ITUPublications

International Telecommunication Union

Standardization Sector

ITU Focus Group Technical Report

(12/2023)

ITU Focus Group on metaverse

Embedding safety standards and the user control of personally identifiable information (PII) in the development of the metaverse

Working Group 6: Security, Data & Personally identifiable information (PII) Protection

Technical Report ITU FGMV-11

Embedding safety standards and the user control of personally identifiable information (PII) in the development of the metaverse

Summary

Technical Report ITU FGMV-11 develops three key areas of a rights-based approach to embedding ethics and safety standards and user control of PII in developing the metaverse that build conceptually on each other:

– Data control and agency of users in relation to their service and platform provider,

– Human rights test governing workflow design as well as the conduct of service and platform providers as that conduct relates to their public stakeholders, and

– Principles for the development of safety standards in line with the SDGs that can effectively govern user conduct within the metaverse spaces such providers offer.

The report further maps out key lenses in which these three areas interact with one another, with platform design considerations, and other stakeholders. It also offers a practical use-case on an open source and decentralized protocol demonstrating how technical infrastructure can enable user control of PII.

Keywords

Ethics, human rights, metaverse, personal data, protocols, safety, social graph, standards.

Note

This Technical Report is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1.0 of the ITU Technical Report on "*Embedding safety standards and the user control of personally identifiable information (PII) in the development of the metaverse*" approved at the 4th meeting of the ITU Focus Group on metaverse (ITU FG-MV), held on 4-7 December 2023 in Geneva, Switzerland.

Acknowledgments

This Technical Report was researched and written by Sarah Nicole (Project Liberty Institute), Jan Eissfeldt (Wikimedia Foundation), and Neha Vijay (Radix) as a contribution to the ITU Focus Group on metaverse (FG-MV). The development of this document was coordinated by Vincent Affleck (DSIT, United Kingdom), as FG-MV Working Group 6 Chair, and by Radia Funna (Build n Blaze) as Chair of Task Group on building confidence and security in the metaverse.

Additional information and materials relating to this report can be found at: https://www.itu.int/go/fgmv. If you would like to provide any additional information, please contact Cristina Bueti at tsbfgmv@itu.int.

Editor:	Sarah Nicole Project Liberty Institute	E-mail: sarah.nicole@projectliberty.io
Editor:	Jan Eissfeldt Wikimedia Foundation	E-mail: jeissfeldt@wikimedia.org
Editor:	Neha Vijay Radix	Email: nehav@radix.email
WG6 Chair:	Vincent Affleck DSIT United Kingdom	Email: Vincentaffleck2@hotmail.com
Task Group Chair:	Radia Funna Build n Blaze	Email: rfunna@buildnblaze.com

Table of contents

		Page
1	Scope	1
2	References	1
3	Definitions	1
	3.1 Terms defined elsewhere	1
	3.2 Terms defined in this Technical Report	1
4	Abbreviations and acronyms	1
5	Conventions	1
6	Allowing data control through open protocols for an ethical metaverse	1
7	Human rights test	3
8	Principles for safety standards	4
9	Conclusion	5
Bibliography		7

Technical Report ITU FGMV-11

Embedding safety standards and the user control of personally identifiable information (PII) in the development of the metaverse

1 Scope

This Technical Report develops three key areas that build conceptually on each other:

– Data ownership and agency of users in relation to their service and platform provider,

– Human rights test governing workflow design, as well as the conduct of service and platform providers as that conduct relates to their public stakeholders, and

– Principles for the development of safety standards in line with the SDGs that can effectively govern user conduct within the metaverse spaces such providers offer.

2 References

None.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Technical Report

None.

4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

CSAM	Child Sexual Abuse Material
DSNP	Decentralized Social Networking Protocol
PII	Personally Identifiable Information
UGC	User Generated Content

5 Conventions

None.

6 Allowing data control through open protocols for an ethical metaverse

As social media and other Internet services evolve into metaverse environments, the issue of protecting users' personal identifiable information and privacy becomes even more critical. Wearable devices will collect an unprecedented amount of data, including sensitive information, surpassing the capabilities of current digital tools. This data will include biometric and geospatial data to enhance immersive experiences. Additionally, avatars will require critical data, such as physiological data, for their creation.

To build confidence, trust, and security in the metaverse, a new generation of technical protocols needs to address the privacy, security, and interoperability issues. Decentralized, open and public protocols should be the foundation of social metaverse development. The decentralized social

networking protocol (DSNP) -an open-source protocol stewarded by Project Liberty- and other similar protocols, have the potential to ensure personal data privacy in metaverse platforms by letting each user possess and control their own social graph. Figure 1 shows the DSNP ecosystem.

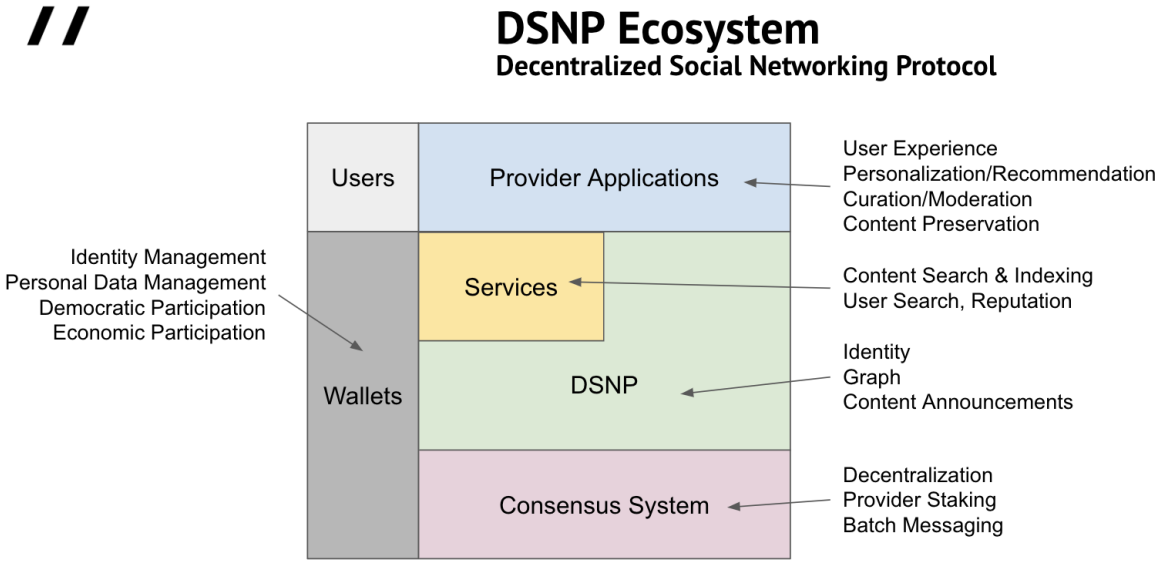

Figure 1 – DSNP ecosystem (Source: [b-DSNP] [b-Project-Liberty])

Table 1 – DSNP and user control of personally identifiable information (PII) use-case

For users to authenticate and access metaverse environments while controlling their PII, they would need to have a DSNP account. This would offer users a seamless and secure decentralized self-sovereign identity solution. Upon download of any wallet from any application store and source, users could generate their unique identity, granting them access to a multitude of services across the Internet, such as metaverse environments. Integration of such a login option could enhance the login process, potentially replacing traditional methods like "Login with Facebook". Additionally, this would facilitate the receipt, storage, and sharing of various credentials, such as KYC and membership details, that could be issued by services to their clientele. Some metaverse services might necessitate specific credentials alongside the decentralized identity for access. Distinguishing itself from centralized counterparts like Facebook or Google, users on metaverse services based on DSNP would retain complete control over their personal identifiable information, obviating the need for web services to maintain centralized personal databases, ensuring general data protection regulation (GDPR) compliance. Personal data would remain decentralized within each user's preferred wallet, eliminating the risk of large-scale data breaches. Users would exercise full agency in determining when, where, and with whom (i.e., platform, service providers) they would share their data. Furthermore, credential issuers would play no role in the verification process, bolstering user privacy and scalability. The system would guarantee maximum reliability, as there would be no risk of server outages. Its decentralized nature would afford resistance to censorship, with no central components susceptible to interference. Lastly, Liberty's architecture would be impervious to external intervention, ensuring that no entity could deactivate an identity.

Table 1 presents a use case of DSNP and user control of PII. Decentralizing the social metaverse means disconnecting the user interface from the underlying data. The users' social aspects (followers, interests, friends, etc.) should be integrated into the Internet itself through protocols rather than being confined to a particular metaverse application. The metaverses interoperability should be functioning on the same principles as the current email system. It relies on open protocols that numerous services can utilize. Despite multiple email applications with varying functionalities and levels of quality, users' contacts can be easily transferred and will invariably function.

An open and public social graph that is operating completely in the user's interest is the first essential piece of infrastructure required to unlock an ethical metaverse. Social network metaverses constructed to provide a universal, global social graph will allow people to interact with friends, family, content-creators and public figures regardless of the metaverse application of their choice. Users should have the agency to choose where their data is kept and located and who is allowed to access it.

The International Telecommunication Union (ITU) is the United Nations specialized agency for information and communication technologies (ICTs). The organization recently established the first Focus Group on metaverse (https://www.itu.int/go/fgmv) to bring together experts from around the world to shape the development of metaverse standardization for the benefit of all.

This technical report was developed in the Task Group (TG) on building confidence and security in the metaverse including PII protection, ethical issues, and standards as part of Working Group 6 (WG6) on security, data and personally identifiable information (PII).

7 Human rights test

The metaverse, both for consumer-facing services and industrial applications, creates novel collaborative and interactive spaces that do not align with jurisdictional boundaries while also providing unprecedented opportunities to gather PII. Therefore, service and platform providers need to create systems, processes, and policies to uphold the human rights of their customers and users while aiming to accommodate differing expectations, laws, and regulations provided by sovereigns. Providers are required to balance these obligations that the sovereign nations and the service providers both have under international agreements and towards their customers and users.

One key area of balance within a reliable, SDG-aligned policy framework informing decision-making, is human rights and, by extension, determining the applicability of laws outside of the provider's own home jurisdiction, usually involving both international legal demands and court orders. These demands are often tied to the jurisdiction of the customers involved but can also encompass other types of challenges. For example, a demand might be based on the geographic location or specific topics that users might have interacted with each other on or shared views about, as well as observability of such issues by third-party customers located in jurisdictions that have concerns about such issues even though none of the users involved on the service provider's platform nor the providers themselves are located in such a jurisdiction.

The technical report proposes the following three-stage test to help providers navigate the challenges, protect the human rights of their customers under applicable United Nations agreements, and provide clear expectations that sovereigns can effectively work with in pursuit of their respective public policy goals:

A services or platform provider in the metaverse shall find an international law applicable and comply with legal demands or court orders outside their jurisdiction if and only if:

- The case is one where the provider cannot effectively challenge the law. The inability to challenge the law includes both the primary law at issue in the case and the ability to challenge geographic jurisdiction.
  - The provider will make that determination either based on a court order, in-house legal advice, or outside expert counsel.
- The case must be one that presents an identifiable risk to the provider or to impacted customers or third parties.
  - Examples of risks include, but are not limited to, risks to customer safety or security, risks of service or platform blocking or similar technical disruption, and/or monetary risks.
- A human rights analysis finds that compliance with the law is in line with international human rights norms.

  - Examples: a case that found that an individual person's right to dignity and privacy outweighed the public interest in certain information would likely align with human rights norms. On the other hand, a case that ordered the removal of public information about a major historical event available in the metaverse service the provider offers would likely not align with human rights norms.

All three tests need to be met for the provider to consider the international law applicable and the legal demands or court orders based on it to be valid, allowing it to cooperate and comply with them. The scope of the test also includes:

- Demands to modify or re-design the platform or service to accommodate national security preferences by:
  - Creating separate instances of the service only accessible to the public of a limited number of jurisdictions;
  - Creating encryption backdoors;
  - Creating or modifying identification and authentication systems and processes.
- Demands from intelligence and security agencies;
- Demands from international organizations; and
- All demands of the release of user data independent of the source without the impacted user's active consent specifically provided for that instance.

8 Principles for safety standards

Building upon the human-centric approach to data control and the SDG-aligned human rights test, the establishment of safety standards in the developing metaverse can significantly assist service and platform providers in navigating this evolving landscape. These principles, drawing from the collective efforts of member states [b-eSafety], academia [b-Perrino], and industry best practices [b-TSPA] hold the potential to effectively address societal challenges, enhance benefits for customers and users, instil confidence for providers in their investments, and contribute to the realization of public policy objectives.

Expanding existing approaches towards the challenges and opportunities of the metaverse, the report proposes the following safety principles:

- Providers are responsible for understanding, assessing, and addressing safety risks to their users and customers throughout the service or platform lifecycle from design to scaled operations and closure.
- Products, services, platforms, and provider practices supporting them safeguard the human dignity [b-UDHR] of users, customers, and subjects of interactions.
- Providers publish their safety objectives as part of regular transparency reports in line with regulatory best practices.

It is useful to exercise the approach for practical use cases, including but not limited to asset digital twins, living being digital twins [b-Björnsson-Borrebaeck-Elander] [b-Evseeva-Erdniev] [b-Gonsard-AbouTaam-Prévost] [b-Hernandez-Boussard-Macklin-Greenspan] [b-Lautenbacher-Niarakis-Helikar] [b-Neethirajan-Kemp], and social media's traditional default use case that is likely the best analogy to safety challenges to govern metaverse e-commerce and user-to-user interactions: user-generated content. Taking the three one by one:

Asset digital twins, such as of wind turbines [b-Haghshenas-Hasan-Osen], hydrocarbon pipelines [b-Li-Gai-Xue], and other industrial assets, pose the perhaps easiest but also least-obvious safety principal challenge. Unlike traditional simulations or models, which focus on aspects of an asset, digital twins emphasize a holistic approach to the asset mapped in its environment and lifecycle. Therefore, their scope encompasses distinct safety concerns that extend beyond cybersecurity. These

concerns can arise from gathering data to construct the clone. Additionally, they stem from the conduct of the asset's operators - the users in this scenario; its customers, whose behaviour might be inferred by studying the digital twin's manifestation in the metaverse, leading to specific societal or military insights; and societies overall if the asset serves a broader purpose creating dependencies beyond its direct customer base. Providers would not only have to review the release of asset digital twin data via the human rights test outlined above, trivial in this use case, but also safeguard and account for the asset digital twin's safety principal implications complementary to the actual physical asset itself and adherence to SDG 9.

Living beings' digital twins pose a range of challenges beyond the scope of assets, especially if the digital twin does not deal with plants or non-human animals in the agricultural industry or comparable contexts but humans. Touching upon the most sensitive of this group of use cases, digital twins of human medical patients, the metaverse provides unprecedented opportunities to aid scientific progress and effective treatment but also reveals notable challenges across all three layers of this technical report.

Metaverse services and platforms need to be able to; secure the safety and privacy of the holistic medical data about the patient that the digital twin contains as well as generates in interacting with, say, a doctor exploring a treatment; empower the patient to shift service provider with all the data and without constraint as outlined in the free protocol section; navigate release of these highly sensitive and personal data sets under the human rights test outlined above; and design and deploy their offerings in line with the safety principles to inspire justified confidence in the technology and the service so enabling a fuller realization of its beneficial potential. Customers who have no confidence in the metaverse would not be using it for services like medical digital twins and it would also be problematic in the light of SDGs 3, 4, 5, and 9.

Finally, safety challenges tied to user-to-user interactions are presented in a well-studied use case with both academic insights [b-Basu] as well as mainstream accounts provided [b-Frenkel-Browning]. Problems focus not only on the traditional content moderation social media has historically focused on since the rise of Trust & Safety in the 2000s [b-TSPA] but also broader conduct of the users, their experience of each other and their metaverse environment. Providers, bound by safety principle 1, cannot offload user or customer safety issues to them but need to study, design, and deploy with preventive intent. Mitigating opportunities not only relate to the user-interface or in-service features but also include platform design, no different than has been well-established across traditional services. For example, child sexual abuse material (CSAM) is a well-studied challenge on platforms that offer non-public interaction opportunities but is not prevalent on even very large platforms that are entirely public due to their transparent architecture with public interest intent and that hence disincentivise the distribution of such deeply harmful and illegal material.

This suggests that metaverse services and platform providers can mitigate potential ethical and safety challenges by designing and building their offerings guided by the specific purpose of their business model. The business model behind the existing social web, making profits out of users' data, is highly harmful for users leading to addiction and polarization. Ethical and safety-designed business models will necessarily shift towards a more open and decentralized operation from platform services providers. The broader the purpose, the higher the risks and hence investment is required to adhere to safety principles. Such decisions also impact a wider range of SDGs, including SDG 5 ensuring societal acceptability in line with a rights-based approach to metaverse development.

9 Conclusion

In summary, as the metaverse emerges from Internet services, protecting user privacy in the face of extensive data collection from wearable devices is critical. This Technical Report expresses how infrastructure is key to ensuring users' data protection. Open, decentralized protocols like the

decentralized social networking protocol (DSNP) offer a solution, enabling users to control their data across metaverse platforms, aligning with UN objectives such as interoperability and accessibility.

The challenge intensifies as the metaverse transcends jurisdictional boundaries, necessitating service providers to navigate complex legal landscapes while upholding human rights. A proposed three-stage test outlines guidelines to ensure compliance with international norms and valid legal demands.

Moreover, emphasizing a human-centric approach, this Technical Report advocates for comprehensive risk assessment and transparent reporting practices. This framework addresses safety concerns across diverse metaverse use cases, aligning with societal goals and ethical business models.

In essence, safeguarding user data, aligning with international norms, and implementing robust safety standards are pivotal for an ethical metaverse. This strategic approach not only tackles immediate challenges but also paves the way for a sustainable, inclusive metaverse ecosystem in line with global objectives.

Bibliography

[b-Basu] Basu, T. (2021). *The metaverse has a groping problem*. MIT Technology Review. Available [viewed 2023-10-09] at: https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem

[b-Björnsson-Borrebaeck-Elander] Björnsson, B., Borrebaeck, C., Elander, N. et al. (2022). *Digital twins to personalize medicine*. Genome Med 12, 4 (2020). Available [viewed 2023-09-10] at: https://doi.org/10.1186/s13073-019-0701-3

[b-DSNP] Decentralized Social Networking Protocol (DSNP). *DSNP WhitePaper.* Available [viewed 2023-11-21] at*:* https://dsnp.org/dsnp_whitepaper.pdf

[b-eSafety] eSafetyCommissioner Australia (2023). *Safety by Design Principles and Background*. Available [viewed 2023-09-10] at: https://www.esafety.gov.au/industry/safety-by-design/principles-and-background

[b-Evseeva-Erdniev] Evseeva, I., Erdniev, A. (2022). *Digital twins in pedagogical modeling*. Applied psychology and pedagogy, p. 95-102. Available [viewed 2023-09-10] at: https://doi.org/10.12737/2500-0543-2022-7-2-95-102

[b-Frenkel-Browning] Frenkel, S., Browning, K (2021). *The Metaverse's Dark Side: Here Come Harassment and Assaults*. New York Times. Available [viewed 2023-09-10] at: https://www.nytimes.com/2021/12/30/technology/metaverse-harassment-assaults.html

[b-Gonsard-AbouTaam-Prévost] Gonsard, A., AbouTaam, R., Prévost, B. et al. (2023). *Children's views on artificial intelligence and digital twins for the daily management of their asthma: a mixed-method study*. Eur J Pediatr 182, 877–888. Available [viewed 2023-09-10] at: https://doi.org/10.1007/s00431-022-04754-8

[b-Haghshenas-Hasan-Osen] Haghshenas, A., Hasan, A., Osen, O. et al. (2023). *Predictive digital twin for offshore wind farms*. Energy Inform 6, 1. Available [viewed 2023-09-10] at: https://doi.org/10.1186/s42162-023-00257-4

[b-Hernandez-Boussard-Macklin-Greenspan] Hernandez-Boussard, T., Macklin, P., Greenspan, E.J. *et al.* (2021). *Digital twins for predictive oncology will be a paradigm shift for precision cancer care*. Nat Med 27, 2065–2066 (2021). Available [viewed 2023-10-09] at https://doi.org/10.1038/s41591-021-01558-5

[b-Lautenbacher-Niarakis-Helikar] Laubenbacher, R., Niarakis, A., Helikar, T. et al. (2022). *Building digital twins of the human immune system: toward a roadmap*. npj Digit. Med. 5, 64. Available [viewed 2023-09-10] at: https://doi.org/10.1038/s41746-022-00610-z

[b-Li-Gai-Xue] Li, B., Gai, J., Xue, X. (2020). *The Digital Twin of Oil and Gas Pipeline System*. IFAC-PapersOnLine Volume 53, Issue 5, Pages 710-714. Available [viewed 2023-09-10] at: https://doi.org/10.1016/j.ifacol.2021.04.162

[b-Neethirajan-Kemp] Neethirajan. S., Kemp, B. (2021). *Digital Twins in Livestock Farming*. Animals (Basel). Apr 3;11(4):1008. Available [viewed 2023-09-10] at: doi: 10.3390/ani11041008. PMID: 33916713; PMCID: PMC8065673

[b-Perrino] Perrino, J. (2022). *Using 'safety by design' to address online harms*. Brookings Institution. Available [viewed 2023-09-10] at: https://www.brookings.edu/articles/using-safety-by-design-to-address-online-harms/

[b-Project-Liberty] Project Liberty Institute, Internal Communication Chart

[b-TSPA] Trust & Safety Professional Association (2023). *Trust and Safety Curriculum*. Available [viewed 2023-09-10] at: https://www.tspa.org/curriculum/ts-curriculum/

[b-UDHR] United Nations General Assembly (1948). The Universal Declaration of Human Rights (UDHR). New York: United Nations General Assembly. Available [viewed 2023-09-10] at: https://www.un.org/en/about-us/universal-declaration-of-human-rights