WORKINGPAPER

Unlocking the Potential

The Call for an International Decade of Data

Dr Stefaan G. Verhulst[1]

---

[1] Dr Stefaan G. Verhulst is Co-Founder of The GovLab and The Data Tank, and Research Professor at New York University.

Introduction

Over the last year or so, global conversations about technology appear to have entered a new era. Since the launch of OpenAI's ChatGPT in November 2022, policymakers, entrepreneurs, and citizens have been preoccupied with the potential of artificial intelligence (AI), and generative AI in particular. AI has dominated the media and captured the public imagination. AI is clearly an important technology with tremendous transformative potential, both positive and negative. Without diminishing any of this potential, it is also clear that the ongoing conversation has often been driven by hype and lacked the necessary nuance.

Additionally, it is often overlooked that data forms the foundation of AI and other recent technological advancements. It is imperative to recognize that without a robust data foundation, the challenges associated with AI and other advanced technologies, including their limitations and unintended consequences, are likely to be exacerbated. A solid data foundation is not only essential for the development and performance of these technologies, but also plays a critical role in addressing the ethical, privacy, security, and other concerns they raise.

AI is undeniably a component and result of the data lifecycle, but it is essential to remember that it is just one piece of the larger puzzle. Data stewardship, quality, responsible re-use, and governance are all crucial aspects of the broader data lifecycle that significantly influence the effectiveness and impact of AI and other technological advancements, such as Internet of Things (and even the so-called Metaverse). Recognizing this holistic view is essential for responsible and sustainable innovation.

In this paper, we therefore call for a reset of the conversation. In particular, our goal is to reiterate the central importance of data – to AI in particular, but more generally to the landscape of digital technology.

What follows serves as a clarion call to the global community to prioritize and advance data as the bedrock for social and economic development, especially for the UN's Sustainable Development Goals (SDGs). We begin by recognizing the existence of significant remaining challenges related to data; encompassing issues of accessibility, distribution, divides, and asymmetries. In light of these challenges, and as we propel ourselves into an era increasingly dominated by AI and AI-related innovation, it becomes imperative to establish a more robust foundation for the stewardship of data, one that, for instance, embodies inclusivity, self-determination and responsibility.

Hence, we advocate for the creation of an **International Decade of Data (IDD),** an initiative aimed at solidifying this foundation globally and advancing our collective efforts towards data-driven progress.

Data, AI, and the SDGs

The SDGs represent a global commitment to address pressing challenges such as poverty, inequality, climate change, and health. Data plays a central and indispensable role in the pursuit of these goals; indeed, data can be characterized as the foundation of the SDGs.

High quality aggregated and disaggregated data is critical in many ways. It enables governments, organizations, and policymakers to identify areas of need, allocate resources efficiently, track outcomes, and hold a variety of stakeholders accountable. Data also helps in targeting interventions to vulnerable populations and regions, ensuring that no one is left behind. Furthermore, data-driven insights support evidence-based policymaking and facilitate international cooperation, for example by providing common metrics for assessing global progress toward sustainable development. For all these reasons, data should be considered not just a tool, but a fundamental catalyst for the SDGs.
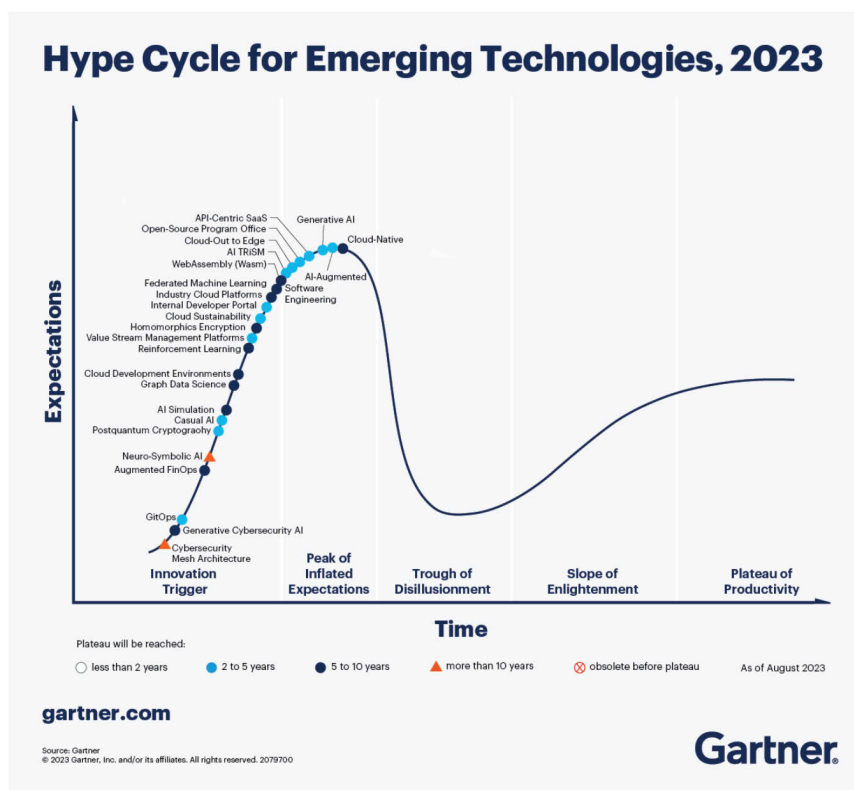
It is easy to lose sight of the centrality of data in the ongoing excitement over AI. The consulting firm Gartner has popularized the notion of a five-stage Hype Cycle that accompanies many new technologies (see Figure 1). In a recent report, it places generative AI at Stage 2 ('Peak of Inflated Expectations'), or at the height of this cycle. One consequence is that AI now dominates a large number of discussions over technology and the public good, including those involving the SDGs. For example, at a recent event in New York, 'Artificial Intelligence for Accelerating Progress on Sustainable Development Goals,' several speakers predicted that "AI could advance progress on nearly 80 percent of the SDGs and their targets."

Of course, such predictions may not be incorrect: AI does have tremendous potential to advance the SDGs. But it is important not to overlook the reality that AI, like most emerging and cutting-edge technologies, is itself fundamentally reliant on a bedrock of data. Indeed, it is precisely the widespread generation of re-usable data that has enabled machine learning, natural language processing (NLP), large language models (LLM), and other elements that are at the core of modern generative AI.

Despite the importance of data in achieving the SDGs and advancing AI and other technologies, there are significant challenges. To name a few:

- **Data inaccessibility and poor distribution:** A significant challenge is the inaccessibility of data or its poor distribution. Many critical datasets are locked behind proprietary systems, limiting access for researchers, policymakers, and organizations that could leverage this data for social and environmental progress.
- **Data divides and asymmetries:** There are vast disparities in data generation, availability, and quality across regions and communities. Developed countries often have better data infrastructure, while developing nations may lack essential data collection and analysis capabilities. This leads to data divides and asymmetries that hinder global progress.

All this requires:

- **Increased data responsibility and innovation:** To bridge these gaps, there is a need for increased data responsibility and innovation. Data holders, whether governments or private entities, should prioritize responsible data sharing and open data initiatives, making essential datasets available to the public.
- **Inclusivity and agency:** A key goal should be to ensure inclusivity and empower individuals and communities to have agency over their data. This involves respecting human rights and providing mechanisms for people, communities, and organizations to determine how their data is used and shared.
- **Equitable distribution of benefits:** To promote fairness, it's essential to ensure that the benefits of data-driven technologies and AI are distributed equitably. This includes ensuring that marginalized communities worldwide also benefit from technological advancements.
- **Community engagement:** Engaging communities and stakeholders in data collection and decision-making processes can help address local needs and preferences, ensuring that data-driven initiatives are tailored to the specific challenges faced by different groups.

It is for all these reasons that we are now calling for an International Decade for Data (IDD) – one that would reaffirm the centrality of data to sustainable development and technological advances and make progress toward responsible and inclusive data access and re-use to meet the SDGs and other societal challenges.

Priorities for An International Decade for Data

A key task in operationalizing the IDD is to set priorities and areas of focus. Priorities are always important in policymaking because resources—time, financial, and human—are limited. Furthermore, the intersection of data and the SDGs occupies a vast surface area. In order to derive maximum benefit from the proposed IDD, it is essential to identify those mechanisms and catalysts that are most resource-effective, and most likely to drive positive transformation with a minimum of negative externalities.

In the section below, we suggest some possible priority areas. These can form the basis for more specific areas of action. We offer these areas in a spirit of suggestion, rather than prescription. In order to determine actual areas of focus, we call for an inclusive, participatory process that would consider various options and select those most likely to be inclusive and impactful in a resource-effective manner. As we explain further, such a process would highlight multi-stakeholder engagement, and would ensure the adequate representation of traditionally marginalized voices and regions, especially the Global South, which is disproportionately affected by the SDGs and their implementation.

Priority Area: Upholding Human Rights in the Data Age

In the context of the IDD, the prioritization of upholding human rights should be a foundational imperative. Human rights, encompassing principles such as the right to privacy, freedom of expression, non-discrimination and equality, right to information, right to remediation and accountability, right to development, right to participation in government, and the rights of the child, should be promoted as pivotal cornerstones in the pursuit of using data to meet the SDGs. It is imperative that any framework devised for harnessing the potential of data in advancing the SDGs places these rights at its core. The IDD must champion values of transparency, inclusivity, and commitment to ethical, fair, and lawful data management practices across the entire spectrum of the data value chain, ranging from data collection, handling, and storage, to its responsible usage.

High-Level Action Areas

- **Advancing human rights-centric data governance**: Develop and champion global standards and best practices that prioritize human rights, including the right to privacy, in data governance.
- **Combatting bias and discrimination:** Encourage fair and unbiased data handling to ensure equal treatment for all, regardless of protected characteristics such as race or gender.
- **Facilitating the right to information:** Support open data initiatives that make government and organizational data accessible to the public, promoting transparency and accountability.

- **Ensuring remediation and accountability:** Establish mechanisms to enable individuals and communities to seek remediation when their rights are violated through data misuse or flawed data-handling practices.
- **Equitable data use for development:** Promote equitable access to, and use of data as a resource for social and economic development, aligning with the right to development.
- **Protecting children's rights in data practices:** Develop principles and guidelines for responsible use of data about and for children.

Priority Area: Establishing a Mutual Commitment Framework for Data Access during Times of Crisis

In an interconnected world facing increasing crises, harnessing data-driven solutions is essential for achieving rapid disaster response. Yet, the potential of data, especially from private sector sources, remains constrained by barriers to access and reuse. Data collaboratives, which enable responsible data sharing among diverse stakeholders, offer a solution.

However, setting up data collaboratives is often time-consuming and fraught with delays, hindering rapid responses to global crises. Obstacles include data use agreements, technical infrastructure, methodology, and clarity on data's potential. Creating a Mutual Commitment Framework (MCF) before crises strike is a practical solution to reduce response times and accelerate data collaborative establishment.

High-Level Action Areas

- **Collectively identify possible use cases** where timely and comprehensive data access is crucial, and engage stakeholders, including data providers, users, governmental bodies, and civil society organizations.
- **Conduct a comprehensive needs assessment** to understand data sharing and processing requirements at times of crisis.
- **Draft a framework** that outlines baseline commitments and guidelines for data collaboration.
- **Solicit feedback** from stakeholders through consultation sessions and workshops.
- **Pilot test the framework** with a diverse group of stakeholders during a real-world use case.

Priority Area: Professionalizing Data Stewardship

In an era where data is increasingly recognized as a valuable asset, it has become evident that the true potential of this resource lies in its responsible and collaborative use. The emergence of data collaboratives, partnerships that open up access to corporate data and expertise for the broader public interest, holds immense promise. These collaborations have the potential to drive innovation, unlock the public good potential of previously isolated datasets, and contribute significantly to societal advancement and the SDGs.

However, the success of data collaboratives is not guaranteed. Through extensive research and analysis of numerous such initiatives, one crucial factor has emerged as the linchpin of their success: the presence of dedicated individuals or teams within data-holding organizations who possess the authority and initiative to proactively initiate, facilitate, and coordinate these collaborations in the public interest. These champions of responsible data use are what we refer to as 'data stewards.'

Data stewardship plays a pivotal role in systematically partnering with external parties, scaling promising data initiatives, and ensuring that data is leveraged effectively and ethically for societal benefit. They bridge the gap between data holders and data users, forming a critical link in the data value chain.

Recognizing the significance of data stewards, the European Commission's 'High-Level Expert Group on Business-to-Government (B2G) Data Sharing' has emphasized their importance. In their final report, they note: "A key success factor in setting up sustainable and responsible B2G partnerships is the existence, within both public- and private-sector organizations, of individuals or teams that are empowered to proactively initiate, facilitate and coordinate B2G data sharing when necessary. As such, 'data stewards' should become a recognized function."

Furthermore, the report acknowledges the need to establish a network or community of practice centered around data stewardship. This network would serve as a platform for sharing knowledge, best practices, and experiences, further enhancing the effectiveness of data stewards in their mission to unlock the potential of data for the greater good.

High-Level Action Areas

- **Develop guidelines and best practices for data stewardship:** We must create comprehensive guidelines and best practices that define the role and responsibilities of data stewards. These guidelines will serve as a roadmap for individuals and organizations looking to embrace data stewardship.
- **Develop curriculum and training materials:** Equipping future data stewards with the necessary skills and knowledge is crucial. We must invest in the development of curriculum and training materials to educate and empower these champions of responsible data use.
- **Establish an international network of data stewards:** Building on the European Commission's recommendation, we should work towards establishing a global network or community of practice for data stewards. This network will facilitate cross-border collaboration, knowledge exchange, and the collective pursuit of responsible data use on a global scale.

Priority Area: Advancing Data Accessibility through Synthetic Data and Privacy-Enhancing Technologies

One crucial priority is to advance data accessibility through innovative approaches such as synthetic data, sandboxes, and privacy-enhancing technologies. These technologies are pivotal in addressing the challenges of data scarcity, privacy concerns, and inclusivity in data-driven initiatives aimed at achieving the SDGs.

In particular, synthetic data i.e. artificially generated data that replicates the statistical properties and patterns of real-world data while safeguarding privacy and confidentiality, should be considered in more depth. This approach allows for the creation of datasets that mimic real data without the associated privacy risks.

Embracing synthetic data as a priority area in the IDD can revolutionize the data landscape in several ways, such as:

- **Addressing data scarcity:** Synthetic data helps in overcoming data scarcity, which is particularly pronounced in regions like the Global South.
- **Privacy preservation:** Generating synthetic data can help preserve privacy as it enables the development of AI and analytics projects without compromising individual privacy.
- **Bias mitigation:** Synthetic data can be utilized to address bias issues in existing datasets, leading to better representation of populations and more equitable outcomes.
- **Cost-effectiveness:** Generating synthetic data is often less expensive compared to collecting large datasets.
- **Supporting 'Open Data' initiatives:** Synthetic data can help facilitate access to publicly available online information that can be used for any purpose at little or no cost.

High-Level Action Areas

- **Research and development:** Invest in research and development efforts to enhance the generation and utilization of synthetic data.
- **Data democratization:** Promote the use of synthetic data to democratize access to relevant information and models.
- **Capacity-building:** Offer training and capacity-building programs to educate stakeholders about synthetic data generation, utilization, and other privacy preserving tools and sandbox practices.
- **Governance frameworks:** Develop governance frameworks and guidelines for the responsible use of synthetic data.

Priority Area: Digital Self-Determination

One of the paramount priorities for the next decade is the realization and operationalization of Digital Self-Determination (DSD). This principle underscores the vital role that data subjects

should play in managing their information in ways that benefit themselves and society. In an era where data reuse is increasingly necessary, the implementation of DSD becomes pivotal in building the social license required for broader and more responsible data reuse.

International laws that champion individual freedom and human agency are founded on the belief that humans should autonomously make decisions that affect their life choices. However, these principles seem inadequately operationalized in the digital landscape. The increasing data and agency disparities challenge the tenets of self-determination, which advocate for citizens and communities as active decision-makers.

These emerging challenges, combined with the untapped potential of responsibly reused data for public benefit, necessitate innovative strategies for data and digital technology management. A human-centric approach, where citizen and community empowerment in the digital space forms the basis of data governance, is essential. This approach should manifest at individual, collective, and societal levels, transcending geographical boundaries.

In response to these digital tensions, the principle of DSD emerges as a solution. DSD seeks to empower individuals, communities, and organizations by extending the traditional notion of self-determination to acknowledge the fundamental role of agency in the digital era. DSD surpasses conventional notions of sovereignty and ownership, recognizing that genuine digital empowerment goes beyond mere consent mechanisms; it involves trustworthy and responsible data relationships. DSD is not only about preventing the misuse of data but also about harnessing its full potential.

High-Level Action Areas

- Develop foundational and operational frameworks for digital self-determination, including processes, policies, professions, and infrastructures.
- Operationalize DSD through real-world use cases that illustrate its value and dimensions in everyday life.
- Formulate international guidelines to support trustworthy data innovations based on DSD principles.
- Facilitate dialogue with diverse stakeholders to promote international adoption of DSD principles.
- Develop a platform to maintain a global dialogue on DSD, foster global advocacy, and represent DSD in international events.

The International Decade of Data and the United Nations

Given the centrality of data to the SDGs, it is essential to have the United Nations drive the IDD.

Initiate and Announce the International Decade Data

1. **Establish a multi-stakeholder 'IDD Task Force':** Begin by forming a dedicated Task Force under the auspices of the UN Secretary General for the IDD, consisting of experts, UN agency representatives, policymakers, and stakeholders from civil society, academia, and the private sector. This Task Force could provide the necessary leadership and guidance within the UN system and facilitate broad public engagement to determine priority areas.

2. **Engage UN Member States:** Initiate a coordinated outreach effort to engage UN Member States in the IDD process. Encourage countries to appoint national focal points for the IDD within their diplomatic missions to the UN.

Connect and Align with Existing Mandates, Efforts and Actors

3. **Align with UN mandates:** Define the objectives and scope of the IDD to align explicitly with the United Nations' overarching mandates, including the 2030 Agenda for Sustainable Development and the SDGs.

4. **Global Digital Compact integration:** Emphasize the IDD's synergy with the Global Digital Compact, recognizing that responsible data governance is a fundamental pillar of a sustainable and equitable digital future. Promote the integration of IDD principles and initiatives into the Compact's framework.

5. **Global Digital Compact signatories:** Encourage signatories of the Global Digital Compact to explicitly include IDD-related commitments in their Compact pledges. Ensure that responsible data management, equitable data access, and ethical AI principles are woven into the fabric of the Compact's principles.

6. **Summit of the Future agenda inclusion:** Include the IDD as a prominent item on the agenda for the forthcoming Summit of the Future. Highlight the critical role of data in shaping the future of society, the economy, and sustainable development. Stress that responsible data practices are central to addressing global challenges and achieving the SDGs.

7. **High-level IDD session:** Organize a dedicated high-level session within the Summit of the Future, specifically focused on the IDD. Invite world leaders, UN officials, data experts, and representatives from diverse sectors to discuss the IDD's significance and its alignment with the global digital future.

8. **IDD commitment pledges:** Encourage governments and stakeholders attending the Summit of the Future to make formal commitment pledges to support the IDD. These commitments should outline concrete actions, resources, and timelines for advancing responsible data practices and leveraging data for the SDGs.

Leverage Existing Mechanisms

9. **UN consultative mechanisms:** Leverage existing UN consultative mechanisms, such as the High-Level Political Forum on Sustainable Development (HLPF), to gather input and commitments from Member States regarding their contributions to the IDD.

10. **UN working groups:** Engage and leverage existing UN working groups that are focused on data governance, responsible data use, and AI ethics.
11. **Integration with UN programmes:** Integrate IDD initiatives with existing UN programmes and agencies, such as the UN Global Pulse, UNICEF, UNDP, UN-OCHA, UN DESA, and others, to maximize synergy and avoid duplication of efforts.

Assure Funding and Provide Assistance

12. **UN funding and resources:** Mobilize resources and funding from UN Member States, international organizations, and UN agencies to support the IDD's implementation. Ensure these resources are channeled efficiently and transparently.
13. **Technical assistance and capacity-building:** Promote technical assistance and capacity-building programmes tailored to the needs of UN Member States, especially those in the Global South, to enhance their data capabilities and compliance with responsible data practices.
14. **UN Member State commitments:** Encourage UN Member States to make formal commitments to the IDD and incorporate IDD principles and initiatives into their national development agendas.

Conclusion

We live in a paradoxical world of both unprecedented opportunity and unparalleled risk. Data is in many ways at the core of this paradox: it has the potential to catalyse rapid progress toward the SDGs, or, if handled poorly, could increase social divisions and marginalization, and cause or exacerbate a number of other harms. How we approach data – the frameworks we use to collect and use it – is therefore central to defining the parameters of our collective future. Virtually every important social, economic, and political challenge we face – from climate change to inequality to war and the risk of pandemics – intersects with data.

Given all of this, we reiterate the need for a proposed International Decade for Data – a decade that would reaffirm the importance of data for policymakers and other stakeholders, highlight the importance of approaching data through a responsible and inclusive framework, and in doing so, accelerate progress toward the SDGs and a variety of other public goods.

Technology moves fast. We call on the global community to act promptly in creating this International Decade for Data, and in establishing its core priorities. We cannot afford to wait; hanging in the balance is a world where data drives socioeconomic progress, ensures equity and self-determination, and upholds human rights.

This is a world well worth striving for.