Research Article

International Journal of Blockchain Applications and Financial Technology

The Information and Communication Technologies Expose Our Personal Identity and violate our Personal Identity

Yair Oppenheim

Tel Aviv University, The Lester and Sally Antin Faculty of Humanities, School of Philosophy, Linguistics and Science Studies, Israel

[*]Corresponding author: Yair Oppenheim, Tel Aviv University, The Lester and Sally Antin Faculty of Humanities, School of Philosophy, Linguistics and Science Studies, Israel.

Submitted: 03 November 2025 **Accepted:** 09 November 2025 **Published:** 15 November 2025

https://doi.org/10.63620/MKIJBAFT.2025.1009

Citation: *Oppenheim, Y. (2025). The Information and Communication Technologies Expose Our Personal Identity and violate our Personal Identity. I Jou of Bloc App nd Fina Tech, 1(2), 01-10.*

Abstract

The proliferation of Information and Communication Technologies (ICTs) exposes and recombines multiple facets of personal identity—biometric/passport, numeric cognitive, affiliative, functional social, and communicative—through large-scale data collection, IoT sensor monitoring, online profiling, and cross-database linkage. This paper offers a conceptual synthesis of how these processes erode anonymity and the right to limited access to the self (following Gavison), and formally reviews prevailing privacy frameworks—k anonymity, ℓ diversity, t closeness, and differential privacy. Using linkage attack reasoning and illustrative tabular examples, we show that record level anonymization provides only partial protection: multiplicity of identities does not preserve anonymity when AI enabled link ability can fuse disparate profiles; k anonymity is brittle under adversarial background knowledge; ℓ diversity and t closeness mitigate specific risks but often trade away utility and still fail in sparse distributions; and differential privacy bounds individual influence on statistical outputs yet does not directly prevent harms arising from indirect inference or real time transactional profiling. We argue for a system-level privacy by design paradigm that prioritizes data minimization, segmentation, and enforced unlikability; selectively applies differential privacy with policy-tuned ε; adopts multi-layer identity management with default pseudonymization; and institutionalizes transparency, meaningful consent, and continuous privacy risk monitoring. Reframing identity/anonymity protection from record redaction to ecosystem defenses against linkage and composition better aligns privacy safeguards with contemporary socio-technical infrastructures.

Keywords: Communication Technologies, Social function, Humanity, Logical Obstacles.

Introduction

The philosophical question of human identity has concerned humanity since the days of Plato[1]. It has been phrased in many ways[2], such as “Who am I?”, or the Socratic “know thyself”, or Paul Gauguin’s “Where do we come from? What are we? Where are we going?”. Identity is a vague and complex concept. In this article, I make a distinction between “selfhood” and “identity”, which I define as the outer representation of one’s inner self. According to the liberalist approach, selfhood is a product of one’s unique individuality[3], for even though every individual is made up of the same universal building blocks[4], such as common human physiology, genetics[5], and cognitive abilities, the particular combination of those, together with every person’s unique history, is what creates each individual’s unique selfhood. Therefore, a person’s identity is also unique, as it represents that person’s unique selfhood. This does not mean to say that identity is static and unchanging, only that it can always be associated with one unique individual. An individual can have multiple identities, each of which either represents a different part (or parts) of the same “core self”, or an altogether different self or selves. The different identities are time- and context-dependent, and not necessarily distinct from each other. Some of them may coexist simultaneously, while others replace previous identities.

The new category of identities that was born in the age of ICTs can be collectively called “online identities”. Those are identities one creates or assumes in the virtual world. Every person who is connected to the Internet has at least two identities at the same time: their offline identity and their online identity. Online identities can be divided into two sub-categories: identities one creates for oneself on social media and in various Internet applications, and user identities (profiles) created by large Internet companies to use as a business aid. Online identities created by Internet companies do match the choices, experiences, and actions of specific real-life individuals, and change as the Internet company collects more and more information about the individual in question. Many times, e.g., on Tinder, Amazon, Facebook and other similar applications, they are added to the online identity the individual themself knowingly created.

Despite the fact that an individual can have more than one online identity at the same time, and that these may be very different from one another, they can still be associated with and traced back to one physical person with a continuous memory and experience. In the age of ICTs, this is done by means of constant monitoring of people through IoT technologies. Neither of those systems alone is adequate for making the associations required to trace all the different identities back to a specific person, but using several systems to cross-reference data can reveal the one individual behind the multiple online identities. Information collection and surveillance are practiced all over the world by government organizations as well as by commercial companies using commercial applications.

The collected data undergo “profiling” – creation of a data pool about an individual, an organization, or a physical or Internet entity. In this process, data from different sources are analyzed using various algorithms, including deep learning, to make a profile of that individual or entity. The profile is then used to predict various behavior patterns, such as commercial preferences, social habits, and even political opinions, with a high level of accuracy. Personal information that relates to every aspect of our lives is being collected, stored, analyzed, and shared as an inherent part of this surveillance process. After the information is processed, it becomes a commodity to be traded between various consumers of information, the government, and the private sector alike.

The profiling that exposes a person's identity makes anonymity in the ICTs age almost impossible. The inability to protect personal identity and anonymity in the sense of limited access to the self, as defined by Ruth Gavison. The right to limit access to oneself is constantly threatened in the age of ICTs: We can be identified in public spaces, and the information collected about us makes it possible to get to (or at) us physically, by phone, text or email. All the three elements of limited access listed by Gavison – secrecy, anonymity and solitude – are becoming increasingly harder to achieve despite various technological efforts to manage the challenges. In this appendix, I will review the inherent technological flaws of the most common privacy protection methods:

Identity

Identity is one of the components of personal privacy . In this article, I use Raffaele Rodogno’s definitions of the various “identity questions” to define which information is carried by the identity component of privacy.

(1) Passport Identity : Carries basic bureaucratic information such as identification number, passport number, and/or social security number. Until the age of ICTs, this information had been stored in analog format or in designated digital databases such as population registers and tax authority registers; today, it is replicated and disseminated among multiple interconnected databases. One can use passport identity information as a unique identification key to connect different databases. We also have new biometric means of unique identification, such as face recognition, DNA, and identification by IP (Internet protocol) address of a computer or mobile phone.

Here’s a clean, content-only spec for a Passport Identity Content. The list's attributes of Passport Identity are:

(a). Core Biographic (Required): given name, middle name, family name, full name in native script (optional -e.g., Hebrew/Arabic/Cyrillic), date of birth, place of birth city, place of birth country code (e.g., “ISR”), sex

(b). Passport Document Data (Required for a Passport Profile): document type (e.g., "P" (passport), passport number, given name, family name, birth date, issuing state, issuing authority (optional -e.g., Ministry/Agency name or code), issuance date, expiry date, mrz line1, mrz line2(optional -machine-readable zone, as printed).

(c). Biometric References (Optional, Content Fields only): facial image, fingerprint templates (optional - if present in the document), iris templates (optional), biometric notes (e.g., standards/capture context (text)).

(d). National/Bureaucratic Identifiers (Jurisdiction-Specific, Optional): national ID number (e.g. population register number), social security number (where applicable), tax ID number (where applicable), personal number (field found in some passports).

(e). Derived/Convenience Attributes (Optional): age years, age over 18, age over 21, age over 65, document has expired, days to expiry, nationality codes (for dual nationality).

(f). Contact & Residence (Not from the Passport Itself; Optional if Needed): residence address, email (optional) , phone (optional) .

**(g). Cross-system Linkage Identifiers (Optional, Non-Pass-

[1] *Luciano Floridi, "The Informational Nature of Personal Identity", Minds and Machines Vol. 21 (4) (November 2011), 549-566.*

[2] *Ilya Levin, "Cyber-physical Systems as a Cultural Phenomenon", International Journal of Design Sciences and Technology Vol. 22 (2016), 67-80.*

[3] *Lior Rabi, Omes ha-individualiyut – ha-shorashim shel ideal ha-individualiyut ha-moderni [The Burden of Individuality: The Origins of the Modern Ideal of Individuality] (in Hebrew) (Haifa: Pardes, 2009), 255.*

[4] *Jorge J. E. Garcia, Individuality: An Essay on the Foundation of Metaphysics (New York: State University of New York Press, 1988), 234.*

[5] *Similar to the way every human’s DNA is made up of the same four basic units, but each person’s complete genome is unique.*

port): issuer internal ID (code used by the issuing organization), registry record ID (population/tax registry record ID), external system IDs,

(h). Administrative Metadata (Helpful for Records Management): record ID (internal unique key), schema version, created at, updated at, source ("passport-scan" | "population-register" | "self-report" |...), notes (free-text remarks if needed).

(2) Numerical Identity: Identity based on one's cognitive abilities. Information about our cognitive abilities is recorded and stored in IQ tests and report cards from preschool to university. Before the age of ICTs, it was stored in analog formats or in designated digital databases of schools and military systems; nowadays, it can be accessed in all those databases using passport identity information.

(3) Attribution Identity: Before the age of ICTs, information related to attribution identity was mostly stored in databases controlled by official entities, such as population registers, tax authority registers, banks and social welfare institutions. Those databases were isolated and mostly static islands of information updated only in case of major life events like marriage, divorce, childbirth and death. In the age of ICTs, the monitoring of our actions, location, shopping habits and other patterns of behavior allows for aggregation of all these data as information in interconnected big data pools.

(4) Social Function Identity: Until the age of ICTs, information about this type of identity had hardly been stored or aggregated. Small parts of it were stored in population registers or appeared in passport identity information (e.g., gender and ethnicity). Today, ICTs allow people to assume different identities in different applications, and sometimes more than one identity in the same application, and to present different faces depending on the place, time, and other circumstances. On the other hand, the constant monitoring of people's activities by means of ICTs allows for aggregation of data in interconnected big data pools, and AI technologies are used to build profiles based on that data in many Internet-based applications. The advanced AI technologies combine the different profiles into a single nearly complete identity, which can be associated with a specific individual, violating their personal privacy. There is an "arms race" of sorts between individuals' ability to assume multiple distinct identities and the technological capability to unify them, voiding the individual's efforts. Currently, the technological capability to make a link between the same person's different identities seems to prevail over the human attempt to maintain multiple identities.

(5) Attachment Identity: This type of identity resides almost entirely in the sphere of deep personal privacy; any information about it can exist outside our body and mind only if we choose to share it – usually only with confidants.

Numerical Profiling Example

Here I present the ability to build a profile for Passport identity profile, I propose a general profile model. In the profiling process, the entity that builds the specific Identity Profile collects all relevant information regarding a specific person and stores it in databases. I don't argue that the actual databases are organized in the order presented above. But I argue that if the identity information was collected, we can order it according to the methodology I presented here. Under this assumption, I'll present a numeric representation of the Identity Profile. This numerical will enable us to replace the Identity issues in the age of ICTs age to quantitative issues. The attributes given name, family name, and date of birth are common for all sub-identity types and are always considered as key attributes with the unique document number. ca_i – Is a unique catalog number for attribute ca_i[18] for i ={1,2,3}

Passport Identity Profile Numerical Representation.

There are 8 subtypes of the Passport identity profile. How ever the only the first two subtypes are required; the others are optional. Therefore, I'll present below in detail only these two subtypes.

Let 1 ≤ j ≤ 2 be the index for the sub–Passport Identity Content (e.g. j=1 means Core biographic).

Core Biographic Numeric Profile

Let be a_i^1 $1 \le i \le 8$ (e.g. a_1^1 is the attribute given name)

ac_i^1 – Is a unique catalog number for attribute a_i^1

$$a_i^1 = \begin{cases} 0 & \text{there is no collecting data for attribute } i \\ 1 & \text{there is collecting data for attribute } i \end{cases}$$

Let be ar_i^1 $1 \le i \le 8$ (e.g. ar_1^1 means that the attribute given name is required)

$$ar_i^1 = \begin{cases} 0 & \text{that attribute } i \text{ is optional} \\ 1 & \text{that attribute } i \text{ is reuired} \end{cases}$$

$$a_i^1 * ar_i^1 = \begin{cases} 0 & \text{that attribute } i \text{ is or optional or has no data collecten} \\ 1 & \text{that attribute } i \text{ is reuired and has data collecting} \end{cases}$$

For Core Biographic attributes, the only attribute, full name in native script is optional therefore, $0 \le MP'_1 = \sum_i^8 a_i^1 \le 8$. We can say that if MCB' (the measure for Core Biographic profile) is equal 8 we have the full Core Biographic profile.

$0 \le MP_1 = \sum_i^8 a_i^1 * ar_i^1 \le 7$, If $MP_1 = 7$, we have the full required profile for Core Biographic.

The given name, family name, and birth date are the core attributes of Core Biographic. These attributes enable with high-probability match within the population (especially with an uncommon name plus a full date of birth), and these attributes serve as a "key" for linking records across databases, thereby enriching the attacker with additional information. Therefore, MP_1

[6] *Floridi, "The Informational Nature of Personal Identity".*

[7] *Marya Schechtman, "Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View", in: Narrative and Understanding Persons, ed. Daniel Hutto (Cambridge: Cambridge University press, 2007), 162-167.*

[8] *As exposed by Edward Snowden in June 2013. See: Michael Calderone, "Washington Post Began PRISM Story Three Weeks Ago, Heard Guardian's 'Footsteps'", Huffpost, 7 June 2013. https://www.huffpost.com/entry/washington-post-prism-guardian_n_3402883 (Accessed 6/7/2013)*

[9] *Mireille Hildebrandt, "Defining Profiling: A New Type of Knowledge?", in: Profiling the European Citizen, eds. Mireille Hildebrandt and Serge Gutwirth (Dordrecht: Springer, 2008), 17-45.*

[10] *An Internet entity is a virtual entity, such as a Facebook user, while a physical entity is a real-life person or machine.*

= 3 is the core value of this measure.

Passport Document Data (Required for a Passport Profile)

Let be a_i^2 $1 \le i \le 11$ (e.g. a_1^2 is the attribute document type)

ac_i^2 – Is a unique catalog number for attribute a_i^2

$$a_i^2 = \begin{cases} 0 & \textit{there is no collecting data for attribute i} \\ 1 & \textit{there is collecting data for attribute i} \end{cases}$$

Let be ar_i^2 $1 \le i \le 8$ (e.g. ar_1^2 means that the attribute document type)

$$ar_i^2 = \begin{cases} 0 & \textit{that attribute i is optional} \\ 1 & \textit{that attribute i is reuired} \end{cases}$$

$$a_i^2 \cdot ar_i^2 = \begin{cases} 0 & \textit{that attribute i is or optional or has no data collecten} \\ 1 & \textit{that attribute i is reuired and has data collecting} \end{cases}$$

For Passport Document Data attributes, the issuing authority, the mrz line1, mrz line2 attributes, are optional therefore, $0 \le MP_2' = \sum_i^8 a_i^2 \le 11$.We can say that if MPD' (the measure for Passport document data profile) is equal 8 we have the full Passport document data profile.

$0 \le MP_2 = \sum_1^8 a_i^2 \cdot ar_i^2 \le 8$. If $MP_2 = 8$, we have the full required profile Passport document data.

The passport number, given name, family name, and birth date are the core attributes of Passport document data.

These attributes enable with high-probability match within the population (especially with an uncommon name plus a full date of birth), and these attributes serve as a "key" for linking records across databases, thereby enriching the attacker with additional information. Therefore MP_2 = 4 is the core value of this measure.

As a conclusion of the discussion above, we can say that MP = Minimum $\{MP_1, MP_2\} = 3$.. That means that to achieve k-anonymity , privacy security should be ≥ 3.

Anonymity

Anonymity is one of the components of personal privacy .As Michael Birnhack has stated, anonymity is a concept that can be hard to define. The literal meaning of the word "anonymous" is "nameless", and it usually means doing something without identifying oneself

Helen Nissenbaum describes anonymity as the inability to get to (or at) a person, i.e. to associate elements of information or things done in the physical world with a specific individual. She thus extends the concept of anonymity, which before the age of ICTs meant conducting oneself without revealing one's name (e.g., a literary work whose author is unidentified, an anonymous donation, or conducting oneself in a foreign city where nobody knows one's name), i.e., preventing access to one's passport identity.

According to Andreas Pfitzmann and Marit Hansen, anonymity of a subject means that the subject is not identifiable within a set of subjects (the anonymity set). This is a further extension of Nissenbaum's definition because it also includes subjects that are not individual humans. This form of anonymity is weaker because it only means that a subject is unidentifiable within a given set, not unidentifiable in general. There is no general protection of the subject's identity, only the inability to associate a specific identity with a specific subject. This definition overlaps with Nissenbaum's in the sense of preventing physical access to the subject. The immediate conclusion from Pfitzmann and Hansen's definition is that anonymity is possible only if the subject is part of a group or a category; the bigger the set of subjects one is part of, the greater one's anonymity.

In the age of ICTs, the claim to anonymity has been extended to:

- Sending electronic messages (e.g., emails) without the sender's (and sometimes also the addressee's) name being revealed;
- Participation in virtual interactions (chat rooms, online gaming, online dating services, e-commerce) without one's real name being known to other participants;
- Purchasing goods and services online using virtual currency (e.g., bitcoin) and without giving one's name;
- Ability to visit any website without having to divulge one's identity.

Anonymity also means the ability to conduct oneself under a false identity (online or offline), which cannot be associated with a particular real-life individual. If before the age of ICTs you could remain anonymous by preventing access to your passport identity, now it is much more difficult: a few seemingly unrelated pieces of information that are aggregated in data pools, shared and analyzed by AI software can suffice to reveal the identity of any individual. Those pieces of information may include your shopping habits, web surfing patterns, or the places you visit. In other words, through various fragments of indirect details, it is quite possible to physically reach a person – to knock on their door and demand money, or sue them for something they have done or said. According to Bruce Schneier, in the age of ICTs it is impossible to remain anonymous against the surveillance and monitoring conducted by governments, commercial companies and other entities because in the physical world, one's actions are recorded by IoT sensors, while in the virtual world, all the patterns of one's online behavior – including web surfing patterns, tweets and interactions – are constantly monitored; thus, ICTs are making anonymity impossible. Anonymity is never a game of one: there is the individual who wishes to remain anonymous, and the other who wishes, for whatever reason, to breach that anonymity. Most often, it is a game of three: the individual, their confidant(s), and the other.

Before the age of ICTs, anonymity was a social norm protected by laws that were part of the Fordist culture. The age of ICTs has extended the concept of anonymity and made it harder to achieve. It has challenged the social norms that supported anonymity, especially anonymity in the public sphere. Anonymity as part of personal privacy is an individual's shield against incapacitating surveillance by the government or by other individuals. It allows free self-expression and independence of thought,

[11] Gavison, Ruth. "Privacy and the Limits of Law". The Yale Law Journal, Vol. 89, No. 3 (January 1980): 421-471.

[12] Oppenheim, Yair. (2024) Personal Privacy in the Age of the Internet. Spines, 37-39

[13] Rodogno, Raffaele. "Personal Identity Online". Philosophy & Technology Vol. 25 (3) (2012): 309-328.

For a more detailed explanation of each "identity question", see Chapter 6. Here, I will focus on the elements of information carried by each type of identity.

[14] The passport identity is the core component of the profiling

feeling, and action. In commerce, it protects one from those who might try to manipulate one or to take advantage of one's weaknesses and whims.

To summarize it the anonymity is: Any piece or pieces of information that can, alone or when combined, breach our anonymity, i.e., our ability to remain unidentified in the public sphere. According to Prof. Ruth Gavison, , limited access to one consists of three independent and irreducible elements: secrecy (information about one), anonymity (attention to one), and solitude (physical access to one).

Technological and Logical Obstacles to Identity and Anonymity Protection

Here, I present the inability to protect personal identity and anonymity in the sense of limited access to the self. The right to limit access to oneself is constantly threatened in the age of ICTs: we may be under surveillance by IoT devices, we are being profiled, we can be identified in public spaces, and the information collected about us makes it possible to get to (or at) us physically, by phone, text, or email. All three elements of limited access listed by Gavison – secrecy, anonymity, and solitude – are becoming increasingly harder to achieve despite various technological efforts to manage the challenges. In this paragraph, I review the inherent technological flaws of the most common privacy protection methods:

K-Anonymity

K-anonymity is a formal model designed to protect the anonymity of users whose data are stored in a data pool(s). The basic concept of k-anonymity is that any given individual whose information is contained in a data release should be indistinguishable from at least k-1 other individuals whose data also appear in the release. For example, k = 2 means that a given individual's data cannot be distinguished from the data of at least one other person. The formal definition of k-anonymity:

- Let T be a table that contains personal data;
- Let RT be a table derived from T;
- Let A_1, A_2 ……., A_n be a list of attributes;
- Let a_1, a_2 ………, a_l be quasi-identifiers (QI) included inthe list of attributes, so that for every i, $A_l = a_l$ and l < n;

Database T can be said to satisfy k-anonymity if and only if each sequence of values in Table RT appears with at least k occurrences in RT.

Naturally, some attributes are unique identifiers, such as ID number, phone number, and lately – biometric data. All these are components of our passport identity. Other attributes are quasi-identifiers, i.e., they are associated with a specific person, but are not unique to that person, such as address, date of birth, weight, education, etc. There are several ways to modify table T, which explicitly contains all the attributes of a specific person, so that it would satisfy k-anonymity, including removing unique identifiers, shortening quasi-identifiers or replacing them with ** characters, etc.

A practical Rule of Thumb for K-Anonymity

There is a rule of thumb that flags when k is too large relative to the number/nature of your QI (quasi-identifiers), meaning it likely loses important utility. The more QIs you include—or the richer they are, the more generalization you need to reach k-anonymity, which hurts usefulness.

Rule of Thumb

Let:

- d= number of QI attributes you actually include.
- For each attribute j:
- C_j= number of distinct (raw) values;
- t_j= minimum granularity you want to preserve after anonymization (e.g., at least 16 age buckets, at least 20 ZIP groups).
- If

$k > \min_j (\frac{C_j}{t_j})^d,$

then, to satisfy k, you'll be forced to coarsen at least one QI below its target granularity t_j→ high chance of **losing important information.**

Why this works: To reach equivalence classes of size k, the effective generalization per attribute is roughly $k^{1/d}$. If $k^{1/d} > C_j/t_j$ for any j, that attribute must drop below the granularity you wanted.

Quick Example

QI={BirthYear,ZIP3}.

C_{year} =80(years),wantt_"year" =16(5-yearbuckets).

C_{ZIP3} =100,wantt_"ZIP3" =20.

Threshold=$\min \{(80/16)^2, (100/20)^2\} = \min \{5^2, 5^2\} = 25.$

So k>25 will likely cause over-generalization beyond your targets.

Tip: very low-cardinality QIs (e.g., Gender = 2) often shouldn't drive this test; either exclude them here or set $t_j = C_j$ so they don't distort the threshold.

Picking kin Practice

1. Risk Target: choose a max re-identification risk R(e.g., 5%) and set k≥1/R (so k≥20).
2. Utility Check: try several k values (e.g., 5,10,15,20,25,30) and measure information-loss metrics (e.g., NCP/Precision, Discernibility Metric) or downstream model performance (accuracy/AUC drop). Choose the smallest k that meets the risk bound and your utility thresholds (e.g., NCP ≤ 0.2 or ≤10% accuracy drop).

Let us consider the following example:

[15]It could be many mail addresses for one person
[16]It could be many phone numbers for one person

Table 1: K-Anonymity Example Example of k-anonymity, where k=2 and QI={Race, Birth, Gender, ZIP}.

ID	Race	Birth	Gender	ZIP	Problem
t1	Black	1965	m	0214*	short breath
t2	Black	1965	m	0214*	chest pain
t3	Black	1965	m	0213*	hypertension
t4	Black	1965	f	0213*	hypertension
t5	Black	1964	f	0213*	obesity
t6	Black	1964	f	0213*	chest pain
t7	White	1964	m	0213*	chest pain
t8	White	1964	m	0213*	obesity
t9	White	1964	m	0213*	short breath
t10	White	1967	m	0213*	chest pain
t11	White	1967	m	0213*	chest pain

The most important point is that the method of k-anonymity only prevents the distinction between individual people (identification) based on their quasi-identifiers. The individuals' sensitive attribute remains unscrambled by definition to keep the data meaningful, and that is exactly one of this method's weaknesses. The table in the above example satisfies k = 2 regarding the The most important point is that the method of k-anonymity only prevents the distinction between individual people (identification) based on their quasi-identifiers. The individuals' sensitive attribute remains unscrambled by definition to keep the data meaningful, and that is exactly one of this method's weaknesses. The table in the above example satisfies k = 2 regarding the quasi-identifiers, but fails to satisfy k-anonymity if the Problem column is considered as well: for example, t1 has no equivalent. In practice, maintaining privacy by making sure data releases satisfy k-anonymity is failing for the following reasons: The underlying assumption is that databases are separate and cannot be combined. However, in practice, there is almost always a way to fuse databases and thus nullify the anonymity. Let us take, for example, the following private table PT and linked table LT:

Table 2: PT

Race	BirthDate	Gender	ZIP	Problem
black	9/20/1965	male	02141	short of breath
black	2/14/1965	male	02141	chest pain
black	10/23/1964	male	02138	painful eye
black	8/24/1965	female	02138	wheezing
black	11/7/1964	female	02138	obesity
black	12/1/1964	male	02138	chest pain
white	10/23/1964	male	02138	short of breath
white	3/15/1965	female	02139	hypertension
white	8/13/1964	male	02139	obesity
white	5/5/1964	male	02139	fever
white	2/13/1967	male	02138	vomiting
white	3/21/1967	male	02138	back pain

Table 3: LT

Race	BirthDate	Gender	ZIP	Problem
black	1965	male	02141	short of breath
black	1965	male	02141	chest pain
black	1965	male	02138	painful eye
black	1965	female	02138	wheezing
black	1964	female	02138	obesity
black	1964	male	02138	chest pain
white	1964	male	02138	short of breath
white	1965	female	02139	hypertension

[17] *Marya Schechtman Marya, "Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View". In: Narrative and Understanding Persons, ed. Daniel Hutto, 155-178. Cambridge: Cambridge University Press, 2007.*

[18] *We can conduct that [illegible] = {1,2,3}*

white	1964	male	02139	obesity
white	1964	male	02139	fever
white	1967	male	02138	vomiting
white	1967	male	02138	back pain

If we combine the two tables, everybody's full dates of birth will be revealed, breaching the k = 2 anonymity of LT.

- Another way to breach the anonymity of a data release is by having prior knowledge of the problem or the attributes of the person in question. For example, if we look at the following medical release:

Table 4

DSM diagnosis	Gender	ZIP	Date of birth
HIV	*	Center	199*
HIV	*	Center	199*
Bulimia nervosa	*	Center	199*
Bulimia nervosa	*	Center	199*

Assuming that the ratio of males to females is 50:50, and knowing that the prevalence of HIV in young men (under 35) living in the Central region is 20%, while the prevalence of bulimia nervosa is 2%, if we find out that Dan – a man in his 20s living in the Central region – has a medical condition, we can infer with a high level of certainty that he has HIV.

Conclusion

Because the vast majority of data releases that contain personal information are included in at least one Internet database, and the people who analyze them usually have some background knowledge that helps them overcome the quasi-identifier scrambling, there is almost always a way to breach k-anonymity and violate one's privacy by combining different databases and using prior knowledge.

L-Diversity

In the previous section, I reviewed the inherent flaws of k-anonymity and showed how it fails to guarantee the anonymity (privacy) of data in large data releases due to insufficient protection of the sensitive data. This understanding gave birth to the principle of l-diversity – an extension of k-anonymity that measures the diversity of sensitive values for each column in which they occur. According to the formal definition, a data release has l-diversity if for every set of rows with identical quasi-identifiers ("equivalence class" or "block"), there are at least l well-represented values for each sensitive attribute. For example, the below table has k-anonymity of k = 4, but no l-diversity in the third block (lines 9-12), where the medical condition is the same for all four entries.

However, l-diversity cannot help protect data anonymity if there is insufficient diversity in the sensitive attribute to begin with, i.e., if there is insufficient distribution of the sensitive attribute in a given population, or if l-diversifying the dataset compromises the integrity of the data. Let us consider, for example, the following table:

Table 5

	Non-Sensitive			Sensitive
ID	Zip Code	Age	Nationality	Condition
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

[19] *On paragraph 4, elaborate it*

Table 5: HIV (sensitive data)

Date of Birth	Zip	HIV (sensitive data)
9*	Center	Negative
9*	Center	Negative
9*	Center	Negative
9*	Center	Negative
8*	North	Negative
8*	North	Negative
8*	North	Negative
8*	North	Positive

In this example, l-diversity cannot provide adequate protection of identity and anonymity because there is only one person from the North region born in the 1980s who is HIV-positive, which means that person can be easily identified. To meet the requirement of l = 2, we would have to change the "HIV" value in one of the lines 5-7 from "negative" to "positive", but that would misrepresent all the people in that class as 50% HIV-positive (even though the actual prevalence in the general population is 1%), presenting a serious risk to their identity security.

In this example, l-diversity cannot provide adequate protection of identity and anonymity because there is only one person from the North region born in the 1980s who is HIV-positive, which means that person can be easily identified. To meet the requirement of l = 2, we would have to change the "HIV" value in one of the lines 5-7 from "negative" to "positive", but that would misrepresent all the people in that class as 50% HIV-positive (even though the actual prevalence in the general population is 1%), presenting a serious risk to their identity security.

T-Closeness

The principle of t-closeness was designed to address the limitations of l-diversity. The basic requirement is that the distribution of a sensitive attribute in any equivalence class be close to the distribution of the attribute in the overall table, before it was k-anonymized. According to the formal definition, an equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t. A table is said to have t-closeness if all equivalence classes have t-closeness.

All three abovementioned principles are based on the redaction of unique identifiers and scrambling of quasi-identifiers, and can help protect the identity of personal information provided that an individual cannot be identified based on their quasi-identifiers alone.

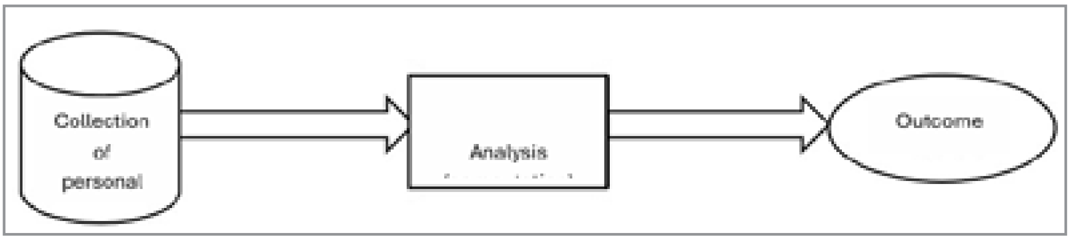
One of the known limitations is that making semantic changes to a database might lead to information loss and therefore compromise the usefulness of the data release: for example, if the anonymization process changes the correlation between the distribution of symptoms and the distribution of medical conditions, the data release cannot be used for medical inference.

Another limitation is the false assumption of a lack of background knowledge. Most people have extensive background knowledge which can be used to narrow down the possibilities and, when combined with AI algorithms, overcome these methods of personal privacy protection.

Differential Privacy

Differential privacy (DP) is a mathematical framework designed to address the inherent limitations of the information identity and anonymity protection methods . It can provide a strong security of identity and anonymity by allowing data to be analyzed without revealing sensitive information about any individual in the dataset.

According to Kobbi Nissim, the intuitive idea that removing identifiers guarantees protection of information privacy is often wrong. Science and practice have proven that the routinely applied statistical tools and methods of informational privacy protection (illustrated by the figure below) provide less than sufficient privacy protection.



Nissim gives the following example: NIH collects DNA from people with a certain medical condition and publishes minor allele frequencies for the control group, which includes 100,000 people.

[20] Oppenheim, Y. (2024) Personal Privacy in the Age of the Internet. Spines, 37-39

[21] Helen Nissenbaum, "The Meaning of Anonymity in an Information Age", The Information Society, Vol. 15 (2) (May 1999), 141-144.

[22] Rodogno, "Personal Identity Online", 310.

[23] Marit Hansen and Andreas Pfitzmann, "A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management", 9. https://www.researchgate.net (Accessed 24/1/2021)

5%	3%	12%	3%

Given that the frequencies in the overall population are:

3%	3%	11%	3%

If an individual's genome is:

1	0	0	1

We can infer with a high level of certainty that the person is not in the control group.
According to Nissim, it is a mathematical fact, regardless of privacy policy, that any useful analysis of personal data must leak some information about individuals, and that the leakages accumulate with multiple analyses/releases. That is true, he argues, for any statistical privacy protection method.

The purpose of differential privacy is to address this issue. Differential privacy means that sensitive personal identity information is slightly modified by the addition of carefully crafted random "noise" so that the inclusion of the said personal information in the analysis does not present a privacy threat. Analysis is said to satisfy differential privacy if the inclusion or non-inclusion of one's personal identity information does not affect one's privacy, e.g., one's chances of getting arrested, denied insurance, shamed socially, etc.

Thus, differential privacy can guarantee adequate protection of information identity if the computation (analysis) process meets the requirements. However, it does not provide adequate protection in the following cases:

- Differential privacy usually cannot protect against the invasion that make use of indirect information, which means it does not matter whether one's data is included in the dataset or not.
- Differential privacy is not applicable when one has to provide specific information that needs to be acted upon – for example, when one makes an online transaction or surfs the web, so it cannot prevent profiling and use of identity information for profit by the Internet companies.
- Despite these limitations, differential privacy has many applications; among other things, it can be used in data mining to balance the protection of information identity against the accuracy of decision tree induction in data mining.

In this paragraph, I have given an overview of today's most common methods of ensuring identity and anonymity of information in big data pools, their strengths, and limitations. Even though each of those methods provides some protection of information of identity and anonymity, they fail to provide an adequate solution to the identity and anonymity challenges posed by ICTs. Therefore, I believe , that we must look for other solutions in the form of a new paradigm of personal identity and anonymity which would be more compatible with the spirit of the age of ICTs.

Conclusions

In this article, I examined how Information and Communication Technologies (ICTs) expose personal identity and erode anonymity and the right to limited access to the self. It focuses on large-scale data collection, IoT sensor monitoring, online profiling, and cross database linkage. The paper distinguishes among several forms of identity (biometric/passport, numeric cognitive, affiliative, functional social, and communicative) and shows how AI and connected data repositories can fuse them into an almost complete profile of a single individual. It also reviews four common privacy frameworks—k anonymity, ℓ diversity, t closeness, and differential privacy—and highlights their limitations under high connectivity and extensive background knowledge.

Key Conclusions

- Multiplicity of identities does not guarantee protection: database link ability, biometrics, and network identifiers enable fusion of disparate identities and erode anonymity.
- k anonymity alone is insufficient: background knowledge and linkage attacks can reveal sensitive attributes even when direct identifiers are removed.
- ℓ diversity and t closeness mitigate certain risks but can fail in edge cases with sparse sensitive distributions and may degrade data utility.
- Differential privacy bounds an individual's influence mathematically, yet does not directly prevent harms driven by indirect inference or real-time online interactions (e.g., commercial profiling).
- A system-level paradigm is required—shifting from identifier suppression to defenses against linkage, composition, and adversarial background knowledge.

Implications

Do not rely solely on masking direct and quasi-identifiers; address the attack surface created by cross dataset link ability.
Gavison's triad (secrecy–anonymity–solitude) is strained by continuous physical and online monitoring.
Biometric infrastructures, MRZ, face recognition, and IP address signals operate as powerful linkage keys across data islands.

Practical Recommendations

- Privacy by default and privacy by design: strict data minimization, segmentation, and enforcement of unlikability; prefer federated analytics where feasible.
- Apply differential privacy to statistical releases and dashboards, tuning ε to explicit risk utility targets.
- Adopt multi-layer identity management: pseudonymization by default; avoid propagating passport/MRZ identifiers into workflows that do not require hard identification.

References

1. Machanavajjhala, A., Gehrke, J., & Kifer, D. (2006). ℓ-Diversity: Privacy beyond k-anonymity. In Proceedings of

[24] *Bruce Schneier, Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World (New York: W. W. Norton and Company, 2016), 4-42.*
[25] *Ruth Gavison, "Privacy and the Limits of Law", 433.*

the 22nd International Conference on Data Engineering (ICDE'06).

2. Calderone, M. (2013). Washington Post began PRISM story three weeks ago, heard Guardian's "footsteps". HuffPost. https://www.huffpost.com/entry/washington-post-prism-guardian_n_3402883
3. Floridi, L. (2011). The informational nature of personal identity. Minds and Machines. 21: 549-566.
4. Garcia, J. E. J. (1988). Individuality: An essay on the foundation of metaphysics. State University of New York Press.
5. Gavison, R. (1980). Privacy and the limits of law. The Yale Law Journal. 89: 421-471.
6. Hansen, M., & Pfitzmann, A. (2008). A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management (Version 0.9). Retrieved January 24, 2021, from.
7. Hildebrandt, M. (2008). Defining profiling: A new type of knowledge? In M. Hildebrandt & S. Gutwirth (Eds.), Profiling the European citizen (pp. 17-45). Dordrecht: Springer.
8. Levin, I. (2016). Cyber-physical systems as a cultural phenomenon. International Journal of Design Sciences and Technology. 22: 67-80.
9. Hildebrandt, M. (2019). Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. Theoretical Inquiries in Law. 20: 83-121.
10. Nissim, K. (2020). Differential privacy: Why, how, and where to? Presented at Privacy in Challenging Times: The 8th Technion Summer School on Cyber and Computer Security.
11. Nissenbaum, H. (1999). The meaning of anonymity in an information age. The Information Society. 15: 141-144.
12. Oppenheim, Y. (2024). Personal privacy in the age of the Internet (pp. 37-39, 21). Spines.
13. Rabi, L. (2009). Omes ha-individualiyut – ha-shorashim shel ideal ha-individualiyut ha-moderni [The burden of individuality: The origins of the modern ideal of individuality]. Haifa: Pardes.
14. Rodogno, R. (2012). Personal identity online. Philosophy & Technology. 25: 309-328.
15. Schneier, B. (2016). Data and Goliath: The hidden battles to collect your data and control your world (pp. 4-42). New York: W. W. Norton & Company.
16. Schechtman, M. (2007). Stories, lives, and basic survival: A refinement and defense of the narrative view. In D. Hutto (Ed.), Narrative and understanding persons (pp. 162-167). Cambridge: Cambridge University Press.
17. Sweeney, L. (2002). k-Anonymity: A model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems. 10: 557-570.

[26] Latanya Sweeney, "k-Anonymity: A Model for Protecting Privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5) (October 2002), 557-570.

[27] bid., 6-7.

[28] Sweeney, "k-Anonymity: A Model for Protecting Privacy", 9.

[29] Ibid., Sweeney, "k-Anonymity: A Model for Protecting Privacy", 11.

[30] Ashwin Machanavajjhala, Johannes Gehrke and Daniel Kifer, "ℓ-Diversity: Privacy Beyond k-Anonymity", 22nd International Conference on Data Engineering (ICDE'06), 3-7 April 2006, 2.

[31] Hildebrandt, Mireille. "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning". Theoretical Inquiries in Law, vol.20 (1) (2019): 83-121.

[32] Ibid., 4.

[33] Kobbi Nissim, "Differential privacy: Why, How, and Where to?", delivered at: Privacy in Challenging Times: The 8th Technion Summer School on Cyber and Computer Security (September 2020).

[34] Nissim, "Differential privacy: Why, How, and Where to?", 8-15.

[35] bid., 16.

[36] Nissim, "Differential privacy: Why, How, and Where to?", 63-78.

[37] Arik Friedman and Assaf Schuster, "Data Mining with Differential Privacy", delivered at: Privacy in Challenging Times: The 8th Technion Summer School on Cyber and Computer Security (September 2020).

[38] Oppenheim, Y. (2024) Personal Privacy in the Age of the Internet. Spines, 223-230