CR. Starting from the schema, the PII Controller can generate a "proposed CR JSON", by adding values to all the relevant fields. The "proposed CR JSON" should be presented to the PII Principal in an appropriate way (human readable and possibly using one of the developed modules). PII Principal can give his consent, after which a "CR JWT" is generated which both the Controller and Principal can save (to different variants of storage).



Kantara specification of the consent receipt is used, which is a de facto standard. Other JSONSchema can be used, depending on the need of the user.

Consent Receipt Suite also allows for storing the Consent receipt in Swarm [13] decentralized storage, where the same version of the file can be accessed by both involved parties using their Fairdrop accounts, establishing a single source of truth. This method of storage removes the need to synchronize Consent receipts between the parties.

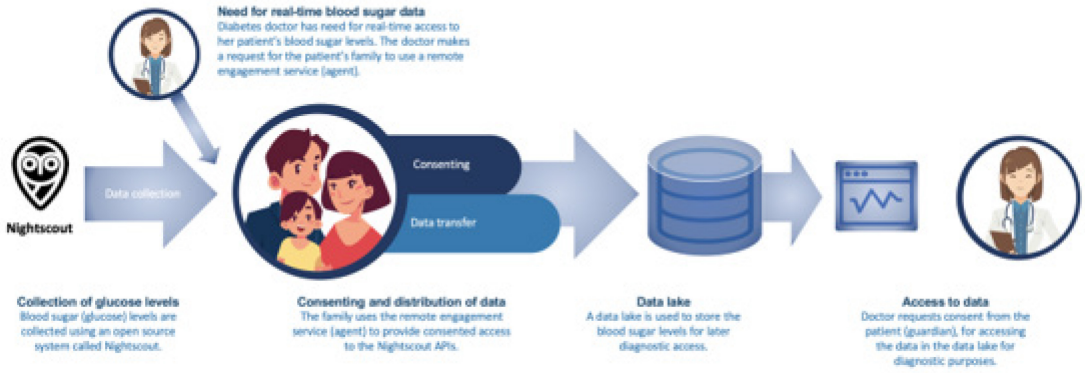
The Consent receipts can be signed by both parties via their private keys and the transactions recorded to a blockchain, to prove the acts of asking/giving of consent.

2.13 Reference implementation: Dynamic API authorization and data access using Verifiable Credentials (Case HUS Child's Diabetes consenting)

Traditional API authorization mechanisms work well when the data sources in use are known, and consent for using the data can be applied and verified through traditional centralized means. However, when an end user is using a data source that is not previously known, and where explicit consent to access data is required because of strict regulation,

there is a challenge to how this can be done while retaining control of the data at the data subject's hands.

In the use case of providing Consented API access for Diabetic Child's Blood Sugar level, the approach is to implement a decentralized Verifiable Credentials (VC) based consenting model. In this model, the patient (or guardian) is able to provide the healthcare provider, a cryptographically verifiable credential, that contains both the consent information, as well as the data source location information. Using this credential, the healthcare enterprise system is able to prove that they have a consent from the data subject. Additionally, when the personal data is gathered, there is a second set consent required by the clinical staff to access the information, enabling end users with a way to control access to their data from an external system.



3 Logging

3.1 Logging in IHAN

IHAN is about Human Centric Fair Data Ecosystem where giving and changing the state of consent related to data is an essential enabler. Contracts and agreements between the End User and IHAN service and data providers must be reflected in the logging so that the End User can verify that these arrangements have been respected.

In order to gain End Users' confidence in the system, they must have proof that their will about data has been honored. In practice this means that a reliable and chronological track record of events always must be available for inspection. Without immutability and proof of respecting End Users' given consents there is a low level of trust in data economy.

Identity and consent management combined with undisputed and immutable logs provide the fundamentals for transparent trust in circumstances where trust is not self-evident.

Logging in IHAN will build the framework to answer questions for fair data economy implementors and users.

- How can I as a service provider prove I had the end user's explicit consent for processing their personal data?
- How can I as an end user see who has accessed my personal data?
- How can I as a data provider prove that I have shared personal data with the end user's explicit wishes?

Logging in IHAN context is not focused on technical or application / service internal logging. The focus is to log user data-related events: What user data is transferred between different stakeholders? How can the trace of events be tracked in a meaningful way? How is the immutability of the logs ensured?

3.2 In Scope

This document describes the background and principles of logging in IHAN context for IHAN compatible component design purpose. It is not a comprehensive technical requirements list since it varies between different business domains. It describes the IHAN context aspects as guidelines to be considered in component design. Defining the detail level requirements for implementing the component in the boundaries set by the guidelines remains the task of the component development team.

All state changes on End user data must be logged alongside the party responsible for the change. Physical logging covers the changed content and logic of the commands that lead to the changes in the content.

The approach in this document is on defining the dimensions relevant for IHAN compliant logging to set the expectation level. Dimensions are followed by the use cases found by the work stream organizations, received through external comment rounds, pilot projects and from other IHAN CEN-CENELEC work streams. These use cases result in requirements that IHAN compliant services, applications and components need to consider during the design phase.

3.3 Standardization Needs

Transparent identity and consent management is critical for IHAN compatible trust generating architecture. Logs are the witness through which all operations related to them

can be verified. Standards accelerate component development in a new business area by giving a framework and guidelines on top of which they can be designed and implemented.

3.4 Out of scope

Defining exact data models, APIs and endpoints are out-of-scope. Individual services will have their detail level requirements built on top of a framework and guidelines established in this document.

3.5 Background information

The past decade has been characterized in the digital services industry by the increased use of personal data. The uncontrolled hoarding of personal data has led to an increase of privacy regulation like GDPR, CPA, DPA, HIPAA and CBPR.

A fundamental requirement of all privacy regulation is the increased transparency of who has accessed an individual's personal data and how the data is purportedly used. One way of achieving this is through logging of any personal data-related events. In the IHAN context, we are most interested in data and events around personal data use. Some categories with logging aspects around personal data are:

Log Categories	Description
Access logs	Who has accessed the data
Consent logs	Collection of consents
Event logs	Events around accessing personal data, requesting, issuing, editing and revoking consent
Data sharing notifications	Notification on lawful sharing of data without explicit consent
End User Consents to Contracts and Agreements	See dedicated chapter

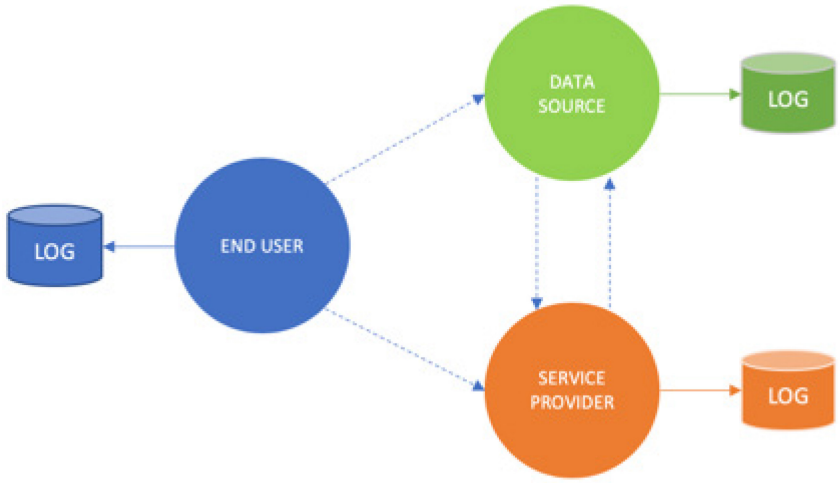
3.6 Architectural models

In case there are multiple logs upheld of the same event (e.g. data source and data subject's operator both hold their own logs), there should be either a syncing or verification method between the logs to resolve potential non-consensus situations.

Logging can be done centrally inside a single role, in case which that party needs to provide all the logging needs for that use case. A decentralized approach would allow distribution of logs between roles, while maintaining logical connections between them. Another approach to distributed logging holds the log in a decentralized storage and makes the immutable log accessible to all parties, making only one version of the log needed. Swarm [13] presents an example of this kind of architecture.

Centralized model

In a centralized model, each role keeps their own log and is centrally responsible for its integrity.



Decentralized model

In a decentralized model, a globally accessible distributed database, such as a distributed ledger or blockchain, is used to maintain a single source of truth for logging.

3.7 Use cases

Use cases define context further processed into the requirements below. They help to shape the IHAN architecture still in the draft phase.

3.8 Data Operator

A Data Operator manages some aspects of the exchange of data on behalf of other entities (companies, organisations, or individuals). A Data Operator may concentrate on managing the technical transfer of data, on operating the consent management process, on negotiating deals for secondary use of data, on calculating and distributing the income from licensed data between entities, or perhaps on all of these aspects.

3.9 Distributed Identity Agent

A distributed identity agent is software that can be used for creating secure messaging channels with other similar agents. An agent can be used on behalf of an End User or an organization for exchanging digitally signed messages, such as verifiable credentials. An agent can receive loggable events, such as reception or revocation of a consent.

3.10 Data Provider

A data provider holds information about the End User and offers interfaces for retrieving data. The data provider concentrates on verifying that the data requester has proper consent and/or authorization to retrieve the data. A data provider can be, for example, a simple API service.

The data provider is responsible to the End Users and authorities to uphold personal data with the utmost care. A data provider's interests are in proving that data is shared and processed according to one of the six legal bases (see the table below) defined in GDPR, and that they have carried out their due diligence when somebody requests access to an End User's data.

The following information should be logged:

Aspect	Notes / Example
What operation was performed?	Catalog of possible operations for data modelling and distributed logging purposes
Which component/system performed the operation?	Catalog of components in the system for data modelling purposes
Who initiated the operation?	Identifier of the authenticated individual or organization, if such identification can be performed, and is in scope of the service.
Which component/system received information about the End User?	Catalog of components in the system for data modelling purposes
What End User information was handed over?	Metadata structure / elements transferred
When was the operation performed?	Timestamp synchronization in a distributed system
Did the operation succeed?	Verification method
Which consent was used?	Catalog of consents used in the system.

In short: the data provider should log data access and what basis that access has had.

NOTE: An IHAN application or service may require the use of non-IHAN compliant data providers. Due to regulatory or other reasons, access to the logs of these components may be restricted or prohibited by the data operator [14].

3.11 End User

The End User is interested in having transparency regarding his/her personal data. Who has accessed the data and on what grounds? What kinds of consents of delegations of authority has one given, and how they are used?

At least the following data is logged in the End User's personal log:

Aspect	Notes / Example
Identity changes (Operation?)	• For example, a new Identity Record is created, i.e. a new identity provider (3rd party) and credentials are linked to the Personal Identity Wallet, or an existing one is removed or modified • Catalog of allowed transitions for identity in the system
Definition of identity in the context	• Identity definitions provided in the section describing the End User identity in the IHAN context
Service changes	• For example, a new Service Provider is added to the Personal Service Directory • Catalog of allowed transitions for service in the system
Service usage	• For example, the End User uses a service provided by a Service Provider • Catalog of services in the system
Consent issuance	• When an End User provides a new consent for accessing or fetching personal data
Data usage	• For example, a new Data Access Record is created to be used with a selected Data Provider or a Service Provider uses a Consent to access data for a Data Provider • Metadata structure / elements accessed
Revocation of consent	• a previously issued consent is revoked for data users

3.11.1 End User Consents to Contracts and Agreements

Agreements and terms of services are important in the IHAN context. Traditional EULAs are replaced with TOS that describe the relationship between End User, Service Provider and Data Provider(s) in a meaningful way. The content of the TOS and user's agreement to it needs to be logged to verify afterwards that the End User has gone through it during the initial state of the given consent. TOS defines what data will be collected and used.

All changes to TOS and the End User's agreement to the changes to it must be logged. If a data provider wants to share End User data with other service providers, the user's consent has to be logged. If the End User instructs some data to be forgotten by the data source or service provider, as per GDPR he / she is entitled to, the activity needs to be logged.

3.11.2 End User Consent Wallet

A Consent Wallet is an independent app or a service where individuals can store their given consents (based on a Consent Receipt issued or acknowledged by the Data Controller), inspect the use of data based on the consent, and manage the state of the consent.

See, for instance, the Kantara Consent Receipt specification [1] and the related use cases.

A Consent Wallet may be implemented, for instance, as a web service, as a mobile application, or as a combination of the two.

“Fairdrop as Consent Wallet (Datafund)” section in this document describes a decentralized application that uses Consent Suite [16] (Sitra pilot project).

3.12 Service Provider

Service providers are the most active parties in the data sharing network as they need to apply for the consent and then prove the consent to gain access to the data provider’s data.

The service provider has a legal requirement to prove that they have a basis for accessing the personal data of the End User. The data sharing process has the same requirements as presented in the table for the data provider.

3.12.1 Service Invocations

When a service is invoked, the consent for the invocation and the data (structures) access must be logged.

3.12.2 Inspect the Flow of Data

Data from data sources may require processing, like transformation, merging and filtering to be harmonized into a meaningful structure. The logic of these processes must be transparent and the flow of data through the process logged. In case of erroneous processing, the logs must reveal the point in the process where the logic has failed.

3.12.3 Contract-based Events

Business models in IHAN may be based on the number of events or transactions executed according to a contract between the end user and service provider. In this kind of scenario, logs must provide an undisputed basis for billing.

3.13 Requirements

The requirements presented in this chapter follow the principles defined by the Internet Engineering Task Force [17]. The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

3.13.1 Data Operator

The Data Operator MAY log data related to technical transfer of data.

The Data Operator MAY log data related to End User consent management process.

The Data Operator MAY log data related to contracts for the secondary usage of the data.

The Data Operator MAY log data related to distributing the income from licensed data between entities.

The Data Operator MAY interface with Data Providers to aggregate personal data access logs of the End User.

3.13.2 Agent

The Agent SHOULD have a logging system that can react to event triggers.

The Agent MAY expose an internal API endpoint for trusted internal systems to receive triggers.

The Agent MAY expose an internal webhook system for throwing event triggers requiring logging.

The Agent MUST trigger a logging event when it receives personal data related requests.

The Agent SHOULD log all given credentials

The Agent SHOULD inform all parties if a credential has been revoked

3.13.3 Data Provider

The Data Provider MUST log all personal data related access requests.

The Data Provider MAY provide an interface to End User to access the personal data request logs.

3.13.4 End User

The End User MUST log all changes to their identity.

The End User MUST log all changes to Personal Service Directory.

The End User MUST log all consent issuance and revocation operations.

3.13.5 Derived from End User Consents to Contracts and Agreements

The End User MUST log all content of the TOS and changes to them and End User agreement to them.

If the End User instructs some data to be forgotten by the data source or service provider, it MUST be logged.

3.14 Service Provider

3.14.1 Derived from GDPR

GDPR gives the End User rights to control his / her own data and consents. The processing organization should be able to demonstrate proof of consent and allow individuals to review previously given consents and withdraw it if necessary. A well-defined framework for persisting and exposing logs of activity also helps to protect the rights and interests of the Data Controllers and Data Processors.

Article 30 (EU GDPR "Records of processing activities") [4] describes the requirements for logging (recording) of the processing activities that are relevant in the IHAN context.

3.14.2 Derived from Service Invocations

The Service Provider MUST log the End User’s consent for the service invocation.

3.14.3 Derived from Inspect the Flow of Data

The Service Provider SHOULD log possible data transformation and processing, if executed for the data retrieved from the data source.

3.14.4 Derived from Contract-based Events

The Service Provider MUST explain the event and transaction-based billing mechanism to the End User in the contract between the End User and the service provider so that the End User can verify the validity of the billing.

The Service Provider MUST log all events and transactions on End User data.

The Service Provider SHOULD provide an API endpoint to event and transaction log to End Users for verifying the number of events and transactions used as basis for the billing

3.15 Data Model

3.15.1 Derived from Consent Wallet

The Consent Receipt SHOULD expose the initial state of the Consent.

The Consent Receipt SHALL include the address, and required credentials, if any, of the API to access the logs of the Data Controller related to the consent (Log API).

The Consent Receipt MAY expose a Push API, which the Consent Wallet can register to, in order to get push notifications of changes to the log.

The Consent Receipt MAY expose a Pull API, which the Consent Wallet can poll on demand or at set intervals to receive information on the changes to the log.

3.16 Reference implementations

3.16.1 Distributed consenting for API access

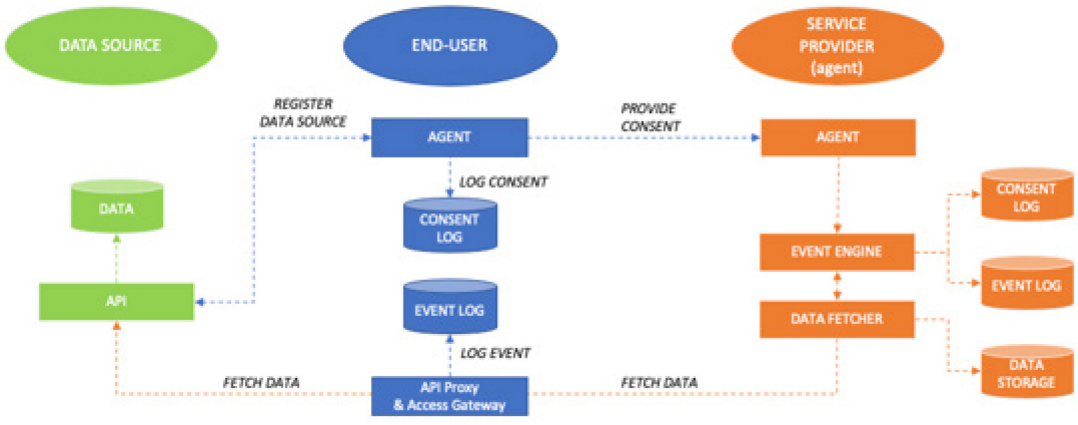
Case HUS, Finnish Olympic Committee & Finnish Defence Forces

The use case enables consenting of external data source (API) access, using a distributed consenting mechanism. In the use case, the End User uses an agent application to govern digital connections to service providers and connect them with external data sources. The End User uses these secure digital connections to provide consents for accessing APIs on a third-party data source.

The use case has two types of logs:

1. **Consent logs** store information about the consent and the subjected data source

2. **Event logs** are used to track any events that have to do with the consent. These include use of consent, revocation of consent, changes in consent, etc.



The image above describes the use case functionality.

1. **End user** registers **data sources** which access can be consented to **service providers**

2. **End user** provides a consent credential to **Service provider**, which includes also access credentials. Both parties log the consent.

3. **Service provider** uses the consent credential to retrieve data from the data source. Service provider and end user log the usage event.

The End User wants to

- keep a record of all who have been given consent to access their data sources
- keep a record of all who have used the consent to access the data sources
- keep a record of all created credentials that handle personal data
- Keep a record of all revocations for the given credentials

The Service provider wants to

- Keep a record of the consents they have received from End Users
- Keep a record of all revocations for the held credentials (e.g. consent)
- Keep a record of all operations relating to the credentials (e.g. consent)
- Keep a record of all operations relating to fetching and using of the data

3.16.2 Kela Kanta Services

Kanta [15] produces digital services for the social welfare and healthcare sector. These services benefit citizens as well as social welfare and healthcare service providers. The users of the services include citizens, pharmacies, and healthcare and social welfare services.

There are three types of database logs maintained in the Kanta patient data repository:

- event log (technical log)
- use log (logging of custodians' use of their own data saved in Kanta)
- log of disclosed data (disclosure of data between organizations from Kanta)

Logs in the prescription service are the same excluding the "log of disclosed data" because of the difference in custodianship of the data. The logs are not accessible outside of Kanta. My Kanta pages (OmaKanta), however, show information about the disclosure of patient data on the organizational level.

The Kanta centralized consent management system currently consists of information, consent and refusal/denial documents how patient's data can or cannot be shared from Kanta between organizations. Those documents are basically forms (CDA R2) where the patient/user has given consent to share patient data from Kanta services according to the legislation and that he/she understands what this means. There is an interface (SOAP) in Kanta to check what kinds of documents are saved in the consent management system for the patient.

4 References

CEN-Cenelec Level of Guidelines for CWA Documents

Although a CWA is developed outside the normal CEN/CENELEC technical body structure, it is important to ensure the coherence of all the different CEN/CENELEC deliverables in order to protect the credibility of European standardization. Therefore a CWA shall not conflict with a European Standard.

[1] https://kantarainitiative.org/

[2] https://kantarainitiative.org/confluence/display/infosharing/Blinding+Identity+Taxonomy

[3] https://w3c.github.io/did-core/

[4] https://eur-lex.europa.eu/eli/reg/2016/679/oj

[5] https://www.iso.org/standard/45123.html

[6] https://ipen.trialog.com/wiki/ISO#NWIP_TS_Privacy_technologies_.E2.80.93_Consent_record_information_structure

[7] https://odca.online/odca_poster.pdf

[8] http://docs.oasis-open.org/coel/COEL/v1.0/cs02/COEL-v1.0-cs02.pdf

[9] https://www.iso.org/standard/45169.html

[10] https://en.wikipedia.org/wiki/OpenEHR

[11] http://schema.org/

[12] https://media.sitra.fi/2018/11/14144842/261018-ihan-blueprint-2.0.pdf

[13] https://ethersphere.github.io/swarm-home/

[14] http://mydata2016.org/2016/07/04/mydata-operators/

[15] https://www.kanta.fi/documents/20143/141748/Kanta-services+brochure_en.pdf/3b31a750-aa16-fa0d-0353-5bf65171ec6f

[16] https://github.datafund.io/

[17] https://www.ietf.org/rfc/rfc2119.txt