BUILDING ETHICS INTO PRIVACY FRAMEWORKS FOR BIG DATA AND AI

Executive Summary

Big data, new technologies, and new analytical approaches, if applied responsibly, have tremendous potential to be used for the public good. Big data's greatest value for global development lies in harnessing the power of real-time and predictive analytics for smarter decision making, anticipatory approaches to managing risk, and new ways to measure social impact.

At the same time, big data amplifies risks to privacy, fairness, equality, and due process. Large-scale data collection can expose characteristics and behaviors of individuals, lead to biased decision-making based on unrepresentative or inaccurate data samples, and lack transparency, preventing individuals from exercising due process rights. Organizations have to weigh the benefits of processing data against associated risks, but they often lack a recognized systematic framework to deploy. Also, understanding the complexities of big data analytics requires the engagement of professionals who are versed in an interdisciplinary body of knowledge covering law, technology, ethics, engineering, risk management, and design.

The United Nations Global Pulse and the International Association of Privacy Professionals explored these issues and discussed their many aspects at a jointly hosted event: "Building a Strong Data Privacy and Ethics Program: From Theory to Practice," held in May 2017 in New York (hereafter referred to as the "UN GP/IAPP event").

The event brought together humanitarian and development organizations, including the World Food Programme (WFP), the United Nations Development Programme (UNDP), the United Nations High Commissioner of Refugees (UNHCR), the United Nations Children's Fund (UNICEF), and the International Committee of the Red Cross (ICRC); NGOs; representatives of private sector corporations such as MasterCard, TrustArc, Nielsen, and IBM; as well as privacy regulators from across the world. The presence of stakeholders from different sectors presented an opportunity to view big data analytics not only from the perspective of routine uses of big data for marketing and other business-related purposes, but also in the context of humanitarian and development causes, such as using data to track the emergence and progress of pandemics, to prioritize resource allocation for sustainable development, or to target disaster relief to the most vulnerable populations.

This report provides an overview of how organizations can operationalize data ethics, drawing on the discussions at the UN GP/IAPP event as well as on additional research about data ethics and privacy best practices in a world of big data analytics.

It demonstrates how organizations deploying data analytics and artificial intelligence (aka AI) can reflect ethics considerations in their decision making, borrowing tried and true operational tactics from the field of information privacy. Such steps may include, for example: building a multi-disciplinary team across departments to practice ethics "on the ground"; conducting ethics assessments for new big data projects to consider their personal and societal impacts, while consulting with external and internal ethics working groups; and building programs that are scalable and flexible, which depend on factors such as the societal context of a big data project and the organizational structure of the entity performing it.

Big Data Insights for the Public Good

"The member states of the United Nations set ambitious sustainable development goals [aka SDGs] that do not allow trade-offs between prosperity, the health of the planet, and social progress. We need data to reach these goals and to transform society. However, this data is largely produced by people, often without their knowledge, collected by machines, and owned by the private sector," noted Director of UN Global Pulse Robert Kirkpatrick during his opening remarks at the UN GP/IAPP event.

Similarly, United Nations Special Rapporteur on the Right to Privacy Joe Cannataci, in his most recent report to the UN General Assembly, while noting the risks that come with the use of large data sets, recognizes that "there is broad agreement that big data can produce social benefits including personalized services, increased access to services, better health outcomes, technological advancements and accessibility improvements." There is, he said, a "need to make sense of 'big data,' leading to innovations in technology, development of new tools and new skills."[1]

Numerous pilot and research projects have already shown the feasibility of using data from various sources like mobile phones, social media, financial transactions, or satellite imagery to support the SDGs. For example, following a typhoid outbreak, UN Global Pulse's data science lab in Kampala produced a series of weekly data visualization reports[2] from health centers across Uganda, with interactive maps at district, sub-county, and individual health facility level. The visualizations revealed clusters of the infectious disease, aiding in the allocation of medicine, medical personnel, and health centers.

Private sector companies have similarly started to experiment with using big data for public good. In 2017, the mobile industry association, GSMA, launched a Big Data for Social Good Initiative that currently comprises 19 mobile operators including Vodafone, Telefonica, Telenor, KT, Airtel, and Orange[3]. Twitter has partnered with the UN, through Global Pulse, and is providing access to its data and tools to support sustainable development and humanitarian action.[4] A number of other private sector entities, including financial service providers and retailers, have partnered with the UN to use their data for social impact.

Many participants at the UN GP/IAPP event also noted that society as a whole needs to better realize the potential of big data for public good. Terrell McSweeny, former Commissioner for the U.S. Federal Trade Commission, noted there is "incredible potential for big data analytics to provide benefits to people," such as through increasing educational attainment, supporting access to credit, providing health care solutions tailored specifically to individuals' needs and characteristics, and even helping to build a more diverse workforce. Knowledge gained from large-scale data analytics can greatly add to the public knowledge base to also benefit scholars and researchers.

These applications of big data, coupled with AI and machine learning, can bring widespread benefits. As noted by the UN Secretary-General, "Artificial intelligence can help analyze enormous volumes of data, which in turn can improve predictions, prevent crimes and help governments better serve people. But there are also serious challenges – and ethical issues at stake. There are real concerns about cybersecurity, human rights and privacy – not to mention the obvious and significant impact on the labor market."[5]

Applications of AI can help people with disabilities better navigate everyday challenges, provide workforce support where needed, or reduce traffic and workplace accidents. They can help physicians better diagnose patients or improve delivery of care by reducing

1   Joseph Cannataci (Special Rapporteur on the right to privacy), Report of the Special Rapporteur on the right to privacy, U.N. Doc. A/72/43103 (Oct. 19, 2017).

2   Data Visualization and Interactive Mapping to Support Response to Disease Outbreak, https://www.unglobalpulse.org/projects/mapping-infectious-diseases.

3   See additional examples in Omer Tene & Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw J. Tech & IP 239 (2013).

4   https://www.unglobalpulse.org/news/twitter-and-un-global-pulse-announce-data-partnership.

5   The UN Secretary General's video message to the Artificial Intelligence for the Global Good Summit, https://www.un.org/sg/en/content/sg/statement/2017-06-07/secretary-generals-video-message-artificial-intelligence-global-good.

waiting times. However, they can also be used as weapons to target sensitive population groups and to cause individual harm.

To mitigate some of the risks, the privacy and ethical challenges associated with big data use should be addressed during the data collection and project development phase, as opposed to reactively, after the fact. At the same time, as noted by Teki Akkueteh, former Data Protection Commissioner of Ghana at the UN GP/IAPP event, any risks to individual privacy should be weighed against the benefits of using technology to save lives and improve livelihoods. According to her, "in Africa [for example] people need basic economic rights guaranteed to them. Therefore, it is important for data protection to work in such a way as to assure the achievement of the sustainable development goals."

This is in concert with the findings of the GSMA's "The State of Mobile Data for Social Good Report," which found, "[W]hile there are certain risks and costs associated with using mobile data to produce social good insights, there may also be risks and harms associated with a failure to include this and other new data sources to inform policy, humanitarian response and other development interventions."[6]

Ethics, AI and Big Data

While privacy protection has been a longstanding tool in ensuring responsible and accountable use of personal data, participants at the UN GP/IAPP event noted that in a world of AI and big data analytics, principles for responsible use of data must extend beyond privacy norms.

In today's world, a comprehensive solution for realizing big data and AI benefits for the greater good requires a combination of technical, governance, legal, and ethical responses. This calls for a multidisciplinary approach that draws on the expertise of major players in these distinct, yet complementary fields.

Recently, both private and public sector organizations have started to consider ethics as an additional element for mitigating risks associated with the use of big data. In 2017, the United Nations Development Group, a forum comprising more than 35 UN agencies, came together to craft an approach to big data that is based not only on privacy, but also on ethical and moral obligations concerning data use in development and humanitarian contexts, which it published in its UNDG Guidance Note on "Data Privacy, Data Protection and Ethics: Big Data for the achievement of the 2030 Agenda."[7]

UNICEF builds ethical considerations[8] into its data collection policies by adhering to mechanisms for review, such as internal and external review boards that work to identify anticipated or actual ethical issues that could arise during data collection, and by offering basic ethics training for researchers.

Data ethics can be defined as the branch of ethics that studies and evaluates moral problems related to data use, as explained by the participants of the Data Ethics session at the ITU AI for Good Summit in Geneva in June 2017. "Data ethics must address the whole conceptual space,[9]" including a "diverse set of ethical implications of data science within a consistent, holistic and inclusive framework."

7   https://undg.org/document/data-privacy-ethics-and-protection-guidance-note-on- big-data-for-achievement-of-the-2030-agenda/.

8   Children and the Data Cycle: Rights and Ethics in a Big Data World, https://www. unicef-irc.org/publications/907-children-and-the-data-cyclerights-and-ethics-in-a- big-data-world.html.

9   Luciano Floridi and Mariarosaria Taddeo, What is data ethics?, http://rsta. royalsocietypublishing.org/content/roypta/374/2083/20160360.full.pdf.

6   The State of Mobile Data for Social Good Report (June 2017), https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/06/Mobile-Data-for-Social-Good-Report_29June.pdf.

Data ethics are necessary but insufficient: Incorporation of privacy by design is likewise essential, as applications that employ big data and AI operate with less human supervision. The latter raises new risks for mission critical technologies. A data ethics program could, however, be a core component of privacy by design. In this vein, Ann Cavoukian, who coined the term "privacy by design," has now also proposed seven principles of "AI Ethics by Design"[10] in addition to her privacy by design principles.[11]

Renowned ethicist and Oxford University professor Luciano Floridi defines[12] the field of data ethics as "the branch of ethics that studies and evaluates moral problems related to data, algorithms and corresponding practices … , in order to formulate and support morally good solutions (e.g. right conducts or right values)." Big data and AI require consideration of human rights concepts beyond individual privacy, extending assessment to ethical implications of data use not only on individuals but also on groups of people. In areas like humanitarian response, where fundamental principles of privacy like accuracy are inherently challenged and consent is not a panacea given that benefits are associated with risks, ethical and moral obligations must come into play in addition to well established privacy principles.

The Perils of Inaction

One of the ethical concerns associated with big data is the price of failing to use it for the public good in cases when the possibility to do so responsibly exists.

---

10   See, "AI Ethics by Design," https://www.ryerson.ca/pbdce/papers/.

11   Ann Cavoukian, Privacy by Design, The 7 Foundational Principles, https://www.ipc. on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf.

12   See, Luciano Floridi and Mariarosaria Taddeo, "Introduction: What is data ethics?," http://rsta.royalsocietypublishing.org/content/374/2083.

Learning from medical ethics: A case study

Following a series of highly publicized abuses of human subjects in research experiments in the United States, including the infamous Tuskegee Experiment and the Stanford Prison Experiment, the medical field recognized the need to incorporate ethics into data use frameworks. The disclosures prompted a study called the Belmont Report, which resulted in what is known as "the Common Rule," codified in U.S. federal regulations that apply to 15 government agencies.

The Common Rule incorporates three central ethical principles, which underlie ethical frameworks in many countries around the world. The first, respect for persons, pertains to individual dignity and autonomy and is applied through the concept of informed consent. The second principle, beneficence, requires weighing the risks to people associated with a research activity against the benefits to people and society. The third principle, justice, considers who gets chosen to be a research subject and what population segments stand to benefit — or be harmed — by the research results.

By adopting these ethical pillars, the medical field in the U.S. addressed a major problem afflicting human subject research at the time, namely that engagement with research subjects had shifted away from regarding them as humans capable of suffering harm to viewing them as mere objects or data points.

Because data analytics today may avoid any personal interaction with the data subjects themselves, many big data research projects present similar risks. And because data analytics is now deployed ubiquitously across industries — not just in health care — opportunities for abuse have multiplied exponentially.

Within the UN system, Global Pulse has been working to harness the power of big data and AI to accelerate progress towards the SDGs in a way that is privacy protective, inclusive and fair. Other international organizations, such as the World Health Organization, World Food Programme, International Organization of Migration, The UN High Commissioner for Refugees, the Food and Agriculture Organization, International Telecommunication Union, UN Development Programme, International Committee of the Red Cross, UNICEF, and UNESCO, to name only a few, are likewise already doing innovative work with big data and AI to support the achievement of the SDGs in ways that prioritize protection of privacy.

The ICRC "Handbook on Data Protection in Humanitarian Action"[13] describes how humanitarian relief efforts such as "mapping or identifying" patterns of events in emergencies or situations of violence can be greatly enhanced when data analytics are used to recognize general crisis patterns or identify individuals and communities in need. The UN Global Pulse Advisory Group, in its report "Big Data for Development: Towards Responsible Governance," suggests that big data analytics raises concerns about risks of harms, fueled by the lack of proper guidance, especially to address the risks in development and humanitarian contexts and balance them with data rewards. This often prevents effective harnessing of data when it is most critical[14].

Privacy is a fundamental human right, and data use must be privacy protective. However, any right, including the right to privacy, should be assessed in light of all human rights, such as the right to life, food, shelter, health, and education. As noted by the UN Special Rapporteur on the right to privacy, "All of [the human] rights are important and commitment to one right should not detract from the importance and protection of another right. Taking rights in conjunction wherever possible is healthier than taking rights in opposition to each other."[15]

Robert Kirkpatrick summed it up at the 2018 ITU AI for Good Global Summit: "As global efforts to develop new frameworks around the responsible use of emerging technologies begin to take shape, it is imperative that they address not only the human rights implications of 'misuse,' but also those of the 'missed use.'" In considering the risks of data use, therefore, professionals must take into account not only the potential risks and harms that could result from improper action, but also those that could arise from inaction.[16]

"Misinformed" Decision-Making and Other Consequences of Unethical Use

While bearing great potential for public good, data analytics may also lead to poor decision making, and unethical data practices can produce harmful results. Algorithms that fail to correct bias or discrimination, or data sets built on inaccurate information, could lead to flawed results that may deprive people of crucial opportunities or services[17] — for example, excluding minority demographics from ads for housing or employment. Big data can present ethical issues and social dilemmas arising from poorly considered use of technology. The unintended consequences of big data are a constant risk that should be accounted for in the design process of projects and programs.

13   International Committee of the Red Cross, Handbook on Data Protection in Humanitarian Action (Christopher Kuner and Massimo Marelli, eds., 2017), https://brusselsprivacyhub.eu/publications/dataprotectionhandbook.html.

14   UN Global Pulse, "Big Data for Development: Towards Responsible Governance,", available at https://www.unglobalpulse.org/news/big-data-development-and-humanitarian- action-towards-responsible-governance-report.

15   Report of the Special Rapporteur on the right to privacy, Joseph A. Cannataci. 8 March. A/HRC/31/64, pp. 6, 10. Annex II. A more in-depth look at Open Data & Big Data.

16   See Omer Tene & Gabe Maldoff, The Cost of Not Using Data: Balancing Privacy and the Perils of Inaction, Journal of Law, Economics & Policy (forthcoming 2018).

17   UN OCHA, Building data responsibility into humanitarian action, https://www. unocha.org/publication/policy-briefs-studies/building-data-responsibility-humanitarian-action.

As the ICRC Handbook explains, the use of data analytics can lead to misleading and inaccurate results, and can justify "arbitrary and automated decisions that do not take case-specific particularities into consideration." Sometimes big data projects facilitate effective surveillance through digital footprints or violate privacy through re-identification of individuals who may have participated in a dataset on the promise of anonymity.

There may be a need for a codified set of data ethics because datasets can no longer be considered static archives; rather they can be reused to generate new insights and consequences for individuals. For example, as described in the Risks Harms and Benefits Assessment Tool[18] of UN Global Pulse, a new data set created by an algorithm may by itself become a risk.

While the risks may create unaccounted harms for humans, many concerns over legal impediments and ethical restrictions also diminish productive collaboration between researchers, public sector organizations, and private sector businesses for the greater social good. This, in turn, risks depriving society of an opportunity to live better and safer — another unaccounted harm on its own.

Using Old Data Through New Means

When researchers interact with data gathered for other purposes, ethical issues might arise. These ethical issues should be weighed in light of the value of these data uses to society as a whole.

Speaking at the UN GP/IAPP event, Daniel Goroff, VP and program director at the Alfred P. Sloan Foundation, emphasized the need to have clearly defined research goals with demonstrated potential to advance

the public good in order to justify privacy sacrifices."[19] However, some participants at the event also noted that, in some cases, researchers may have to begin the process with feasibility studies based on aggregate statistics, for example, to determine whether more detailed research could potentially lead to scientifically significant outcomes. Attendees nevertheless agreed that incentives to give up privacy rights in exchange for research that may eradicate a serious disease or lead to more efficient humanitarian assistance are much stronger than those supporting research to improve product placement.

No doubt, data research that requires privacy compromises must be justified by a benefit to the data subject or society at large. Ethics in big data require engaging in a risk analysis weighing the potential value to society versus individual rights and interests. Codified by the UNDG Guidance Note on Big Data Privacy, Ethics and Data Protection, such risk analysis should embody a utilitarian ethic where "any potential risks and harms should not be excessive in relation to the [likely] positive impacts of data use."[20]

Programming Behaviors

When decisions are automated, they might raise ethical considerations, in particular with respect to the decision-making methodology. A common example involves the application of the age-old "trolley problem" for self-driving cars, which will inevitably encounter hard choices affecting human lives. Depending on what

18   https://www.unglobalpulse.org/privacy/tools

19   Also see Daniel Goroff, Jules Polonetsky & Omer Tene, Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data, Annals of the American Academy of Political and Social Science 675 (January 2018).

20   UNDG Guidance Note on Big Data for Achievement of the 2030 Agenda: Data Privacy, Ethics and Protection. https://undg.org/document/data-privacy-ethics-and-protection-guidance-note-on-big-data-for-achievement-of-the-2030-agenda/; also see Jules Polonetsky, Omer Tene & Joseph Jerome, Benefit-Risk Analysis for Big Data Projects, Future of Privacy Forum White Paper, September 2014, https://fpf.org/wp-content/uploads/FPF_DataBenefitAnalysis_FINAL.pdf.

machines are designed to do, moreover, questions arise[21] regarding displacement of the human workforce; the deployment of machines in certain arenas, such as armed conflict; and the level of intelligence society is willing to accept in machines.

In light of the myriad ethical issues raised by data analytics and AI, professionals working with organizations using big data should have a basic understanding of data ethics and tools for incorporating it into decision making. Big data draws on the fields of physics, computer science, and applied mathematics, disciplines that "have not been required to practically grapple with ethics requirements, [and therefore] they often lack access to pedagogical resources about research ethics that are widespread in other fields."[22] Hence, there is a growing need to make data ethics a requirement in the education of not just students of the humanities but also technologists and those who use data to make decisions.[23]

Tools for Thinking about Data Ethics — and Privacy

Data ethics involves organizations investing resources to build internal programs to weigh and balance benefits and risks to individuals from big data and AI uses. According to Emory University ethicist Paul Root Wolpe, data ethics addresses considerations of bias (often inherent in the choices made when selecting data subjects, building a data set, and deciding on research methods); data

ownership and the rights to control data use; and power imbalances between an organization collecting data and individual data subjects. It is true that, at minimum, awareness of unintended consequences can mitigate potential sources of bias. One of the tools recommended by the Institute of Electrical and Electronics Engineers to reduce bias is for "designers [to] take on an interdisciplinary approach and involve relevant experts or advisory group(s) into the design process … when designing for dynamically vulnerable populations."[24]

Data ethics also incorporates privacy. This encompasses the protections an organization puts in place around how the data is collected, used, and shared. For example, data analytics in the context of a humanitarian effort may create a conflict between personal privacy interests on the one hand — reflected in the purpose limitation and data minimization principles — and the interests of the broader community or society on the other hand, served by observing patterns and predicting outcomes.

When data is gathered from individuals in response to a particular incident, it may later become useful in analysis to evaluate trends for purposes of, say, providing adequate emergency response and disaster relief, or matching lost persons with their families. Data protection principles caution against any further processing that might lead to identifying an individual or leaking sensitive data about him or her, as such disclosures could lead to harmful personal consequences. Yet, data analysis could at the same time empower humanitarian efforts to reduce or eliminate future harm to numerous individuals by allowing for conflict intervention or rushing food aid to those on the brink of famine. Reconciling the clash between these competing public goods — individual privacy protection vs. societal benefit — is at the heart of data ethics.

21   Burton, et al., Ethical Considerations in Artificial Intelligence Courses, https:// arxiv.org/ pdf/1701.07769.pdf?imm_mid=0ed017&cmp=em-data-na-na-newsltr_ ai_20170206.

22   Jacob Metcalf, Emily F Keller, and Dannah Boyd, Perspectives on Big Data, Ethics, and Society, The Council For Big Data, Ethics, And Society (2016), http://bdes.datasociety.net/ wp-content/uploads/2016/05/Perspectives-on-Big-Data.pdf.

23   Dennis Hirsch & Keir Lamont, The Age of the Cyber Pro, IAPP White Paper, 2018, available at https://iapp.org/media/pdf/resource_center/Age-of-the-Cyberpro.pdf.

24   The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1. IEEE, 2016. http:// standards. ieee.org/develop/indconn/ec/autonomous_systems.html.

In 2016, UN Global Pulse developed a Risk Assessment Tool — a compliance mechanism that includes elements of both privacy and data ethics for understanding and managing risks, harms and benefits associated with big data in development and humanitarian contexts. The guidelines to using the tool encourage organizations to include, where reasonably practical, a representative of the individuals or groups who are potentially affected by data use in the deliberations. Mila Romanoff, Privacy Specialist at UN Global Pulse, explained that involvement of representatives of affected populations or consideration of their specific context, culture, social layer etc., provides a better ground for an ethical decision in situations that are not clear cut.

Privacy professionals have been deployed on the ground[25] across jurisdictions and industries to implement privacy programs and raise awareness throughout the data and technology space. As Deirdre Mulligan of UC Berkeley School of Information explained at the UN GP/IAPP event, this has for example involved engaging management of private businesses to see privacy as a strategic issue, thus incorporating privacy into daily decision-making processes, and more profoundly, into a company's DNA.

At the UN GP/IAPP event, participants discussed how privacy professionals incorporate ethics into their data governance programs. Fortunately, while answers do not come easily, building a process to generate the right questions about data ethics — and to put them in front of the right professionals — shares many features with traditional privacy programs. As Bonnie Green from the World Food Programme put it, ethics and privacy should be "mainstreamed" together: "We have risk assessments and humanitarian principles. We should continue to use them as ways to assess and incorporate privacy and ethics, together." Considerations of ethics are also included in the WFP Guide to Personal Data Protection and Privacy.[26]

Model Data Ethics Structures

Various existing frameworks can supplement organizations' decision-making processes to address the newly emerging field of data ethics. These frameworks offer different perspectives and governance options to organizations with weighty data ethics questions or large-scale processing of sensitive data.

Internal Frameworks

Organizations can choose from a variety of tools to implement data ethics considerations. Many tools are derived from more traditional frameworks for privacy impact assessments (often called PIAs, or DPIAs, for data protection impact assessments, in Europe especially). The Information Accountability Foundation's Comprehensive Data Impact Assessments (or CDIAs)[27] address uses of big data not clearly authorized by either consent or legislation. The various CDIA forms can be tailored to specific countries and serve as a framework to address gray areas of data use that may not be illegal but are unethical.

Some organization also task their internal privacy working group with data ethics considerations. Several organizations have established interdepartmental privacy working groups to consider how to properly and effectively handle privacy challenges with big data. Rather than establishing a separate working group for data ethics, it may be efficient to add data ethics to the remit of those groups in order to take advantage of their structure and expertise.

---

25   Kenneth Bamberger and Deirdre Mulligan, Privacy on the Books and on the Ground, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1568385.

26   https://docs.wfp.org/api/documents/e8d24e70cc11448383495caca154cb97/ download/

27   IAF, Comprehensive Data Impact Assessment Master Project, http://informationaccountability.org/comprehensive-data-impact-assessment-master- project.

Internal frameworks can be helpful in promoting "inward facing goals." Professor Jacob Metcalf, who runs a National Science Foundation-funded multisite project, Pervasive Data Ethics for Computational Research, defines an "inward facing goal" as an ethical framework that "provides guidance when an existing inexplicit norm or value is insufficient." After completing a cross discipline of numerous ethical codes, Metcalf noted that codified internal frameworks are "beneficial for creating generalized rules for individuals and organizations that have responsibilities for important human goods." Further, internal frameworks can establish role-specific guidelines and standards of behavior for academics, research students, and clients.[28]

Another approach could be modeled on the idea of an internal ethics board, similar to an academic internal review board (often called an "IRB"). A multidisciplinary board weighs the potential benefits of a project against any attendant risks. This system is an example of a practical approach at the intersection of privacy and data ethics. Here flexibility is valued "to accommodate the fast moving nature of research projects, and the importance of steering away from the attempted application of one strict set of standards to all departments and teams."

Researchers "conduct analysis of a wide array of data sources, from massive commercial or government databases to individual ... postings publicly available online, with little or no opportunity to directly engage human subjects to obtain their consent or even inform them of research activities."[29] In such situations, where obtaining data subject consent is not feasible or practical, an internal IRB can be one of the risk mitigation tools.

IRB-like mechanisms could significantly improve the inclusion of ethical, moral and human rights concerns in decisions about data analysis. This may be complemented by a need to consult the affected individuals or groups for a more informed decision.

UN Global Pulse's Risks, Harms, Benefits Assessment[30] suggests the use of "a diverse team comprised of the project leader as well as other subject matter experts, including — where reasonably practical — a representative of the individuals or groups of individuals who could be potentially affected." The aforementioned ITU AI for Good Summit produced recommendations to include historically underrepresented groups in the development process of AI systems, thereby ensuring AI responds to the broadest society needs and considerations.

External Frameworks

External ethics review boards may be appealing for small and medium-sized businesses[31] that lack the scale required to sustain an internal ethical review board. Additionally, an external ethics review board could help develop ethical standards and best practices, providing an accessible knowledge base. Finally, external frameworks can promote "outward facing goals," described by Professor Metcalf as goals to "protect vulnerable populations, protect the reputation and trust of the profession by providing a basis for public expectations, [and allow for an] evaluation of the profession."[32]

28 Jacob Metcalf, Ethics Codes: History, Context, and Challenges, The Council for Big Data, Ethics, and Society, (2014) http://bdes.datasociety.net/wp-content/ uploads/2016/10/EthicsCodes.pdf.

29 Omer Tene and Jules Polonetsky, Beyond IRBs: Ethical Guidelines for Data Research, 72 Wash. & Lee L. Rev. Online 458 (2016).

30 UN Global Pulse, Tools: Risk, Harms and Benefits Assessment, https://www. unglobalpulse.org/privacy/tools.

31 Tene and Polonetsky, supra, n. 30, available at https://scholarlycommons.law.wlu. edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1044&context=wlulr-online.

32 Jacob Metcalf, Ethics Codes: History, Context, and Challenges, The Council for Big Data, Ethics, and Society, (2014) http://bdes.datasociety.net/wp-content/ uploads/2016/10/EthicsCodes.pdf.

One example of how external review boards could function in practice comes from the idea of Consumer Subject Review Boards (CSRBs) proposed by Ryan Calo.[33] The concept involves forming outside panels that can review organizations' proposed projects and suggest adjustments to meet industry standards. These review boards could be industry-specific and composed of experts from diverse practices so that multiple viewpoints are represented in the evaluation process. Though external IRBs may have limits regarding how specific they can be in evaluating risk and rewards due to the independent nature of the boards and presumed volume of requests they will receive, they could serve as a useful tool to promote transparency and provide useful feedback.

One example of such an external advisory group in the humanitarian and development sector is the UN Global Pulse Privacy Advisory Group, comprising regulators, experts from the private and public sectors, civil society and academia. The Group was established as part of Global Pulse's advocacy and policy work around accelerating responsible data innovation for social good within the UN system. Its experts provide input on various big data initiatives for social good, including on privacy policies, guidelines and tools. Another example is the EDPS Ethics Advisory Group. As explained by Depphine Harou of the European Data Protection Group, this group explores "the ethical dimensions of data protection and the relationships between human rights, technology, and business models."

There are additional examples of cross-industry initiatives examining ethical implications of AI. The Partnership on AI to Benefit People and Society (aka, "The Partnership") was established to develop and share best practices within the AI industry. The Partnership includes large tech companies such as Google, Facebook, Amazon, IBM, Microsoft, and Apple. The Institute of Business Ethics, a non-profit professional group based in London, offers a range of services from certifications to ethical training and tools to implement on an individual and group basis. The European Business Network for Corporate Social Responsibility is similar, but has a broader approach, rather than ethics-specific.

Other proposals for external frameworks are distinct from the traditional consulting board. One approach suggests delegating data access entirely to an external body. Administrative Data Research Facilities (known as ADRFs) store and regulate access to already-existing data sets. When an organization or researcher wishes to use a data set, they must submit a request to the facility. The facility then evaluates the request, guided by the principle that the benefits to policy-making must outweigh the costs to individuals. The facility has freedom to determine if data sets will be available to the applicant and in what form — for example, after de-identification, via a data enclave, or only after signing a series of non-disclosure agreements. One example of this approach is the Kilts Center for Marketing, a partnership between the University of Chicago Booth School of Business and the Nielsen Company. The center "makes comprehensive marketing datasets available to academic researchers around the world," using data donated by Nielsen.

Finally, an outside ethical advisory board could be based in a geographic area. One example is the City of Seattle's Community Technology Advisory Board (known as the CTAB), which comprises local community members and subject-matter experts along with representatives from different-sized and -focused businesses. The board meets monthly and discusses best approaches to new and emerging technologies and any conflict or risk that arises with use of older technology or data. The minutes of these meetings, contents, and reports are then made public.[34]

33   Ryan Calo, Consumer Subject Review Boards, 66 Stanford Law Review Online 97 (2013); also see Omer Tene & Jules Polonetsky, Privacy And Big Data: Making Ends Meet, 66 Stanford Law Review Online 25 (2013).

34   City of Seattle Open Data Risk Assessment, FPF Report, January 2018, available at https://fpf.org/2018/01/25/examining-the-open-data-movement/.

Operationalizing Data Ethics

Accountability is the backbone of any data management program, including where data analytics — and therefore ethical issues — are involved. Transparency is an important element of accountability, including in AI, "because transparency builds trust in the system, by providing a simple way for the user to understand what the system is doing and why." In order to maintain transparency, the IEEE recommends the development of new standards that describe measurable, testable levels of transparency, so that "systems can be objectively assessed and levels of compliance determined."[35]

One mechanism for enterprise-level accountability is the market discipline imposed by reputation-harming consequences of poor publicity and consumer complaints. In Japan, Professor Hiroshi Miyashita explained, companies that behave unethically are "named and shamed" and must issue public apologies to make up for their mistakes and save face. Of course, a better method is to prevent ethics missteps before they occur. To do this, organizations must develop data ethics frameworks and adhere to them throughout the data flow process.

When deploying an effective data ethics structure, organizations should begin by asking a series of questions designed to tailor existing frameworks to their organization. The participants in the discussions at the UN GP/IAPP event suggested that organizations consider three factors in operationalizing ethics: 1) a flexible approach to creating a privacy and ethics framework; 2) data ethics leadership within the organization; and 3) establishing tools for ethics impact assessments or risk assessments that incorporate ethics, to consistently evaluate company-wide ethics approaches.

A flexible and holistic approach

Both data ethics and privacy are contextual and require a flexible approach tailored to an organization's mission, structure, and management style, as well as to the sensitivity of the data involved.

Privacy structures differ greatly between different organizations and can vary across departments within the same organization. Sales and marketing teams, focused on revenue generation, have a different perspective from engineers who are constantly trying new ideas, or lawyers who attempt to mitigate legal risk. Each of these departments requires an adapted privacy framework and different training to ensure relevance and a complete approach to organizational privacy. Although a flexible approach that is tailored to each individual organization is practical, the IEEE advocates that every framework should at its core "be designed and operated in a way that respects human rights, freedoms, human dignity, and cultural diversity."[36]

Similarly, how to deploy ethics frameworks will vary between departments. While all employees should be aware of the overarching ethics goals and standards of the organization, specific training must be flexible and role-based to be most effective. Further, a uniform ethics framework will not be universally effective across all organizations. Different types of data require different standards of care. As Christina Peters, IBM's Associate General Counsel, put it at the UN GP/IAPP event, "the right model for operationalizing a privacy and ethics program depends on context" and "one size is not going to fit all."

Hillary Wandall, General Counsel and Chief Data Governance Officer at TrustArc, added that when she worked at Merck "ethics and ethical decision-making were at the heart of the privacy program." She

35 The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1. IEEE, 2016. http:// standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

36 The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1. IEEE, 2016. http:// standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

explained that at Merck, a research-driven pharmaceutical company, data collected (with informed consent) for one purpose might prove useful in another, and it wasn't satisfactory to simply say, "the data is anonymous." The company worked to develop a comprehensive data management program that applied accountability throughout the organization from the highest level of the company to every department where personal data was used.

JoaAnne Stonier, Chief Data Officer at MasterCard, acknowledged another challenge to data management, which is the urge to "do good" with data when a company is approached by an academic researcher or non-profit with a social justice mission. At such times, organizations with large data sets need to have a protocol that can be followed by employees as well as management: "Both of them could be violating certain laws in some jurisdictions, let alone ethics or human rights concerns." The IEEE[37] recommends, when applicable, "that designers and developers alike document changes to the systems in their daily practice … with the highest level of traceability to document changes and behaviors in the system."

The need for a recognized protocol to determine how to ethically treat requests for data sets is thus another key component of a data ethics framework that would need to be assessed by circumstance.

At the AI For Good Global Summit, participants in the "Ethics of AI" session proposed a set of recommendations regarding the ethical development of AI. Some of the recommendations included: a) defining a transparent open data policy; b) designing ethics evaluation procedures; c) developing educational tools that could communicate the capabilities and limitations of AI to policy makers, the general public, and business executives; and d) engaging in a multi-faceted dialogue with various stakeholders to

improve the design process, awareness and attitude towards the applications of AI for social good.

Leadership

Without leadership, privacy and ethics protocols may never be developed, distributed, and enforced throughout the organization.

Chief privacy officers often take up the mantle of consumer protection and advocacy on privacy, so it stands to reason that there needs to be someone who does the same for data ethics. IBM's Peters stressed the importance of integrating data privacy into a data governance strategy, within an overall business strategy. Successful privacy and ethics programs require management "buy-in," meaning a willingness to invest time and resources into the programs. This initial investment pays off by avoiding costly privacy violations and empowering employees to make informed privacy decisions and streamline the response process. For example, IBM's data governance program uses a software-enabled tool to help employees throughout the multi-national organization answer routine privacy questions.

Creating this institutional framework and developing a privacy- and ethics-sensitive culture requires leadership. And "leadership," as the ethicist Wolpe put it, "makes all the difference." Privacy and ethics issues have to "pervade the organization."

The data ethics leader should function with relative independence. For example, the model of the data protection officer is described in the European Union's General Data Protection Regulation as a form of internal auditor with a data subject ombudsman-like role. Another example of a more formal governance mechanism is of an internal ethics leader, an external structure for ethical guidance, or a combination of the two.

37 IEEE, Embedding Values into Autonomous Intelligent Systems, https://standards. ieee.org/develop/indconn/ec/ead_embedding_values.pdf.

An internal ethics leader should be able to communicate with all departments within the organization and should be kept informed at each stage of the data gathering and use process. In this way, this position must function similarly to a chief privacy officer or perhaps even be the privacy lead.

If an organization decides to use the expertise of an external structure for ethical guidance, the organization will still require an internal employee or corporate structure to properly incorporate guidance from the external structure. For most organizations, the chief privacy officer is currently the most likely internal conduit for this operation.

Teamwork

Data ethics should be assessed at each point in the data life cycle (development, storage, data collection, analysis, etc.). To that end, every employee should have basic data ethics training. Additionally, representatives of each relevant department should be consulted when making data decisions. When proposing a new data research project, for example, the data ethics leader should consult with the programmers who would create the data collection mechanism, the researchers proposing the project, public relations and communications representatives who will present findings and field criticisms, and any other departments involved.

In both internal and external discussions, input from differing perspectives is essential. Elizabeth Buchanan, an ethicist at the University of Wisconsin-Stout, noted[38]: "[I]t's really important for boards to have either a data scientist, computer scientist, or IT security individual. … It's not just thinking about someone getting upset about questions on a survey."

---

38 Wired, Scientists Are Just as Confused About the Ethics of Big-Data Research as You (May 20, 2016), https://www.wired.com/2016/05/scientists-just-confused- ethics-big-data-research/.

Ethics Impact Assessments

Effective privacy programs incorporate an evaluation at every new stage: development, research, access, storage, sharing, and testing. One key example of this is in the privacy impact assessments referenced earlier. PIAs serve to routinely evaluate an organization's procedures for collecting, storing, analyzing, and disposing of sensitive data. Separate PIAs are recommended for every part of the data life cycle. They serve as a way to ensure best practices, evaluate risks, prevent missteps, and establish that the organization takes privacy concerns seriously.

PIAs allow organizations to effectively conduct risk management, ensuring compliance with privacy requirements, identifying mitigation measures, and effectively classifying the impacts on the use of the data. By classifying both the positive and negative impacts, organizations have an additional opportunity to assess ethical considerations and ensure stronger privacy compliance.

An ethical equivalent that directly addresses ethical goals — an ethical impact assessment or a CDIA, as earlier referenced — could serve as a useful tool as well. This could be achieved by incorporating ethics into a PIA-style process. The UN's Romanoff noted that "human rights" — an aspect of data ethics — is a consideration built directly into the privacy impact assessments of UN Global Pulse. "Part of having the right framework for responsible big data use in place," she said, "is having everyone in the organization recognize and be aware of the potential risks linked to data use. Considering ethics in risk assessment can provide additional insights on the negative and positive impacts of data use as well as its non-use." Building ethics and human rights questions into risk assessments may, for many organizations engaged in data analytics, be more effective than developing a separate parallel track for ethical reviews.

Human rights and sector-specific considerations

Finally, when forming a data ethics framework, organizations must take into account sector-specific requirements or guidance. These standards may come from experts in a particular field, from industry associations, or from professional organizations within a field.

Ethical approaches to big data and AI should include protection of human rights. Decisions to use or not to use big data and AI can have an adverse impact on the protection of fundamental rights. By employing a human rights-based approach to big data and AI, adverse impacts can be mitigated.

Building the future privacy — and ethics — program

Privacy professionals need a broad range of skills — technical, ethical, and legal.

Privacy lawyers need to possess awareness of privacy, technology, and ethics. Technologists require legal, ethical, and privacy knowledge. Ethicists need privacy, legal, and tech awareness.

Though professional training and certifications can assist in providing knowledge and skills in needed areas, educational institutions and businesses must recognize the need for a comprehensive approach to privacy. Several institutions are already combining tech, ethics, and law in innovative ways to create multidisciplinary professional programs built to meet modern workforce needs.

Professor, ethics, and human rights expert Lorna McGregor suggested at the AI for Good Global Summit that AI and humans have different forms of reasoning. "However, autonomous decision making should not provide a cover for human responsibility and error," she said. "Humans should retain control over AI systems and ensure internal and external oversight." Each actor at each stage of development and application of AI should remain responsible, thereby allowing specific actors within their respective sectors to better determine the ethical use of AI.

Various data ethics frameworks should have common features to ensure a uniformly high ethical standard of data practices. However, these frameworks will be most effective if they are flexible enough to be tailored for each specific company and organization, adjusting for a company's size, resources, subject matter area, and impact on data subjects. Overall, ensuring ethical parity between commercial applications and deployment of data for social use is crucial to safeguarding the ethical development of new technologies across all sectors and will lead to greater trust in technology and acceptance from the general public.

Conclusion

Data ethics serves to address areas of consumer protection and data stewardship that are beyond the reach of the traditional privacy practices. A model privacy program's existing structure, however, serves as an ideal foundation for adding ethical checks and balances.

With the proper investment in ethics leadership and awareness training, along with adding ethics assessments and frameworks to existing privacy programs, organizations will go a long way toward finding the appropriate balance between protecting consumer information and trust, and using big data to its fullest potential and for the greater good.

The UN GP/ IAPP event brought together experts from various fields, acknowledging the fact that there is a need to approach data use from a multidisciplinary perspective and that collaboration between various stakeholders is necessary and fundamental for big data to be a transformative force for good.

The Secretary-General of the United Nations recently in his Web Summit 2017 speech noted that "to avoid the mistakes of the past" and to "maximize the potential of the enormous advantages that the innovations we are discussing can provide to our world," it is important to build collaboration and dialogue between stakeholders. Developing stronger frameworks for data privacy and ethics in general will also, no doubt, require efforts from different experts in numerous sectors. And no doubt we should remember that, "just as misuse of AI may lead to harm, non-use of AI may allow preventable harms to occur."[39]

39 Information and Communication for Development (ICT4D 2018) Data Driven Development, Section on Big Data for Social Good, Digital Ethics, M. Romanoff, M. Luengo Oroz, R. Kirkpatrick, P. Biggs, F. Vacaleru (World Bank, 2018- pending).