Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized
Public Disclosure Authorized

ID4D
IDENTIFICATION FOR DEVELOPMENT

PRACTITIONER'S NOTE

DIGITAL ID AND THE DATA PROTECTION CHALLENGE[2]

October 2019

Inclusive and trusted identification (ID) systems are crucial tools for achieving sustainable development, including the World Bank Group's twin goals of ending extreme poverty and of boosting shared prosperity and greater equity in the developing world.[3] Indeed, the essential role that identification plays in development is explicitly recognized in Sustainable Development Goal (SDG) Target 16.9, to "provide legal identity for all, including birth registration" by 2030.[4]

Traditionally, proof of identity has been provided through physical documents, such as birth certificates, passports, or ID cards. As the world becomes increasingly digitized, the next generation of ID systems use new technologies to provide digital proof of legal identity for in-person and remote transactions. These digital ID systems can help achieve multiple development goals, but also create challenges for digital privacy and data protection. This note describes these risks and then presents concrete steps to mitigate them while harnessing the full potential of digital ID for development.

Digital ID for development

In addition to helping achieve SDG Target 16.9 directly, digital ID systems that provide proof of legal identity can support multiple rights and development goals—such as financial and economic inclusion, social protection, healthcare, and education-for-all, gender equality, child protection, agriculture, good governance, and safe and orderly migration—through:

- **Empowering individuals and facilitating their access to rights, services, and economic opportunities that require proof of identity.** This includes social services, pension payments, banking, formal employment, property rights, voting, and more.

---

1 This note was prepared by Julia Clark and Conrad Daly as part of the Identification for Development (ID4D) Initiative, under the supervision of Vyjayanti Desai. This note benefited greatly from the inputs and reviews of World Bank Group staff including David Satola and Jonathan Marskell, as well as feedback from Kanwaljit Singh (Bill & Melinda Gates Foundation), CV Madhukar (Omidyar Network), and David Symington (Office of the UN Secretary General's Special Advocate for Inclusive Finance for Development).

2 Much of the material in this Briefing Note draws from the *ID4D Guide for Practitioners*, available at http://id4d.worldbank.org/guide.

3 World Bank. 2017. *Principles on Identification for sustainable Development: Toward the Digital Age*. Washington, DC: World Bank Group. http://id4d.worldbank.org/principles.

4 UN General Assembly. 2015. A/RES/70/1. *Transforming our world: the 2030 Agenda for Sustainable Development*. SDG 16.9. https://www.refworld.org/docid/57b6e3e44.html.

- **Strengthening the transparency, efficiency, and effectiveness of governance and service delivery.** Digital ID systems can help the public sector reduce fraud and leakage in government-to-person (G2P) transfers, facilitate new modes of service delivery, and increase overall administrative efficiency.[5]
- **Supporting private sector development.** In addition to the public sector, digital ID systems can also help private companies reduce operating costs associated with regulatory compliance (e.g., eKYC), widen customer bases, generate new markets, and foster a business-friendly environment more broadly.[6]
- **Enabling the digital economy.** Combined with trust services like e-signatures, digital ID systems facilitate trusted transactions, streamline "doing business," and create opportunities for innovation—providing a core platform for the digital economy.

Digital ID and data protection—What are the risks?

While digital ID systems can support multiple development goals, they also create risks to digital privacy and data protection. While such risks are inherent to any ID system, digitization can exacerbate their scale and frequency. These risks may have serious, often immeasurable, consequences for people, and therefore require appropriate protections.

Box 1. Data protection and privacy international good practice standards

Digital ID systems raise data privacy concerns because they collect personal data. Building upon existing principles[7], **the European Union's (EU) General Data Protection Regulation (GDPR)[8] sets a new, international good-practice standard for data protection and privacy.** Here, we extract some working definitions.

Data privacy differs from the fundamental right to privacy—commonly defined as the "right to be let alone"[9]—and should be understood as the **appropriate and permissioned use and governance of personal data.** In ID systems, data privacy does *not* necessarily mean that all data is kept secret at all times. Rather, it means that data should only be accessed, processed, or shared by and with authorized users for pre-specified purposes that have been agreed in advance. **Data protection—**which includes the legal, operational, and technical methods and controls for securing information and enforcing rules over access and use—is therefore fundamental to ensuring data privacy.

Not all data merits the same level of protection. **Personal data** refers to "any information relating to an identified or identifiable natural person" (GDPR Article 4). An **identifiable natural person** (or "data subject") is defined as a natural person "who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (GDPR Article 4). **Sensitive personal data** (or "special categories of data") refers to "personal data that, by their nature, are particularly sensitive in relation to fundamental rights and freedoms and merits specific protection as the context of their processing could create significant risks to a person's fundamental rights and freedoms."[10] They include data consisting of racial or ethnic origin, political opinions, religious or philosophical beliefs, health, life or sexual orientation, as well as biometric and genetic data. (GDPR Recital 51).

5 World Bank. 2018. *Public Sector Savings and Revenue from Identification Systems: Opportunities and Constraints*. Washington, DC: World Bank Group. http://id4d.worldbank.org/research.
6 World Bank. 2018. *Private Sector Economic Impacts from Identification Systems.* Washington, DC: World Bank Group. http://id4d.worldbank.org/research.
7 *E.g.,* U.S. Federal Information Processing Standards, OECD Privacy Principles, ISO/IEC, and PbD standards (see below).
8 Regulation (EU) 2016/679 of 27 April 2016 (GDPR).
9 See Warren, Samuel and Louis Brandeis. 1890. "The Right to Privacy". 4 Harvard Law Review. p. 193 and Cornell, Anna Jonsson. 2016. "Right to Privacy", Max Planck Encyclopedia of Comparative Constitutional Law.
10 In order to facilitate readability, the term "personal data" will be used throughout without distinguishing between notions of "personal data", "personally identifiable information" (PII), and "sensitive personal data".

Data and privacy risks related to personal information

Any activity that collects, stores, or processes personal data raises certain risks, including, but not limited to:

- **Security breaches:** Physical or cyberattacks on data in transit or at rest.
- **Unauthorized disclosure:** Inappropriate transfer of data between government agencies, foreign governments, private companies, or other third parties.
- **Exposure of sensitive personal information:** Disclosing sensitive personal information (*e.g.,* biometrics, religion, ethnicity, gender, medical histories) for unauthorized purposes.
- **Function creep:** The use (and even sharing) of data for purposes beyond those for which consent was given.
- **Identity theft:** Identity theft in the digital world can lead to consequences that are at least as serious as those in the "real", physical world, and, given the global, decentralized nature of the internet, damages that are often more difficult to repair. In a digitized world, impersonation can be undertaken by just about anyone.[11, 12]
- **Surveillance risks:** The ability to correlate identifying information across databases (*e.g.,* via facial recognition) increases surveillance risks, particularly where biometrics are involved.[13]
- **Discrimination or persecution:** Identity attributes might be used to discriminate or persecute particular people or groups. Relatedly, reputational attacks (professional and personal) can be launched with significantly greater ease—and to significantly greater effect.[14]
- **Unjust treatment:** Incomplete or inaccurate data can lead to mistakes or unjust treatment.

Digitization: raising the stakes

While the above-discussed risks are present in any ID system, **digital ID systems may augment both the risks and the harms** beyond traditional, paper-based systems because they enable:

- **Ever-more massive data security breaches:** Consolidation of data increases the impact of data breaches,[15] while also making such databases more attractive targets.
- **Easy destruction of digital records:** Digitization allows for the easy (or mass) deletion of data—as anyone who has had their phone wiped may readily attest to. Without appropriate data safeguards, entire records—and therefore individuals—might be made to "disappear" For instance, in healthcare, the use of electronic health records (EHRs) has raised concerns that loss of documentation integrity could compromise patient care, coordination, reporting and research, and even allow fraud and abuse.[16]
- **Easy copying of digital records:** When, in 1971, the so-called "Pentagon Papers" were stolen and leaked, one of the most significant obstacles was the physical copying and subsequent collation of some 7,000 pages. By contrast, in 2015, the so-called "Panama Papers" involved some 11.5 million digital files.[17]
- **Exposure of "hidden"-but-connected personal data:** Automatic data processing, as supported through AI and machine learning, makes possible discovery of vast arrays of patterns and other information, such as by connecting disparate information about a person from disparate sources or using metadata about individuals or groups.[18]

---

11 Gercke, Marco. 2007. "Internet-Related Identity Theft", Council of Europe Discussion Paper. https://rm.coe.int/16802fa3a0.

12 World Bank; United Nations. 2017. *Combatting Cybercrime: Tools and Capacity Building for Emerging Economies.* Washington, DC: World Bank Group. http://www.combattingcybercrime.org/.

13 Barber, Gregory. 2019. "San Francisco Bans Agency Use of Facial-Recognition Tech". Wired.com. https://www.wired.com/story/san-francisco-bans-use-facial-recognition-tech/.

14 *Ibid.*

15 Cameron F. Kerry. 2017. "Why protecting privacy is a losing game today—and how to change the game", Brookings Institute. https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/.

16 Arrowood, D, E Choate, E Curtis et al. 2013. "Integrity of the Healthcare Record: Best Practices for EHR Documentation", Journal of AHIMA. pp.58-62. https://library.ahima.org/doc?oid=300257#.XNt9so5JE2w.

17 See Chokshi, Niraj. 2017. "Behind the Race to Publish the Top-Secret Pentagon Papers", New York Times. https://www.nytimes.com/2017/12/20/us/pentagon-papers-post.html and Harding, Luke. 2016. "What are the Panama Papers? A guide to history's biggest data leak", The Guardian. https://www.theguardian.com/news/2016/apr/03/what-you-need-to-know-about-the-panama-papers.

18 *See, e.g.,* "Behind the Data: Investigating Metadata", Exposing the Invisible. https://exposingtheinvisible.org/guides/behind-the-data-metadata-investigations/.

Security benefits of going digital

At the same time that the digitalization of identification systems creates or augments certain risks, however, it also presents new opportunities and technological means for greater protection. Specifically, digital ID systems may offer:

- **More accurate identification and authentication.** As digital ID systems leverage computer processing and advanced technologies, they can offer a higher level of assurance and accuracy than manual, paper-based authentication processes that are subject to human error and discretion. Doing so increases trust, reduces costs, and supports sustainable, flexible systems.
- **Improved data integrity.** Although digital ID systems present new security risks they can—by adopting the data protection measures described above—also better assure the integrity and use of collected data compared with paper-based records systems that can be easily destroyed, damaged, or altered. Furthermore, automated, tamper-proof transaction logging provides an auditable records of data processing, thereby improving accountability and helping to address security breaches.
- **Better and more nuanced data privacy guarantees.** Digital technology enables new privacy-enhancing features that were previously not possible. In systems using non-digital credentials, transaction typically involves presenting a physical ID card to a service provider, and therefore revealing all the displayed information (e.g., presenting a physical credential as proof-of-age reveals additional information, such as full name, date-of-birth and, often, address). Digital technology can help resolve this issue through digital credentials that obscure or selectively present only the data necessary.
- **Increased agency and control.** New technologies and design strategies give individuals greater control over their personal data, including access portals that allow users to verify accuracy of their data and monitor data usage, and which automate data-breach notifications. Further, emerging digital ID ecosystems provide users with greater choice of ID providers.

Protecting data and privacy in a digital world

Data privacy and security measures should be integrated throughout the ID lifecycle—that is, data protection must become an organizational norm. This requires a "privacy-and-security-by-design approach"[19] that builds upon the following foundational principles:

1. ***Developing proactive—not reactive—systems*** that take a preventative not approach;
2. ***Making privacy the default setting***, rather than requiring affirmative action;
3. ***Embedding privacy into the technical design*** from the start rather than retrofitting it;
4. ***Construing privacy in a positive-sum manner*** ("win-win"), and not as a zero-sum ("either/or");
5. ***Developing end-to-end security*** with a view to full-lifecycle protection;
6. ***Building-in visibility and transparency*** and keeping systems open and accountable; and
7. ***Keeping the system user-centric***, with an eye to respecting user data privacy.

In practice, implementing a privacy-and-security-by-design approach require a series of complementary controls:

- **Legal controls,** including comprehensive legal and institutional frameworks safeguarding data and assuring user rights, especially their consent to use and control of personal data;
- **Management controls** for monitoring and oversight;
- **Operational controls** that promote security awareness, training, and detection; and
- **Technology controls** that limit and protect the processing of personal data and ensure the physical and virtual security of systems that process personal data.

19 First conceptualized by Ann Cavoukian as "Privacy by Design" or PbD. See Cavoukian, Ann. 2011. Privacy by Design. https://iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf.

Data protection begins with a comprehensive legal framework

A legal and institutional framework requires a series of interlocking instruments. To begin, data protection and privacy need to be enshrined in cross-cutting laws and principles, such as those enumerated in Box 2. These legal instruments should include explicit applications to ID systems and be policed by strong, high-capacity, institutional actors (e.g., data protection agencies). Larger policy instruments, such as national cybersecurity strategies, help to assure whole-of-government approaches and should apply across actors. Enabling laws should also be technology neutral, and not require legislative revision to adapt to technological progress.

Certain groups—such as ethnic, racial, or religious minorities—may also face particular concerns regarding the collection and use of data that indicates their group identity, and which could be used to profile or discriminate against them. Practitioners should carefully consider risks to these groups from collecting sensitive information and adopt sufficient legal and procedural protections against discrimination.

Box 2. Principles for processing personal data

According to the UN Personal Data Protection and Privacy Principles[20] personal data should be:

1. ***Processed in a fair and legitimate manner***, taking into account the person's consent and best interests, as well as larger legal bases.
2. Processed and retained ***consistent with specified purposes***, taking into account the balancing of relevant rights, freedoms and interests.
3. ***Proportional to the need,*** by being relevant, limited and adequate to what is necessary to the specified purposes.
4. ***Retained only for the time necessary*** for the specified purposes.
5. ***Kept accurate*** and up-to-date in order to fulfill the specified purposes.
6. Processed with due regard to ***confidentiality***.
7. Secured by ***appropriate safeguards*** (organizational, administrative, physical, technical) and procedures should be implemented to protect the security of personal data, including against or from unauthorized or accidental access, damage, loss or other risks presented by data processing.
8. Processed with ***transparency to the data subjects***, as appropriate and whenever possible.
9. ***Only transferred given appropriate protections*** to a third party.
10. ***Done accountably***, with adequate policies and mechanisms in place to adhere to these Principles.

Designing systems that implement data protection principles

While legal frameworks are vital to protecting personal data in ID systems, they must be put into practice with organizational, management, and technology safeguards. ID systems must translate laws and regulations into their technical and operating specifications, including limits on data collection and usage. This includes the use of operational controls—e.g., detailed operational manuals, staff training, physical and cybersecurity measures, etc.—and privacy-enhancing technologies (PETs). These technologies and controls work to implement privacy principles through various strategies, including minimizing data processing; hiding, separating, or aggregating personal data; informing individuals and giving them control over data use; and enforcing and demonstrating compliance with legal requirements (see Table 1).

20 Although developed specifically to govern data processing with UN organizations, these principles embody international good practice. See UN High-Level Committee on Management. 2018. *UN Personal Data Protection and Privacy Principles*.

Table 1. Examples of PETs and operational controls

	Strategy	Example solutions (not exhaustive)
Data-oriented	**Minimize** the collection and processing of personal data to limit the impact to privacy of the system	• Collecting and sharing minimal data • Anonymization and use of pseudonyms when data is processed
	Hide personal data and their interrelationships from plain view to achieve unlinkability and unobservability, minimizing potential abuse	• Encrypt data when stored or in transit • End-to-end encryption • Key management/key obfuscation • Anonymization and use of pseudonyms or tokenization for data processing • "Zero semantics" or randomly generated ID numbers • Attribute-based credentials (ABCs)
	Separate, compartmentalize, or distribute the processing of personal data whenever possible to achieve purpose limitation and avoid the ability to make complete profiles of individuals	• Tokenization or pseudonimization by sector • Logical and physical data separation (e.g., of biographic vs. biometrics) • Federated or decentralized verification
	Aggregate personal data to the highest-level possible when processing to restrict the amount of personal data that remains	• Anonymize data using k-anonymity, differential privacy and other techniques (e.g., aggregate data over time, reduce the granularity of location data, etc.)
Process-oriented	**Inform** individuals whenever their data is processed, for what purpose, and by which means	• Transaction notifications • Data breach notifications
	Give individuals tools to **control** the processing of their data and to implement data protection rights and improve the quality and accuracy of data	• User-centric identity services • Attribute-based credentials
	Enforce a privacy and data protection policy that complies with legal requirements	• Role-based access control with two-factor authentication • Remote access • Physical and cyber-security measures
	Demonstrate compliance with the privacy policy and applicable legal requirements	• Tamper-proof logs • Audits

Source: Table adapted from the ID4D Practitioner's Guide (www.id4d.worldbank.org/guide). Original framework adapted from https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design to fit the ID system context. Note: this table is meant to be illustrative of common privacy-enhancing technologies and operational controls, but it is not exhaustive.

In addition to designing systems with these operational and technical controls, practitioners should consider additional policy measures to identity and mitigate key data privacy and security risks, including:

- **Pro-active consultation and communication:** Frequent engagement with the public and civil society is crucial for identifying and mitigating data protection threats and building trust in the system. Practitioners should implement outreach and education campaigns early-on to consult with the public on privacy and data protection issues and ensure effective and transparent communication about the purpose and use of these systems and the protections they offer. Where threats or breaches are identified, they should be treated promptly and transparently.
- **Identifying risks to be mitigated through a privacy impact assessment (PIA):** Conducting a PIA is recommended to evaluate the impact of the ID system on personal privacy and data and articulate how various controls will help mitigate these risks.
- **Undertaking threat modeling exercises:** Before finalizing the design of an ID system and beginning procurement, practitioners should undertake a threat modeling exercise to assess potential internal and external threats throughout the identity lifecycle (see Table 16 for examples of potential vulnerability at different stages of data processing). This is crucial not only for the security of the system, but to ensure uptake—people are less likely to participate in an ID system if they fear that their data will be misused or mismanaged.

Protecting data is a cornerstone of good practices for ID

The ten *Principles on Identification for Sustainable Development*[21] (see Table 2) enshrine many of the legal, operational, and technical controls discussed above, such as the need to protect users' data privacy and assure their control of their personal data from the design stage (Principle 6); to ensure that data is accurate (Principle 3); to develop a comprehensive legal framework (Principle 8); and create mechanisms for independent oversight, grievance redress, and enforcement (Principle 10). Combined with other measures to ensure that ID systems are inclusive, designed in an interoperable and sustainable way, and meet the needs of a variety of users, proactive data protection measures are therefore essential for building ID systems that can meet development goals in the digital era.

Table 2. Principles for developing ID systems

PRINCIPLES	
INCLUSION: UNIVERSAL COVERAGE AND ACCESSIBILITY	**1.** Ensuring universal coverage for individuals from birth to death, free from discrimination. **2.** Removing barriers to access and usage and disparities in the availability of information and technology.
DESIGN: ROBUST, SECURE, RESPONSIVE AND SUSTAINABLE	**3.** Establishing a robust—unique, secure, and accurate—identity. **4.** Creating a platform that is interoperable and responsive to the needs of various users. **5.** Using open standards and ensuring vendor and technology neutrality. **6.** Protecting user privacy and control through system design **7.** Planning for financial and operational sustainability without compromising accessibility
GOVERNANCE: BUILDING TRUST BY PROTECTING PRIVACY AND USER RIGHTS	**8.** Safeguarding data privacy, security, and user rights through a comprehensive legal and regulatory framework. **9.** Establishing clear institutional mandates and accountability. **10.** Enforcing legal and trust frameworks through independent oversight and adjudication of grievances.

21 These Principles have now been endorsed by more than 20 organizations. See World Bank. 2017. *Principles on Identification for Sustainable Development: Toward the Digital Age*. Washington, DC: World Bank Group. http://id4d.worldbank.org/principles.

Conclusion

Digital ID systems can offer new possibilities for achieving sustainable development goals—*if they are inclusive and trustworthy*. When designed appropriately, digital ID systems can be more secure than analogue systems, with stronger, more intelligent, and more easily monitorable data protection measures, which in turn offer better guarantees of data privacy.

Taking advantage of these benefits, however, requires purposeful preventative action and an ongoing commitment to identifying and mitigating potential threats. This involves adopting a privacy-and-security-by-design approach from the beginning—not as an afterthought—starting with the development of a legal and institutional framework. This framework should provide extensive data protection and privacy guarantees through the application of international principles and effective oversight and be supported by comprehensive and complementary organizational and technical controls. Only by taking data protection seriously will digital ID systems live up to their transformative potential.

About ID4D

The World Bank Group's Identification for Development (ID4D) Initiative uses global knowledge and expertise across sectors to help countries realize the transformational potential of digital identification systems to achieve the Sustainable Development Goals. It operates across the World Bank Group with global practices and units working on digital development, social protection, health, financial inclusion, governance, gender, legal, and among others.

The mission of ID4D is to enable all people to access services and exercise their rights by increasing the number of people who have an official form of identification. ID4D makes this happen through its three pillars of work: thought leadership and analytics to generate evidence and fill knowledge gaps; global platforms and convening to amplify good practices, collaborate, and raise awareness; and country and regional engagement to provide financial and technical assistance for the implementation of inclusive and responsible digital identification systems that are integrated with civil registration.

The work of ID4D is made possible with support from the World Bank Group, Bill & Melinda Gates Foundation, the UK Government, the Australian Government and the Omidyar Network.

To find out more about ID4D, visit id4d.worldbank.org. To participate in the conversation on social media, use the hashtag #ID4D.