PULSE ON THE PRINCIPLES

Principles *for* Digital Development
PAPER SERIES

Responsible Data

When the Principles for Digital Development were created, the information and communications technology for development (ICT4D) community was primarily focused on the design of digital tools. As the field advanced, attention began to focus more on the data that is collected through those digital tools and devices as well.[1]

Massive amounts of data are produced by global actors. With progressively sophisticated data collection tools and technology, governments and nongovernmental organizations (NGOs) are building their digital capacities and digital ecosystems to support a wide range of services, including data tracking, transparency, and accountability purposes. Additionally, the increased use of mobile devices and the internet by people in low- and middle-income countries around the world has created growing volumes of data, and this has only increased since the arrival of the COVID-19 pandemic.

Not only has the amount of data increased, but its uses have become more sophisticated, and the ways of combining data sets have evolved. Likewise, social impact work has deepened our understanding of how data can lead to harm, especially to vulnerable people and groups.

Responsible data (RD) is about dealing with the unintended consequences of working with data and ensuring that no one is harmed in the process. According to the RD community, "Responsible data is about: 1) prioritizing people's rights to consent, privacy, security and ownership when using data in social change and advocacy efforts and 2) implementing values and practices of transparency and openness."[2]

Because data is such a critical aspect of digital development, the Principles for Digital Development must clearly address where data considerations can arise in their practical application. This paper sets out some of the areas where RD does and does not come into play regarding the Principles and makes some proposals for broadening our understanding of RD concerns.

ABOUT THE PULSE ON THE PRINCIPLES PAPER SERIES

The Digital Principles were created in 2008 by practitioners primarily working for Global South organizations on donor-funded digital projects taking place in the majority world. In the years since, the community of organizations endorsing the Principles has grown to include academic institutions, multilateral organizations, donors, NGOs, and private-sector companies. The Principles have become a core part of how we understand responsible digital development practice—as a standard, a capacity-building framework, a project design practice, and a curriculum. Yet the Principles don't cover everything we now know to be good practice, and their design focus at the project level can mean they miss important longer-term or systemic considerations. This series applies the nine Principles to different ethical considerations to help practitioners ensure that their work takes these critical areas into account and to encourage community conversations about the future development and improvement of the Principles.

[1] Klein, M. (2020). "The Digital Principles Are Rooted in Collaboration and Primed for Growth." https://digitalprinciples.org/cross-post-the-digital-principles-are-rooted-in-collaboration-and-primed-for-growth/

[2] See https://responsibledata.io/.

ABOUT THE DIGITAL PRINCIPLES

The Principles for Digital Development are nine living guidelines that are designed to help integrate best practices into technology-enabled programs and are intended to be updated and refined over time. They include guidance for every phase of the project lifecycle, and they are part of an ongoing effort among development practitioners to share knowledge and support continuous learning. The Digital Principles were created in a community-driven effort, the result of many lessons learned through the use of information and communication technologies (ICTs) in development projects.

Responsible Data and the Principles for Digital Development

Many of the Principles address the issue of responsible data in various ways. They advocate for keeping the best interests of end users and individuals in mind, and they emphasize the importance of data minimization, data security, and restricted data access to reduce the risk of data misuse, unauthorized sharing, and data breaches. They note that risks are contextual and that a risk-benefit analysis should always be conducted when appropriate. They also note the importance of transparency in terms of how data will be collected, used, shared, and stored. However, the Principles do not focus specifically on the ethics of collecting and using digital data.

RESPONSIBLE DATA AND THE PRINCIPLES AT A GLANCE

Design With the User

Design With the User focuses on designing technology with the user in mind. It does not address the importance of involving users and data subjects in decisions about the use of data collected via digital platforms and tools. Working with users on both the design and data collection aspects of a project can help practitioners identify potential risks and harms emanating from data during rollout, enable better-informed strategies to mitigate risk, and help data subjects make better decisions about whether or not to consent to the use of their data.

Understand the Existing Ecosystem

While *Understand the Existing Ecosystem* is a core part of responsible data, it does not currently highlight the need for context analysis in order to understand all aspects of the data ecosystem, particularly power dynamics. This includes socio-economic, cultural, legal, and political aspects of data use.

Build for Sustainability

Build for Sustainability focuses on programs, platforms, and digital tools, not building sustainable data systems. Creating a sustainable data system requires skills and capacities to collect and clean robust, relevant, and verifiable data sets; budgets, staff, and systems to maintain, retain, or destroy data; and a focus on local skills and knowledge to manage data over time.

Address Privacy & Security

Address Privacy & Security falls short because it is procedural guidance that misses the wider ethical implications of data collection and use. While a procedural approach to data protection is critical, an ethical approach helps answer bigger-picture questions in the design phase. It allows for the consideration of short- and long-term implications and tradeoffs implicit in the adoption of digital platforms and tools, as well as the collection of increasing amounts of data.

Building on Responsible Data in the Principles for Digital Development

Responsible data is a challenge for NGOs and government agencies because it is a complex concept that requires awareness, knowledge, capacity, and resources—things that these organizations often don't have. Responsible data collection and use requires weighing positives and negatives and making hard decisions after thoroughly exploring tradeoffs. Responsible data practices require those with greater levels of power to assume responsibility for the safety and well-being of less powerful individuals and groups whose data they are collecting and managing, particularly when the data could be used or managed in ways that could lead to risk or harm.

> Responsible data practices require those with greater levels of power to assume responsibility for the safety and well-being of less powerful individuals and groups whose data they are collecting.

When the Principles address data, they focus on procedural, legal, and operational data management and the need to be data driven. Responsible data goes beyond this to specifically cover the ethical aspects of data collection, management, use, and sharing; designing for safety and digital safeguarding; and understanding the power dynamics of data.[3] Below are some areas where we could expand on our current understanding of the Principles in practice to manage digital data in safer and more ethical ways that mitigate digital harm to vulnerable populations.

1. GO BEYOND ADDRESSING PROCEDURAL PRIVACY AND SECURITY

Collecting, processing, and holding data must be done responsibly. This is even more pressing when working with marginalized or vulnerable individuals and groups, since they have less power and recourse when data use leads to harm, or when data use, management, or mismanagement puts them at risk.

For the *Address Privacy & Security* principle to more fully cover a responsible data approach, it should go beyond legal and procedural data privacy and security. Whereas a legal framework considers, "What do we want to do, and can it be done legally with limited liability?" a responsible data approach asks, "Should this be done, regardless of whether or not it's legal?" Whereas a procedural privacy and security framing asks, "What do we want to do and how can we make sure it's secure and private?" a responsible data approach asks, "Should we do this at all, regardless of how private and secure it is?"

As part of a responsible data approach, practitioners should lay out the potential risks to the individuals and groups who will use their tools or platforms and whose data will be captured and then design for their safety. For data subjects to truly give informed consent, there must be a meaningful alternative to participation.

2. RESPONSIBLY REUSE DATA WHEN BEING DATA DRIVEN

Reuse and Improve and *Be Data Driven* are two principles that should be expanded to address responsible data. As new forms and sources of data emerge, as well as new uses for data, it's important to find ways to responsibly use and reuse existing data and new kinds of data for better decision-making and improved impact. We also need to ensure that we are not introducing harm when we access and use nontraditional data sources.

When accessing call detail records (CDRs), data harvested from social media, drone or satellite data, geolocation and mobility data, and data coming in through sensors, practitioners need to clearly assess the ethical implications, consent and ownership issues, and potential harms of these types of data. This also applies to the use of administrative data, open data, and data that is acquired from outside sources. These data sets can be of low validity and can carry a variety of biases that can lead to harm. They can also be non-representative of some people and groups, including the most

[3] The Engine Room's Responsible Data Forum https://www.theengineroom.org.

vulnerable.[4] Newer uses of data such as machine learning and automated decision-making can create risk or harm in the short or long term, due to the ways they enable profiling of certain individuals and groups and because they leave out certain individuals or groups, data sets, or context about the data.

It is critical that practitioners work to understand how bias enters data sets and to eliminate, reduce, or, at the very least, acknowledge bias in large data sets. They must then determine how to ensure that data will not be used in unjust or unethical ways or used to profile specific demographic groups in ways that lead to harm, oppression, or further inequity. This may include involving experts who can provide in-depth analysis of the potential benefits and harms resulting from data capture, use, or sharing when it involves advanced data techniques such as predictive analytics, use of historical or proxy data, or the combining of multiple data sets.[5]

Examples of private and secure data management that could still be problematic

- **Contact tracing.** In 2020, Apple and Google partnered on a COVID-19 contact tracing application as a way of controlling the spread of COVID-19 in the United States and potentially beyond. The application was considered technically adequate in terms of privacy and security. However, some privacy professionals determined that the potential efficacy of the application did not justify the amount and type of personal and sensitive data that the application would amass. Many also had concerns about potential scope creep of these troves of data and the potential for reuse of data or the technology for other purposes.[6]

  In 2020, Singapore introduced TraceTogether, a COVID-19 contact tracing application. The Singapore government assured citizens that COVID-19 data would "never be accessed unless the user tests positive" for the virus, and personal data would be substituted with a random permanent ID and stored on a secured server. However, in January 2021, the Singapore government confirmed local law enforcement will be able to access the data for criminal investigations. There is concern that citizens' location data is stored in a way that is harmful to their personal privacy, freedom of movement, and their right to free association. The government's plans to make the TraceTogether app mandatory for all citizens also raise issues about creating an extensive surveillance network.[7]

Questions that still need to be asked as governments tap tech to contain coronavirus

Some compromise in personal privacy has been deemed necessary in countries such as Singapore, Taiwan, and South Korea that have turned to technology to aid in contact tracing and movement monitoring, but there are questions citizens should still ask to protect their cyber wellbeing.

- **Digital identity.** The use of biometrics for digital identification has numerous benefits. Access to various services can be tied to a digital ID, aid agencies can reduce fraud and duplication using biometric digital ID, recipients of aid can reduce the time spent proving eligibility for benefits and services, and digital ID can enable smoother and more secure access to financial services. These services can be secure and private to the extent that the law requires, but concerns have been raised over data sharing of biometric or digital ID with private-sector companies by governments and international nongovernmental organizations (INGOs) without the involvement or consent of the individual. There are also concerns about UN agencies sharing refugee biometric data with country governments without the involvement or consent of the individual or the possibility of nonstate actors in fragile or conflict settings accessing these biometric identifiers. Both of these scenarios could result in individuals being targeted and harmed. Since digital ID, including biometrics, is increasingly the norm in many countries and is often mandatory for access to benefits and services in refugee settings, dealing with data security and privacy must be a priority.[8, 9]

[4] Noble, S. (2018). Algorithms of Oppression. New York University Press.

[5] D'Ignazio, C. and Klein, L. (2020). "The Numbers Don't Speak for Themselves." https://data-feminism.mitpress.mit.edu/pub/czq9dfs5/release/3

[6] Renieris, E. M. "When Privacy Meets a Pandemic." March 23, 2020. https://onezero.medium.com/when-privacy-meets-pandemic-fbf9154f80b3

[7] Yu, E. "Singapore police can access COVID-19 contact tracing data for criminal investigations." January 1, 2021. https://www.zdnet.com/article/singapore-police-can-access-covid-19-contact-tracing-data-for-criminal-investigations/

[8] The Engine Room. (2018). "Biometrics in the Humanitarian Sector." https://www.theengineroom.org/wp-content/uploads/2018/03/Engine-Room-Oxfam-Biometrics-Review.pdf

[9] Hersey, F. "Verify my life: Could a Nigerian problem lead to a global trust solution? (or Fuel a Two-Tiered Society)" August 5, 2020. Biometric Update. https://www.biometricupdate.com/202008/verify-my-life-could-a-nigerian-problem-lead-to-a-global-trust-solution

3. UNDERSTAND THE DATA ECOSYSTEM AND MITIGATE HARMS

While it is not possible to fully eliminate all risk, it is possible to manage it and mitigate its harmful impacts. The *Understand the Ecosystem* principle moves us toward this goal but could be taken further to urge practitioners to understand the data ecosystem and incorporate clear processes and mechanisms for managing risk associated with data collection and use. This could include the following:

- **Emphasize safety, safeguarding, and protection as central to the goals of our work.** A responsible data approach considers the big picture and how any initiative that involves ICT tools and data contributes to the well-being of those with whom we are working or whose data we are capturing, using, sharing, and retaining. Data that practitioners collect should never compromise the safety or well-being of the populations we engage with. Using a responsible data lens helps ensure that the well-being of individuals and groups is centered in physical, emotional, political, and cultural terms; that agency actions are not harming affected people or groups; and that our collection of data does not inadvertently lead to harm perpetrated by those within or outside of our organizations.
- **Assess the potential for harm, negative impact, or unintended consequences** in the short and long terms across the full data lifecycle, including whether the proposed data collection and use are necessary and proportionate to the expected benefits. A responsible data approach ensures that a diverse group of stakeholders, including representatives of the groups whose data is being collected, are involved in this assessment and that there is a meaningful alternative to participation. Responsible data also requires that mitigation plans to reduce the potential for harm to acceptable levels are designed and implemented. When potential harms to individuals or groups are greater than potential benefits and there are no clear mitigation strategies, data should not be collected.

4. ADDRESS RISKS FROM OPEN DATA

While there are tremendous benefits from working openly and using open source, in the case of open data, there are some risks that practitioners need to take into consideration if they are to be responsible with data use.

- **Consider the risks with open data and interoperable data sets.** Data interoperability is a goal that many organizations are working toward because it will allow them to unlock the massive potential of data use for improving development outcomes. Open data can reduce the burden of data collection on individuals and put data to better use with less cost. However, opening data can lead to greater risk of re-identification of individuals if precautions are not taken. Concerns have also emerged about the potential for re-identification of data that has been anonymized when large data sets are combined and as technology to re-identify individual records becomes more and more sophisticated. While the Principles do note that it's important to consider privacy and security when opening data, they could be expanded to offer more guidance, considering the technological advances that have emerged in the past several years that make it easier to re-identify data.
- **Assess and address potential risks to groups and individuals.** The Principles and most data legislation are focused on the privacy and security of *individual* or *personal* data. While this is extremely important, advances in data science, artificial intelligence, and machine learning can lead to entire groups being profiled and potentially excluded, targeted, or harmed based on data that indicates ethnicity, sexual orientation, gender identity and expression, political leanings, religion, citizenship or refugee status, geolocation, or other characteristics or behaviors.[10]

5. BE COLLABORATIVE WITH RESPONSIBLE PARTNERS

Partnership and collaboration are key aspects of the Principles. Therefore, it is imperative that our collaborators share our values when it comes to data responsibility. This could be spelled out more clearly in the *Be Collaborative* principle.

In the proposal or partnership development phase, the Principles could encourage us to conduct due diligence on platforms, data processors, and partners. This ought to include an assessment of

[10] OCHA. "Exploring the Mosaic Effect on HDX Data Sets," July 22, 2020. https://centre.humdata.org/exploring-the-mosaic-effect-on-hdx-datasets/

how a partner has used data in the past and whether it could have harmed vulnerable or historically marginalized groups or enabled data to be used by others for targeting, surveillance, or harm. Due diligence should also include whether the partner has had any critical data incidents and what its response was to any data breaches or the misuse or harmful use of data.

6. BE TRANSPARENT AND ACCOUNTABLE FOR DATA

The Principles focus on the design, inception, and implementation of digital development projects. They could more explicitly encourage building transparency and accountability into the design of interventions. For both ethical and legal reasons, practitioners need to inform people about why we collect their data and how we will use and share it. The Principles could also say more about the importance of data governance for ensuring responsible data, including policies, practices, participation, and decision-making processes related to data.

Practitioners collect digital data in both traditional, linear ways (e.g., an enumerator uses a digital device to collect data from someone, and that data is managed by one agency) and in less direct ways (e.g., various kinds of personal, sensitive, and behavior data, as well as metadata, are collected via online or mobile platforms, shared, and used by multiple organizations). Additionally, multiple organizations may have responsibility for a person or group's data in the current ecosystem. Individuals, groups, and entire communities may be unaware that their data is being collected and/or unclear about how and why their data is used, shared, and stored, and for how long.

Because digital data use is complex and it can be difficult to explain to people how their data will be used, current models of collecting consent are inadequate. Practitioners should inform and consult those whose data is being captured and used as much as they can. Additionally, they should research the lawful basis for data collection that provides the most transparency and accountability.[11]

Some key elements of transparent and accountable data governance include:

- **Define roles and responsibilities along the data lifecycle.** Different people and organizations will have specific roles in data collection or access, data analysis and use, data transmission and storage, data sharing, and data retention or destruction. These roles and responsibilities should be laid out in clear, written agreements that have signoff from the various parties. Agreements should include who makes decisions about these processes; who implements the decisions; and who can access, use, share, or otherwise make decisions about technology or data. These roles and responsibilities should be clear and transparent to stakeholders, with accessible mechanisms for raising complaints and consequences for not following what has been agreed upon.
- **Be transparent with users about what will happen with their data.** Agencies need to explain to users what will happen to their data along the full lifecycle and ensure that they have users' consent or have determined another lawful basis for transparently collecting and processing their data. Key aspects that need to be made transparent to users include:
    - What information is being collected, who is collecting it, and how it's being collected.
    - Why information is being collected, how it will be used, how long it will be retained, and with whom it will be shared.
    - What the effect will be on the individuals concerned and whether the intended use will likely cause individuals to object or complain.
    - What the users' rights are related to their data and whether communities can access and use the data.
    - Whether users can expect any feedback or response related to the data they have provided and when.
- **Ensure transparent and accessible mechanisms for complaints.** When agencies collect data, they need to transparently communicate how people can make complaints, withdraw their consent, correct the data being held about them, or remove their data completely from a system. Agencies should also be sure that they have a plan for how to manage and inform users, partners, and relevant authorities (if required) of any breach, loss, or unintended use of data.

[11] Hayes, B. and Marelli, M. "Facilitating innovation, ensuring protection: The ICRC Biometrics Policy," October 18, 2019. https://blogs.icrc.org/law-and-policy/2019/10/18/innovation-protection-icrc-biometrics-policy/

Join the discussion and debate

This is our current thinking on how the Principles and the guidance around them should evolve to address responsible data. However, we know we have our own blind spots and may be missing crucial perspectives. Therefore, we want to hear from you.

Specifically, we're seeking answers to four questions:

1. **How can we emphasize ethical collaboration** in the design phase and include due diligence on platforms, data processors, and partners?
2. **What are the best ways to conduct** risk-benefits and risk-harms assessments?
3. **How can we better hold accountable** those who capture, manage, share, and store data and ensure they take care and avoid potential harm?
4. **How can the sector stay up to date with new data approaches** to ensure that we don't introduce harm when we work with data?

WAYS TO ENGAGE

- **Listen:** Subscribe to our podcast, Pulse on the Principles, or join one of our webinars.
- **Comment:** Join the conversation on our Forum or on Twitter at @digiprinciples using #digitalprinciples.

These Principles are yours. We want them to evolve to be as useful as possible for you. This is an invitation to help us shape them.

Further resources and guidance

- Oxfam Great Britain. Responsible Data Management Training Pack
- UNICEF. Responsible Data for Children
- CaLP. Cash and Voucher Assistance Data Responsibility Guide (coming in January)
- USAID. Considerations for Using Data Responsibly
- Center for Humanitarian Data. Guidance Note: Humanitarian Data Ethics
- Girl Effect. Digital Safeguarding Tips and Guidance
- CARE. Responsible Data Maturity Model
- The Engine Room. Responsible Data Forum
- Center for Humanitarian Data. Guidance Note: Data Incident Management
- Global Pulse. Privacy Assessment Tool
- UK Information Commissioner's Office. Lawful Basis for Processing
- ICRC. Data Protection in Humanitarian Action Handbook
- Digital Empowerment Foundation. Data Rights for Communities
- CGIAR Platform for Big Data in Agriculture. Responsible Data Guidelines for Research
- Data Feminism. Preview — Data Feminism by Catherine D'Ignazio