

Insight Report

Data Collaboration for the Common Good

Enabling Trust and Innovation Through Public-Private Partnerships

Produced in Collaboration with McKinsey & Company

April 2019



World Economic Forum
91-93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland
Tel.: +41 (0)22 869 1212
Fax: +41 (0)22 786 2744
Email: contact@weforum.org
www.weforum.org

© Creative Commons License
This work is licensed under a Creative Commons
Attribution-NonCommercial 4.0 International
License.

© 2019 World Economic Forum. All rights
reserved. No part of this publication may be
reproduced or transmitted in any form or by any
means, including photocopying and recording, or
by any information storage and retrieval system.

Contents

Letter from the Forum	5
Letter from the Community	6
Executive summary	7
Section 1: Understanding the landscape	8
Strengthening trust to achieve impact	9
Section 2: Establishing a new lens	12
1. Stakeholder alignment	12
2. Responsible data governance	14
3. Insight generation and validation	16
4. Insight adoption	19
5. Economic sustainability and scalability	21
Section 3: Catalysing impact	24
Appendix: Fit-for-purpose technology archetypes	26
Acknowledgements	28
Endnotes	30

Letter from the Forum



Derek O'Halloran, Head, Future of Digital Economy and Society, Member of the Executive Committee, World Economic Forum



Anne Toth, Head of Data Policy, World Economic Forum

As we look at the diversity across the many stakeholders of the Fourth Industrial Revolution, three core and interdependent concepts about what we want our future digital environment to look like emerge. Our digital future must be inclusive, trustworthy and sustainable.

The following report looks at the manner in which private-sector data can be used, in combination with public-sector data, for this vision of a shared digital future. It provides a new lens to better understand how we can balance the need to maximize the power of data to deliver transformative change while at the same time protecting against emerging risks that could cause irreversible harms to vulnerable populations.

As the report demonstrates, there is now abundant evidence that public-private data collaboration can deliver a measurable impact. The results can be seen in faster decision-making during natural disasters and disease outbreaks, better insights on addressing the complex challenges related to poverty, health and employment and more precise indicators to measure the achievement of the Sustainable Development Goals (SDGs).

Yet with this encouraging progress is a troubling deficit of trust. Sadly, the trust among and between individuals, institutions and industry is at a low ebb. This systemic lack of trust threatens to derail the innovative possibilities of using private sector data for the common good. While there are material concerns related to the technical and operational dimensions of public-private data collaboration, strengthening trust across the entire data collaborative life cycle is the top priority to catalyse meaningful progress.

This is a challenge of maintaining balance between competing concerns. While the importance of protecting against the risks of data misuse cannot be overstated and is very much at the centre of the public debate right now, the avoidable harms that result from the non-use of data also need to be accounted for in the cost-benefit analysis. The cost of poor decision-making due to insufficient and inaccurate data can result in the loss of human life and property.

We hope you enjoy reading this report and find the tools and approaches it advances to be both pragmatic and valuable.

Letter from the Community



Chang-Gyu Hwang, Chairman and Chief Executive Officer, KT Corporation

Artificial intelligence (AI), machine learning, and blockchain – these are just a few of the technologic advances that have transformed how we collect and analyse data over the past few years. In the private sector, corporations have quickly incorporated these innovations into their daily work and even embarked on data collaborations with other companies. By combining diverse datasets and by sharing data insights, each entity gains access to information that complements or adds to its own research. With better source material, the resulting insights about markets, customers and competitive strategies become even more powerful.

But corporations are not just gaining a commercial edge through their collaborations – they’re also supporting the common good. Already, businesses across sectors have entered collaborations aimed at improving disaster relief, driving economic development in emerging markets and increasing public health. While these partnerships have delivered impressive gains, they could benefit further from more consistent engagement with the public sector. There have been several successful collaborations involving government agencies but nowhere near the volume needed to unlock the full value these partnerships could represent. The powerful combination of public- and private-sector data could help resolve some of the most pressing and long-standing social issues, including the problems that the UN’s SDGs are designed to tackle.



JoAnn Stonier, Chief Data Officer, Mastercard

Many corporate and government leaders refrain from sharing their information due to a trust deficit across institutional boundaries as well as the challenges of ensuring the privacy and security of information. Some fear that potential partners will misuse the information or share it in ways that violate confidentiality; compounding the problem, rules governing data collaborations are often absent or uncertain. Without appropriate guidance, some leaders believe they have little recourse if their counterpart steps out of line.

Although it may be tempting to say that data collaborations are too risky, the cost of inaction is immense. If companies and governments keep working in isolation, they will make slow progress in eradicating poverty and other social ills. That’s not a situation society can afford to face. Instead, it’s time for leaders across the public and private sector to co-create a policy and data governance framework that strengthens trust and data practices in more pragmatic and sustainable ways, while encouraging a healthy dose of innovation. By “going slow to go fast” we can collectively improve how organizations use data collaboration for both the common good and commercial gain – and that’s a win for us all.



Nicolaus Henke, Global Leader of Digital and Analytics and Chairman of QuantumBlack, McKinsey and Company

Executive summary

As the digital technologies of the Fourth Industrial Revolution continue to drive change throughout all sectors of the global economy, a unique moment exists to create a more inclusive, innovative and resilient society. Central to this change is the use of data. It is abundantly available but if improperly used will be the source of dangerous and unwelcome results.

When data is shared, linked and combined across sectoral and institutional boundaries, a multiplier effect occurs. Connecting one bit with another unlocks new insights and understandings that often weren't anticipated.

Yet, due to commercial limits and liabilities, the full value of data is often unrealized. This is particularly true when it comes to using data for the common good. While public-private data collaborations represent an unprecedented opportunity to address some of the world's most urgent and complex challenges, they have generally been small and limited in impact. An entangled set of legal, technical, social, ethical and commercial risks have created an environment where the incentives for innovation have stalled. Additionally, the widening lack of trust among individuals and institutions creates even more uncertainty. After nearly a decade of anticipation on the promise of public-private data collaboration – with relatively few examples of success at global scale – a pivotal moment has arrived to encourage progress and move forward.

In response, the World Economic Forum's Trustworthy Data Initiative has spearheaded an in-depth exploration of the contributing factors for catalysing progress in the domain of public-private data collaboration. Focusing on the multidimensional challenge of strengthening trust, a diverse community of commercial, government, academic and civil society leaders have participated in a series of global workshops and summits. The resulting outcome is a pragmatic framework for balancing two competing concerns: the imperative to innovate and the need to protect against emerging risks.

The following report reflects the synthesis of an in-depth review of case studies, expert interviews and global workshops with prominent members of the practitioner community. They point to the need for a more holistic, iterative and outcome-based understanding of public-private data collaboration. The findings point to five areas for leaders to focus upon to strengthen trust. Leaders need to: 1) ensure that all relevant stakeholders are committed to shared outcomes; 2) operationalize the principles of responsible data governance; 3) deliver insights that are achievable, accurate, fair and explainable; 4) support both senior leader decision-makers and front-line users with the skills and resources to use data; and 5) establish sustainable economics to ensure long-term impact.

Figure 1: Critical enablers of public-private data collaboration



Section 1: Understanding the landscape

As the technologies of the Fourth Industrial Revolution continue to evolve, the role of data has become indisputable. Described as the “lifeblood” of the 21st-century economy, rapid innovations across the data life cycle have created an unprecedented moment to turn data into meaningful insight.



Data holds great promise as a transformative resource for social good.



JoAnn Stonier, Mastercard

“With the adoption of sound and ethical data practices, all actors – governments, the corporate sector, university researchers and non-governmental organizations (NGOs) – can voluntarily work together to remove barriers and increase incentives to unlock the full power of data through sharing and analysis,” notes JoAnn Stonier, Chief Data Officer, Mastercard.

Yet, with the promise of public-private data collaboration also come deeply held concerns on the significant risks of misuse. Data is a dual-use technology. It can be used both as a way to address global challenges and as a means to heighten existing inequities. Progress will demand balance. Ensuring that risks are fully managed while supporting collaborative innovation requires new frameworks that are holistic, agile and pragmatic.

Organizations are now linking and connecting diverse datasets at an accelerating pace to create value – and this is one of the primary factors shaping today’s global economy. From 2017 to 2019, the number of companies forming data-related partnerships rose from 21% to 40%. A growing number of business competitors are also deciding to connect their data – rising from 7% to 17%.¹ Overall, McKinsey estimates that connecting data across institutional and geographic boundaries could create roughly \$3 trillion annually in economic value by 2020.²

As it relates to applying commercial data-sharing practices in the context of humanitarian and development challenges (in areas such as poverty, public health, environment and sustainable agriculture), there is also growing momentum. Collaboratively leveraging private-sector data has been widely recognized as an essential element for achieving the 2030 SDGs.³ “Data collaboratives represent a new form of multilateral collaboration where participants from different sectors – including private companies, research institutions and government agencies – can exchange data to help solve public problems,” says Stefaan Verhulst, co-founder and Chief Research and Development Officer of the Governance Laboratory. “They will be essential vehicles for harnessing the vast stores of privately held data towards the public good.”

Given the wide range of ways that data collaboratives can take shape, the scale of their intended outcomes and the types of data involved, this report will focus primarily on collaborative efforts at the global level. Its general focus examines the entangled set of technical, operational and governance challenges in accessing large-scale, commercially controlled personal data for the common good.

From a stakeholder perspective, this report will primarily focus on the commercial entities that function as the data holders from the supply side (particularly from the mobile network operator, financial services, healthcare and social media sectors) and from the demand side the needs of large-scale international organizations and the United Nations System. In this light, it is important to note that the report’s focus on accessing large-scale private-sector data is a promising yet relatively new area of discussion. As it relates to the global development and humanitarian sectors, an equally daunting set of challenges lies in bringing together small amounts of non-standardized “public-to-public” data from dozens of NGOs and organizations to create a common picture of needs and responses.⁴

A review of more than 200 use-cases conducted by the UN has identified three clusters of public-private data collaboration activity:⁵

- **Humanitarian action and crisis response.** These collaborations support activities related to the prediction, preparation, prevention, response and recovery from natural and human-made disasters and crises. For instance, the combined analysis of social-media trends, telecommunications data and proprietary records from humanitarian groups can facilitate the generation of more accurate, robust and actionable insights to facilitate more effective relief efforts in natural disasters or crises.
- **Global development.** These collaborations involve researching long-term social, economic and political issues. They provide insights into citizens’ behaviour that allow stakeholders to establish more effective programmes and make better decisions. Also, cross-analysing seemingly disparate datasets from multiple sectors and industries can help uncover hidden insights and patterns critical to solving highly complex challenges.
- **Official statistics.** These collaborations create indicators and measurements that can serve as a proxy for more traditional data-collection methods to support national statistical offices. This, in turn, can help macro policy-makers monitor and evaluate the impact of their policies, allow for more accurate and timely indicators and support the work of local, national and international statistical agencies with a focus on achieving the SDGs.

Figure 2: The various types of data collaboratives

Through its **Data Cycle** analysis, the GovLab has identified more than 150 examples of data collaboratives on a global basis. Listed below are the various forms they may take:

- *Data cooperatives*. Corporations and other important data holders group together to link and connect data resources.
- *Prizes and challenges*. Corporations make data available to qualified applicants who compete to develop new apps or discover innovative uses for the data.
- *Research partnerships*. Corporations share data with universities and other academic organizations, giving researchers access to consumer datasets and other sources of data to analyse social trends.
- *Intelligence products*. Shared (often aggregated) corporate data is used to build a tool, dashboard, report, app or another technical device to support a public or humanitarian objective.
- *Application Programming Interfaces (APIs)*. APIs allow developers and others to access data for testing, product development and data analytics.
- *Trusted intermediary*. Corporations share data with a limited number of known partners. Companies generally share data with these entities for data analysis and modelling, as well as other value chain activities.

Strengthening trust to achieve impact

While growing evidence shows the value of public-private data collaboration, the challenges and risks remain daunting. As the Global Partnership on Sustainable Development Data notes, “access to data remains a great challenge due to real or perceived barriers”.⁶ Interconnected issues related to security, privacy, commercial risk, cross-border data flows, reputational concerns, due process and regulatory uncertainty all serve to create an environment that operates at a slow and deliberate pace.

Underlying these concerns is a profound and widening lack of trust among individuals, institutions and governments. The 2019 Edelman Trust Barometer points to the plummet in trust in a variety of ways.⁷ The cost of damage caused by hackers, malware and data breaches is projected to double from \$3 trillion to \$6 trillion by 2021.⁸ The growing public outcry against the private sector on the misuse of personal data (as well as concerns on the limitations of industry self-regulation) have supported the call for more effective and meaningful privacy and data-protection regulations globally.⁹ The rise of the “Data Justice” movement – which seeks to ensure fairness in the way people are made visible, represented and treated in the production of digital data – continues to gain momentum.¹⁰

Addressing the trust deficit has become a top priority at the most senior levels. Prime Minister Shinzō Abe of Japan has called for data governance to be a vital pillar of discussion at

the 2019 G20 Summit. The UN Secretary-General has warned about the current “trust deficit disorder” and has highlighted – through the Secretary-General’s High-Level Panel on Digital Cooperation – the need for greater trust and the importance of shared vocabularies to address this challenge.¹¹

Strengthening trust will require a number of coordinated actions related to the economics, operations and governance of public-private data collaborations. Without this cooperation, they will face difficulty balancing tensions between the need to protect data and the opportunities to innovate in its use. At the forefront of this governance challenge is the need for global frameworks that harmonize the requirement for local data-protection regulation with the need for data innovation at scale. Balancing these competing concerns, while navigating an evolving global data-policy landscape, will require approaches that are agile, interoperable and iterative.

Ensuring leaders have the appetite and patience to continually iterate and “fail fast” will also be vital for building sustainable and trustworthy data collaboratives. As the 2019 Edelman Trust Barometer demonstrates, people are shifting their trust to relationships within their control, most notably their employers.¹² The 2019 global survey notes the increasing internal pressures for chief executive officers to actively engage and take actions that both increase profits and improve socioeconomic conditions in the community in which they operate.¹³

Figure 3: The changing data life cycle and its impact on policy

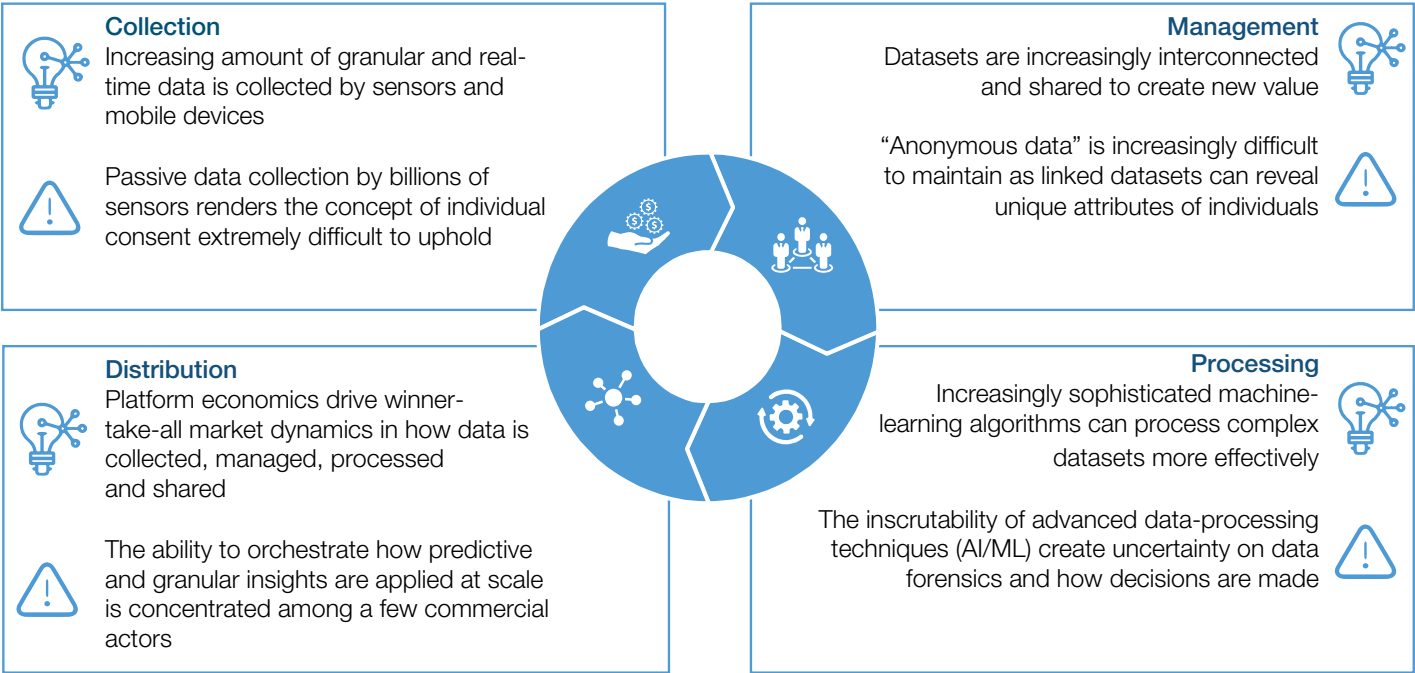


Figure 4: Balancing the value and risk dimensions of data collaboration



Despite the challenges ahead, the cost of inaction is unacceptable. While the risks and reasons for not using private-sector data for the common good are many, the cost of non-use – the failure to innovate in the use of new data sources – poses additional concerns. In that light, there is a reframing on the global dialogue to address both the risks of the *misuse* of data as well as its *missed use*.¹⁴ “While our data has been collected and used for more than a decade in ways we can’t determine, it also *hasn’t* been used to improve official statistics, public-service delivery or early warning for disaster response,” notes Robert Kirkpatrick, Director of UN Global Pulse. “We are paying a daily – and largely invisible – opportunity-cost in the ‘missed use’ of data for the public good.

“

The lack of data innovation is resulting in a failure to protect the public from preventable harms.¹⁵

“

Robert Kirkpatrick, UN Global Pulse

Voicing concerns about the consequences of moving too fast and the consequences of “permission-less innovation” is the call for more dialogue on ways to identify emerging harms and more adaptive ways to manage those risks.¹⁶ “When a new challenge arises, there is a tendency to focus on building new data tools. However, it is equally important – and arguably much harder – to figure out how to use data responsibly to improve our humanitarian response,” says Sarah Telford, Head of the UN OCHA Centre for Humanitarian Data.

The principle of “first do no harm” is widely held among many in the humanitarian sector, where the call for proportionate checks and balances to ensure irreversible harms do not occur is gaining strength. “If we don’t slow down, we run the

risk of short-term gains and long-term failures,” notes Linda Raftree, data privacy advocate and founder of MERL Tech. “Efforts to ensure privacy through more deliberative and iterative design approaches could reduce possible harms and lead to better outcomes that improve trust over time.”

So how can organizations strengthen trust and implement effective public-private data collaboration given the range of entangled risks and challenges?¹⁷ While there are no easy answers to this question, aligning on shared taxonomies can serve as an initial step for diverse stakeholder communities to pursue common goals in concrete ways. Accordingly, the World Economic Forum has arrived at six dimensions of trust:¹⁸

- **Security** involves having the people, processes and tools required to ensure that the confidentiality, integrity and availability of data will be upheld and protected throughout the life cycle against malicious attacks, unintended accidents and naturally occurring “acts of God”.
- **Accountability** involves processes to ensure that network stakeholders are held responsible for upholding accepted standards and agreements so that relationships remain reliable and predictable.
- **Transparency** involves giving stakeholders meaningful ways to understand relationships, intent and outcomes. To achieve this, individuals need accessible and understandable information on how relationships are structured and how data is being used. Additionally, transparency requires that organizations have the capacity and oversight to ensure that all outcomes from a data collaboration are accurate and that biases (intentional or unintentional) are not systemic.
- **Auditability** involves the creation of feedback loops for externally checking, verifying and monitoring data flows across an array of stakeholders and jurisdictions. This element builds trust among parties by facilitating transactions and enabling efficient dispute resolution.
- **Equity** involves ensuring that value is apportioned fairly and that outcomes are unbiased. Regulators and consumer-protection agencies are typically focused on ensuring fairness and are increasingly vigilant about the potential for data to hurt vulnerable populations, disempower individuals or create an unfair playing field for competitors.
- **Ethics** guide stakeholders through potentially ambiguous, uncertain or highly context-dependent decisions. Unethical (or even illegal) activities can permanently damage trust among collaboration partners. For data collaborations, one core ethical principle should involve protecting the rights, aspirations and intentions of vulnerable populations.

Figure 5: Six dimensions of a trustworthy system



Section 2: Establishing a new lens

Through a series of global workshops, expert interviews and use-case analyses, the World Economic Forum has established an evidence-based framework designed to identify specific areas for strengthening trust. The approach's emphasis on iterative alignment, consistent

communication and comprehensive governance is intended to help stakeholders understand and respond in pragmatic and practical ways. This holistic approach identifies five areas for strengthening trust and catalysing action.

Figure 6: Critical enablers of public-private data collaboration

- Achieving stakeholder alignment at the outset of a partnership
- Establishing responsible data governance
- Delivering insights that are accurate, unbiased and explainable
- Providing decision-makers with the tools, processes and support to act on new insights
- Ensuring long-term economic sustainability



1. Stakeholder alignment

Ensuring stakeholders commit to intended outcomes

The first step towards an effective data collaboration is for all relevant stakeholders – including government, industry, civil society, NGOs and individual data producers – to align on a shared value statement and to gain assurances that there is a long-term commitment by all parties. As part of the process for seeking stakeholder assurance and commitment, it is essential to balance the diverse incentives. As such, applying the concept of user-centred design can help to ensure that individuals are empowered and that the underlying needs and aspirations of vulnerable populations are identified. In addition, given the power asymmetries inherent in many public-private data collaboratives, actively engaging the communities of vulnerable populations and those at risk of becoming “digital invisibles” is particularly important.

From a trust perspective, a primary concern that arises during this phase centres on the use-case's potential impact on data privacy and security risks. Commercial data holders

need a baseline understanding of who will serve as the data custodian, the intended use of the data and the specifics of who has access rights and permissions. Vetting the privacy capacities of the requesting party can help shape the nature of the engagement.¹⁹ Additionally, getting clarity on funding (who is paying for what) can be highly valuable for establishing overall accountability among the stakeholders of a given collaborative.

While concerns around the potential impact on privacy and security are centrally important, reputational concerns can also play a significant role – especially among large institutions with a tendency to make decisions based on worst-case scenarios. It is also important for stakeholders to acknowledge the political and social context for how a data product may be used in practice.

As stakeholders build alignment, they will also strengthen their interpersonal relationships. It will help if each group tries to understand its partner's perspective. For instance, many private organizations have been repeatedly asked for their data – often for use in ventures that were vague

or ill conceived. NGOs and governments can help win the confidence of their business colleagues by explaining the defined outcome they are trying to achieve and the implications of not having access to private-sector data. Understanding the regulatory constraints that businesses face when sharing data, as well as the time and energy it takes employees to prepare data for its analysis, can also build interpersonal trust.

Likewise, corporate leaders need to be aware of certain gaps in technical skills and expertise that are typically found within the public sector. Demonstrating their willingness to invest time and energy to educate non-technical experts can go a long way to strengthen trust. Patience from the business side in regards to operational inefficiencies and bureaucratic decision-making from the public sector can also help build trust.

Strengthening trust through stakeholder alignment	
Public-sector concerns about the private sector	Private-sector concerns about the public sector
<ul style="list-style-type: none"> – Privacy and data protection – Lack of transparency on deeper private-sector intentions – Vendor lock-in and price-gouging concerns – Reputational and ethical concerns related to other market activities or clients – Uncertainty on the long-term commitment of private-sector actors to achieve project goals – Powerful actors can shape and change the governance structure on their terms 	<ul style="list-style-type: none"> – Privacy and data protection – Inadequate expertise and resources to adequately ensure the security of the data infrastructure – Limited capacities for data management and analytics – Relatively slow, inefficient and bureaucratic operations from public-sector actors – Influence of political agendas on how final outputs are actually used – Lack of acknowledgment from public sector on the investment of time and resources
Responding to the challenges	
<ul style="list-style-type: none"> – Conduct rigorous due diligence to ensure commitment and resource availability – Identify and balance competing incentives and needs – Commit to a clear use-case with measurable outcomes – Establish an operational plan with auditing and ethical oversight 	
Tools for strengthening trust	
<ul style="list-style-type: none"> – The Principles for Digital Development (DigitalPrinciples.Org) – The Human-Centered Design Toolkit (IDEO.org) – The GovLab Academy Canvas (NYU GovLab) – Data Impact Assessments and Oversight Models (The Information Accountability Foundation) – Due Diligence Checklist and Questionnaire for Prospective Technology Partners (UN Global Pulse) 	

There are four areas that can help strengthen trust during the stakeholder alignment phase:

Conduct rigorous due diligence to ensure commitment and resource availability. Rigorous and mutual vetting of the various stakeholders is essential to ensure commitment and resources are in place. Particular attention should be paid to ensure that the input of front-line workers is incorporated into the decision-making processes of senior leaders. Actively engaging with vulnerable populations is also critically important, as the needs, risks and aspirations of those at risk of becoming “digital invisibles” are often overlooked. Lastly, establishing an open and flexible process for defining and balancing limits on scarce resources

(available time and funding) is necessary to minimize implementation risks.

Identify and balance competing stakeholder incentives. Transparency of stakeholders’ questions and concerns about their counterparts can ensure mission alignment from the start. With a shared-value proposition, stakeholders are better encouraged to deliver on the collaboration’s purpose.

Commit to a clear use-case with measurable outcomes. The collaboration must have a clear goal with defined outcome measurements to guide participants. Stakeholders should have some flexibility in working towards their objectives and should be free to pursue new directions if results indicate another path is more promising.

Establish an operational plan with auditing and ethical oversight processes. Translation of stakeholders' incentives, requirements, objectives and commitments into an operational plan will establish execution requirements and demonstrate clear accountability on roles, responsibilities and duties of care, while providing a contingency plan to manage ethical or other operational uncertainties that may arise.

Putting trust into practice: KT Corporation and KCDC

The data collaboration between Korea Telecom (KT) Corporation and Korea Centers for Disease Control and Prevention (KCDC) to mitigate the MERS (Middle East respiratory syndrome) disease outbreak offers an example of how stakeholders can effectively achieve alignment. To show commitment and build trust, KT engaged in a series of smaller projects with KCDC leadership. These efforts helped allay public-sector concerns about the long-term dedication of their private-sector counterpart. In turn, KCDC helped government authorities coordinate their activities as they developed critical legislation to address private-sector concerns around data privacy and consumer protection. More specifically, the legislation called for the destruction of any mobility data used during the collaboration after a designated period. This guideline helped earn the trust of KT's leaders and subscribers, while improving transparency in the data collaboration. KT's effort to improve public health via its Epidemic Readiness platform has encouraged the participation of other major mobile operators to participate in the initiative.

2. Responsible data governance



Establishing a set of principles, processes and tools that holds stakeholders accountable to support the legal, fair and just use of data

With a defined use-case and measurable outcome indicators in place, identifying the required datasets and accountable data governance model is the next step. This phase aims to establish an approach that is legal, fair and just in the use of data.²⁰ In this phase, adherence to regulatory compliance requirements is necessary but often insufficiently addressed. The scope of governance concerns extends beyond privacy and data protection to include a wider set of issues related to the agency of individuals and safeguarding against group harms in the use of demographically identifiable information (DII).^{21,22}

From a technology perspective, the accelerating rate of change in the advanced use of data continues to fuel an underlying sense of uncertainty – if not distrust – among stakeholders.²³ As noted in the UN OCHA working draft

Data Responsibility Guidelines: “The technical tools and methods for managing data have evolved faster than the institutional policy instruments and guidelines that govern their use.”²⁴ The construct of “responsible data governance” provides a foundation for both the guiding principles and the concrete actions for establishing a trustworthy public-private data collaborative. These principles are not intended to replace applicable law; rather, they are informed by multiple existing guidelines including decades of internationally accepted norms for privacy and data protection known as the Fair Information Practice Principles (FIPPs). The FIPPs may not be perfectly suited to today's massively distributed and interconnected information networks; however, they are broadly applied in existing laws in many countries, offering the advantage of a common starting point for data strategies and analyses.

Operationally, the foundation of responsible data governance rests on outcome-based data-policy frameworks. Data that is not defined as personal or sensitive is generally not covered by data-protection laws. Yet it is routinely used in data collaboratives.^{25,26} As such, there is a need for comprehensive risk-based assessments at the project level to identify likely and severe data-related harms at the individual, community and societal levels that are not addressed by relevant laws. There is also a need to have inclusive design processes (i.e. privacy-by-design and ethics-by-design) at each stage of the collaborative life cycle so that effective data governance is not an afterthought. It is more cost effective and efficient to address these issues early in the process.

The issues affecting stakeholder trust in this phase are multilayered and complex.²⁷ Along with increasing anxieties about the advanced use of data, the lack of clear regulation is another concern – particularly when it comes to the use of group data (as distinct from personally identifiable information). The use of anonymous group data holds unique challenges as it can be combined with other datasets to reveal attributes and activities of individuals and/or communities. Anonymized and/or pseudonymized location data can be used to re-identify individual records within an aggregated dataset.²⁸ This risk can be most acute in the humanitarian context, where non-personal data (i.e. data on critical infrastructure such as hospital locations) is routinely collected. Safeguarding against the use of non-personal data to enable activities that constitute human rights violations, exploitation of vulnerable populations, discrimination, exclusion and/or misinformation should not be overlooked during data-impact assessments for the data collaborative.

In practice, the challenges of balancing complex data governance concerns have often resulted in many early-stage public-private data collaborations deploying virtual “sandboxes”, where data is accessed from the secure infrastructure of one of the sponsoring organizations (usually the business entity) by a limited set of authorized individuals. An example of this can be seen in the Japanese government’s “Investments for the Future Strategy” of 2017

– one aspect of this strategy was the creation of a regulatory sandbox “for carrying out business trials regardless of current regulations”.²⁹ A second example from Japan is an open data platform “to promote innovative products and services”.³⁰ To achieve these ends, the Japanese government began certifying “big-data banks”, tasked with compiling and consolidating data from the private and public sectors to be made available for corporations, non-profits and research institutions.

Strengthening trust through responsible data governance	
<i>Public-sector concerns about the private sector</i>	<i>Private-sector concerns about the public sector</i>
<ul style="list-style-type: none"> – Limited engagement by individuals and vulnerable populations on data ownership, consent and specified purpose of the data collaborative – Repurposing of data (and insight derivatives) for commercial purposes beyond the original scope – Disproportionate focus on privacy harms to individuals at the expense of potential group harms in the use of location and demographically identifiable data – The potential for bias, resulting from incomplete datasets, and a lack of acknowledgement of the needs of “digital invisibles” – Lack of local engagement and redress for individuals 	<ul style="list-style-type: none"> – Lack of data infrastructure and management capabilities among public-sector actors – Lack of data literacy and defined processes for what constitutes responsible data governance – Accidental disclosure of proprietary data – Limited clarity on liabilities when aggregated data is held by third-party intermediaries operating in multiple countries
Responding to the challenges	
<ul style="list-style-type: none"> – Implement a responsible data-sharing protocol to operationalize ethical data-sharing principles at the organizational level – Build a secure, resilient and fit-for-purpose governance and technical infrastructure – Establish external review boards to address legal ambiguities and ethical uncertainties – Turn principles into practice 	
Tools for strengthening trust	
<ul style="list-style-type: none"> – Key Actions and Outputs for Data Responsibility (UN OCHA) – Ethical Data Impact Assessments and Oversight Models (The Information Accountability Foundation) – Risks, Harms and Benefits Assessment Tool (UN Global Pulse) – Data Protection in Humanitarian Action (ICRC) – UN Personal Data and Privacy Principles (United Nations) – Communicating with Communities (Internews) – Information Sharing Protocol Template (UN OCHA) 	

There are four responses that can help strengthen trust during the responsible data governance phase:

Implement information sharing protocols and guidelines to balance security, privacy and ethical considerations in the use of data. After agreeing upon a use-case and the scope and volume of required datasets, stakeholders can create an ecosystem map with data-flow diagrams to demonstrate what data processes are occurring, when and by which individuals. Articulate the value of required data and how it relates to the specified purpose of the collaborative effort. Establish a comprehensive data-impact assessment to identify severe and likely data risks and to balance competing interests and potential impacts (negative and positive) across the broader ecosystem.

Build a secure, resilient and fit-for-purpose governance structure and technical infrastructure. It is essential to clearly articulate stakeholders' technical and governance requirements to guide the design of a fit-for purpose technical infrastructure that meets the use-case demands. These requirements can range from addressing concerns about data security, data-storage location, access rights and regulatory restrictions. A thorough understanding of the requirements, strengths and trade-offs of various technical design archetypes will equip stakeholders to make the best implementation decisions.

Establish external review boards to address legal ambiguities and ethical uncertainties. An external oversight body can provide guidance on ensuring fairness, local engagement and transparency. It can also assist with decisions about the use of data and the intended outcomes. These review boards can be a helpful counterweight to the internal power dynamics that shape how decisions are made.³¹

Turn principles into practice. Data collaborations need clear standard operating processes and pragmatic frameworks that enforce rigorous security measures and protect data subjects. Pragmatic frameworks that guide decision-makers through the complex reputational, social, legal and ethical trade-offs within a data collaborative are also required. These concrete processes and frameworks should be communicated to all stakeholders and at all levels of an organization. Contracts and agreements are also essential tools for ensuring that stakeholders are held accountable to governing principles and that they understand the front-line risks, realities, expectations, roles, penalties, duties and responsibilities as they embark on data-collaboration efforts.

Putting trust into practice: Mastercard and Harvard

A collaboration between Mastercard and the Harvard Center for International Development (CID) demonstrates how business leaders can address risks (perceived or real) as they participate in establishing a responsible data environment. To help understand and improve economic conditions in Colombia, the CID was able to use insights on local spend patterns based on Mastercard's aggregated and anonymized transaction data. To create a trustworthy environment for these data insights, Mastercard required robust privacy and data-protection controls. First, the insights were created on Mastercard-controlled technical infrastructure and only pre-approved CID researchers were granted access to the resulting reports. Second, CID and Mastercard agreed upon the types of analyses the researchers could conduct at the outset of the collaboration. For instance, the CID team prepared a report for Colombia's Ministry of Economic Development that shared the findings from their research, but not the underlying insights themselves. Their collaboration helped generate a deeper understanding for improved policy decision-making regarding tourism-related economic growth and foreign investment in Colombia. At the same time, their responsible data governance effectively managed privacy and security risks and held stakeholders accountable.

3. Insight generation and validation



Delivering insights that are accurate, fair and explainable

The next phase of a successful data collaboration is where raw data is converted into useful information. The process involves three stages: acquiring the necessary data inputs, processing and validating the inputs and assessing the derived packaged insights.^{32,33}

As it relates to trust, the concerns at this phase are multilayered: the data inputs should be legitimately collected, complete and accurate, the data processing should be reliable, replicable and interpretable and the derived and packaged outputs should be valid, fair and interoperable within a defined context. As a 2019 survey of data practitioners within the humanitarian sector notes, one of the top challenges from the demand side is ensuring the quality of the data.³⁴ There may be technical constraints related to collecting, cleaning and assessing the quality of data. In addition, there may be capacity issues or capability gaps. For instance, individuals may not know what quality data looks like or how to access it. Along with the cleanliness and completeness of the input data, there are also concerns with sample bias. If the input data is skewed, the derived insights may result in unintended consequences.

Along with the completeness and quality of the data, it is also critically important to have individuals with both deep domain knowledge and the data-science skills to both identify important context-dependent questions and to act on them with relevant data and analytic techniques. These “data bilinguals” are immensely valuable at all phases of the collaborative process.³⁵

At this stage, it is important to recognize the context in which new data products and services will be used. Will they be used as a “new way for answering old questions” or will they help to discover new questions that “no one knew to ask”? The path for strengthening trust will vary depending on the approach. When the data product serves as a “new input for old questions”, data quality and compatibility with existing analytic procedures is paramount for enabling institutional trust. Because the incumbent decision-makers generally have existing data-governance processes in place, questions related to legitimacy of purpose, institutional accountability and data rights are generally addressed. For example, in the use of geospatial imagery to create population estimates, there are significant concerns about the compatibility of these new data sources and integrating them with existing survey and census data.³⁶ When advanced Facebook image recognition algorithms were used to process the raw pixels of commercial satellite imagery to create new population density estimates (in a few countries where existing maps were outdated and inaccurate), a feedback loop was established with external and local entities (including national census authorities, national mapping agencies, household survey campaigns and earth observation groups) to normalize and validate the new findings. As concerns related to the data’s collection methodology, provenance and completeness were addressed, the focus shifted to identifying new opportunities to employ the new population assets in additional ways.³⁷

When it comes to the discovery of “new questions we didn’t know to ask”, the scope of trust-related issues expands. By definition, this path is intended to generate new and novel insights, so it has a greater potential to disrupt relationships with the status quo. Strengthening trust in these contexts requires unpacking an entangled set of under-measured and unknown “black box” anxieties.³⁸ While questions about data collection, provenance and completeness are still highly relevant, there are additional issues related to advanced data analytics enabled by AI or machine learning. For instance, stakeholders may have concerns about the reliability, replicability and validity of advanced-analytics algorithms.³⁹ How does the processing work? How reliable will the models be over time? How uniquely dependent is the model on its training data?

When it comes to the bias of algorithms, there is no such thing as an unbiased algorithm. Algorithms reflect the values of those who create them at every step of the process, including designing, building, testing and refining the model. Likewise, the statistical correlations contained within the underlying training data can lead to disparate impacts. Algorithms predict the future based only on past events contained within the training data.⁴⁰ To ensure the long-term reliability, accuracy and fairness of finished data products and services, data collaborations should establish robust feedback loops to ensure that the products and services do not deteriorate over time. In some instances, “helper algorithms” can be used to monitor automated outputs and offer advanced warning of model performance deterioration.

There are also concerns of “relational trust” where the engagement of new actors (particularly those from the private sector) raises questions of legitimacy, fairness, interpretability and data rights ownership. For example, when innovative data sources (e.g. social media, mobile phone data, retail transactions) are used to establish credit scoring, it is important for individuals to not only have access to their scores, but also to have a meaningful understanding of how their score was created, how it was applied by which decision-makers and how errors can be corrected. In this regard, legal requirements in the European Union and the United States for “reason codes” point to an interesting model that explains to individuals how complex data indicators about them are created and used by others to make decisions. Overall, it is important to note that the development of this phase is in its early stages – understanding trustworthy approaches for the governance of *packaged data insights* (versus the governance of the data inputs) is an area where additional conversations and deliberations are needed. Greater engagement with the Data Ethics and Fair Accountable Transparent ML (FAT/ML) communities is a priority.

Strengthening trust through insight generation and validation

<i>Public-sector concerns about the private sector</i>	<i>Private-sector concerns about the public sector</i>
<ul style="list-style-type: none"> – Inability to determine the quality and completeness of input data – Lack of concern for intended and unintended algorithm biases – Limited incentives to collaborate with commercial competitors to ensure the completeness of data inputs – Inability to explain how advanced data analytics operate in layperson’s terms – Lack of meaningful engagement for individuals regarding redress and remediation 	<ul style="list-style-type: none"> – Lack of control over the governance of packaged data inputs and commercial liabilities if data products get into the wrong hands – Lack of expertise in understanding data outputs and the limitations of decision-making – Little understanding of technical challenges and time taken to create data outputs – Fear that IP investments for developing models will be appropriated by governments with no compensation – Erosion of commercial “data-as-a-service” profit margins by demanding low prices for highly granular insights
Responding to the challenges	
<ul style="list-style-type: none"> – Verify the provenance, completeness and accuracy of data inputs – Ensure the interpretability of derived data insights – Establish feedback loops to check the reliability and legitimacy of outputs – Implement responsible governance processes for the distribution of data products 	
Tools for strengthening trust	
<ul style="list-style-type: none"> – Principles for Accountable Algorithms (FAT/ML) – Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability (AI Now Institute) – Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods (Upturn, Omidyar Network) – Digital Decisions (Center for Democracy and Technology) 	

There are four responses that can help strengthen trust during the insight generation and validation phase:

Verify the provenance, completeness and accuracy of data inputs. Stakeholders can introduce cross-functional metadata standards with common attributes to understand the origins, completeness and accuracy of datasets. This will help aid in minimizing bias and model performance testing.

Ensure the explainability of algorithms and derived data insights. While the mechanics of any advanced analytical method or model can be difficult to explain, there are certain measures stakeholders can take to ensure the explainability and trustworthiness of their process. Investing in interpretability measures on the front-end – such as designing a model with fewer features or visually mapping critical data inputs – will make outputs easier to understand for data engineers and end users alike.⁴¹

Establish feedback loops to check the reliability and legitimacy of outputs. Stakeholders can implement transparent decision-making processes and feedback loops to gain the trust of end users and data subjects. These mechanisms will encourage input from local experts who are closer to the data collaboration’s application, while encouraging the collective and iterative refinement of insights.

Implement responsible governance processes for the distribution of data products. Governance is necessary to monitor how data-derived products and services are packaged and used. Stakeholders can ask themselves the following questions: What decisions and actions will the packaged data products and services support? How should they be published or presented? How often should they be updated to protect against model deterioration?⁴²

Putting trust into practice: ‘Open Algorithms’ (OPAL) in Senegal and Colombia

In 2017, an OPAL-driven public-private data collaboration involved multiple groups: Data-Pop Alliance, Imperial College London, MIT Media Lab, Orange and the World Economic Forum – in conjunction with Sonatel, Telefónica and the governments of Senegal and Colombia. It shows how stakeholders can navigate the challenges of insight generation and validation.⁴³ Deployment of OPAL started in mid-2017 in Colombia and Senegal.

The intent of the open-sourced OPAL platform is to serve as a decentralized means of accessing private-sector data globally. The data stays within the premises of the private company and only an authorized set of third-party queries (the open algorithms) request the pseudonymized data. What results is a safe question-and-answer system where the questions are validated in advance by a board of advisers comprised of experts and local members of the community.

With OPAL, national statistics officers can draw new data inputs into existing statistical models for the creation of more timely and accurate population density maps and poverty indices. The challenges OPAL faces lie in the ability of incumbent national statistics offices to adopt a standard data product. Ensuring that the new products are “better, faster and cheaper” than existing methodologies requires greater evidence. There are also concerns about the completeness of the input data as current pilots reflect the baseline data of only one mobile operator in the region. While support from regulators has been strong enough to encourage this new form of collaboration, one of the unanswered questions is if there is enough commercial incentive to collectively engage all of the operators. The real (and perceived) risks from mobile operators of disclosing commercial insights to a competitor are material.

4. Insight adoption



Ensuring decision-makers and front-line users act on insights

Insight adoption, the fourth step, is where time invested during the planning and stakeholder alignment phases will pay dividends. Barriers to insight adoption most often arise due to challenges implementing data products and/or a lack of alignment around the monitoring and evaluation of the success of new decision-making processes. “The number-one reason why innovative decision support systems aren’t adopted is that people don’t fully understand what they need,” notes Samuel V. Scarpino, Chief Strategy Officer at Dharma Platform.



Months of iterative design at the early stages of a project can be critically important.



Sam Scarpino, Dharma Platform

The barriers to insight adoption are dependent on factors at both the individual and organizational level. At the individual level, many people resist using new data products and services because the quality is low. Is the product simple, sticky and intuitive to use? Will individuals gain genuine utility by using the tool? Is there effective product support? Are the developers prepared to both listen and respond to feedback from users in the field? Are individuals ready to “unlearn” the way they currently make decisions and use the new tool? Doubling down on answering these baseline questions can help to ensure long-term adoption.

At the organizational level, it is essential to show why the new insights are important to adopt. It is also important to show why change is needed within the organization and the cost of inaction. Leaders should also note the time frame over which the changes must occur, identify the senior champions for the initiative and the reputational cost to the organization (and its senior leaders) if things go wrong.

Changing formal decision-making processes within an organization requires a large number of personnel. In fact, capability building was cited as the top priority in a 2019 global survey of more than 1,200 humanitarian practitioners who work with data.⁴⁴ In terms of capacity building, one of the most frequently voiced needs is for individuals with a blended set of technical, organizational and communications skills who can build trust between the technical data community and non-technical leaders of an organization. These “data integrators or bilinguals” can help navigate organizational power dynamics, recognize decision rights and factor in the repercussions of poor decisions. They can serve to address organizational-change management issues, address last-mile implementation challenges and support the need for ongoing capability building.



Closing the gap between data experts and senior decision-makers shortens the time from when raw data is collected to when better decisions are made.



Sarah Telford, UN OCHA Centre for Humanitarian Data in the Hague

Strengthening trust through insight adoption

<i>Public-sector concerns about the private sector</i>	<i>Private-sector concerns about the public sector</i>
<ul style="list-style-type: none"> – Lack of effective support to explain the product and the ability to adjust it – Overly complex user experience requiring too much effort – Ineffective communication about the value of applying the new insight and the cost of not using the new approach 	<ul style="list-style-type: none"> – Limited incentives for middle management to adopt tools based on long-time experience and reputational risks of bad decisions – Lack of effective ways to introduce change management processes into bureaucratic decision-making processes – Limited ability of public-sector actors to deliver last-mile implementation requirements for front-line decision-makers
Responding to the challenges	
<ul style="list-style-type: none"> – Identify top-down and bottom-up champions for change – Establish two-way product-support communication channels – Deliver sustained commitment to build capacities – Invest in last-mile implementation 	
Tools for strengthening trust	
<ul style="list-style-type: none"> – Strategies for Setting Up Data Teams (Internews) – Change Management Toolkit (UC Berkeley) – The Human-Centered Design Toolkit (IDEO.org) 	

There are three areas of focus to help strengthen trust during the insight-adoption phase:

Identify top-down and bottom-up champions for change. Organizations can build a dedicated cohort of champions across the institution from the top down and bottom up. Word-of-mouth and peer-to-peer exchange on the benefits of using the new tool can drive adoption from the bottom up. Linking job performance evaluations to how often individuals have used a new tool to make decisions can serve as an effective top-down measure.

Establish two-way product support communication channels. It will be essential to explain and continuously improve new products. Organizations should also have robust and responsive product-support channels to answer specific questions and receive feedback from the field.

Deliver sustained training to build capabilities. Invest the time and technologies to train users in how the tool works, its limits and how it fits into existing decision-making processes. Placing clear limits on the amount of time that those beyond the implementing organization will need to provide for product support and training can also help eliminate hidden costs and implementation risks. While initial product support may come from outside the adopting organization, a transition plan needs to be in place for how product support and training will be managed over time.

Putting trust into practice: Facebook and NetHope in Puerto Rico

A practical example of how stakeholders in a data collaboration have helped change front-line worker behaviours based on generated insights is epitomized by the Facebook Disaster Maps Program, which has generated maps for hundreds of disasters and has been actively used by more than 40 organizations since 2017. For example, when Hurricane Maria hit Puerto Rico in 2017, Facebook collaborated with NetHope to help aid workers understand where connectivity services were needed. Facebook Disaster Maps – which are designed to be practical and easily distributed to both decision-makers and front-line users – generate data on connectivity, population movement and services requested by disaster-affected communities within 24 hours of a natural disaster. NetHope was able to use Facebook network-coverage maps, which provided an unprecedented view of connectivity, to install 99 connectivity sites across the island where they were needed most. Many of these sites provided public Wi-Fi spots for local citizens to receive and share news, connect with friends and family and help organize the next steps in their recovery. Insights were implemented on the ground supported by a team at Facebook and other collaborating agencies, including the American Red Cross, which all shared a Slack channel to ensure that questions were answered quickly and maps could be put into action in real time.

5. Economic sustainability and scalability

Establish long-term economic viability

The fifth step focuses on the economic sustainability and scalability of the data collaboration. As defined by the GSMA, “A sustainable solution is one which allows for the robust, repeatable and replicable use of data across different geographies and use-cases, underpinned by a secure source of funding which enables continuity in the supply and analysis of the data, to generate actionable insights.”⁴⁵ The question of sustainable economics for data collaboratives is, in many ways, the question du jour. Given that many existing collaboratives were initially underwritten by donors (or started as “data-philanthropy” donations from the private sector), the question of sustainable economics has historically been less of a priority for data collaboratives.

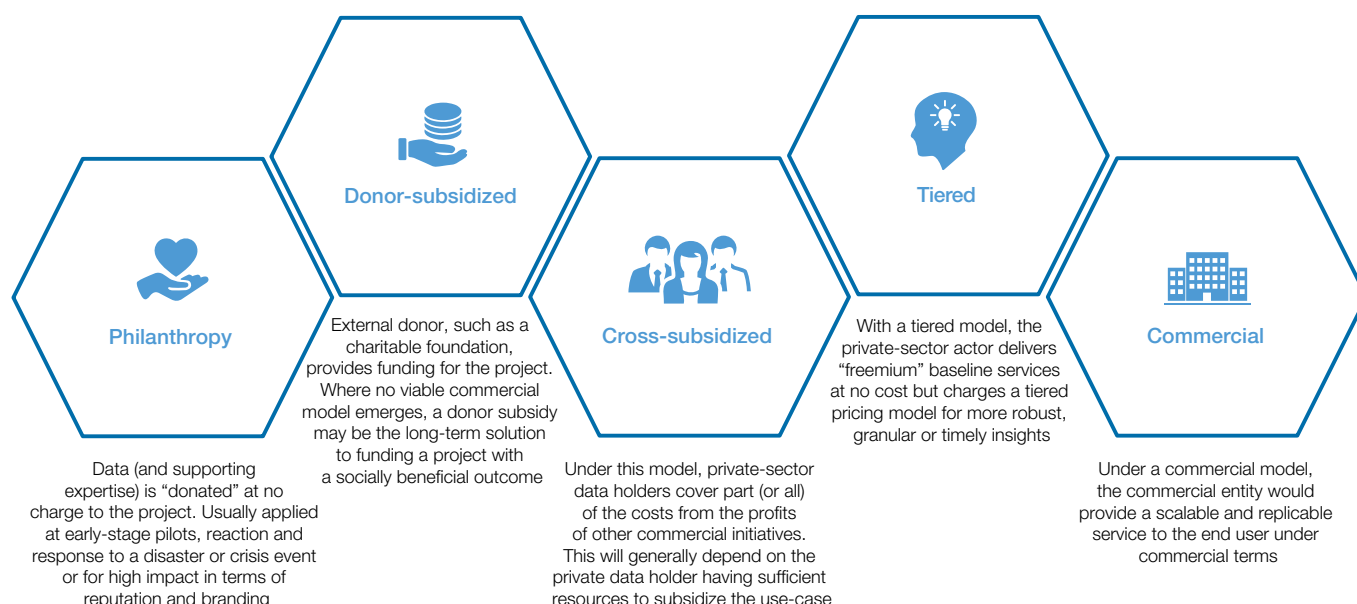
As early-stage data collaboratives mature, however, the question of how to establish sustainable economics has promoted the need for a more holistic framing of the challenge. Building sustainable economics introduces complexities from both the demand and the supply side of the equation. As noted earlier, both the regulatory uncertainty as well as the operational and capacity-building concerns represent material obstacles for establishing a more structured, cohesive and organized data for the common-good marketplace.⁴⁶ It is impossible to prescribe a single strategy for building economically sustainable data collaborations, since each partnership differs in terms of stakeholder governance structures, objectives and other features. These variations make it difficult to determine which economic model works best.

For decision-support tools tailored to natural disasters, emergency response or humanitarian relief, the non-commercial nature of this context places the question of sustainable economics generally in the area of donor-subsidy and government funding. In terms of scaling this type of collaboration, the global reach of many corporations can provide the opportunity to expand into new geographies. KT Corporation, for example, introduced its Global Epidemic Prevention Platform to sub-Saharan Africa through partnerships with the Kenyan and Ghanaian governments using their corporate subsidiaries within those regions.

For services that provide periodic updates – such as delivering new indicators for the SDGs, census-taking or policy evaluation – the economics may be based on various subscription models. Recurring subscription fees could be cross-subsidized by donors, or “freemium” models could exist where coarsely grained data products are offered for free while access to more granular products incurs a fee.⁴⁷

For services designed to provide routine and regular services for larger development concerns, the economic models lend themselves to more traditional economic structures and approaches. While still in the very early stages, models based on data cooperatives, data trusts and platform economics (where a network of actors are compensated based upon their collective ability to deliver a measurable outcome) also hold great promise. The sustainable agriculture and precision medicine domains are early frontrunners in this regard.⁴⁸

Figure 7: The GSMA's five types of data for social-good business models⁴⁹



Given the current differential in trust among stakeholders, concerns about sustainable economics include both familiar and newly emerging threats and vulnerabilities. At one level there are the “traditional” concerns about private-sector price-gouging, vendor lock-in and vendor stability. From the perspective of the business sector, there are familiar concerns related to a lack of willingness to pay, worries that baseline operational costs will not be covered or that governments will simply demand that the data be handed over for free, thus killing off any incentive to invest. Beyond the early-stage economic models of data philanthropy and donor-subsidized models, there are general business concerns that there are insufficient short-term returns, that margins are unstable (governments could arbitrarily establish pricing limits) and that cross-border data-flow restrictions would inhibit the ability to scale.

There is also a deeper and more polarizing set of economic concerns based on how the predictive “behavioural data exhaust” will be employed downstream and monetized over time.⁵⁰ In general, these anxieties stem from the current lack of trust individuals have with many of the large social-media platforms and internet companies and their concerns on data-extractive business models.⁵¹ Underlying these conversations are concerns about data ownership, the privatization of public data, the accountability (and auditability) of highly complex data-supply chains, the lack of meaningful engagement of individuals and anti-trust concerns related to data monopolies. Resolving these concerns is beyond the scope of this paper, but given the importance of using private-sector data for the global good, the need to have economic models that address these deeper concerns is central for long-term sustainable progress.

Strengthening trust through economic sustainability and scalability

<i>Public-sector concerns about the private sector</i>	<i>Private-sector concerns about the public sector</i>
<ul style="list-style-type: none"> – Price-gouging and vendor lock-in – Little transparency about “real costs” (e.g. staffing, technology) – Lack of clarity on data ownership, downstream monetization and consent of individuals – Anxieties about the commercial incentives of a concentrated set of highly powerful platform actors and the power asymmetries inherent in data monopolies 	<ul style="list-style-type: none"> – Lack of long-term sustainability of donor-subsidized pilots – Lack of willingness and ability of the public sector to pay for services and cover expenses – Not jeopardizing existing pricing models; ability to experiment with new approaches – Lacking clarity on the secondary reuse of data and the ability to monetize insights from data collaboration

Responding to the challenges

- Evaluate a variety of funding mechanisms matched to project objectives
- Identify opportunities to expand and scale across industries and geographies
- Encourage commercial opportunities that use the notion of data trusts and the engagement of community co-ops, credit unions and trade associations

Tools for strengthening trust

- Big Data for Social Good Toolkit: Sustainable Business Models (GSMA)
- Designing, Building and Scaling Data Trusts (The Open Data Institute)
- Data Trusts, Ethics, Architecture and Governance for Trustworthy Data Stewardship (Web Science Institute)

There are three areas of focus to help strengthen trust during the economic sustainability and scalability phase:

Evaluate a variety of funding mechanisms matched with the project objectives. Stakeholders need to establish a sustainability strategy as early as possible. While data philanthropy and donor subsidies are essential for early-stage pilots (as well as humanitarian and crisis-related situations), they should explore innovative pricing and platform economics within controlled marketplaces.

Identify opportunities to expand and scale across industries and geographies. To minimize tensions surrounding economic sustainability and scalability, stakeholders first require a thorough and clear understanding of the collaboration's intended path to scale.

Encourage commercial opportunities that draw on the notion of data trusts and the engagement of community co-ops, credit unions and trade associations. Explore the notion of data trusts as both a means to provide ethical, architectural and governance support for trustworthy data processing and as a way to reapportion the value of data products and services back to individuals and local data producers.

Putting trust into practice: Climate Corporation

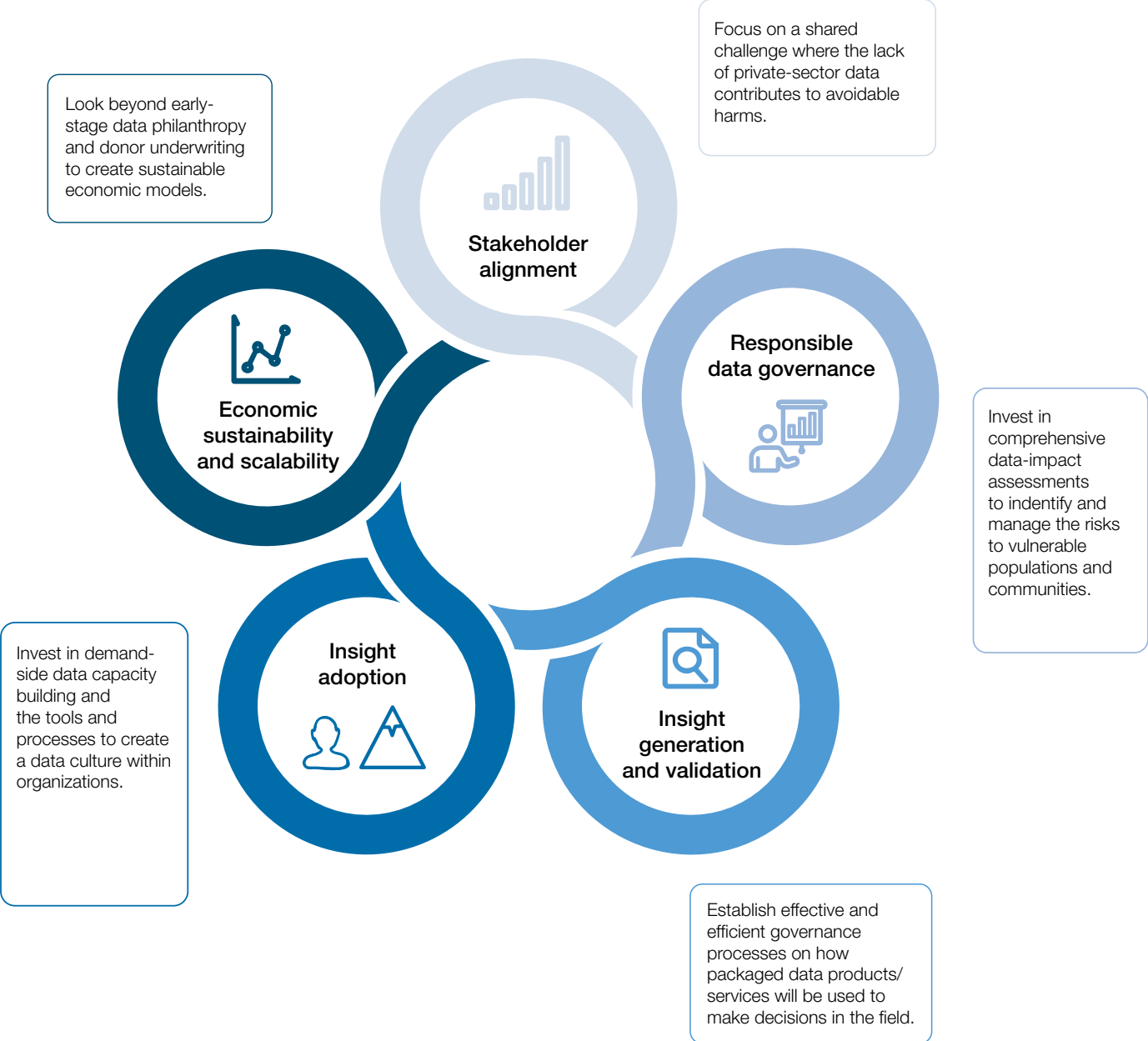
The Climate Corporation data collaboration shows how stakeholders can ensure economic sustainability and scalability. To tackle the trust deficit among stakeholders, the parent company Bayer (formerly Monsanto) gave farmers full ownership rights to the data they generate. The resulting “Climate FieldView” app has emerged as one of agriculture’s leading farm-software platforms, enabling farmers to analyse their data in one place. Centring on individual ownership of diverse data, the collaboration grew into a sustainable line of business.

Section 3: Catalysing impact

If data is the foundation of our digital society, then access to it will be vital for building the world we want. Addressing our most complex global challenges, making the best decisions during times of crisis and monitoring progress on the SDGs all require access to the most granular, timely and complete sets of data available. That won't be possible without robust public-private data collaboration.

A world of robust public-private data collaboration will also require a continued commitment to strengthening trust. It's true that the technologies of the Fourth Industrial Revolution are transformative, yet the real revolution lies in their ability to connect diverse groups of individuals and institutions in unprecedented ways. The framework discussed in this report provides leaders with a new approach on how to encourage positive change by both strengthening relational trust and unlocking new forms of technical innovation.

Figure 8: Key priorities across the public-private data collaboration lifecycle



Of all the considerations for leaders to reflect upon, perhaps the most important is the need to embrace a dilemma. What's the cost of doing nothing versus the price of running too fast? What's the impact of *not* using the best data to understand and address our most complex global challenges?

“

If you're not seeing the paradox, you're not seeing the reality.

“

Scott David, University of Washington

In times of crisis, bad decisions and delays will occur without having access to the most accurate, timely and complete sets of data. What opportunities will be lost if we can't innovate and discover things we literally can't imagine?

Conversely, how do we ensure that first we do no harm? Who gets to define the limits? Who gets access to the predictive models that can determine the future of entire populations? Who holds the liability when the advanced use of data – with inputs from multiple data holders – causes irreversible harm? Who gets to hit the stop button when an innovative use of data only amplifies existing inequities and makes things worse?

The answer will lie in how we frame the questions. Balancing the competing tensions of public-private data collaboration is not a problem to be fixed. It's an ongoing challenge to be managed. The pathway to progress will require the ongoing commitment of critical financial, technical and political resources. It will demand patience when things fail, the resilience to start anew and the drive to scale when things succeed. It will take leadership to address old problems with new data-driven insights and courage to discover new questions that no one knew to ask.

What's next?

While embracing the deep dilemmas of data collaboration is vitally important, so is the need to take action and deliver impact. In that spirit, below are some concrete opportunities for leaders to support that came out of the World Economic Forum's Global Dialogue on Public-Private Data Collaboration.

Humanitarian action and crisis response: epidemic readiness

Access to commercially controlled population movement data is vital for readiness and response efforts for epidemics and natural disasters. The World Health Organization's EPI-BRAIN initiative is a first-of-its-kind sustainable, shared, accessible and integrated data-innovation ecosystem to reduce the impact of infectious disease outbreaks through forecasting and predictive analytics. One dimension of the initiative will be the need for trustworthy and accountable access to population movement data at scale. This multidisciplinary and multistakeholder community will be launching in 2019.

Global development: data literacy and stewardship

Increasing data-literacy capacities on both the supply and demand sides is centrally important to ensure a robust data-collaboration environment emerges that is efficient, impactful and trustworthy. By strengthening the function of data stewardship among private-sector actors which would relate to, but be separate from, data privacy and security, a community of practice can serve to advance the use of commercial data for the common good. The Data Stewards Network (datastewards.net) is a unique community in this regard and is working to build open repositories of case studies and methodologies.

Official statistics: contracts for data collaboration

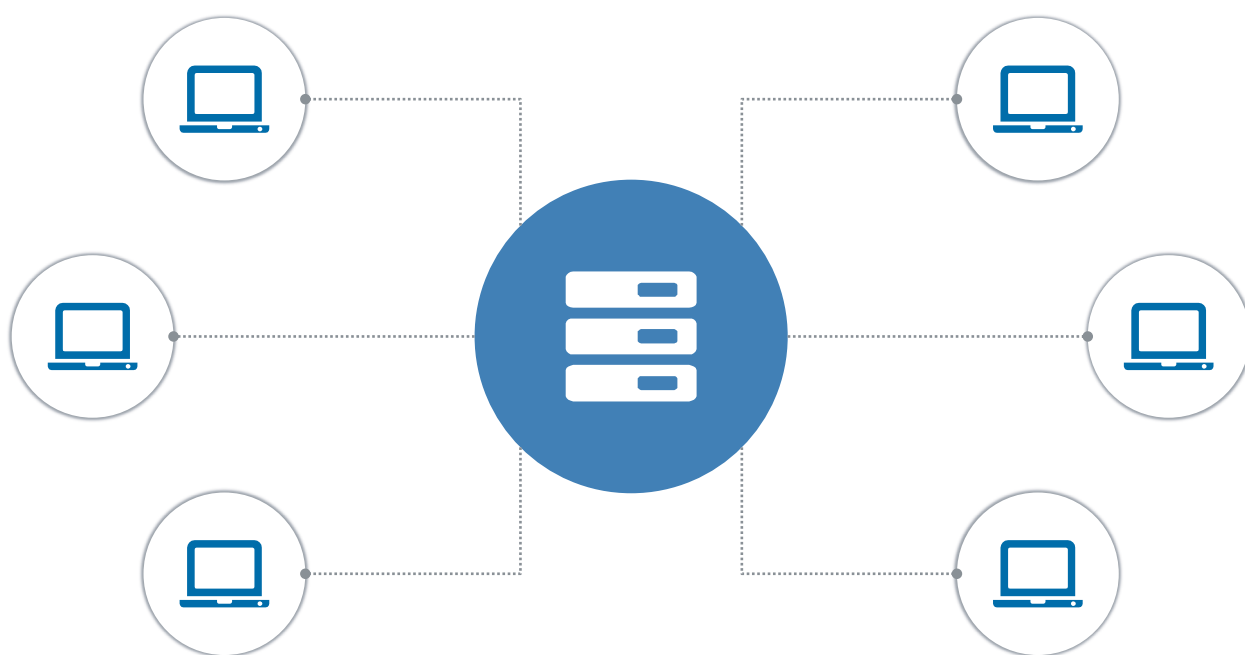
One of the main friction points in establishing these collaboratives are the knowledge gaps and transaction costs in drawing up contractual data-sharing agreements. The [Contracts for Data Collaboration](#) (C4DC) is an initiative launched in 2019 to address the inefficiencies of developing contractual agreements by making available a toolkit for non-legal experts and a shared repository of contractual clauses taken from existing data-sharing agreements. Charter members of the C4DC include the GovLab at NYU, UN SDSN Thematic Research Network on Data and Statistics (TReNDS), the University of Washington and the World Economic Forum.

Appendix: Fit-for-purpose technology archetypes

Successfully implementing a public-private data collaboration depends on a fit-for-purpose technology architecture that can help address technical and governance-related risks. A brief description of the mechanisms, advantages and trade-offs for different design archetypes can strengthen trust in a collaborative solution.

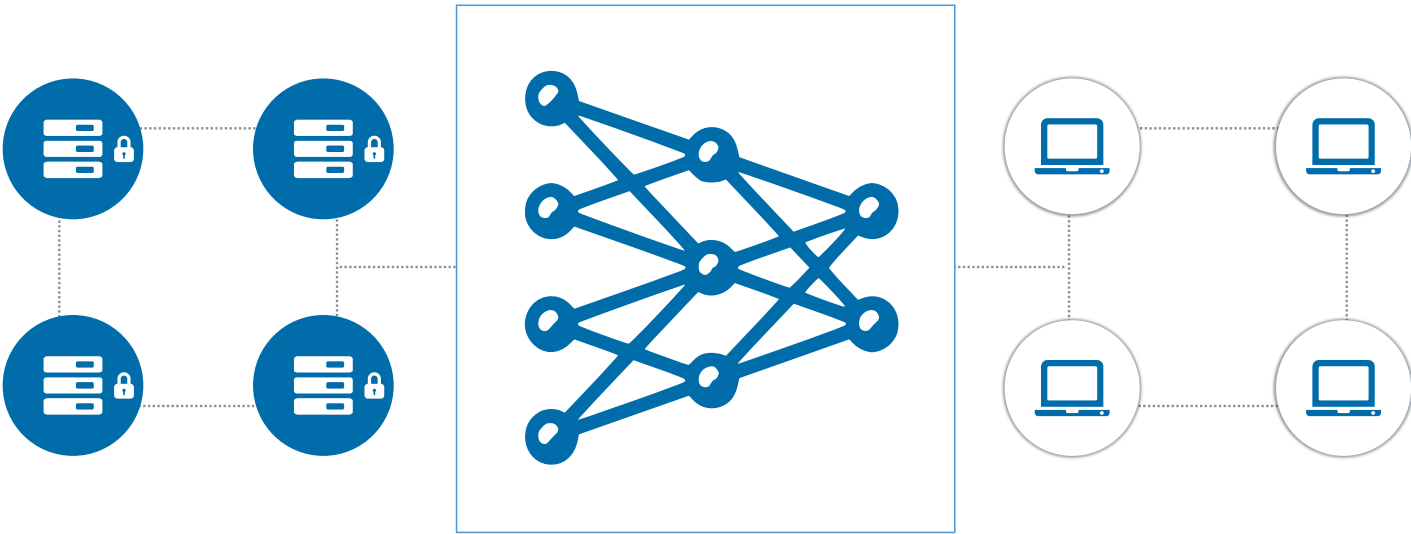
Centralized technology archetype. Stakeholder data is consolidated and housed in one location. This archetype is suitable for public-private data collaborations where centralized control of the collaboration's consolidated dataset is required. The need for a centralized archetype may also stem from the use of legacy data systems and standards that are not interoperable with other stakeholders' systems or regulatory requirements restricting a stakeholder from transferring data. The centralized archetype can defend against a range of technical and business concerns such as data misuse (using the data for purposes other than the shared goal), the absence of sufficient security and data quality controls and dissimilar operational licences or standards. This approach has advantages in its relatively low implementation and maintenance requirements and its ability to support exploratory analyses on the consolidated datasets. However, this archetype has certain trade-offs such as the inherent scalability limit to the number of potential participants, limited stakeholder control over proprietary data and the potential power imbalance due to a single stakeholder housing and managing the consolidated data.

Figure 9: Centralized technology archetype



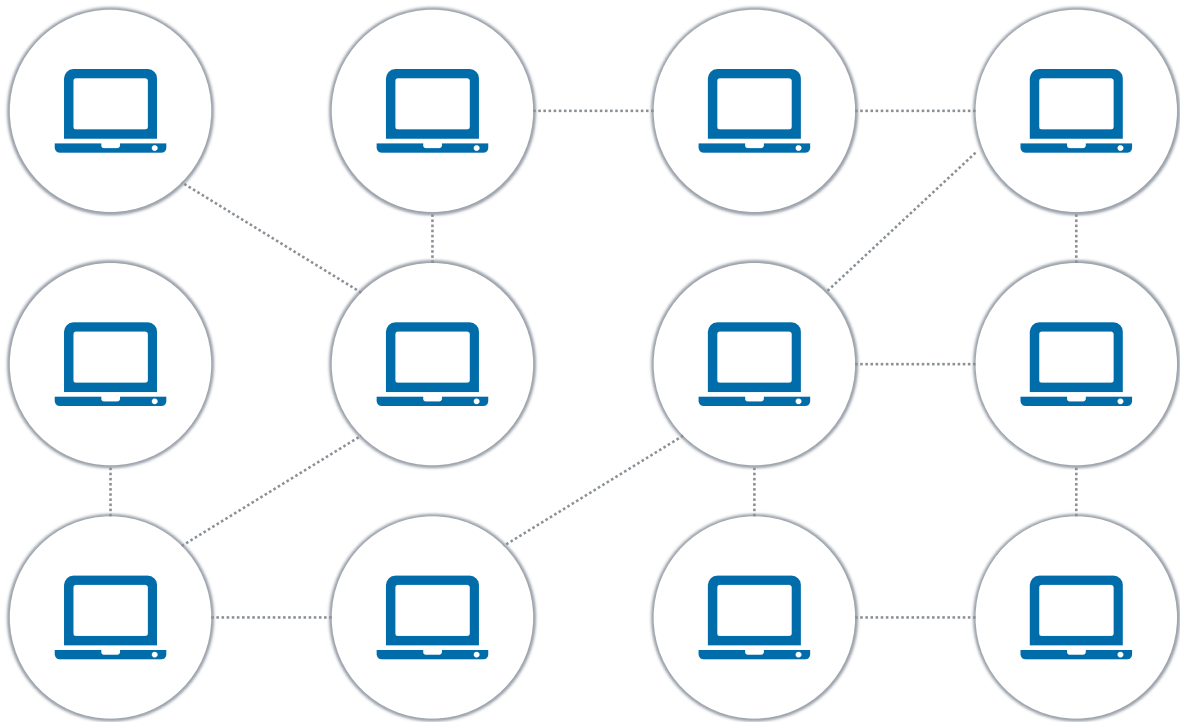
Federated technology archetype. Within a federated archetype, a predefined dataset resides within the secure infrastructure of the various data holders. Access to the data is provided via secure APIs and only to a limited and authorized set of agreed-upon queries. The raw data never leaves the control of the individual data holders, nor is it transferred across sovereign borders or to a central repository. During the design phase of a federated system, stakeholders align on the data required, level of granularity and format for the agreed-upon queries, thereby limiting the breadth of insights that can be derived from the distributed datasets. This approach is best suited when access to sensitive data is required and when both local security and regulatory compliance measures are mandatory but the need for global scale is also present. The relatively slower implementation speed, a higher maintenance cost and uncertainty on data-processor liabilities are some of its limitations.

Figure 10: Federated technology archetype



Distributed technology archetype. The distributed archetype employs distributed ledger and blockchain technology so that the risks of third-party intermediaries can be addressed. The distributed archetype is well-suited for public-private data collaboration efforts where a single “source of truth” (with unique timestamps that cannot be altered or destroyed) is needed. Some of the major advantages include a higher fault tolerance and redundancy, cost efficiency and security (due to the use of encrypted communication protocols). This approach can be complex to implement with high maintenance costs.

Figure 11: Distributed technology archetype



Acknowledgements

Lead authors

William Hoffman, Project Lead, Data Policy, World Economic Forum
Austin Boral, Project Collaborator and Seconded, McKinsey & Company
Didunoluwa Olukoya, Senior Business Analyst, McKinsey & Company

Knowledge partners

Nicolaus Henke, Global Leader, Digital and Analytics; Chairman, Quantum Black, McKinsey & Company
Raphael Bick, Partner, McKinsey & Company
Khaled Rifai, Partner, McKinsey & Company
Marcus Roth, Partner, McKinsey & Company
Tom Youldon, Associate Partner, McKinsey & Company

At the World Economic Forum

Jeremy Jurgens, Managing Director, Head of Centre for Global Industries and Strategic Intelligence, World Economic Forum
Derek O'Halloran, Head, Future of Digital Economy and Society, Member of the Executive Committee, World Economic Forum
Ryan Morhard, Project Lead, Global Health Security, World Economic Forum
Richard Samans, Managing Director, Head of Policy and Institutional Impact, World Economic Forum
Anne Toth, Head of Data Policy, Member of the Executive Committee, World Economic Forum

Project advisers

Sam Scarpino, Chief Data Scientist, Dharma.AI
Matthew Harrington, Global Chief Operating Officer, Edelman
Laura McGorman, Policy Lead, Data for Good, Facebook
Erik Wetter, Co-Founder and Chairman, Flowminder and Stockholm School of Economics
Stefaan Verhulst, Co-founder and Chief Research and Development Officer, GovLab
Vivian Tan, Vice-President, Strategy and Transformation, Kaiser Permanente
Chang-Gyu Hwang, Chairman and Chief Executive Officer, KT Corporation
JoAnn Stonier, Chief Data Officer, Mastercard
Emme Deland, Senior Vice-President, Chief Strategy Officer, New York Presbyterian
Shwetha Shetty, Vice-President, Corporate Strategy, SAP
Jessica Espey, Senior Adviser, Sustainable Development Solutions Network
Scott David, Executive Director, Information Risk Research Initiative, University of Washington
Jos Berens, Data Policy Officer, UN OCHA Centre for Humanitarian Data

Project and workshop participants

Liam Condon, President, Crop Science Division, Bayer
Rose Shuman, Founding Partner, BrightFront Group
Claudia Juech, Executive Director, Cloudera Foundation
David Muse, Chief Executive Officer, Dharma.AI
Katy Pullen, Head of Operations, EMEA, Dharma AI
Kate Wilson, Chief Executive Officer, Digital Impact Alliance
Justin Blake, Global Lead, Executive Positioning and Family Enterprises, Edelman
Joe Tropiano, Executive Editor, Edelman
Chaya Nayak, Research + Data for Good, Facebook
Linus Bengtsson, Executive Director, Flowminder
Claire Melamed, Executive Director, Global Partnership for Sustainable Development Data
Jenna Slotin, Senior Director for Policy & Strategy, Global Partnership for Sustainable Development Data
Andrew Young, Knowledge Director, GovLab
Brian Sikora, Senior Director, Analytics and Business Intelligence, Kaiser Permanente
Sang Ick Moon, Director, Sustainability Management Unit, KT Corporation
Haeyeon Jung, Manager, BigData Initiative TF, KT Corporation
Hun Seok Ko, Director, BigData Initiative TF, KT Corporation
Seung Woo Ban, Manager, Sustainability Management Unit, KT Corporation
Gilberto Caldart, President, International, Mastercard
Andrew Reiskind, Senior Vice-President, Lead Privacy and Data Protection Counsel, Mastercard
Linda Raftree, Lead Organizer, MERL Tech
Lauren Woodman, Chief Executive Officer, NetHope Inc.
David Kenny, Chief Executive Officer, Nielsen
Vikram Somaya, Executive Vice-President, Chief Data Officer, Nielsen
Peter Austin, Privacy and Civil Liberties Engineer, Palantir
Hayden Dahmm, Analyst, Sustainable Development Solutions Network
Jay Neuner, Communications Manager, Sustainable Development Solutions Network
Piotr Marczewski, Head of Corporates, Thomson Reuters
Robert Kirkpatrick, Director, UN Global Pulse
Mila Romanoff, Data Privacy and Data Protection Legal Specialist, UN Global Pulse
Stuart Campo, Team Lead, Data Policy and Senior Fellow, UNOCHA Centre for Humanitarian Data
Sarah Telford, Founder and Centre Head, UNOCHA Centre for Humanitarian Data
Rima Qureshi, Executive Vice-President, Chief Strategy Officer, Verizon Communications
Haishan Fu, Director, Development Data Group, World Bank
Holly Krambeck, Senior Transport Economist, World Bank
Laura Tuck, Vice-President for Sustainable Development, World Bank

Endnotes

1. Henke, Nicolaus, et al., *The Age Of Analytics: Competing in a Data-Driven World*, McKinsey and Company, December 2016, <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world> (link as of 3/4/19).
2. Manyika, James, et al., *Open Data: Unlocking Innovation and Performance with Liquid Information*, McKinsey Digital, October 2013, www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information (link as of 3/4/19).
3. United Nations, *A World That Counts: Mobilising the Data Revolution for Sustainable Development*, 2014 <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf> (link as of 3/4/19).
4. Telford, Sarah, *Opinion: Humanitarian World Is Full of Data Myths. Here Are the Most Popular*, Devex, 25 January 2018, www.devex.com/news/opinion-humanitarian-world-is-full-of-data-myths-here-are-the-most-popular-91959 (link as of 3/4/19).
5. UN Global Pulse and GSMA, *The State of Mobile Data for Social Good Report*, 2017, https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/06/Mobile-Data-for-Social-Good-Report_29June.pdf (link as of 3/4/19).
6. Global Partnership for Sustainable Development, *Mobile Data for Social Impact*, 2019, <http://www.data4sdgs.org/index.php/news/mobile-data-social-impact> (link as of 3/4/19).
7. 2019 Edelman Trust Barometer, Edelman, January 2019, <https://www.edelman.com/trust-barometer> (link as of 3/4/19).
8. World Economic Forum, *Our Shared Digital Future*, 2018. The report notes that in the first half of 2018, more than 4.5 billion data records were compromised by malicious actors, resulting in an annual loss of over \$400 billion to the global economy, see www3.weforum.org/docs/WEF_Our_Shared_Digital_Future_Report_2018.pdf (link as of 3/4/19).
9. Recent updates to national data policy regulatory frameworks relevant to data collaboration include the EU's General Data Protection Regulations, Singapore, Kenya and Canada, US.
10. Taylor, Linnet, *What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally*, Big Data and Society, July–December 2017, <https://journals.sagepub.com/doi/10.1177/2053951717736335> (link as of 3/4/19).
11. UN High Level Panel on Digital Cooperation, *Meeting Summary, Consultation Insights and Next Steps*, February 2019, <https://digitalcooperation.org/meeting-summary-consultation-insights-next-steps/> (link as of 3/4/19).
12. 2019 Edelman Trust Barometer, Edelman, January 2019, <https://www.edelman.com/trust-barometer> (link as of 3/4/19).
13. 2019 Edelman Trust Barometer, Edelman, January 2019, <https://www.edelman.com/trust-barometer> (link as of 3/4/19).
14. Kirkpatrick, Robert, *Unpacking the Issue of Missed Use and Misuse of Data*, blog, 2019 <https://www.unglobalpulse.org/news/unpacking-issue-missed-use-and-misuse-data> (link as of 3/4/19).
15. World Economic Forum, *Rethinking Personal Data: A New Lens for Strengthening Trust*, May 2014, http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf (link as of 3/4/19).
16. Zuboff, Shoshana, *The Age of Surveillance Capitalism*, PublicAffairs, 2019.
17. World Economic Forum, *Rethinking Personal Data: A New Lens for Strengthening Trust*, May 2014, http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf (link as of 3/4/19).
18. World Economic Forum, *Data Policy in the Fourth Industrial Revolution*, November 2018, <https://www.weforum.org/whitepapers/data-policy-in-the-fourth-industrial-revolution-insights-on-personal-data> (link as of 3/4/19).
19. Digital Impact Alliance, *Unlocking MNO Data to Enhance Public Services and Humanitarian Efforts*, February 2018, https://digitalimpactalliance.org/wp-content/uploads/2018/02/DIAL_D4D-Report_2018.pdf (link as of 3/4/19).
20. Information Accountability Foundation, *Ethical Data Impact Assessments and Oversight Models*, 2019, <http://informationaccountability.org/wp-content/uploads/Model-Ethical-Data-Impact-Assessment-January-2019-002.pdf> (link as of 3/4/19).
21. Zuboff, Shoshana, *The Age of Surveillance Capitalism*, PublicAffairs, 2019.

22. The Center for Humanitarian Data and OCHA, *Working Draft Data Responsibility Guidelines*, March 2019, <https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf> (link as of 3/4/19).
23. World Economic Forum, ComScore and McKinsey and Company, *Shaping the Future Implications of Digital Media for Society: Valuing Personal Data and Rebuilding Trust*, January 2017, www3.weforum.org/docs/WEF_End_User_Perspective_on_Digital_Media_Survey_Summary_2017.pdf (link as of 3/4/19).
24. The Center for Humanitarian Data and OCHA, *Working Draft Data Responsibility Guidelines*, March 2019, <https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf> (link as of 3/4/19).
25. The Center for Humanitarian Data and OCHA, *Working Draft Data Responsibility Guidelines*, March 2019, <https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf> (link as of 3/4/19).
26. World Economic Forum and Pact, *Civil Society in the Fourth Industrial Revolution: Preparation and Response*, January 2019, http://www3.weforum.org/docs/WEF_Civil_Society_in_the_Fourth_Industrial_Revolution_Response_and_Innovation.pdf (link as of 3/4/19).
27. World Economic Forum, comScore and McKinsey and Company, *Shaping the Future Implications of Digital Media for Society: Valuing Personal Data and Rebuilding Trust*, January 2017, http://www3.weforum.org/docs/WEF_End_User_Perspective_on_Digital_Media_Survey_Summary_2017.pdf (link as of 3/4/19).
28. DeMontjoye, Yves-Alexandre, *openPDS: Protecting the Privacy of Metadata Through SafeAnswers*, 9 July 2014, <https://doi.org/10.1371/journal.pone.0098790> (link as of 3/4/19).
29. *Council on Investments for the Future (The Prime Minister in Action)*, Prime Minister of Japan and Cabinet, japan. kantei.go.jp/97_abe/actions/201705/30article6.html (link as of 3/4/19).
30. Ikeda, Masaru, *Japanese Government Launches Public Data Portal, Aiming to Keep up with Open Data Trend*, The Bridge, December 2013, thebridge.jp/en/2013/12/japanese-public-data-portal-launch (link as of 3/4/19).
31. The Information Accountability Foundation, *Ethical Data Impact Assessments and Oversight Models*, 2019, <http://informationaccountability.org/wp-content/uploads/Model-Ethical-Data-Impact-Assessment- January-2019-002.pdf> (link as of 3/4/19).
32. Rieke, Aaron, et al., *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, February 2018, https://www.omidyar.com/sites/default/files/file_archive/Public%20Scrutiny%20of%20Automated%20Decisions.pdf (link as of 3/4/19).
33. World Economic Forum, *Rethinking Personal Data: A New Lens for Strengthening Trust*, May 2014, http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf (link as of 3/4/19).
34. UN OCHA Centre for Humanitarian Data, *Data Literacy Survey*, 2019, <https://centre.humdata.org/wp-content/uploads/2019/03/Final-Survey-Results.pdf> (link as of 3/4/19).
35. The 100 Questions.org, <https://the100questions.org/> (link as of 3/4/19).
36. Big Data UN Global Working Group, *Using Big Data for the Sustainable Development Goals*, UN Department of Economics and Social Affairs – Statistics Division, <https://unstats.un.org/bigdata/> (link as of 3/4/19).
37. Big Data UN Global Working Group, *Using Big Data for the Sustainable Development Goals*, UN Department of Economics and Social Affairs – Statistics Division, <https://unstats.un.org/bigdata/> (link as of 3/4/19).
38. Brin, David, as cited in World Economic Forum, *Data Driven Development Pathways for Progress*, 2015, http://reports.weforum.org/data-driven-development/?doing_wp_cron=1554206442.1454079151153564453125 (link as of 3/4/19).
39. Fairness Accountability Transparency and Machine Learning, 2018 Conference Papers, <http://www.fatml.org/schedule/2018/page/papers-2018> (link as of 3/4/19).
40. Center for Democracy and Technology, *Digital Decisions*, <https://cdt.org/issue/privacy-data/digital-decisions/> (link as of 3/4/19).
41. Williams, Mike Lee, *New Research on Interpretability*, Cloudera Fast Forward Labs Blog, August 2017, <https://blog.fastforwardlabs.com/2017/08/02/interpretability.html> (link as of 3/4/19).
42. GSMA, *Technical Considerations: Big Data for Social Good Digital Toolkit*, 2019, https://bigdatatoolkit.gsma.com/resources/Big-Data-for-Social-Good_Technical-Considerations-EP3.3.pdf (link as of 3/4/19).

43. *Open Algorithm for Better Decisions*, OPAL, <https://www.opalproject.org/> (link as of 4/4/19).
44. UN OCHA Centre for Humanitarian Data, *Data Literacy Survey*, 2019, <https://centre.humdata.org/wp-content/uploads/2019/03/Final-Survey-Results.pdf> (link as of 3/4/19).
45. GSMA, *Scaling Big Data for Social Good: The Need for Sustainable Business Models*, September 2018, https://bigdatatoolkit.gsma.com/resources/BD4SG_The_need_for_sustainable_business_models.pdf (link as of 3/4/19).
46. Digital Impact Alliance, *Unlocking MNO Data to Enhance Public Services and Humanitarian Efforts*, February 2018, https://digitalimpactalliance.org/wp-content/uploads/2018/02/DIAL_D4D-Report_2018.pdf (link as of 3/4/19).
47. Digital Impact Alliance, *Unlocking MNO Data to Enhance Public Services and Humanitarian Efforts*, February 2018, https://digitalimpactalliance.org/wp-content/uploads/2018/02/DIAL_D4D-Report_2018.pdf (link as of 3/4/19).
48. GSMA, *Scaling Big Data for Social Good: The Need for Sustainable Business Models*, September 2018, https://bigdatatoolkit.gsma.com/resources/BD4SG_The_need_for_sustainable_business_models.pdf (link as of 3/4/19).
49. GSMA, *Technical Considerations: Big Data for Social Good Digital Toolkit*, 2019, https://bigdatatoolkit.gsma.com/resources/Big-Data-for-Social-Good_Technical-Considerations-EP3.3.pdf (link as of 3/4/19).
50. Zuboff, Shoshana, *The Age of Surveillance Capitalism*, PublicAffairs, 2019.
51. 2019 Edelman Trust Barometer, Edelman, January 2019, <https://www.edelman.com/trust-barometer> (link as of 3/4/19).



COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744

contact@weforum.org
www.weforum.org