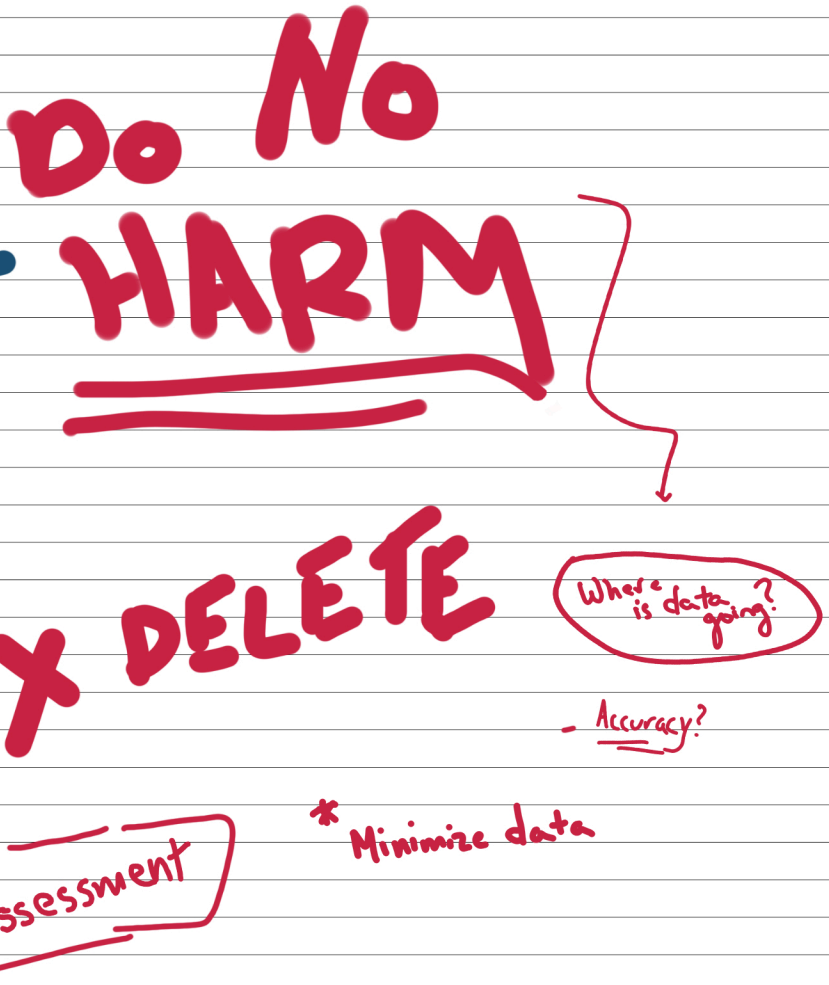
RESPONSIBLE DATA FRAMEWORKS:

IN THEIR OWN WORDS



JUNE
2018

CENTER FOR DEMOCRACY & TECHNOLOGY

CENTER FOR DEMOCRACY & TECHNOLOGY

Responsible Data Frameworks: In Their Own Words

June 2018

Nongovernmental organizations focusing on public interest, development, and humanitarian aid have sought to collect a vast amount of personal information. Some of this is the byproduct of viewing more data as an important tool for providing better, more targeted services and to better serve their charges; organizations have also been forced to collect more information to demonstrate accountability and results to funders. In order to address these trends, there has been a significant push in recent years to policies, procedures, and protections for using data about and on behalf of beneficiaries.

Nonprofit public interest organizations tend to have missions that prioritize societal good. This makes them potentially well-positioned to develop responsible data collection and use policies that are state-of-the-art in terms of upholding ethical principles and respecting and protecting data subjects, including vulnerable populations and beneficiaries of aid. But many nonprofit organizations are not equipped with the resources, training, or expertise needed to implement sophisticated legal, technical, and ethical compliance regimes or to understand how certain data collection, use, and sharing activities could put the communities they serve at risk. Understanding how data – by itself – can create the risk of liability for organizations is an additional challenge.

A growing effort within academia and civil society aimed at responsible data governance has led to the development of principles and guidance for how data should be collected, used, and shared in ways that maximize value and minimize harm to beneficiaries and other vulnerable individuals. According to the U.S. Agency for International Development (USAID), "responsible data" attempts to recognize the tensions among privacy protection and data security, use of data for decision-making, and transparency and openness considerations.[1] As a result, it is something of a broad concept involving the entire lifecycle of information and the interests of beneficiaries;[2] the Responsible Data Forum, which has played a key role in promoting the concept, explains further that responsible data duties include (1) ensuring people's rights to consent, privacy, security, and ownership; (2) protecting information processes, including collection, analysis, storage, presentation, and reuse of data; and (3) respecting values of transparency and openness.[3]

---

[1] USAID, An Introduction to USAID's Responsible Data Work (Jan. 24, 2018), https://www.usaid.gov/digital-development/responsible-data/introduction.

[2] Linda Raftree, How to Develop and Implement Responsible Data Policies, ICTworks (Nov. 21, 2016), https://www.ictworks.org/how-to-develop-and-implement-responsible-data-policies/.

[3] Responsible Data Forum, About, https://responsibledata.io/about/.

Resulting guidance on data governance tends to be similarly multi-pronged, capturing overarching ethical questions, long-standing fair data practices, institutional accountability mechanisms, and, more recently, the types of privacy and security controls instituted by private industry and mandated by data protection regulations.[4] USAID is currently engaged in a detailed literature review of existing of responsible data practices,[5] but a recent survey from SIMLab suggests that there is a baseline level of confusion among organizations as to what responsible data governance means.[6]

"Resulting guidance on data governance tends to be similarly multi-pronged, capturing overarching ethical questions, long-standing fair data practices, institutional accountability mechanisms, and, more recently, the types of privacy and security controls instituted by private industry and mandated by data protection regulations." ----------

To map the work that academics, civil society, and government agencies have done to develop principles for responsible data use in the nonprofit sector,[7] this paper reviews 18 frameworks. Many were largely based on two sets of foundational principles: the Fair Information Principles (FIPs) and the ethical frameworks around human subjects research embodied in the Belmont Report. Part I of this literature review discusses these foundational principles. Part II discusses how responsible data frameworks combine FIPs-based data protection principles and research ethics principles to form a baseline framework for responsible data. As one of the first examples of an organization attempting to implement a transparent data protection policies, we use Oxfam's Responsible Data Policy as an exemplar of how organizations are embracing responsible data governance.

Part III reviews 18 data use frameworks and organizes their principles into six common themes that exist across the frameworks. These six themes, which sometimes overlap, are:

[4] For example, a collaborative team is currently working to develop responsible data practices for USAID, and the effort intends to address ways to:

- Mitigate privacy and security risks for beneficiaries and others;
- Improve performance and development outcomes through use of data; and
- Promote transparency, accountability and public good through open data.

A description of the effort is available at: https://lindaraftree.com/2017/02/06/responsible-data-case-studies/.

[5] USAID, *supra* note 1.

[6] Laura Walker McDonald & Kelly Church, Good Data Collaborative Consultation Report (Nov. 1, 2017), http://simlab.org/resources/dogooddata/.

[7] We reviewed 20 different frameworks and accompanying materials/documents as part of this review.

1. *respect for individual rights and autonomy, which includes concepts such as consent and access to one's personal information;*
2. *fairness or justice, as in distribution of resources;*
3. *beneficence and the necessity of assessing the risks and benefits of collecting or using data;*
4. *FIPs-based privacy and data protection principles, including data minimization;*
5. *transparency and accountability for information practices; and*
6. *information security.*

Part IV briefly discusses international issues addressed or absent in the frameworks.

This review suggests that a common lexicon has emerged with respect to the principles that constitute responsible data governance. The challenge ahead is what this should mean in practice. Future resources must be less high-level and more context-dependent and organization specific, warranting the need for scalable strategies and details into what strong accountability mechanisms might look like. These may need to be tailored to either identified or describable risks, which may require imbuing nonprofit organizations with a broader conception of risks from data collection and use than is currently understood.

"The challenge ahead is what this should mean in practice. Future resources must be less high-level and more context-dependent and organization specific, warranting the need for scalable strategies and details into what strong accountability mechanisms might look like." ----------

I. Foundations

The responsible data movement brings together overarching ethical concerns about robust collection of sensitive data about vulnerable populations with data protection frameworks. As a result, the principles and strategies that emerge embody long-standing elements that appear in the FIPs and in ethical requirements for human subjects research as articulated in the foundational 1979 Belmont Report. This section explains each of these frameworks briefly in turn. It is worth acknowledging at the forefront that ethical frameworks and other codes of conduct frequently emerge in response to major scandals or questionable activities.[8] Both the FIPs and the ethical framing that led to the U.S. Federal Common Rule were products of public concerns and headline-grabbing stories about how individuals

[8] Jacob Metcalf, Ethics Codes: History, Context, and Challenges, Council for Big Data, Ethics, and Society (2014), http://bdes.datasociety.net/council-output/ethics-codes-history-context-and-challenges/.

could be mistreated and have their own information used against them.[9] Nonprofit public interest organizations now face this challenge.

A. Fair Information Principles (FIPs)

The Fair Information Principles emerged out of congressional investigations into U.S.-government surveillance activities and post-Watergate support for government reform.[10] The U.S. Department of Health, Education, and Welfare (HEW) held a series of meetings that considered the impact of the computerization of information on privacy, and ultimately included a series of five recommendations that came to underlie more expansive definitions of the FIPs embraced by international organizations and foreign governments.[11]

For example, in 1980, the Organization of Economic Cooperation and Development (OECD) codified its Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, which detailed eight principles.[12] These principles are: *[next page]*

[9] For example, the Tuskegee syphilis study was a catalyst for the Common Rule. First begun in the 1930s, hundreds of African-Americans were enrolled in a study on the long-term implications of syphilis. The study was conducted without adequately informing the participants about the study or its real purpose. Indeed, the men were actively misled, and researchers did not offer proper treatment needed to cure their illness. Public outrage ensued after an Associated Press story in 1972. Centers for Disease Control, U.S. Public Health Service Syphilis Study at Tuskegee, https://www.cdc.gov/tuskegee/timeline.htm (last visited Feb. 10, 2018).

[10] U.S. Dep't of Homeland Sec., Privacy Policy Guidance Memorandum (2008), https://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf.

[11] U.S. Dep't of Health, Education, and Welfare, Records, Computer, and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems (1973) https://aspe.hhs.gov/report/records-computers-and-rights-citizens.

[12] In 2013, the OECD updated its guidelines but retained the eight foundational principles: http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

i. ***Collection Limitation Principle:*** *There should be limits to the collection of personal data, and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.*
ii. ***Data Quality Principle:*** *Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete, and kept up-to-date.*
iii. ***Purpose Specification Principle:*** *The purposes for which personal data are collected should be specified no later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes, or such others as are not incompatible with those purposes, and as are specified on each occasion of change of purpose.*
iv. ***Use Limitation Principle:*** *Personal data should not be disclosed, made available, or otherwise used for purposes other than those specified in accordance with the Purpose Specification Principle except: (a) with the consent of the data subject, or (b) by the authority of law.*
v. ***Security Safeguards Principle:*** *Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification, or disclosure of data.*
vi. ***Openness Principle:*** *There should be a general policy of openness about developments, practices, and policies with respect to personal data. Means should be readily available to establish the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.*
vii. ***Individual Participation Principle:*** *An individual should have the right: (a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to them; (b) to have communicated to them, data relating to them within a reasonable time; at a charge, if any, that is not excessive; in a reasonable manner; and in a form that is readily intelligible to them; (c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and (d) to challenge data relating to them and, if the challenge is successful to have the data erased, rectified, completed or amended.*
viii. ***Accountability Principle:*** *A data controller should be accountable for complying with measures which give effect to the principles stated above.*

Organizational conversations about how they intend to use information should address these eight key principles. Elements and various formulations of each of these principles are found in most privacy or data protection laws around the globe, including the forthcoming European Union's General Data Protection Regulations (GDPR).

While there is broad agreement on the foundational substance of the FIPs, statutory implementations vary, and there is frequently a degree of flexibility with respect to how data controllers and data processors may comply with the FIPs depending upon context, industry sector, or type of data.[13] Best practices across industries and data types also tend to address each of these principles.

This flexibility has made the FIPs adaptable over time. Any organization can ostensibly consider how each of these practices apply to their activities, but it also leaves this framework open to the criticism that it establishes only a minimum standard of practice. Some of these principles may also be increasingly aspirational in a data-driven world. (Current disputes over the FIPs include implementations that may over-emphasize the effectiveness of individual notice-and-choice to protect privacy, or, on the other hand, focus on use limitations or considerations of context that can disempower individuals.)

It is also important to recognize that even the OECD envisions that its Privacy Guidelines will be put into effect though a number of different mechanisms. These could include institutional programs, procedures, and personnel that are tailored to an organization's activities, integrated throughout the organizational culture, and include risk assessments, appropriate safeguards, and ongoing monitoring and revisions.[14]

B. Ethical Frameworks Emerging from the Belmont Report

Ethical codes for research involving human subjects emerged against the backdrop of highly publicized medical research scandals, including the infamous, decades-long Tuskegee Syphilis Study and the Stanford Prison Experiment. Issued in 1979, the Belmont Report built upon existing medical ethics guidance to create a framework that continues to govern human subjects research.[15] The report identified three key principles: (1) respect for persons, (2) beneficence, and (3) justice.

---

[13] Robert Gellman, Fair Information Practices: A Basic History, at 19 (2012-2017), https://bobgellman.com/rg-docs/rg-FIPshistory.pdf; see also, White House Consumer Privacy Bill of Rights and the notion of "Respect for Context."

[14] OECD Privacy Guidelines ¶ 15.

[15] U.S. Dep't of Health, Education, and Welfare, The Belmont Report (1979), https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html.

- ***Respect for Persons:*** *Researchers are to respect the basic dignity and autonomy of their subjects. This generally requires that research subjects provide informed consent. In turn, informed consent has three elements: information, comprehension, and voluntariness. Relevant information must be presented to the research subjects in a comprehensible format and then voluntarily agree to participate. Other values such as privacy and autonomy may also be captured by this principle*
- ***Beneficence:*** *Researchers are to demonstrate beneficence by balancing the benefits of research and data use against any potential harm. Researchers and project reviewers are tasked with engaging in a multi-step cost-benefit analysis, weighing the risks of a wide variety of potential harms, including psychological harm, physical harm, legal harm, social harm, and economic harm.*
- ***Justice:*** *Researchers are to respect justice by ensuring that the value of research is accrued across society. Justice manifests itself in considerations of procedures and outcomes in selecting research subjects and ensuring fair distribution of the project's burdens and benefits.*

The Belmont Report also encouraged the development of independent review by Institutional Review Boards that would ensure that these principles were considered, research subjects were carefully selected, and that federal research funding could be made dependent upon adherence to ethical standards.[16]

More recently, the 2012 Menlo Report[17] was commissioned in response to new questions about the ethics of information and communications technology (ICT) research, adding a fourth principle which calls for the consideration of law and public interest. This principle asks researchers to engage in further legal due diligence, additional transparency, and accountability to account for the "expansive and evolving yet often varied and discordant, legal controls relevant for communication privacy and information assurance."

[16] *See* Metcalf, *supra* note 3.
[17] U.S. Dep't of Homeland Sec., The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research (2012), https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf.

II. Responsible Data Baselines: The Oxfam Responsible Data Policy

New and misunderstood risks of discrimination, profiling, tracking, and exclusion presented by black box data analytics and "big data" threaten the self-determination and personal autonomy of vulnerable populations. As a result, there has been considerable debate within civil society and public interest organizations as to whether a new set of overarching principles for data ethics is necessary.[18] Yet our review of existing responsible data approaches revealed that elements of both existing data protection and research ethics frameworks appear repeatedly. At a high level, data frameworks can address questions of "fairness," consent, and personal autonomy, risk assessments, open collaboration, and procedural issues around privacy and/or confidentiality and technical data security. The "Digital Impact Toolkit" classifies these themes into four buckets: (1) pluralism, (2) privacy, (3) consent, and (4) openness.[19] The Oxfam Responsible Data Policy, which is one of the more prominent frameworks in this effort, further fleshes out these concerns. Its five "backbone" rights are clearly informed by existing ethical and data management principles:

- ***Right to be Counted and Heard:*** *This principle largely embraces the principle of "justice" and emphasizes the need for nonprofit organizations to take into account special considerations for vulnerable and marginalized populations. This is also promoted through efforts to ensure that data is accurate, up-to-date, and relevant, which also captures the OECD "data quality" principle.*
- ***Right to Dignity and Respect:*** *The components of this right in the Oxfam policy addresses repeated concerns in the Belmont Report that measures be put in place to minimize disproportionate burdens on individuals. This also captures some of the values behind collection and use limitations that exist in privacy frameworks including the OECD guidelines. This right further invokes elements found in the nascent notion of "Respect for Law and Public Interest" in the Menlo Report that requires further attention be paid to local laws and overarching public policy, which frequently manifests itself in notions of respecting the context of interactions, individual expectations, and societal norms.*
- ***Right to Make An Informed Decision:*** *This right attempts to provide guidance on some of the longstanding and growing challenges around consent (and accurately explaining the purpose for which information could be used) that exist in data protection debates, as well as in the Belmont Report's call to respect personal autonomy.*

[18] Andrew Woods, Do Civil Society's Data Practices Call for New Ethical Guidelines? (2016), https://medium.com/the-digital-civil-society-lab/do-civil-societys-data-practices-call-for-new-ethical-guidelines-2a135cde239a.

[19] Digital Impact Toolkit, https://digitalimpact.io/digital-data/four-principles/. These four principles also encompasses discrete questions involving notice, consent, collaboration, and data privacy while balancing other values including diversity and inclusion.

- ***Right to Privacy:*** *The Oxfam policy discusses a right to privacy that focuses on keeping information confidential and technically secure, as well as minimizing the collection of data to reduce risks. It also introduces concepts such as anonymization of data, which is generally outside of the scope of data protection regulation.*
- ***Right to Not Be Put at Risk:*** *This right is a more detailed statement of the "do no harm" maxim, which is expressed in the Belmont Report's beneficence principle. This also requires organizations to engage in risk mitigation efforts, which calls for both training and understanding what beneficiaries and categories of their information are especially sensitive. Understanding sensitivity requires organizations to have a broad-based view of what could constitute harm to an individual, which is encompassed by physical, psychological, and political harms in the Oxfam policy.*

These rights synthesize concepts embedded in privacy and ethical guidelines. A comparison and mapping of these principles follows:

Oxfam Responsible Data Policy	**Menlo/Belmont Report**	**OECD Privacy Guidelines**
Right to Be Counted and Heard	Justice – each person deserves equal consideration . . . selection of subjects should be fair.	Data Quality Principle
Right to Dignity and Respect	Respect for law and public interest – engage in legal due diligence; be transparent in methods and results; and be accountable for actions.	Accountability Principle Openness Principle Collection Limitation Principle Use Limitation Principle
Right to Make an Informed Decision	Respect for persons – participation . . . is voluntary and follows from informed consent; treat individuals as autonomous agents . . .	Openness Principle Purpose Specification Principle
Right to Privacy		Accountability Principle Security Safeguards Principle
Right to Not Be Put at Risk	Beneficence	Collection Limitation Principle Sensitive Data Considerations

III. Common Themes

We reviewed 18 frameworks, and the principles they articulated fell into six sometimes overlapping categories: (1) respect for individual rights and autonomy, which includes concepts such as consent and access to one's personal information; (2) fairness or justice; (3) beneficence and the necessity of assessing the risks and benefits of collecting or using data; (4) FIPs-based privacy and data protection principles, including data minimization; (5) transparency and accountability; and (6) data security. Because these categories and concepts tend to overlap (for example, many privacy principles are also conceptualized as respect for the individual), others may group or label them differently.

A. Respect for individual autonomy

Nearly every framework in some way articulated the importance of adopting data practices that respect the individual autonomy and/or rights of the data subjects. As discussed in Part I, a core tenet of the Belmont and Menlo Reports is "respect for persons," which requires ensuring that participation in research studies is voluntary and individuals are treated as autonomous agents. Respect for persons is intertwined with notions of informed consent and human dignity, as well as other affirmative rights for individuals.

Data protection principles explicitly include affirmative rights for individuals. For example, the OECD Privacy Guidelines elaborate on the concept of "individual participation" and the right of individuals to "access [their] personal data and have the data erased, rectified, completed or amended."[20] The UK Information Commissioner's Office, moreover, has explained that European data protection laws (including the GDPR) give each individual a set of positive rights with respect to their data, including a right of access, a right to object to processing that is likely to cause damage or distress, and a right to have inaccurate personal data rectified, blocked, erased, or destroyed.[21]

Consent

Frequently, individual autonomy is associated with consent requirements. Several of the frameworks include consent as a necessary component of respect for individual autonomy. The Oxfam policy, for example, emphasizes the "right to make an informed

[20] OECD Privacy Guidelines ¶ 13.
[21] UK ICO, Principle 6: Rights of Individuals, https://ico.org.uk/for-organisations/guide-to-data-protection/principle-6-rights/.

decision." The Digital Impact Principles get at the voluntariness or effectiveness of consent by asking:

> *Were the people from whom the data was gathered actively asked to share their data? Did they actively agree to let you use their data for the purposes for which you intend to use it? Were they given the chance to say "no," without penalty from your organization? Can they get their information back from you if they want it?*

Note that consent provisions raise the question of whether beneficiaries can be penalized for declining to share data. Responsible data approaches ought to give careful consideration to this concern.

Consent is frequently framed in terms similar to that of the Belmont Report. For example, the InterAction Working Group guidance notes that autonomy requires that "participants must be given the opportunity to make an informed decision about their potential participation, which entails three elements: *information*, *comprehension*, and *voluntary participation*."[22] The Cash Learning Partnership's guidance on protecting beneficiary privacy drills down further into what the requirements for consent are, stating that individuals should be informed about (1) the nature of the data being collected, (2) with whom it will be shared, (3) who is responsible for its security, and (4) the chance to question the use made of their data and have an opportunity to withdraw from having their personal data used for the any of the above.[23]

As discussed later, consent is also an important privacy-protecting principle, and its framing as a mechanism for ensuring autonomy can overlap with requirements that individuals have "choices" to protect their privacy.

Respect for cultural norms and power differentials

The InterAction Working Group also embraces notions of dignity alongside autonomy, which appears to ask for more than merely informed consent. Its guidance notes that respect for individual autonomy also entails developing an understanding of cultural

[22] Interaction Protection Working Group, Data Collection in Humanitarian Response: A Guide for Incorporating Protection at 3 (emphasis added).

[23] Cash Learning Partnership, Protecting Beneficiary Privacy: Principles and operational standards for the secure use of personal data in cash and e-transfer programmes, http://www.cashlearning.org/downloads/calp-beneficiary-privacy-web.pdf.

norms and including beneficiaries as "equal participants in the data collection process rather than as victims"; moreover, organizations and their staff are directed to be "cognizant of the concept of power and the power differentials between the collector and the participant."[24]

B. *Justice / Fairness*

About half of the frameworks reference the concept of "justice," though this is frequently embodied by larger considerations of "fairness" around data collection and use. What is meant specifically by the notion of fairness remains somewhat vague, but it includes, at a minimum, that organizations consider the impact on vulnerable populations and avoid unjust discrimination.

"What is meant specifically by the notion of fairness remains somewhat vague, but it includes, at a minimum, that organizations consider the impact on vulnerable populations and avoid unjust discrimination."

Non-discrimination

Several frameworks directly or indirectly define fairness as "non-discrimination." Recall that the Belmont and Menlo Reports described justice as the requirement that selection criteria for inclusion in a study are applied equally to each person, that the burdens of research are allocated equitably across impacted subjects, and that the benefits of research are fairly distributed according to individual need, effort, societal contribution, and merit. The InterAction Working Group guidance echoes the concern for equal consideration of research subjects or beneficiaries, stating that "participant selection must be fair, unbiased, and conducted on the basis of scientific principles accord to the objectives of the data collection . . . . Once the general selection parameters are determined, the selection thereafter must be equitable and fair."[25]

Consideration of vulnerable populations

A common justice-related concern is protecting vulnerable populations and individuals, which generally includes in responsible data contexts, at a minimum, women, children,

[24] Interaction Protection Working Group, Data Collection in Humanitarian Response: A Guide for Incorporating Protection at 2.
[25] *Id.*

the disabled, and those impacted by conflicts and natural disasters. A broader conception of vulnerable populations may also be appropriate. We note that this concept also overlaps somewhat with calls to engage in risk assessments and to consider data sensitivity.

This is a common concern in data use frameworks. The Oxfam principles include a "right to be counted and heard," which emphasizes taking into account special considerations for vulnerable populations. The UN Office for the Coordination of Humanitarian Affairs (OCHA) proposes that organizations must specifically manage risks to vulnerable populations, cautioning that the inappropriate collection of sensitive data impacts individuals and communities and warns that its use can make people more vulnerable as a result.[26] Even the broad-based Principles for Digital Development suggests that the development of solutions be both useful for and sensitive to the most marginalized populations.[27]

C. Beneficence and risk-benefit assessment

The principle of beneficence requires organizations to maximize probable benefits and minimize probable harms. Most public interest organizations operate under the general maxim of "do no harm," but operationalizing this charge requires organizations to consider a wide array of data risks and then establish frameworks to mitigate these concerns. Thus, the concept of "beneficence" appears in most responsible data frameworks in the form of risk assessment principles and guidance.

Any well-meaning activity can pose risks to individual beneficiaries. "Even data that seemingly have nothing to do with people might impact individuals' lives in unexpected ways," cautioned a collection of prominent researchers, scholars, and data ethicists in their proposed *Ten Simple Rules For Responsible Big Data Research*.[28] Responsible organizations will, at minimum, begin any collection or use of data by asking themselves the following high-level questions to assess risk involved with data:

- *Could this data point be exploited for evil, and how?*

[26] UN OCHA, Building Data Responsibility Into Humanitarian Action (2016), https://docs.unocha.org/sites/dms/Documents/TB18_Data%20Responsibility_Online.pdf.
[27] Principle 1, Design with the User, Principles for Digital Development, http://digitalprinciples.org/design-with-the-user/.
[28] Ten simple rules for responsible big data research (2017), www.journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399.

- *Do the potential exploiters have the resources, desire, and knowledge to use it for evil?*
- *Can the good that releasing this data does outweigh that potential evil?*[29]

Resources generally include additional detail as to how to engage in a benefit-risk assessment. For example, Understand Risk provides a four-step framework for understanding – and assessing – data risks.[30] Organizations must first *understand* the data – where it comes from, the context in which it was collected, generated, or shared, and its intended uses and potential benefits. This requires organizations to engage in an assessment of anticipated benefits and risks, including what data constitutes "actionable information" for bad actors and what could set off that threat, and then to conduct a data inventory and implement appropriate security countermeasures. Organizations must be creative and identify potential ways and "risk-producing scenarios" in which risk could materialize.[31]

Data sensitivity

A key concept for understanding and assessing risk is the sensitivity of data held by the organization, particularly in light of the context and effect on vulnerable populations. The Oxfam policy's "right not to be put at risk" requires organizations to understand what beneficiaries and categories of information are especially sensitive. UN Global Pulse goes further to insist it will "employ stricter standards of care while conducting research among vulnerable populations and persons at risk, children and young people, and any other sensitive data."[32] It defines sensitive data to include, at minimum:

- *race or ethnic origin;*
- *political opinions;*
- *trade union association;*
- *religious beliefs or other beliefs of a similar nature;*
- *physical or mental health or condition (or any genetic data);*

[29] GeeksWithoutBounds, Responsible Humanitarian & Disaster Response Project Lifecycle (2014), http://gwob.org/wp-content/uploads/2014/07/responsibleprojectlifecycles.pdf.
[30] Sarah Telford and Stefaan G. Verhulst, Understand Risk, A Framework for Understanding Data Risk, https://understandrisk.org/a-framework-for-understanding-data-risk/.
[31] *Id.* Specific risks highlighted include: "your organization's data being correlated with other data sources to expose individuals; your organization's raw data being publicly released; and/or your organization's data system being maliciously breached." *Id.*
[32] UN Global Pulse, Privacy and Data Protection Principles, http://www.unglobalpulse.org/privacy-and-data-protection-principles.

- *sexual orientation;*
- *the commission or alleged commission of any offense;*
- *any information regarding judicial proceedings;*
- *any financial data; and*
- *any information concerning children, individual(s), or group(s) of individuals who face any potential risk of harm.*[33]

Understanding sensitivity requires organizations to have a broad-based view of what could constitute harm to an individual, including physical, psychological, and political harms.

D. Privacy Principles & FIPs

Practically every responsible data framework includes privacy principles in some way, and while privacy considerations may be mitigated via consent mechanisms or other methods of respecting individual autonomy, privacy is generally operationalized using some version of the FIPs.

At least three of the frameworks we reviewed – the Oxfam Responsible Data policy, the UN Global Pulse framework, and the Digital Impact Principles – include an umbrella principle labeled "privacy" that emcompasses some combination of practices involving data minimization/collection limitation, de-identification, and data use limitations, as well as ensuring the confidentiality of data. The notion of confidentiality is occasionally used alongside or instead of the term privacy.[34] The InterAction Working Group Guidance also uses the term "confidentiality" instead of privacy, but is ultimately describing FIPs concepts like purpose specification and use limitation:

> *Respect for the participant through protecting the individual and family's privacy is essential to the process. Privacy should be ensured in that there is a defined timeframe for the participation and that it is conducted under circumstances and in a location determined to be appropriate for the participant. Confidentiality pertains to the treatment of information that the individual has disclosed in the interview process with the expectation that it will not be divulged or disclosed in*

[33] UN Global Pulse, Data Innovation for Development Guide, Data Innovation Risk Assessment Tool, www.unglobalpulse.org/sites/default/files/Privacy%20Assessment%20Tool%20.pdf.
[34] Digital Humanitarians, Guidance for Incorporating Big Data Into Humanitarian Operations (2015).

> *a manner inconsistent with the way in which the participant was originally informed.*[35]

Discussions of the context of data collection and individuals' relationships with data controllers have come to dominate privacy debates,[36] and discussions of "context" also arise in some responsible data frameworks. For example, the *Ten Simple Rules For Responsible Big Data Research* emphasizes the importance of contextual privacy:

> *Recognize that privacy is more than a binary value: privacy is contextual and situational, not reducible to a simple public/private binary.* Just because something has been shared publicly does not mean any subsequent use would be unproblematic . . . privacy depends on the nature of the data, the context in which they were created and obtained, and the expectations and norms of those who are affected.[37]

"At least three of the frameworks we reviewed include an umbrella principle labeled 'privacy' that emcompasses some combination of practices involving data minimization/collection limitation, de-identification, and data use limitations, as well as ensuring the confidentiality of data." ----------

Data Minimization

At least seven of the frameworks we reviewed include some type of "data minimization" principle.[38] Data minimization was operationalized by promoting some combination of

[35] Interaction Protection Working Group, Data Collection in Humanitarian Response: A Guide for Incorporating Protection at 4.

[36] Helen Nissenbaum, Privacy In Context: Technology, Policy, and the Integrity of Social Life (2010); *See also* Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy, Executive Office of the President, at 15 (2012); FTC, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Business and Policymakers 38-39 (2012), available at http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-changerecommendations/120326privacyreport.pdf.

[37] Ten simple rules for responsible big data research (2017), http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399 (emphasis added)

[38] Additionally, data minimization tip sheets have also been produced that are adjacent to some of these frameworks; see elan, Data Minimization Tip Sheet, http://elan.cashlearning.org/wp-content/uploads/2016/05/Data-minimization-tip-sheet.pdf.

limiting the collection of personal information, de-identifying or anonymizing information, setting retention limits, and establishing disposal procedures.

Collection limitation

There was broad agreement across frameworks that organizations should not collect or retain personal information unless determined as necessary. The Digital Impact Principles go as far as to ask whether the organization has "collected the least possible amount of data to accomplish [its] goals." Other frameworks contain blanket cautions: "Data should not be brought into operations simply for the sake of data. It should only be incorporated if it can be used to solve a problem . . . ."[39] Guidance from the Sunlight Foundation suggests limiting the collection of sensitive information and instructs, "where sensitive information is needed for decision making, evaluate whether that information can be gathered without written documentation" such as via verbal confirmation.[40]

Retention limits and disposal

Another important principle under the data minimization umbrella is the idea that data should not be held for longer than necessary and that organizations should eventually dispose of or destroy unnecessary personal data. Data protection guidance from the UK Information Commissioner's Office provides that personal data must not be retained for longer than is necessary for the purpose it was obtained for. Data controllers must

> *review the length of time [they] keep personal data; consider the purpose or purposes [they] hold the information for in decid[ing] whether (and for how long) to retain it; securely delete information that is no longer needed for this purpose or these purposes; and update, archive[,] or securely delete information if it goes out of date.*[41]

The Cash Learning Partnership framework specifically includes "disposal" as a principle:

[39] Interaction Protection Working Group, Data Collection in Humanitarian Response: A Guide for Incorporating Protection at 14.
[40] Sunlight Foundation, Protecting Data, Protecting Residents, at 3 (2017).
[41] UK ICO, Principle 5: Retaining Personal Data, https://ico.org.uk/for-organisations/guide-to-data-protection/principle-5-retention/.

> *Organizations should not hold beneficiary data for longer than is required unless they have clear, justifiable and documented reasons for doing so otherwise data held by the organization and any relevant third parties should be destroyed.*[42]

Other guidance also instructs organizations to "think carefully" about how to "delete records."[43] Civil society groups, including the authors of this literature review, have acknowledged the importance of developing policies and practices around data disposal and deletion, but organizations collecting data have been slow to develop such practices, and effectively deleting or "destroying" data can require significant technical expertise.

De-identification

At least three frameworks – the UN Global Pulse guidance, *10 Simple Rules*, and the Digital Impact Principles – specifically mention de-identification as an important process, principle, or component of protecting beneficiary policy. The UN Global Pulse Framework states, "We do not attempt to knowingly and purposefully re-identify de-identified data, and we make all reasonable efforts to prevent any unlawful and unjustified re-identification."[44] The challenge for responsible data frameworks is to recognize that even de-identified information may be re-identifiable, and even aggregated statistics may pose serious risks or privacy harms if they reveal that certain communities suffer from stigmatized diseases or social behavior more than other groups.[45]

Use Limitations & Purpose Specification

The terms "use limitation" and "purpose specification" are often used together or even interchanged in frameworks. As discussed above, they generally stand for the principle that organizations collecting data should (1) make the intended purpose of collection clear to the data subject at the time of collection and (2) not use the data for purposes that are incompatible or inconsistent with that stated purpose. Purpose limitations

[42] Cash Learning Partnership, Protecting Beneficiary Privacy: Principles and operational standards for the secure use of personal data in cash and e-transfer programmes, at 8, http://www.cashlearning.org/downloads/calp-beneficiary-privacy-web.pdf.
[43] New Philanthropy Capital, Protecting Your Beneficiaries, Protecting Your Organization, at 6 (2015), http://www.thinknpc.org/publications/safe-use-of-personal-data/.
[44] UN Global Pulse, Privacy and Data Protection Principles, http://www.unglobalpulse.org/privacy-and-data-protection-principles.
[45] Ten simple rules for responsible big data research (2017), http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399.

ideally ensure that organizations use personal data in a way that meets, and does not violate, the reasonable expectations of the data subject.

Data quality

At least five of the frameworks included a principle that data should be accurate, complete, and up-to-date. Although data quality and accuracy do not necessarily serve privacy, data quality is a FIP because it ensures that individuals are not misrepresented in their data or that adversarial action is not taken against them on the basis of inaccurate data, for example, out-of-date credit information. Data accuracy may be particularly important in the public interest sector where organizations are, for example, delivering aid to individuals based on the data the organization has.

Privacy by design

Embraced by data protection regulators globally, privacy by design is the notion that privacy and data protection compliance should be deliberate, systematic, and "baked" into the outset of data projects. (In addition to having a data security and engineering component, the U.S. Federal Trade Commission has suggested that privacy by design has a "process" component that includes personnel, procedures, and controls.[46]) The ICO has described privacy by design as "an approach to projects that promotes privacy and data protection compliance from the start."

> *The ICO encourages organisations to ensure that privacy and data protection is a key consideration in the early stages of any project, and then throughout its lifecycle. For example when: building new IT systems for storing or accessing personal data; developing legislation, policy or strategies that have privacy implications; embarking on a data sharing initiative; or using data for new purposes. We would like to see more organisations integrating core privacy considerations into existing project management and risk management methodologies and policies.*[47]

[46] "Privacy By Design and the New Privacy Framework of the U.S. Federal Trade Commission," Remarks of Commissioner Edith Ramirez at the Privacy by Design Conference, Hong Kong (June 13, 2012), https://www.ftc.gov/sites/default/files/documents/public_statements/privacy-design-and-new-privacy-framework-u.s.federal-trade-commission/120613privacydesign.pdf.

[47] UK ICO, Privacy By Design, https://ico.org.uk/for-organisations/guide-to-data-protection/privacy-by-design/.

More recently, academics and civil society groups have been incorporating "privacy by design" as a responsible data principle.[48] The Cash Learning Partnership framework includes the principle of "protect[ion] by design":

> *Organisations should protect by design the personal data they obtain from beneficiaries either for their own use or for use by third parties for each case or e-transfer programme they initiate or implement.*[49]

E. Transparency and accountability

Every responsible data framework we reviewed included some principle(s) around transparency and accountability. These concepts are often tied together, as transparency is seen as a necessary precondition for holding organizations – or data controllers – accountable.

Transparency and openness

Transparency is often operationalized as informing data subjects about what data is collected and how it is used. The OECD guidance states that "[m]eans should be readily available of establishing the existence and nature of personal data, and the main purposes of their use." According to the ICO's guidance:

> *Transparency is always important, but especially so in situations where individuals have a choice about whether they wish to enter into a relationship with you. If individuals know at the outset what their information will be used for, they will be able to make an informed decision about whether to enter into a relationship, or perhaps to try to renegotiate the terms of that relationship.*[50]

Transparency is sometimes referred to as "openness," which can also encompass the idea of open data or "making data and findings available [to the public]," a concept which is more relevant to organizations using big data or engaged in data research, which are subjects somewhat outside the scope of this literature review. The Digital

[48] *See* Data Maturity Framework, Center for Data Science and Public Policy (2016), http://dsapp.uchicago.edu/resources/datamaturity/.
[49] Cash Learning Partnership, Protecting Beneficiary Privacy: Principles and operational standards for the secure use of personal data in cash and e-transfer programmes, at 7, http://www.cashlearning.org/downloads/calp-beneficiary-privacy-web.pdf.
[50] UK ICO, Principle 1: Processing personal data fairly and lawfully, https://ico.org.uk/for-organisations/guide-to-data-protection/principle-1-fair-and-lawful/.

Impact Principles state that "[a]pproaching your data efforts with an assumption of openness will require you to plan for participant consent and build in privacy protecting strategies."[51]

*"Every responsible data framework we reviewed included some principle(s) around **transparency** and **accountability**. These concepts are often tied together, as transparency is seen as a necessary precondition for holding organizations – or data controllers – accountable."* ----------

Accountability

Ensuring organizations accept accountability for their practices is at the core of responsible data efforts. The Menlo Report, which emphasized ethical considerations in ICT research, added as a foundational principle the need to engage in legal compliance and that organizations "be accountable for actions."[52] The frameworks we reviewed generally agree that organizations and individuals should be held accountable for complying with the law, complying with their own policies, and acting in the public interest.

Compliance

Accountability is commonly conceptualized as compliance with applicable laws and policies. According to the UN OCHA framework, organizations must be "responsible for determining what legal and ethical standards apply to proposed applications of data in specific contexts, and for adhering to these to prevent potential violations of laws and rights."[53] As a result, organizations are charged with navigating an often confusing landscape of laws and regulations governing the collection and use of data, which is particularly difficult when organizations work internationally. However, it is important to note that compliance only serves data subjects and the public insofar as the law protects them. Legal compliance is rarely sufficient on its own to amount to responsible data use.

[51] Digital Impact Toolkit, https://digitalimpact.io/digital-data/four-principles/.
[52] U.S. Dep't of Homeland Sec., The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research (2012), at 18, https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf.
[53] UN OCHA, Building Data Responsibility Into Humanitarian Action (2016), https://docs.unocha.org/sites/dms/Documents/TB18_Data%20Responsibility_Online.pdf.

Oversight

The Sunlight Foundation's framework for responsible municipal data management recognizes oversight of data sharing as an important accountability principle. The framework instructs municipalities to consider the following:

- *Inventory all policies and practices which result in the sharing of information on individuals' citizenship or other sensitive status;*
- *Publicly document all policies, practices, and requests which result in the sharing of information about individuals' citizenship or other sensitive status;*
- *Create policies which limit individual employees' discretion on data-sharing;*
- *Create a municipal oversight body to ensure that the city's protocols for data protection are adequate, well-observed, and legal.*[54]

Third Parties and Collaboration Partners

Two United Nation guidance documents also highlight the need for ensuring the accountability of third-party collaborators.[55] The UN Global Pulse framework requires collaborators to act "in compliance with relevant law, data privacy and data protection standards and the United Nations' global mandate."[56]

F. Data security

Data security is repeatedly emphasized as an essential component of responsible data use. The Oxfam Responsible Data Policy conceptualizes security as the "right to not be put at risk," and at least two-thirds of the frameworks we reviewed include data security as a principle and reiterate the importance of protecting beneficiary data. However, the contours of this requirement are unclear. The OECD has stated that "personal data should be protected by reasonable security safeguards."

Regulators and data collecting organizations alike have struggled with what "reasonable security" means in practice.[57] While there are some general best practices that have

[54] Sunlight Foundation, Protecting Data, Protecting Residents, at 3 (2017).
[55] UN Global Pulse, Privacy and Data Protection Principles, http://www.unglobalpulse.org/privacy-and-data-protection-principles; *see also* UN Development Group, Guidance Note on Data Privacy, Ethics, and Protection (2017), https://undg.org/document/undg-guidance-note-on-big-data-for-achievement-of-the-2030-agenda-data-privacy-ethics-and-protection/.
[56] *Id.*
[57] *See, e.g.*, Fed. Trade Comm'n, Start with Security (2015), https://www.ftc.gov/system/files/documents/plain-language/pdf0205-startwithsecurity.pdf.

emerged around password protection and data transmission, flexible standards are useful because what is "reasonable" continues to evolve along with technology and more sophisticated methods of gaining unauthorized access. The ICO offers slightly more concrete guidance:

> *Design and organise your security to fit the nature of the personal data you hold and the harm that may result from a security breach; be clear about who in your organisation is responsible for ensuring information security; make sure you have the right physical and technical security, backed up by robust policies and procedures and reliable, well-trained staff; and be ready to respond to any breach of security swiftly and effectively.*[58]

As does the Sunlight Foundation:

> *Improve storage practices: regularly delete sensitive data where retention is not legally required; do not create or retain specialized, personally-identifiable databases of vulnerable groups of residents; where sensitive information is collected, do not store it with less-protecting third parties; encrypt sensitive data and communications to limit the potential for data theft.*[59]

This guidance demonstrates the overlap between data security and risk assessment. For example, the four-part Understand Risk framework discusses security in the context of deploying measures that can prevent potential risks for materializing, which might include data handling procedures, access controls, and personnel training.[60] The frameworks agree that each organization must assess its security risks based on factors such as the type, amount, and sensitivity of data it holds; where the data is stored and how it is processed, transferred, or shared; and the particular vulnerabilities of the data subjects.

[58] UK ICO, Principle 7: Information Security, https://ico.org.uk/for-organisations/guide-to-data-protection/principle-7-security/.
[59] Sunlight Foundation, Protecting Data, Protecting Residents (2017).
[60] Sarah Telford and Stefaan G. Verhulst, Understand Risk, A Framework for Understanding Data Risk, https://understandrisk.org/a-framework-for-understanding-data-risk/.

IV. International Considerations

One significant gap in responsible data literature is how to assess the international transfer and storage of data. While this has emerged as a frequent concern about data protection regulators, it is starkly missing from our review. The ICO's guidance directly addresses these issues and suggests organizations ask the following questions:

- *Do you need to transfer personal data abroad?*
- *Is the transfer to a country on the EU Commission's list of countries or territories providing adequate protection for the rights and freedoms of data subjects in connection with the processing of their personal data?*
- *If the transfer is to the United States of America, has the U.S. recipient of the data provided adequate protection for the transfer of personal data?*[61]

Many public interest organizations – especially those focused on humanitarian aid – will collect data from individuals located in countries outside of the organization's country. This data will likely be transferred across borders and may be stored on servers in multiple different countries with different law enforcement access and surveillance laws and different laws and policies around privacy and security. At present, however, our review suggests that little guidance exists for non-profit organizations managing international data transfer, collection, and storage issues. At the very least, organizations must consider the legal and policy implications any international data collection and transfer they conduct.

"One significant gap in responsible data literature is how to assess international transfer and storage of data. While this has emerged as a frequent concern about data protection regulators, ***it is starkly missing from our review.****"*

V. Conclusion

Our review of existing frameworks and principles reveals that a common lexicon has emerged with respect to what constitutes responsible data principles. The challenge ahead is what this should mean in practice. This will be highly context-dependent and organization-specific, warranting the need for strong accountability mechanisms and scalable strategies. These may need to be tailored to identified risks, which may ultimately require a broader conception of risk than is currently understood.

[61] UK ICO, Principle 8: Sending personal data outside the European Economic Area, https://ico.org.uk/for-organisations/guide-to-data-protection/principle-8-international/.

THIS REVIEW WAS PREPARED AS PART OF THE

CENTER FOR DEMOCRACY & TECHNOLOGY'S

WORK WITH

GOOD DATA COLLABORATIVE,

A PROJECT TO LOOK AT THE STATE OF RESOURCES FOR CIVIL SOCIETY TO USE DATA RESPONSIBLY.

THE PROJECT WAS GENEROUSLY SUPPORTED BY THE DIGITAL CIVIL SOCIETY LAB AT THE

STANFORD CENTER ON PHILANTHROPY AND CIVIL SOCIETY.

CENTER FOR
DEMOCRACY
& TECHNOLOGY