MOVEMENT
for an open web

Commentary on the draft W3C Privacy Principles

Introduction

Privacy is increasingly in the public eye as dominant platforms often seek to disguise discriminatory conduct that harms competition using “privacy fixing” language.

MOW supports the purpose listed in the Abstract of the W3C Privacy Principles (“Principles”), namely “people using the Web would benefit from a stronger relationship between technology and policy" that ought to “guide the development of the Web as a trustworthy platform.”

Disappointingly, many of the definitions and related concepts in the rest of the document do not help advance these noble goals and erode consumer sovereignty. The document repeats unsubstantiated positions from large publishers and internet gatekeepers that claim to protect people’s important privacy rights, but do not.

The following commentary identifies areas of alignment between MOW and these authors, while also highlighting areas where the current language raises serious risks to competition across the Open Web.

Contents

1. Using the privacy argument as an excuse to discriminate

MOW agrees that privacy should not be used as a weapon by large entities to favour their own businesses through discriminatory practices by alleging that end users must be protected from the influence of so called third-parties.[1]

MOW further agrees that larger entities tend to have more power relative to both people and markets, and as such, must be better protected from abuses they may perpetrate given this situation.

> *"One of the ways in which the Web serves people is by protecting them in the face of asymmetries of power, and this includes establishing and enforcing rules to govern the power of data…."*

Unfortunately, the document goes on to state the following.

> *"Information is power. It can be used to **predict and to influence people**, as well as to design online spaces that **control people's behaviour**. The collection and processing of information in greater volume, with greater precision and reliability, with increasing interoperability across a growing variety of data types, and at intensifying speed is leading to an unprecedented concentration of power that threatens private and public liberties."*[2]

It is well known that Google and Apple collect and process vast troves of personal data from individuals' interactions with rival online businesses via their OS, app stores and browsers.[3]

However, the document makes an unfounded claim that information used to "predict" behaviour, or provide desirable information to "influence" behaviour, somehow magically transforms into a force that enables software to remove free will and all decision-making ability by "*control[ing] people's behavior*."

A position MOW supports is summarized by Cory Doctorow, a leading online digital expert:

> *To understand why you shouldn't worry about mind-control rays — but why you* should *worry about surveillance* and *Big Tech — we must start by unpacking what we mean by "persuasion."*
>
> *Google, Facebook, and other surveillance capitalists promise their customers (the advertisers) that if they use machine-learning tools trained on unimaginably large data sets of nonconsensually harvested personal information, they will be able to uncover ways to bypass the rational faculties of the public and direct their behavior, creating a stream of purchases, votes, and other desired outcomes.*
>
> *The impact of dominance far exceeds the impact of manipulation and should be central to our analysis and any remedies we seek.*

---

[1] In fact, MOW asserts that the standards being set through the Principles are being used as an objective justification to anticompetitive behaviour. The European Commission's Horizontal Guidelines recognise (at para 305) that a standard should be treated as binding/mandatory (i.e., as a *de facto*) where "*consumer confidence is essential*" and where they "*tend to favour widespread practices*" and where other competing companies "*would need to comply [with such standards] to sell in the market*". These big brands therefore have scope to abuse the reliance and trust that consumers have in these companies. In this context, consumers place reliance and trust that these brands will offer the best solutions as to user privacy. If a standard is formulated by them, all other players will be compelled in practice to follow the same mandatory standards and thereby influence competition by setting privacy standards that suit their business models more than those of their competitors.

[2] https://www.w3.org/TR/privacy-principles/#intro

[3] This is a consistent theme identified in the CMA's Market Study Interim Report on Mobile Ecosystems, in particular, Chapter 5.

But there's little evidence that this is happening. Instead, the predictions that surveillance capitalism delivers to its customers are much less impressive. Rather than finding ways to bypass our rational faculties, surveillance capitalists like Mark Zuckerberg mostly do one or more of three things:

***Segmenting** [while using sensitive data it can be] "seriously creepy. But it's not mind control…."*

***Deception**… This is pernicious and difficult — and it's also the kind of thing the internet can help guard against by making true information available, especially in a form that exposes the underlying deliberations among parties with sharply divergent views, such as Wikipedia. But it's not brainwashing; it's fraud."*

***Domination**. Surveillance capitalism is the result of monopoly. Monopoly is the cause, and surveillance capitalism and its negative outcomes are the effects of monopoly…. that has allowed companies to grow by merging with their rivals, buying up their nascent competitors, and expanding to control whole market verticals. One example of how monopolism aids in persuasion is through dominance: Google makes editorial decisions about its algorithms that determine the sort order of the responses to our queries."*[4]

The authors of these Privacy Principles fortunately understand the dangers to individuals of increased centralization that technical standards can foster, and the benefits provided by competition.

"While privacy principles are designed to work together and support each other, occasionally a proposal to improve how a system follows one privacy principle may reduce how well it follows another principle."[5]

Thus, when designing standards, we must ensure we understand the risk to

"a person using an essential service provided by a monopolistic platform — and those where people and parties are very much on equal footing, or even where the person may have greater power, as is the case with small businesses operating in a competitive environment."[6]

For any policy to have a chance at improving the situation, it is critical to understand both the problem we are addressing and how well we predict the remedy would address it.

The Principles claim to provide a "trustworthy platform" that supports a "Web for all," which would require it to work for individuals and organizations **of all sizes** with which people interact. Accordingly, if a proposed principle "*only benefits powerful, large entities that control both an implementation and services*," this would be "*harmful to the Web*."[7]

Accordingly, we must analyse each principle listed as to whether it benefits online organizations of all sizes or (even unintentionally) favours larger entities and discriminates against smaller ones.

2. Privacy risks should be taken in context: Who is processing data and how?

Privacy risks should not be assumed out of context. They should be assessed in relation to exactly the data being processed and whether the organization processing this data is linking it to people's identity.

MOW agrees with the document's proposition that expectations and risks to "privacy" must be judged by the factors that "*ultimately depend on the nature, scope context and purposes of the processing and how it is implemented*."[8]

---

[4] https://onezero.medium.com/how-to-destroy-surveillance-capitalism-8135e6744d59
[5] https://www.w3.org/TR/privacy-principles/#balancing
[6] https://www.w3.org/TR/privacy-principles/#privacy-labour
[7] https://github.com/w3ctag/design-reviews/blob/main/reviews/first_party_sets_feedback.md
[8] ICO's Data protection and privacy expectations for online advertising proposals, 25 November 2021.

People have different expectations of privacy at work, at a café, or at home for instance. Understanding and evaluating a privacy situation is best done by clearly identifying:

- Its actors, which include the subject of the information as well as the sender and the recipient of the information flow (note that recipients might not always want to be recipients.);
- The specific type of data in the information flow; and
- The principles that are in use in this specific context.

MOW further agrees that in addition to the "specific context" the "specific type of data in the information flow" can increase or reduce privacy risks to individuals. Processing of sensitive and special category data should be prohibited, unless there is an established legal basis and "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place."[9] When data is sensitive or belongs to the enumerated special categories, privacy regulators recommend using heightened security relative to innocuous data.[10]

However, personal information alone is not the only risk to individuals. The recipient must link this information to an individual's identity to cause a privacy harm. There are of course other online harms that do not depend on knowing an individual's identity, such as fraud, sale of illegal goods or an online virus. However, given this document is focused on defining privacy principles, we must ensure we are addressing this problem and that any proposed remedies are tailored to address it.

Data protection regulators advise on how best to address privacy risks to consumers:

> "*To assess the level of risk, you must consider both the likelihood and the severity of any impact on individuals. High risk could result from either a high probability of some harm, or a lower possibility of serious harm*."[11]

Accordingly, when reducing the risk to privacy we must consider the likelihood of whether a recipient will cause specified harm to an individual, such as re-identification. If the legitimate business recipient receives information not linked to identity and has appropriate organizational and technical measures (such as pseudonymization" to provide the "necessary safeguards" for responsible processing of the personal data), GDPR and other privacy regulations clearly state that there must be a balance between the state of the art, the cost of implementation on businesses that are balanced against the varying likelihood and severity of risk to individuals:

> *"Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.*[12]

---

[9] Article 22, Regulation 2016/679 of the European Parliament and of the Council (EU GDPR)

[10] ICO's Guide to the General Data Protection Regulation (GDPR): "*one of the considerations for determining the appropriate level of security is the sensitivity of the personal data. You may need to consider whether you need additional security measures for special category data*."

[11] Jersey Office of the Information Commissioner, Data Protection Impact Assessment; UK ICO's Data protection impact assessments.

[12] Article 25, EU GDPR.

When organizations implement these important and necessary safeguards as well as declare in writing they will not reidentify the personal information they receive, such as on their privacy policy, this reduces the risks to individuals' privacy.

Fortunately, the authors agree with the above data protection principle as they write one method of ensuring an organization's responsible data processing is by:

> *"Specifying in its privacy statement that these types of data are kept separate and implementing policies and procedures to ensure the data is stored separately. (legal and procedural/compliance measures)"*[13]

3. Consent and free choice given by the end user

Consent should not be made difficult for individuals and the ability for consumers to make independent choices should not be ignored or arbitrarily forgotten by large entities when formulating these standards.

The authors of the Principles unfortunately suggest that consent should be "difficult" which directly contradicts data protection regulations that aim to present consent choices "*an intelligible and easily accessible form, using clear and plain language*."[14] The document provides that:

> *"Consent is comparable to the general problem of permissions on the Web platform. In the same way that it should be clear when a given device capability is in use (eg. you are providing geolocation or camera access), sharing data should be set up in such a way that it requires deliberate, specific action from the person (eg. triggering a form control that is not on a modal dialog) and if that consent is persistent, there should be a vivid indicator that data is being transmitted shown at all times, in such a way that the person can easily switch it off. In general, providing consent should be rare, difficult, highly intentional, and temporary."*[15]

Even more troubling is the fact that the authors suggest that consumers expressing a consent choice should be "temporary." If a consumer freely expresses a global opt-in or a global-opt out, why should technology forget this decision? Indeed, by remembering consumer preferences and signalling this to recipients, their interactions with online users can reduce the "privacy labor" that all parties seek to ensure as we build a Web available to all. The Principles reference the Global Privacy Control[16] which specifically contradicts the notion advocated.

Many regions' data protection regimes require explicit consent mechanisms when collecting or processing sensitive category data or reidentifying a specific individual (whether on a single site or across multiple sites). The authors could enhance their document if they also made this privacy choice more explicit.

> *"In specific cases, people should be able to consent to more sensitive purposes, such as having their identity recognised across contexts or their reading history shared with a company. The burden of proof on ensuring that informed consent has been obtained needs to be very high in this case."*[17]

---

[13] https://www.w3.org/TR/privacy-principles/#example-technical-legal-measures
[14] Article 7, EU GDPR.
[15] https://www.w3.org/TR/privacy-principles/#opt-in-out
[16] https://globalprivacycontrol.org/
[17] https://www.w3.org/TR/privacy-principles/#opt-in-out

In fact, reidentification without explicit consent, even on a single site, would constitute a violation of various data protection regulations. The document would therefore be enhanced by removing the discriminatory implication that only cross-contexts reidentification poses a harm to individuals that opt-in consent helps to reduce.

Elsewhere, the authors seem to contradict themselves by stating individuals should have the right to "withdraw consent" suggesting that they must also have the right to provide it:

> *"The right to object, withdraw consent, and restrict use of data about oneself.*
>
> *People may change their decisions about consent or may object to subsequent uses of data about themselves. Retaining rights requires ongoing control, not just at the time of collection."*[18]

The document could be improved by ensuring the browser manufacturer is not guessing what a "person's true preferences" might be, but instead ensuring users can affirmatively provide their preferences.

> *"Attempts to obtain consent to processing that is not in accordance with the person's true preferences result in imposing unwanted privacy labour on the person, and may result in people erroneously giving consent that they regret later."*[19]

Paternalistically making choices on user's behalf robs them of their autonomy.

In addition to the harm created by privacy standards that usurp the end user's autonomous consent, we must also consider the impact of imposing terms that do not in fact protect the user's privacy (i.e., they are not consumer-friendly) and instances where the end user has not in fact given their consent. For example, the document makes reference to "typical controls that are representative of best practices", including ensuring that:

> "*there exist contractual terms between the first party and third party describing the limited purpose for which the data is being shared."*[20]

Limiting the purposes for which data is transferred consequently limits the amount of data, or in what format certain data is shared between first and third parties. In some cases, this may unnecessarily restrict information being transferred between first and third parties, which would otherwise provide a benefit to the consumer.

4. Reworking purpose of the User Agent and transacting on consumer's behalf

User Agents should neither disintermediate people from organizations, nor enforce their will on individuals and legal businesses.

The authors propose that browsers (and by extension app stores and operating systems) disintermediate people from the organizations they wish to interact with.

> *"The user agent acts as an intermediary between a person (its user) and the web. User agents implement, to the extent possible, the principles that collective governance establishes in favour of individuals."*[21]

The user agent does not act as a self-determining intermediary between users and organizations. Accordingly, browser manufacturers should not usurp the role of governments and begin writing quasi-laws for others, policing

---

[18] https://www.w3.org/TR/privacy-principles/#dfn-right-to-object
[19] https://www.w3.org/TR/privacy-principles/#consent-principles
[20] https://www.w3.org/TR/privacy-principles/#deidentified-data
[21] https://www.w3.org/TR/privacy-principles/#user-agents

rival businesses, nor penalizing rivals for violations of internal company policies. MOW must thus disagree with the authors' statement that user agents "*are expected to enforce these principles*."[22]

"Principles" are aspirational goals, rather than legitimate laws enacted by representative governments. One cannot enforce a principle, but only a law or a contract. Given we agree the user agent is not a party to the transactions among and between users and online businesses, they cannot "enforce" their will over that of either individuals or businesses.

The user agent instead facilitates each individual user's direct interactions with those organizations.

> *"The user agent is expected to align fully with the person using it and operate exclusively in that person's interest. It is not the first party. The user agent serves the person as a trustworthy agent: it always puts that person's interest first."*[23]

MOW agrees that user agents are not a party to interactions among people and organizations they interact with online, and thus should not interfere with individual's exercising their autonomy to make such decisions. Accordingly, MOW disagrees that the user agent should remove choice and control from user decisions, regardless of the browser manufacturer's motivation:

> *"In some occasions, this can mean protecting that person from themselves by preventing them from carrying out a dangerous decision, or by slowing down the person in their decision."*[24]

While it is reasonable for browser manufacturers to surface warnings or additional information to inform individual's decision making, they must ensure their warnings are not intrusive or contain dark patterns to nudge specific behaviour that would render discriminatory results in other lines of business operated by the browser manufacturer (e.g., B2C search services, payment services or B2B ad system services).

To prevent such discriminatory behaviour, user agents must not overstep their role by interfering in legal interactions or data transfers among individuals and online businesses. Should the user agent step in between users and businesses to request additional data for its own business purposes, this would be dishonest behaviour that would violate the user's expectation as to what is the role of a Web browser (aka user agent).

> *"Duty of Protection*
>
> *Protection requires user agents to actively protect their user's data, beyond simple security measures… [T]he user agent must… ensure that only strictly necessary data is collected, and require guarantees from any party that the user agent can reasonably be aware that it is shared to."*[25]

The authors' document could be improved by removing suggestions that software ought to make subjective determinations around user decisions, or state that users are unable to transmit decisions directly with businesses they are interacting with.

> *"Duty of Honesty*
>
> *Honesty requires that the user agent try to give its user information of which the user agent can reasonably be aware, that is relevant to them and that will increase their autonomy**, as long as they can understand it** and there's an appropriate time. This is almost never when the person is trying to do something else such as read a page or activate a feature. The duty of honesty goes well beyond that of transparency that is often*

---

[22] https://www.w3.org/TR/privacy-principles/#principles-for-privacy-on-the-web
[23] https://www.w3.org/TR/privacy-principles/#user-agents
[24] https://www.w3.org/TR/privacy-principles/#user-agents
[25] https://www.w3.org/TR/privacy-principles/#dfn-duty-of-protection

included in older privacy regimes. Unlike transparency, honesty can't hide relevant information in complex legal notices ***and it can't rely on very short summaries provided in a consent dialog.*** *If the person has provided consent to processing of their personal data, the user agent should inform the person of ongoing processing, with a level of obviousness that is proportional to the reasonably foreseeable impact of the processing."*[26] (emphasis added).

The text in bold above suggests organizations should never be able to request consent directly from individuals, and that instead they must request consent from the browser manufacturer, who will determine whether or not to allow such a transaction due to its own subjective determination of the "obviousness" or whether it deems the request "proportional" to the "foreseeable impact of the processing."

This disintermediation of the browser manufacturer's own business for those of its rivals is blatant discrimination. The location of a consent notice (inside a browser or on a website) does not determine how well the information and choice is presented.

While we can agree that standard notices can reduce the "labor" of decision making, the marketplace should be left free to compete on offering users choice, rather than centralizing more control into the hands of the dominant B2C software providers that increasing attempt to control the Open Web. Fortunately, the authors acknowledge the risk of discrimination even if they ignore how the trustworthiness of user agents can ever be evaluated.

"Duty of Loyalty

Because the user agent is a trustworthy agent, it is held to be loyal to the person using it in all situations, including in preference to the user agent's implementer. When a user agent carries out processing that is not in the person's interest but instead benefits another actor (such as the user agent's implementer) that behaviour is known as self-dealing. Behaviour can be self-dealing even if it is done at the same time as processing that is in the person's interest, what matters is that it potentially conflicts with that person's interest. Self-dealing is always inappropriate. Loyalty is the avoidance of self-dealing."[27]

Absent some fair method of evaluating the trust of a user agent versus the trust of a website owner, there is no reasonable basis to prefer the individual signals sent by a user agent over those collected by the website owner on its own property. Indeed, a model where individuals can override default preferences (stored by either a browser or a rival consent management solution) on a site-specific basis must be protected across the Open Web.

5. Privacy Principles Seek to redefine fundamental definitions

a. Definition of "inappropriate" data processing

When proposing Principles aimed at clarifying responsible processing of personal data, the authors of the document should not leave undefined what is "inappropriate". The authors again introduce unsupported statements that suggest that "transparency and choice" required by data protection regulations is "often" a signal to individuals that the processing of data will be "inappropriate."

"Reference to the FIPs survives to this day. They are often referenced as "transparency and choice", which, in today's digital environment, is often an indication that inappropriate processing is being described."[28]

The Privacy Principles would be enhanced by listing enumerated cases of appropriate and inappropriate processing. In relation to digital advertising, which seems recently a particular focus of W3C, the authors could refer to the

---

[26] https://www.w3.org/TR/privacy-principles/#dfn-duty-of-honesty
[27] https://www.w3.org/TR/privacy-principles/#dfn-duty-of-loyalty
[28] https://www.w3.org/TR/privacy-principles/#privacy-labour

enumerated list of processing required for competitive digital advertising produced by the largest brands, agencies in the world in collaboration with Google and Facebook.[29]

a. Redefining "privacy" as receiving unsolicited information

The Privacy Principles should not seek to redefine privacy as receiving unsolicited information. The document would be enhanced by ensuring the authors clearly distinguish information directed to people's directly identifiable identity (e.g., email) versus those that are mass messages to audiences that share similar characteristic (e.g., advertising). The following assertion is problematic:

> "*Receiving unsolicited information that either may cause distress or waste the recipient's time or resources is a violation of privac*y."[30]

By using overbroad language, the authors confuse information an online service might guess is related to a request (e.g., a search result) with digital advertising that is paid for by a marketer and subsidizes people's access to ad-funded digital properties. While Individuals never request specific display advertising from specific brands, there are innovative technologies being developed that would enable people to signal back to brands they wish to pause seeing messages from that brand. This signal has a commercial benefit for marketers who can repurpose their media spend to message to individuals more likely to respond to their advertisement.

The authors are quick to point out that digital properties are often commercial enterprises that must either earn their revenues from advertising or from direct payment by their visitors:

> *"Some services have the user pay for their use in data. These services aren't necessarily retaliating by denying their services to users who refuse to pay with data, but the details are more complex than we've had time to write."*[31]

Unfortunately, the authors elsewhere reemphasize that browser manufacturers should have exclusive control over requesting and signalling user choices to all organizations they interact with online. Such statements do not even explain how organizations receiving the signal can trust when it was set, whether it expressed the "true" decision of a user, or is merely a "signal from the user agent" expressing what the browser manufacture guesses its users might want:

> *"specific action from the person (eg. triggering a form control that is not on a modal dialog)"*[32] [Although it is advisable for browser manufacturers to broadcast to organizations].
>
> *"Relying on a global opt-out signal from* ***the user agent****."*[33] (emphasis added)

We should learn from poor experience of 'Do Not Track', where the signal was turned on by default by a browser manufacturer such that recipients had no evidence this signal represented an informed decision by people.[34]

b. Redefining special categories of sensitive information

The authors do not reference established lists of special categories of sensitive information, but instead seek to redefine this term to refer to a host of disparate distinguishable concepts:

---

[29] *See* Policy Framework for Addressable Media Identifiers.
[30] https://www.w3.org/TR/privacy-principles/#unwanted-information
[31] https://www.w3.org/TR/privacy-principles/#non-retaliation
[32] https://www.w3.org/TR/privacy-principles/#opt-in-out
[33] https://www.w3.org/TR/privacy-principles/#consent-principles
[34] https://www.cnet.com/tech/services-and-software/apache-web-software-overrides-ie10-do-not-track-setting/

"2.3 Sensitive Information

Contributes to correlation, identification, secondary use, and disclosure"[35]

In fact, these terms all carry separate and distinct meanings:

- Correlation is a statistical process that has no greater risk to privacy than the selection of an operating system.
- Identification is about establishing the identity of a specific, directly identifiable individual, which could be via a user-initiated authentication.
- Secondary use is a subsequent process of data, regardless of whether the subsequent process was disclosed.
- Disclosure is about the sharing of information, regardless of sensitivity of the information shared.

The authors confuse the definitions of "privacy" with that of "sensitive information."

> *"A particular piece of information may have different sensitivity for different people. Language preferences, for example, might typically seem innocent, but also can be an indicator of belonging to an ethnic minority. Precise location information can be extremely sensitive (because it's identifying, because it allows for in-person intrusions, because it can reveal detailed information about a person's life) but it might also be public and not sensitive at all, or it might be low-enough granularity that it is much less sensitive for many people."*[36]

The authors ignore regulatory definitions of special category sensitive information and instead substitute subjective judgement of individuals, which cannot be intuited by software. The authors then suggest that the browser manufacturer "consider" the factors that might influence subjective judgement of individuals, after denigrating the ability for users to actively express their preferences.

> *"When considering whether a class of information is likely to be sensitive to a person, consider at least these factors:*
>
> - *whether it serves as a persistent identifier (see severity in Mitigating browser fingerprinting);*
> - *whether it discloses substantial (including intimate details or inferences) information about the the person using the system or other people;*
> - *whether it can be revoked (as in determining whether a permission is necessary);*
> - *whether it enables other threats, like intrusion."*[37]

The authors then repeat their error in confusing identifiers, information, identity, and choice related to the likelihood and severity of risks ("threats") to individuals.

MOW agrees that people ought to have the right to consent to personalized marketing (that is content matching "decision making" that depends on "automated profiling") whether or not the identity of the individual is disclosed to the recipient organization. This is because it is always the *use* of the data (e.g., illegal discrimination) rather than the *existence* of data that causes the harm. MOW further agrees that such harms can be exacerbated when a specific individual's identity (e.g., email) is known to the recipient organization:

> *"For some kinds of decision-making with substantial consequences, there is a privacy interest in being able to exclude oneself from automated profiling. For example, some services may alter the price of products*

---

[35] https://www.w3.org/TR/privacy-principles/#hl-sensitive-information
[36] https://www.w3.org/TR/privacy-principles/#hl-sensitive-information
[37] https://www.w3.org/TR/privacy-principles/#hl-sensitive-information

(price discrimination) or offers for credit or insurance based on data collected about a person. Those alterations may be consequential (financially, say) and objectionable to people who believe those decisions based on data about them are inaccurate or unjust. As another example, some services may draw inferences about a user's identity, humanity or presence based on facial recognition algorithms run on camera data. Because facial recognition algorithms and training sets are fallible and may exhibit certain biases, people may not wish to submit to decisions based on that kind of automated recognition."[38]

c. Redefining "de-identified data" as requiring third party control processing

The Privacy Principles should not seek to redefine "de-identified data" as requiring some third-party to control processing inside an organization.

The authors coin a new term "controlled de-identified data" (CDC) to distinguish this from the important similar term used in data protection regulations: "de-identified" data (that is data which has undergone the appropriate measures to ensure the recipient does not link this data to a specific individual's identity rather than making it anonymous).[39]

> *"Data is de-identified when there exists a high level of confidence that no person described by the data can be identified, directly or indirectly (e.g. via association with an identifier, user agent, or device), by that data alone or in combination with other available information****….***
>
> *We talk of controlled de-identified data when there are strict controls that prevent the re-identification of people described by the data except for a well-defined set of purposes."*[40]

The first distinction the authors introduce is that their term, "CDC", must allow for re-identification in cases subject to "a well-defined set of purposes." This suggests that CDC bears a higher risk than the data protection regulation term of "de-identified" data, where the recipient organization has committed to maintain the data in its pseudonymized state.

The authors add another distinction of CDC from the legal definition of de-identified data, such that even with the same organization, no identifier in a given data set can link to any other data set, and that organization will not share the information with any other recipient outside the organization:

> *"Different situations involving controlled de-identified data will require different controls. For instance, if the controlled de-identified data is only being processed by one party, typical controls include making sure that the identifiers used in the data are unique to that dataset, that any person (e.g. an employee of the party) with access to the data is barred (e.g. based on legal terms) from sharing the data further, and that technical measures exist to prevent re-identification or the joining of different data sets involving this data, notably against timing or k-anonymity attacks.*
>
> *In general, the goal is to ensure that controlled de-identified data is used in a manner that provides a viable degree of oversight and accountability such that technical and procedural means to guarantee the maintenance of pseudonymity are preserved."*[41]

---

[38] https://www.w3.org/TR/privacy-principles/#dfn-right-to-be-free-from-automated-decision-making

[39] ICO's Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance: Introduction to anonymisation (May 2021): "*for the purposes of the re-identification offence, the DPA 2018 refers to 'de-identified' personal data as personal data that has undergone pseudonymisation as defined in the UK GDPR rather than (for example) anonymous information*."

[40] https://www.w3.org/TR/privacy-principles/#deidentified-data

[41] https://www.w3.org/TR/privacy-principles/#deidentified-data

The authors suggest that only CDC can guarantee the "maintenance of pseudonymity", despite their explicit assertion that the organization that might perform the generation of the CDC may use a "well-defined set of purposes" to re-identify this data.

Thus, while acknowledging "legal terms" can "bar" certain data processes (e.g., sharing), in the same paragraph, the authors pretend that the identical measures cannot be used to preserve pseudonymity.

In a similar vein, the authors seek to redefine pseudonymous information (data not linked to identity) as "identification":

> *"Identification is the linking of information to a particular individual, even if the information **isn't linked to that individual's real-world identity** (e.g. their legal name, address, government ID number, etc.)."*[42] (emphasis added)

6. Keeping aims of Privacy Principles focused

a. Incorporating social issues into the Privacy Principles

The Privacy Principles should not focus on social issues that are unrelated to privacy. Yet the authors suggest that web standards can help user agents "control" the activities of individuals interacting with online businesses:

> "*Web standards help with data governance by defining structural controls in user agents and establishing or delegating to institutions that can handle issues of privacy. Governance will often struggle to achieve its goals if it works primarily by increasing individual control instead of acting collectively*."[43]

MOW agrees that "governance" requires collective actions, rather than those dictated by "powerful, large entities that control both an implementation and services." In democratic societies, government is the appropriate body to provide policing functions as well as administer due process in courts related to written laws. In contrast, unwritten laws enforced without oversight by user agents is far closer to tyrannical rule that democracy was designed to overcome.

As stated above, we must tailor solutions to a defined problem rather than use the formulation of the Privacy Principles as a means to tackle challenges that face modern society, and that are more suitably dealt with by public bodies.

b. Similarities within groups not to be confused with human rights

The Privacy Principles should not confuse similarities within groups with individual human rights.

> *"Protecting group privacy can take place at two different levels. The existence of a group or at least its activities may need to be protected even in cases in which its members are guaranteed to remain anonymous."*[44]

While MOW agrees that people should have recourse to address the effects of online harms, such harms are entirely separate and distinct from privacy harms arising where "members are guaranteed to remain anonymous." Accordingly, we should remove such social issues to a separate policy document, given this one is focused on principles relating to an individual's privacy.

---

[42] https://www.w3.org/TR/privacy-principles/#dfn-identification
[43] https://www.w3.org/TR/privacy-principles/#collective
[44] https://www.w3.org/TR/privacy-principles/#group-privacy

> *"When people do not know that they are members of a group, when they cannot easily find other members of the group so as to advocate for their rights together, or when they cannot easily understand why they are being categorised into a given group, their ability to protect themselves through self-governing approaches to privacy is largely eliminated."*[45]

The authors confuse "information" that might link people to be classified as belonging to a group (e.g., visitors to a sports site might be considered interested in sports and hence "sports enthusiasts") with a risk to "privacy" that would require there to have been re-identification or some other privacy-related risk to be described when discussing information that is not directly linked to identity.

> "*For example, based on group-level analysis, a company may know that site.example is predominantly visited by people of a given race or gender, and decide not to run its job ads there*."[46]

There is no human right to demand to see an advertisement. In the case above where advertising might be displayed (or not) based on the context of a newspaper article, there is no threat to privacy. Following this same confusion, the authors then seek to redefine "identifying" as any matching of content that is not universally displayed to everyone:

> *"Identifying someone allows a system to treat them differently from others, which can be inappropriate depending on the context."*[47]

When discussing "group privacy," the authors confuse reidentification risks (e.g., directly-identifiable image of an specific individual) with innocuous information people might generate when navigating around the Web (e.g., visiting digital sports sites):

> *"One common problem in group privacy is when the actions of one member of a group reveals information that other members would prefer were not shared in this way (or at all). For instance, one person may publish a picture of an event in which they are featured alongside others while the other people captured in the same picture would prefer their participation not to be disclosed."*[48]

c. Privacy risks are separate from risks of other online harms

The Privacy Principles draft provides a set of examples to illustrate the privacy rights that technology standards could address:

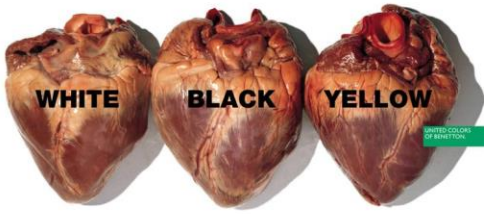
> *"Recipients can have their privacy violated in multiple ways such as unexpected shocking images, loud noises while one intends to sleep, manipulative information, interruptive messages when a person's focus is on something else, or harassment when they seek social interactions."*[49]

To determine the subjectively defined "shocking images" requires software to understand the cultural background and upbringing of the individual visiting a web property. Thus, without invasive collection of each individuals' sensitive category data, this problem does not seem suitable for a web standard to address, unless the W3C believes the culture of its architects ought to impose their own definitions of "shocking" across individuals around the world. Benneton's widely publicized use of shocking images has drawn attention to important social issues, racial discrimination, AIDS, child labour):[50]

---

[45] https://www.w3.org/TR/privacy-principles/#group-privacy
[46] https://www.w3.org/TR/privacy-principles/#collective
[47] https://www.w3.org/TR/privacy-principles/#dfn-identification
[48] https://www.w3.org/TR/privacy-principles/#group-privacy
[49] https://www.w3.org/TR/privacy-principles/#intro
[50] https://web.archive.org/web/20071212034533/http://www.germanlawjournal.com/article.php?id=14

source: https://www.advertisingarchives.co.uk/assets/thumbnails/51/3/42c88351b6b272113468fb857698bfb1.jpg



source: https://www.advertisingarchives.co.uk/assets/thumbnails/141/3/6b08a62e088e66c4bce95c23ab6d295f.jpg

source: https://friendlystock.com/top-ten-controversial-united-colors-of-benetton-ads

Similar subjective issues arise when a technology standard seeks to define "loud," "manipulative," and "harassment." Older individuals may prefer music volume at a higher level or softer level than teenagers. Harassment and illegal dark patterns that deceive or mislead requires a case-by-case determination.

Thus, MOW agrees with the Principles authors that *"because we are all subject to motivated reasoning, the design of defaults and affordances that may impact autonomy should be the subject of independent scrutiny."*[51] MOW further agrees with the authors who suggest that "*people's autonomy with respect to **their own** data" should be "enhanced through appropriate defaults and choice architectures*."

However, this personal "private" property of an individual (i.e., confidentiality and misappropriation of "their own data") must be clearly distinguished from control over their identity (e.g., identity theft or reidentification). Indeed

[51] https://www.w3.org/TR/privacy-principles/#autonomy

"privacy" is often colloquially used to describe very different aspects related to rights around property, identity, and seclusion.[52]

The Principles could be improved if the authors ensured that where they reference "data" they further distinguish when it is private property (e.g., created and owned by the individual), or identity-linked (e.g., whether personal data has been appropriately pseudonymized to render it into a de-identified or anonymous state).

7. Mitigating against higher privacy risks

The Principles should not require organizations to store more data that poses higher risks to individuals.

MOW supports the Principles focus on increasing transparency around data collection and processing:

> *"This consideration extends the TAG's resolution on a Strong and Secure Web Platform to ensure that "broad testing and audit continues to be possible" where information flows and automated decisions are involved."*[53]

Indeed, when multiple supply chain vendors are involved, this can provide enhanced transparency and accountability relative to similar data processing inside a vertically integrated organization, which the absence of an inter-company transfer typically hides such data processing from the view of both individuals and regulators.

However, the Principles can be improved by clarifying ambiguous and overbroad assertions as illustrated in the following statement:

> *"Such transparency can only function if there are strong rights of access to data (including data derived from one's personal data) as well as mechanisms to explain the outcomes of automated decisions."*[54]

Requiring all data derived from "one's personal data" to be stored by all organizations that receive the derived data could actually expose individuals to higher privacy risks. Take, for example, the recipient of a de-identified data set or an anonymous data set containing aggregate data. If each individual were able to interrogate such data sets to learn whether their individual data points were part of the input data set, this would require all recipients to store even more data that poses higher risks to individuals (such as the raw directly-identifiable input prior to the de-identification process). Accordingly, the document can be improved by focusing on ensuring once privacy-by-design techniques have been applied to a data set there are no obligations on organizations to maintain the raw inputs to those privacy-enhancing processes.

8. Separating identity from identifiers or information linked with identifiers

Identity should not be confused with identifiers nor with information linked to identifiers. However, the authors occasionally use the term "identity" to refer to directly identifiable information (such as enumerated by data protection regulations):

> *"In computer systems and on the Web, an identity seen by a particular website… for a person can be:*
>
> - *their name,…*
> - *their phone number,*

---

[52] Joshua Koran, Redefining what we mean by 'privacy' on the ad-funded open web, 13 October 2021.
[53] https://www.w3.org/TR/privacy-principles/#transparency
[54] https://www.w3.org/TR/privacy-principles/#transparency

- *their location data"*[55]

However, the authors also confuse the term of "identity" with both the digital identifiers that are necessary for web properties to operate, and also with mere information linked to identifiers:

> *"In computer systems and on the Web, an identity… can be:*
>
> *…assigned an identifier of some type, which makes it easier for an automated system to store data about that person…. [which] can be:*
>
> *an identification number including those mapping to a device that this person may be using,…*
> *an online identifier such as email or IP addresses, or*
> *[information] factors specific to their physical, physiological, genetic, mental, economic, cultural, or social identity."*[56]

The authors seem to understand the importance to privacy as to whether identity is linked to digital identifiers strings or sensitive category information as they state further:

> *"Strings derived from identifiers, for instance through hashing, are still identifiers so long as they may identify a person."*[57]

Data protection regulations clearly state all the above is "personal data", but that the likelihood and severity of risk to an individual is the determining factor. For example, whether the identifiers are directly linked to identity or have appropriate business and technical measures in place within the recipient organization to keep them distinct from identity (i.e. not re-identify), as well as whether the information linked to pseudonymous identifiers is innocuous or belongs to a special category that poses higher risks.

However, the authors are aware that privacy harms are related to revealing a specific individual's identity rather than merely remembering helpful information (e.g., items in a shopping basket from last visit, frequency capping to prevent annoying overexposure to the same advert):

> *"A privacy harm occurs if a person reasonably expects that they'll be using a different identity on a site, but the site discovers and uses the fact that the two or more visits probably came from the same person anyway."*[58]

9. Standards based on debated assumptions

Standards should be based on debated assumptions. While the following phrasing suggests that user agents should interfere with "recognition" of a user in different "contexts", the authors describe that even this notion is not agreed upon as an appropriate boundary that is directly related to privacy.

> *"There is disagreement about whether user agents may also widen their machine-enforceable contexts. For example, some user agents might want to help their users present a single identity to multiple sites that the user understands represent a single party, or to a site across multiple installations….*
>
> 2.2 Personal Data…

---

[55] https://www.w3.org/TR/privacy-principles/#identity
[56] https://www.w3.org/TR/privacy-principles/#identity
[57] https://www.w3.org/TR/privacy-principles/#identity
[58] https://www.w3.org/TR/privacy-principles/#hl-recognition-same-site

Issue

There is not currently consensus among the authors and contributors to this document about the below principles. Please see this issue for more details."[59]

Moreover, people often wish to authenticate their identity with different organizations, while others wish to receive consistent experiences across different web properties (such as would be supported by default preferences stored inside a browser or with a rival consent management service). While browser manufacturers should help facilitate such activities, they are not free to interfere with them.

"*A user agent should help its user present the identity they want in each context they find themself in.*

Sometimes this means the UA should ensure that one site can't learn anything about their user's behavior on another site, while at other times the UA should help their user prove to one site that they have a particular identity on another site."[60]

The authors inaccurately state that browser manufacturers must prevent information sharing across "sites", regardless of the risk of reidentification. For example, if a user were searching for flights to Paris, another site recommending hotels in Paris might be considered useful. If neither site collects the identity of the user, what is the likelihood or severity of the privacy risk the browser manufacturer is stating as an objective justification for its interference, despite admitting it is not a party to the interactions among the user and these two sites?

Indeed, even the list of "contexts" is ill defined and not agreed upon, but dependent on subjective "assumptions" by browser manufacturers. Surely "standards" should not be based on debated "assumptions" that impose restrictions on user interactions with digital properties.

"*To do this, user agents have to make some assumptions about the borders between contexts. By default, user agents define a machine-enforceable context or partition as*:

- *A set of environments (roughly iframes (including cross-site iframes), workers, and top-level pages)*
- *whose top-level origins are in the same site (but see [PSL-Problems])*
- *being visited within the same user agent installation (and browser profile, container, or container tab for user agents that support those features)*
- *between points in time that the person or user agent clears that site's cookies and other storage (which is sometimes automatic at the end of each session)…*

user agents ***are free to restrict this context*** *as people need*."[61] (emphasis added).

The text in bold above illustrates the contradiction with the authors' earlier statement that the user agent is not a party to the transaction, and hence should not be interfering with user's experience or restricting their autonomy. Indeed, user agents are not "free" to interfere with online transactions as like all organizations, they must abide by laws of the regions in which they operate.

[59] https://www.w3.org/TR/privacy-principles/#identity
[60] https://www.w3.org/TR/privacy-principles/#identity
[61] https://www.w3.org/TR/privacy-principles/#identity

10. Reference laws

The Principles place a burden on the reader to understand what is *required* by law, and what is *recommended* by the authors of guidance in terms of what they consider to be general purpose privacy principles. Each section of the document should clearly state the laws on which it relies, or where there are no laws to support the section in part or in full, to state as much. The reader will then be able to determine what they must do if they wish to observe the legal position in a particular jurisdiction should they wish to treat the law as both a "floor" and a "ceiling".

11. Authors' affiliation

The principal authors of the documents are employed by large entities. Entities whose businesses have an interest in digital advertising. For example;

1. A large publisher that generates the majority of its revenue from subscriptions will benefit if the advertising-funded open web is no longer viable, as more readers will move to their platform. Rivals will find it difficult to compete, leaving them vulnerable to being bought by these large entities.
2. A publishing service that operates many brands might be able to persuade people to authenticate with all these brands and therefore attract more brands to their publishing platform.
3. Platforms that can persuade people to not only authenticate themselves, but also provide directly identifiable personal information, and permission to profile based on content consumed, will be able to offer advertisers a far more compelling proposition than rivals that operate anonymous services.

The motivation of the authors need not be called into question, as they may be ignorant of the foreseeable consequences and the benefits to their employer of any principles they propose. They may not consider the discrimination against smaller rivals as they have not been accountable for profit and loss within a small business.

The position advanced in the Principles benefits most of the authors' employers. As such, the employers would likely support the work and ongoing involvement of the authors, irrespective of their motivation. The standout commercial exception relates to Google, who agreed a voluntary set of commitments with the UK's Competition and Markets Authority which would prohibit further involvement in the Principles unless they align to GPDR.[62]

12. Narrow views in developing the Privacy Principles

The Privacy Principles have been edited by Robin Berjon of the New York Times and Jeffrey Yasskin of Google. Whilst MOW accepts that contributors from smaller organizations have had the opportunity to contribute to developing the document, we stress the importance of ensuring that subsequent iterations and modifications made to the Principles are formulated by a group that is representative of the diversity of W3C Membership, and the wider stakeholders and businesses that operate on the Web. MOW is concerned that going forward, updated versions of the Principles (which will ultimately affect the entire ecosystem) must not solely or primarily reflect the oversized influence of the New York Times and Google. Standards that affect such a wide scope of stakeholders should be

[62] *See* CMA's decision to accept commitments from Google: https://assets.publishing.service.gov.uk/media/62052c52e90e077f7881c975/Google_Sandbox_.pdf; Google's final commitments accepted by the CMA: https://assets.publishing.service.gov.uk/media/62052c6a8fa8f510a204374a/100222_Appendix_1A_Google_s_final_commitments.pdf

developed plurally and without bias. Significantly, many of the groups the Privacy Taskforce Charter state are considered, have in fact not been.[63]

[63] https://github.com/w3ctag/privacy-principles#privacy-task-force-members