COMMERCE DATA ETHICS FRAMEWORK

2022

INTRODUCTION

Data are critical for fulfilling of the U.S. Department of Commerce's mission to create the conditions necessary for economic growth and opportunity for all communities. As "America's Data Agency," the Commerce Department demonstrated leadership through the development of the Commerce Data Strategy[1]. The Strategy provides a roadmap to maximize the positive impact of Commerce data and staff. Implementation of the Strategy will be executed through five interdependent strategic goals, one of which is to 'promote appropriate data use and equitable access' and, more specifically, 'promote data ethics across Commerce by fostering quality, transparency, accountability, and fairness of data policies and practices throughout the data lifecycle.'

To support this goal, the Commerce Data Strategic Action Plan for fiscal years 2021-2022 recommended the development of a Commerce Data Ethics Framework.

This framework makes staff aware and provides guidance for ethical, responsible, and equitable data practices throughout the data lifecycle to fully leverage the value of federal data for mission, service, and the public good by guiding the Federal Government in practicing ethical governance, conscious design, and learning culture[2].

Data ethics are the norms of behavior that promote appropriate judgments and accountability when acquiring, managing, or using data, with the goals of protecting privacy and confidentiality, minimizing associated risks to individuals and society, and maximizing the public good[3]. In this context, the public good is defined as the benefit and well-being of the American people, and it is enhanced by services that are provided to all members of society by the government.

Data ethics cannot be ensured only by technological solutions or by adhering to relevant laws, rules, regulations, and standards. It is essential to advance, not only the understanding of the problem at hand, but also of the data that are an integral part of the problem formulation. Evidence-based decision making—that is, using data to represent known facts to lay a foundation for our reasoning and decision making—is only as trustworthy as the data it is based on.

Data ethics principles include privacy, confidentiality, fairness, objectivity, inclusiveness, transparency, accountability, safety, reliability, security, and trust. It is critical to apply data ethics and diligence in each stage of the data lifecycle, including data collection, storage, transmission, aggregation, analysis, use, sharing, and disposal. The quality of the data and its stewardship determine the reliability, accuracy, and fairness of the outcomes of data-driven decisions.

The following framework aims to be forward-thinking and practical. It makes recommendations and provides guidelines, rather than requirements. Where a law or regulation explicitly applies to an activity discussed in this framework, that law or regulation may supersede the best practices and recommendations of this framework. When implementing any best practices or recommendations from this framework, bureaus should consult their legal counsel to ensure compliance with law.

[1] *See* https:// www.commerce.gov/sites/default/files/2021-08/US-Dept-of-Commerce-Data-Strategy.pdf
[2] *See* https://strategy.data.gov/assets/docs/federal-data-strategy-principles.pdf
[3] *See* https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf

The Commerce Data Ethics Framework is grounded on four pillars, as shown in Figure 1. Through these pillars, the framework establishes best practices for ethical considerations, awareness, and guidance with the goal to empower staff to use appropriate and responsible data practices throughout the data lifecycle and appreciate potential risks and impacts of their data practices.

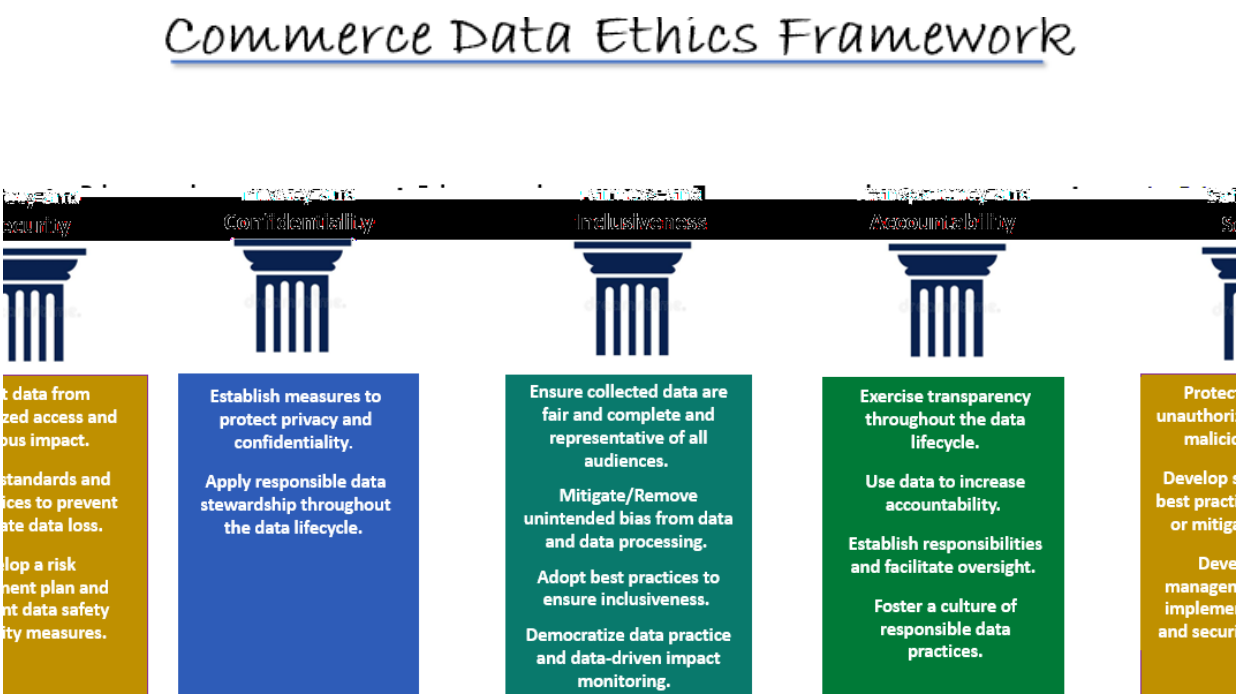

Figure 1. The U.S. Department of Commerce Data Ethics Framework

Below are the definitions and best practices that support the Data Ethics Framework.

PRIVACY AND CONFIDENTIALITY

Definition

Privacy and confidentiality are important in access, ownership, use, and collection of data. Privacy is freedom from unwarranted intrusion into the private lives of individuals and private conduct of businesses. Confidentiality is the state of personal and business information being free from inappropriate access and use.

Best Practices

1.1 Establish measures to protect privacy and confidentiality.

- Ensure data practices comply with confidentiality and privacy laws, rules, and regulations, including the Privacy Act[4], while complying with transparency laws, rules, and regulations, including the Freedom of Information Act (FOIA)[5].
- Follow relevant data governance standards, such as Fair Information Practice Principles[6] (FAIR).
- Identify sensitive data, classify them, and mark them as appropriate, consistent with the National Archives and Records Administration (NARA) regulations and other applicable data marking laws and regulations[7].
- Any data that is collected for research under the Common Rule for the protection of human subjects must comply with all the requirements of the Common Rule[8].
- Define a data usage policy and a mechanism to monitor data access.
- Remember that data subjects own their personal identifiable information (PII) and business identifiable information (BII) and have a right to know about plans for collecting, storing, and using that data, when permitted and consistent with law, including the Privacy Act[9].
- Ensure that the combination of data already available and the data being prepared for release does not enable the re-identification of individuals, businesses, or other entities including government agencies, where laws and regulations prohibit disclosure.
- Only collect the personal and business data to only the information needed to accomplish the authorized purposes.
- Develop a retention plan for all relevant data and retain PII and BII only for as long as is necessary to fulfill the authorized purpose, unless data subjects have given their consent. Be aware that multiple records retention schedules could apply simultaneously.
- Explore and implement where possible data privacy and protection techniques, including anonymization technologies[10], in accordance with laws, rules, and regulations.
- Consider using data protection techniques, such as differential privacy, anonymization, and synthetic data.

1.2 Apply responsible data stewardship throughout the data lifecycle.

- Be specific about the purpose of data practices.
- Appropriately protect privacy and confidentiality when collecting, accessing, transmitting, using, storing, and deleting data, including defining appropriate boundaries and control.
- Ensure that the information collected about individuals is accurate and allow individuals the opportunity to amend or access their own PII, in accordance with the Privacy Act.

[4] *See* https://www.govinfo.gov/content/pkg/USCODE-2018-title5/pdf/USCODE-2018-title5-partI-chap5-subchapII-sec552a.pdf
[5] *See* https://www.foia.gov/
[6] *See* https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/circulars/A130/a130revised.pdf
[7] *See* https://www.archives.gov/about/regulations/regulations.html
[8] *See* https://www.ecfr.gov/current/title-15/subtitle-A/part-27
[9] See https://www.osec.doc.gov/opog/PrivacyAct/PrivacyAct.html
[10] See https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/focus-areas/de-id/tools

- Ensure that control measures are in place to prevent unauthorized access and minimize the severity of possible harms.
- Use data collection methods, including automated and artificial intelligence-driven procedures, that meet legal requirements and ethical norms.
- Recommend that systems, processes, data collections, and data products be built using privacy-by-design principles.
- Encourage long-term data preservation and archiving of data that can be legally retained.
- Restrict data access to those with a lawful government purpose or a business need for information, and who otherwise meet all legal and regulatory requirements.

FAIRNESS AND INCLUSIVENESS

Definition

Fairness and inclusiveness ensure equal and fair treatment regardless of race, gender, age, other protected status, or socioeconomic status.

Fairness is an approach to achieve inclusive representation. It does this by minimizing human bias in research and data collection so that all communities are fairly and objectively represented. Fairness also means mitigating bias and ensuring data projects do not result in unintended effects on social groups and individuals. When evaluating fairness in data applications, cultural, background, legal, economic, political, ethical, and historical factors should all be considered.

Inclusiveness in data ethics means that all relevant people have an equal opportunity to be included in the data collected and in the data's use, and that no one is left out, voices are heard, people have equal access to data, and people can understand the data. Vulnerable communities must be intentionally considered and included.

Best Practices

2.1 Ensure collected data are fair and complete.

- Control and monitor the quality, suitability and objectivity of data collected.
- Be aware that bias can originate at any step of the data lifecycle (creation, sampling, collection, and processing) and consider best practices to mitigate or eliminate it.
- Ensure proper, non-lethal, unbiased, inclusive data collection and curation. Consider the purpose of the data collection and the intended use.
- Enforce data quality standards and governance to ensure that data is complete, inclusive, unbiased, and protected.

2.2 Ensure data are representative of all audiences.

- Collect and process data to include input from all known potential users and the sources of data, to ensure fairness and impartiality in data representation, processing, and use.
- Ensure that datasets are suitable for the intended purpose. This requires attention to several factors, including data quality, statistical methods for mitigating representation issues, processes to account for the socio-technical context in which the application is being deployed, and awareness of the interaction of human factors.

- Proactively gather input from underrepresented communities on data that may have unfair impact and mitigate that impact.
- Make disaggregated data the norm while protecting privacy[11].

2.3 Mitigate or remove bias from data and data processing.

- Raise awareness and provide training about the many forms of bias, including systemic, statistical, computational, and human biases and take steps to reduce it.
- Establish procedures to proactively search for data bias and take steps to mitigate impact, including removal practices to support fairness and inclusiveness.
- Use statistical and data science methods to evaluate results for accuracy and bias.
- Use statistics to correct for bias, including weighting, balancing, and other methods to eliminate sample bias.

2.4 Be transparent about how data are being collected and used.

- Collect data ethically, with notice to and, where required, consent of those involved, in accordance with all applicable policies, regulations, and legal requirements.
- Do not collect data that are prohibited, irrelevant, or unnecessary to accomplish the purpose of the work.
- Enhance access and transparency when using data, including the processing of personal and sensitive data, to support a fair and inclusive process while considering the impact on different groups of people.
- Be aware of the potential use of data in ways that are inconsistent with the purpose for which they were collected and that could lead to biased outcomes.
- Implement a transparent and open as possible process that provides ways to incorporate feedback from end-users and the community, such as protected and confidential avenues, in which data subjects can anonymously report data abuse, misuse, and unintended negative outcomes. This process should be rigorous, efficient, and include an approach to investigate allegations and act against harm, if warranted.

2.5 Adopt best practices to ensure inclusiveness.

- Foster inclusive and informed communication and collaboration through diverse representation around data practices and impact assessment.
- Provide educational services on the role of data governance and its implications for vulnerable communities in plain language, taking into account differences in languages and learning types to be as inclusive and accessible as possible.
- Be transparent about how data will be used, for what purpose, and by whom.
- Establish a collaborative process among stakeholders and maintain transparency throughout the process, including proactive monitoring of unintentional bias and exclusion, and documenting the feedback of relevant stakeholders in communities.
- Offer support and translation services to ensure accessibility and enhance the opportunity to engage with government data.
- Engage with data users to understand their needs.

[11] See https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf

- Ensure that data subjects, vulnerable and underrepresented communities, and relevant stakeholders are involved in data ethics conversations early on.
- Educate yourself about social and historical context for engaging with underrepresented communities, prioritize their needs, and prioritize their presence in all conversations about data practices.
- Provide resources to data subjects to allow them to exercise autonomy, control, and agency over their own data and freely give or withdraw consent for its use, when required and in accordance with laws, rules, and regulations:
  - Provide tools that make it easy to opt-in and opt-out, whenever permitted.
  - Provide educational services that inform data subjects of their rights early in the process.
  - Seek consent, rather than mere consultation, whenever appropriate.

2.6 Democratize data practice and data-driven impact monitoring.

- Provide training to data users on how to understand and use data across the organization.
- Empower data driven culture and data ethics engagement.
- Democratize access to unrestricted, publicly available data at no cost to the user wherever possible.
- Create protected, accessible, and confidential avenues for data subjects to anonymously report incorrect data, data abuse or misuse, and unintended negative outcomes where not limited by statutes.
- Enable transparency to monitor the downstream effects (positive or negative), enable immediate action to reverse harm, and support recourse, in accordance with laws, rules, and regulations.

TRANSPARENCY AND ACCOUNTABILITY

Definition

Transparency is the open disclosure and sharing of information about a project in a complete, clear, intelligible, and easily accessible format.

Accountability is setting and fostering a common expectation by clearly defining the organization's mission, values, and goals while acknowledging responsibilities for actions, decisions, and products. Accountability requires that anyone acquiring, managing, or using data be aware of stakeholders and responsible to them, as appropriate.

Best Practices

3.1 Exercise transparency throughout the data lifecycle.

- Ensure data practices comply with transparency laws, rules, and regulations, including the Freedom Information Act (FOIA).
- Exercise transparency in data practices subject to laws and regulations pertaining to data privacy, protection, and security.
- Share information about the purpose for data collection.

- Use standardized protocols, languages, and schemas for coding, data, metadata, and communications.
- Require data to be in machine-readable, machine-actionable, and non-proprietary formats.
- Make data findable, accessible, interoperable, and re-usable (FAIR) as much as possible.
- Ensure the methods, data, changes, schedules, and regulations for production and publishing are available for inspection in a complete, open, understandable, easily accessible, and free format.
- Be clear about how data are collected and used and ensure data subjects understand the process.
- Clearly state partnerships with other government bodies, private entities, or academic institutions that aid with research, collection, analysis, and other data practices, including data sharing.
- Ensure that rights in data are described in agreements when working in such partnerships.

3.2 Use data to Increase accountability.

- Align operational and regulatory data inputs with performance measures and other outputs to help the public understand the results of federal, state, local, and tribal investments and to support informed decision-making and rulemaking.
- Ensure that analyses are governed by statistical considerations and are free from influence from data providers or users.
- Encourage accountability and transparency when data misuse or abuse has occurred. Projects should be subject to review within the Bureau responsible for the data to determine and correct any issue, and data subjects should be made aware of misuse or abuse as soon as possible, in accordance with laws, rules, and regulations.

3.3 Establish responsibilities and facilitate oversight.

- Set responsibilities of stakeholders across the data lifecycle.
- Implement data and data processing governance that includes safeguarding, monitoring, and oversight processes.
- Require documentation to clearly communicate what datadriven systems can and cannot do.
- Encourage internal and external engagement and feedback to support quality assurance.
- Develop a risk management plan, including instant response in case of unethical data collection practices and biased, unfair impact.

3.4 Apply transparency and accountability best practices.

- Implement tools to validate the intent and impact, as well as granular documentation, monitoring, and evaluation of data practices.
- Provide mechanisms to conduct effective governance and oversight.
- Encourage transparency about the purpose of data collection and research and the attribution of research permits.

3.5 Foster a culture of responsible data use and practices.

- Promote responsible data use and practices, including enhanced participation and collaboration across the organization on data practices and impacts in accordance with laws, rules, and regulations.
- Consider the following in all data practices:
  - Purpose of data collection and use.
  - Stakeholder input on data collection, use, stewardship, and governance.
  - Benefits, costs, burdens, or limitations for individuals and communities.
  - Implications of data use on vulnerable populations.
  - Data security and privacy.

SAFETY AND SECURITY

Definition

Data safety concerns protecting data against unintentional loss and restoring data as necessary.

Data security is the practice of protecting digital information from unauthorized access, corruption, or theft throughout its lifecycle.

Best Practices

4.1 Adopt standards and best practices to secure data.

- Refer to Federal Information Processing Standards (FIPS) guidance for any data that reside on federal information systems[12] and other security standards and guidelines[13].
- Establish best practices for protecting sensitive PII and BII from unauthorized access and comply with applicable statutes and regulations.
- Promote cybersecurity awareness and best practices through robust policies and procedures to prevent unintended consequences such as privacy violation and harm.

4.2 Adopt standards and best practices to ensure data integrity and confidentiality.

- Protect digital data from being corrupted, deleted, or unrecoverable, including from the actions of unauthorized users.
- Establish data governance and protection practices, including organizational standards that comply with federal guidance to mitigate data losses.
- Incorporate data security and recovery measures to limit data loss or ensure prompt, effective recovery.

4.3 Develop a risk management plan to enhance data safety and security measures.

- Develop a risk management plan and implement actions to ensure data safety, data security, and data reliability, in accordance with all federal guidance and standards.

[12] *See* https://www.nist.gov/standardsgov/compliance-faqs-federal-information-processing-standards-fips
[13] *See* https://csrc.nist.gov/publications