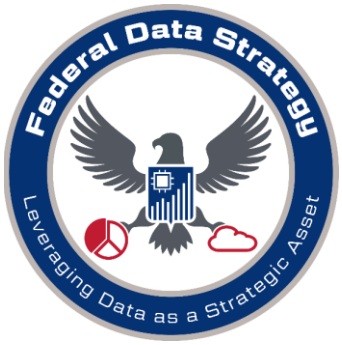
PRESIDENT'S MANAGEMENT AGENDA



Federal Data Strategy

Data Ethics Framework

Special Thanks to Our Contributors

Decisions made with data touch every aspect of American life. The Federal Government uses data to solve problems, develop and deliver services to citizens, defend and secure the nation, and support economic growth. Data's benefits and risks are amplified by the expanding capabilities of digital networks, technology systems, algorithms, and computational methods that enable data to be easily collected, combined, manipulated, and shared.

The Federal Data Strategy, delivered in December 2019, recognized the importance of ethics in its founding principles. When the Federal Data Strategy team created the 2020 Action Plan,[1] they tasked the General Services Administration (GSA) with developing a Data Ethics Framework (Framework) in Action 14 to help federal employees, managers, and leaders make ethical decisions as they acquire, manage, and use data.

To achieve the goal of developing a useful Data Ethics Framework for the Federal Government, GSA sought input from a diverse set of stakeholders to maximize the Framework's value and utility across the federal enterprise. To lead the Framework's development, GSA formed an interagency team comprised of 14 government leaders from across government with expertise in statistics, public policy, evidence-based decision making, privacy, and analytics. In addition, GSA received input on the Framework from the Chief Data Officer (CDO) Council, Interagency Council on Statistical Policy (ICSP), and the Federal Privacy Council (FPC). The resulting Framework will help guide the ethical acquisition, management, and use of data for years to come.

Data Ethics Framework Development Team

- **Trey Bradley**, Strategic Data Initiatives Program Manager, Office of Shared Solutions & Performance Improvement, U.S. General Services Administration (Development Lead)
- **Ken Ambrose**, Senior Advisor for the CDO Council, Office of Government-wide Policy, U.S. General Services Administration
- **Maya Bernstein**, Senior Advisor for Privacy Policy, Office of the Secretary for Planning & Evaluation, U.S. Department of Health & Human Services
- **Ivan DeLoatch**, Executive Director, Federal Geographic Data Committee, U.S. Geological Survey, U.S. Department of the Interior
- **Dave Dreisigmeyer**, Interim Deputy Chief Data Officer, Office of the Under Secretary for Economic Affairs, U.S. Department of Commerce
- **Jeffrey Gonzales**, Research Mathematical Statistician, Economic Research Service, U.S. Department of Agriculture
- **Chris Grubb**, Chief Data Scientist, Center for Analytics, U.S. Department of State
- **Lisa Haralampus**, Director of Records Management Policy and Outreach, Office of the Chief Records Officer, U.S. National Archives and Records Administration
- **Michael Hawes**, Senior Advisor for Data Access and Privacy, U.S. Census Bureau, U.S. Department of Commerce
- **Barry Johnson**, Acting Chief, Research and Analytics Office, Internal Revenue Service, U.S. Department of the Treasury
- **Brandon Kopp**, Research Psychologist, Bureau of Labor Statistics, U.S. Department of Labor
- **John Krebs**, Chief Privacy Officer, Federal Trade Commission
- **Justin Marsico**, Chief Data Officer, Deputy Assistant Commissioner, Bureau of the Fiscal Service, U.S. Department of the Treasury
- **Daniel Morgan**, Chief Data Officer, U.S. Department of Transportation
- **Katerine Osatuke**, Research Director, Veterans Health Administration National Center for Organization Development, U.S. Department of Veterans Affairs
- **Eileen Vidrine**, Chief Data Officer, U.S. Air Force, U.S. Department of Defense

With special thanks for the consistent support and meaningful contributions of the CDO Council, ICSP, and FPC.

Data Ethics Tenets

Federal Data Ethics Tenets help federal data users make decisions ethically and promote accountability throughout the data lifecycle—as data are acquired, processed, disseminated, used, stored and disposed. Regardless of data type or use, those working with data in the public sector should have a foundational understanding of the Data Ethics Tenets. Federal leaders should also foster a data ethics-driven culture and lead by example. The Data Ethics Tenets are:

1 - Uphold applicable statutes, regulations, professional practices, and ethical standards. Existing laws reflect and reinforce ethics. Therefore, data leaders and data users should adhere to all applicable legal authorities. Legal authorities often address historic situations and issues and may not keep pace with the evolving world of data and technology. Organizational leaders are encouraged to maintain up-to-date, comprehensive ethical standards regarding data use and staff are responsible for learning and applying agency guidance appropriately.

2 - Respect the public, individuals, and communities. Data activities have the overarching goal of benefiting the public good. Responsible use of data begins with careful consideration of its potential impacts. Data initiatives should include considerations for unique community and local contexts and have an identified and clear benefit to society.

3 - Respect privacy and confidentiality. Privacy and confidentiality should always be protected in a manner that respects the dignity, rights, and freedom of data subjects. In this context, privacy is the state of being free from unwarranted intrusion into the private life of individuals, and confidentiality is the state of one's information being free from inappropriate access and misuse. An essential objective of privacy and confidentiality protection is to minimize potential negative consequences through measures such as comprehensive risk assessments, disclosure avoidance, and upholding data governance standards. Data activities involving individual privacy should align with the Fair Information Practice Principles (FIPPs).

4 - Act with honesty, integrity, and humility. All federal leaders and data users are expected to exhibit honesty and integrity in their work with data, regardless of job title, role, or data responsibilities. Federal leaders and data users should not perform or condone unethical data behaviors. When sharing data and findings, personnel should report information accurately and present any data limitations, known biases, and methods of analysis that apply. It should also be recognized that no dataset can fully represent all facets of a person, community, or issue. Federal leaders and data users are expected to have humility when presenting data, be open to feedback, and when possible invite discussion with the public. In addition, federal data users should accurately represent their abilities when working with data.

5 - Hold oneself and others accountable. Accountability requires that anyone acquiring, managing, or using data be aware of stakeholders and be responsible to them, as appropriate. Remaining accountable includes the responsible handling of classified and controlled information, upholding data use agreements made with data providers, minimizing data collection, informing individuals and organizations of the potential uses of their data, and allowing for public access, amendment, and contestability to data and findings, where appropriate.

6 - Promote transparency. Individuals, organizations, and communities benefit when the ethical decision-making process is as transparent as possible to stakeholders. Transparency depends on clear communication of all aspects of data activities and appropriate engagement with data stakeholders. Promoting transparency requires engaging stakeholders through easily accessible feedback channels and providing timely updates on the progress and outcomes of data use.

7 - Stay informed of developments in the fields of data management and data science. Advanced technologies provide great benefit to the public sector, but should be deployed with a commitment to accountability and risk mitigation. While traditional data use and analysis can introduce bias, emerging systems, technologies, and techniques require additional awareness and oversight because they can increase opportunities for bias. It is critical to remain informed of developments in the fields of data management and data science, especially as advanced methods impact future data collection, management, and use. In addition, new data innovations (e.g., systems, solutions, computational methods) emerge every day, increasing the importance for federal leaders and employees working with data to keep abreast of market innovations and learn how to ethically use new methods.

Note: The Framework is a "living" resource and to be updated by the CDO Council and ICSP every 24 months.

Content

Overview of Data Ethics Framework

Background

*"Although sometimes described as **the new oil**, because of the way data, and data science, are revolutionizing society just as fossil fuels did earlier, data have unique properties, leading to correspondingly **unique ethical challenges**… Such considerations do not permit simple formulaic answers, since these must be context-dependent and dynamic. Instead, solutions must be principles-based, with higher-level considerations guiding decisions in any particular context."*

- David J. Hand, *Hand Writing: Right, legitimate and proper? The new world of data ethics*[2]

The United States Federal Government is one of the biggest producers and users of data in the world. The exchange of information—or data—with government is necessary for regular tasks, such as renewing a passport, signing up for public assistance programs, and filing taxes. While Americans prize their personal privacy, they are also willing to give information about themselves if it drives a service or concrete public benefit, such as sharing location data to support emergency responders.

Similar to entities worldwide, the functions of the Federal Government rely on data for daily operation, management, and improvement. The ways in which data are collected, linked, analyzed, and shared present tremendous opportunities for government. However, these opportunities are accompanied by significant risks—sometimes unforeseen—to the individuals and communities the government serves. International, state, and local governments, as well as private companies, are making significant strides in the field of data ethics to preserve the benefits and mitigate the risks of data use. The Federal Government plays a critical role in setting the example of ethical oversight by making responsible data decisions and supporting proven values that promote the public good.

Introduction

The Federal Data Strategy, delivered in December 2019, recognized the importance of ethics in its founding principles. The Federal Data Strategy 2020 Action Plan required the development of a Data Ethics Framework (Framework) that is intended to help agency employees, managers, and leaders make ethical decisions as they acquire, manage, and use data. The Framework and its Tenets are a "living" resource and are to be updated by the CDO Council and ICSP every 24 months to ensure the Framework remains current.

The Framework applies to all data types and data uses and incorporates the input and terminology from stakeholders representing many domains, who use different types of data in different ways. The developers of the Framework recognize that some terms may be used differently, depending on the context, type of data being used, and stage in the data lifecycle.

The Framework consists of four parts:

- **About the Data Ethics Framework** outlines the intended purpose and audience of this document.
- **Data Ethics Defined** explores the meaning of the term "data ethics," as background to the Tenets provided in the following section.
- **Data Ethics Tenets** provides the seven Tenets, or high-level principles, for using data ethically within the Federal Government.
- **Data Ethics Tenets in Action** contains use cases demonstrating how the Tenets guide data activities within federal agencies and federally sponsored programs.

Benefits of Data Ethics

The Data Ethics Framework guides the data activities of agencies, providing the foundation for the ethical acquisition, management, and use of data for any federal purpose. Although the ethical challenges that come with data use are many, integrating the Framework's guidance into everyday agency activities will help mature ethical decision-making processes and support benefits across the Federal Government.

Application and use of the Framework drives the following benefits:

- **Consistency.** All federal leaders and data users reference an agreed-upon set of Tenets that help navigate the ethical considerations of data use. Personnel from different domains and fulfilling different roles apply the same foundational ethical considerations.
- **Better, Data-Driven Decisions.** Federal organizations support the use of data-driven decisions for relevant and appropriate purposes. They apply data methods and processes that uncover data limitations, gaps, and biases; facilitate justifiable decisions with data; and communicate known data limitations to promote transparency.
- **Risk Mitigation.** Federal organizations identify, assess, and manage the potential impacts of data activities at each phase of the data and project lifecycle. Federal organizations deploy a proactive approach to data ethics, enabling the effective use of time and resources on their projects and reducing the long-term costs associated with ineffective services and remediation efforts.
- **Increased Transparency.** Federal organizations ensure they document and communicate trustworthy data processes to increase the transparency of their data collection, testing, use, and dissemination activities. Transparency is grounded in clear communication of all aspects of data activities and appropriate engagement with data stakeholders.
- **Consideration of Wider Perspectives.** Federal organizations promote collaboration across internal and external stakeholder groups to better understand data subjects and the impacts of data use. Obtaining a wider perspective also enables data users to better address potential sources of bias.
- **Improved Public Trust.** Federal organizations engender public trust through comprehensive stakeholder engagement, ensuring accountability across the data lifecycle, and reinforcing protocols to protect the privacy, confidentiality, civil rights, and civil liberties of data subjects.

About the Data Ethics Framework

Purpose

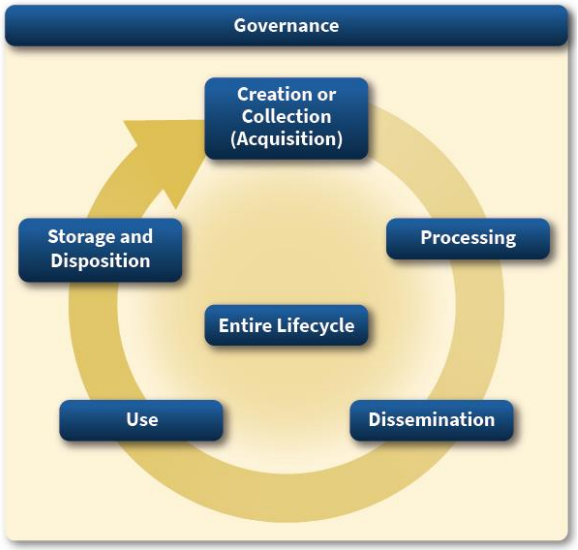
The Framework's purpose is to guide federal leaders and data users as they make ethical decisions when acquiring, managing, and using data to support their agency's mission. The Framework does not include requirements or mandates of its own, but rather provides guidance in the form of Tenets to encourage ethical decision making at all levels of the Federal Government.

Audience

The Framework is intended for use by anyone in the Federal Government who works with or leads work involving data, including employees, contractors, grantees, researchers, and other partners who work on behalf of or as an agent of the government. The ethical use of data is both an individual and organizational responsibility, involving many stakeholder groups that can be internal or external to an agency.

In particular, the Framework is relevant to those who work with data during any stage of the data lifecycle, shown in Figure 1, including collecting, processing, disseminating, using, or storing and disposing of data. This includes:

Figure 1: Data Lifecycle adopted from OMB Circular A-130, Managing Information as a Strategic Resource



- Organizational leaders, including Chief Data Officers (CDOs), Statistical Officials, Evaluation Officers (EOs) heads of agencies, senior executives, and supervisors
- Data practitioners, such as statisticians, data analysts, database professionals, and data scientists
- Program evaluators and operational employees who collect, use, manage, and report data for regular program operations
- Policymakers and those advising decision makers
- Data stewards, including both those who manage data for programmatic purposes and those who manage administrative data, such as human resource employees
- Public relations officials, communications employees, and agency representatives who present information and data to the public
- Data consumers, such as other agencies, communities, or the public
- Research grant recipients across all disciplines

Data Ethics Defined

In the simplest terms, data refer to factual information, such as measurements or statistics, used as a basis for reasoning, discussion, or calculation.[3] ***Data Ethics*** are the norms of behavior that promote appropriate judgments and accountability when acquiring, managing, or using data, with the goals of protecting civil liberties, minimizing risks to individuals and society, and maximizing the public good.

Remaining a leader in data ethics requires individuals, agencies, and cross-agency communities to acknowledge that legal compliance does not guarantee ethical behavior. Therefore, federal leaders and data users must embrace a culture of ongoing discussion, engagement, and learning. Instead of looking at issues from a single perspective, ethical decision making is best achieved by taking a holistic approach and widening the context to weigh the greater implications of data use.

Application of Data Ethics

The Data Ethics Tenets apply to all federal data types and data uses. It is understood that the same dataset may be used at different times for different purposes. No matter the data type or use, federal employees should ensure the protection of privacy, confidentiality, civil rights, and civil liberties during data activities.

For decades, Federal Government laws and regulations[4] distinguished between data uses that can have significantly different ethical impacts (e.g., statistical and non-statistical). This distinction ensures the protection of data, while also enabling the government to better understand the effectiveness of programs and services. The Federal Government recognizes that work with some data, such as weather datasets, requires less scrutiny and is inherently lower risk than information protected by privacy and confidentiality assurances, that require more scrutiny and a full assessment of potential impacts. Each community or agency is ultimately responsible for applying this Framework appropriately based on the context and level of risk for the relevant data. Additional information on this dichotomy is described in Figure 2.

Figure 2: Statistical and Non-Statistical Data Uses[5]

Statistical Data Uses: Statistical activities across the federal enterprise continue to play a vital role in research and program evaluation contributing to policymaking. The design of the Federal Statistical System supports the secure and confidential use of sensitive data exclusively for statistical purposes. Data used for statistical purposes includes describing, estimating, and analyzing the characteristics of a group without identifying the individuals or organizations that make up that group. Statistical data use is governed by a set of enabling legal authorities and protections to reduce collection burden on individuals and businesses, improve the quality of analyses produced by the government, ensure objectivity, and promote the public's trust and confidence in federal use of confidential data while deriving benefits for society.

Non-Statistical Data Uses: Data are often gathered and used to manage the day-to-day work of government services, including for administrative, regulatory, law enforcement, and adjudicative purposes. For example, program administrators often use data to make decisions on an applicant's request to receive funds from a benefits program such as Social Security or the Supplemental Nutrition Assistance Program. Non-statistical data are not governed by the same institutional structures, safeguards, and explicit legal authorities as data used for statistical purposes. For example, using data for daily management of programs requires authorized government personnel to view and make decisions on program delivery using sensitive and confidential data.

Data Ethics Tenets

The Federal Data Ethics Tenets are intended to help federal employees make decisions ethically and promote accountability throughout the data lifecycle. Regardless of data type or use, those working with data in the public sector should have a foundational understanding of the Data Ethics Tenets, and leaders should strive to foster a data ethics-driven culture and lead by example.

Data Ethics Tenets

1. Uphold Applicable Statutes, Regulations, Professional Practices, and Ethical Standards
2. Respect the Public, Individuals, and Communities
3. Respect Privacy and Confidentiality
4. Act with Honesty, Integrity, and Humility
5. Hold Oneself and Others Accountable
6. Promote Transparency
7. Stay Informed of Developments in the Fields of Data Management and Data Science

Each Tenet includes the following:

Each Tenet includes the following actionable guidance for federal leaders and data users to reference.

- **Recommendations for federal leaders and data users** provide actions for ethical data use. It is understood that the responsibilities applied will vary by agency and organization.

Federal Organization Leaders	Federal Data Users
Who? Anyone leading or managing work with federal data **Examples:** Agency leaders (e.g., CDOs, Statistical Officials, EOs, Chief Information Officers, Chief Privacy Officers, Chief Information Security Officers, General Counsels), supervisors, managers, and senior-level personnel	**Who?** Anyone working with federal data **Examples:** Federal employees, contractors, grantees, researchers, and other partners who work on behalf of the government

- **Legal Authorities** provide context for the legal authorities used in the Federal Government.
- **Additional Resources** provide ideas and approaches to data ethics that federal leaders and data users can reference.

The legal authorities and additional resources provided are not exhaustive but are recognized as relevant to federal data ethics.

1. Uphold Applicable Statutes, Regulations, Professional Practices, and Ethical Standards

Data leaders and data users should adhere to all applicable legal authorities, as existing laws reflect and reinforce ethics. Therefore, data leaders and data users should adhere to all applicable legal authorities. Legal authorities often address historic situations and issues, however, and thus may not keep pace with the evolving world of data and technology. Organizational leaders are encouraged to maintain up-to-date, comprehensive ethical standards regarding data use and staff are responsible for learning and applying agency guidance. In addition, if a person works in an area with recognized professional ethical codes of conduct, such as computer science or software engineering, they should be aware of those standards and strive to uphold them. Refer to Figure 3 for additional recommendations, legal authorities, and resources.

Figure 3: Recommendations and Resources for Tenet 1 – Uphold Applicable Statutes, Regulations, Professional Practices, and Ethical Standards

Recommendations for Federal Leaders:

- Identify and clearly communicate the legal authorities, professional codes of conduct, and ethical standards that apply to their organization.
- Develop and maintain agency-level standards and guidance for data ethics.
- Identify and clearly communicate how different types of data within the organization should be handled.
- Recognize the diverse roles of employees working with data throughout the data lifecycle and provide clear guidance and instruction based on the employee's role and level within the organization.
- Provide training and learning opportunities to increase employee knowledge of applicable statutes, regulations, professional practices, and ethical standards.
- Implement mechanisms for reviewing and improving employees' ethical behavior.

Recommendations for Federal Data Users:

- Stay current with data ethics responsibilities and conduct oneself in accordance with the legal authorities, professional codes of conduct, and ethical standards of their organization.
- Perform data activities in accordance with the legal, professional, and ethical standards that apply to their areas of work and the types of data used.
- Understand the policies for handling different types of data within their organization.
- Take training to ensure awareness of the principles of ethical acquisition, management, and use of data as it aligns to their data role and level within the organization.

Legal Authorities:

- All legal authorities cited herein support this Tenet

Additional Resources:

- Office of Government Ethics – 14 General Principles

2. Respect the Public, Individuals, and Communities

Data activities have the overarching goal of benefiting the public good. All other Tenets support this notion, and responsible federal leaders and data users should approach data activities with promoting the "public good" in mind. Responsible use of data begins with careful consideration of its potential and differential impacts. Data initiatives should include considerations for unique community and local contexts, such as for federally recognized Tribes, Alaskan Natives, and local governments, and have an identified and clear benefit to society.

In situations where data on individuals or communities are used to deliver services, data users should engage with impacted stakeholders to better understand those represented by the data and honor guarantees made to data subjects. Refer to Figure 4 for additional recommendations, legal authorities, and resources.

Figure 4: Recommendations and Resources for Tenet 2 – Respect the Public, Individuals, and Communities

Recommendations for Federal Organization Leaders:

- Promote the protection of privacy, civil rights, and civil liberties in their organization's data use.
- Consider the impacts their organization's data activities might have on the public, individuals, and communities and take measures to minimize any negative consequences.
- Establish procedures to mitigate harm when negative consequences are unavoidable.
- Assess stakeholders and implement procedures for the appropriate engagement of those impacted by the organization's data activities.

Recommendations for Federal Data Users:

- Understand that data activities might impact the public, individuals, and communities.
- Promote the public good through data activities.
- Consider the impacts their data activities might have on the public, individuals, and communities and take measures to minimize any negative consequences.
- Take measures to mitigate harm when negative consequences are unavoidable.
- Engage data subjects to better understand and promote their interests when data on individuals or communities are used to deliver services to the public.

Legal Authorities:

- Executive Order: Consultation and Coordination with Indian Tribal Governments (E.O. 13175)
- Federal Information Security Modernization Act (FISMA) (44 U.S.C. § 3551-3558 (2014))
- Foundations for Evidence-Based Policymaking Act of 2018 (Pub. L. 115-435)
- Information Quality Act (IQA) of 2001 (Pub. L. 106-554)
- Intelligence Reform and Terrorism Prevention Act of 2004 (Pub. L. 108-458)
- OMB Memorandum: Improving Implementation of the IQA (M-19-15)
- Tribal Self-Governance Act (25 U.S.C. § 46 (1994))
- Privacy Act (5 U.S.C. § 552a (1974))

Additional Resources:

- National Endowment for the Humanities (NEH) Code of Ethics Related to Native Americans
- Department of Health and Human Services (DHHS) Implementation Guidance on Data Collection Standards for Race, Ethnicity, Sex, Primary Language, and Disability Status

3. Respect Privacy and Confidentiality

Privacy and confidentiality should always be protected in a manner that respects the dignity, rights, and freedom of data subjects. In this context, privacy is the state of being free from unwarranted intrusion into the private life of individuals, and confidentiality is the state of one's information being free from inappropriate access and use. An essential objective of privacy and confidentiality protection is to minimize potential negative consequences, such as the mosaic effect,[6] through measures such as comprehensive risk assessments and disclosure avoidance.

Confidential information obtained by agencies should always be protected by the appropriate access, use, and disclosure limitations. Data activities that involve individual privacy should align with the Fair Information Practice Principles (FIPPs) listed below.

- Access and Amendment
- Accountability
- Authority
- Minimization
- Quality and Integrity
- Individual Participation
- Purpose Specification and Use Limitation
- Security
- Transparency

Figure 5 offers additional recommendations, legal authorities, and resources.

Figure 5: Recommendations and Resources for Tenet 3 – Respect Privacy and Confidentiality

Recommendations for Federal Organization Leaders:

- Provide training to employees on appropriate handling of sensitive data.
- Monitor technological advances that increase or minimize the risk of identification of individuals or entities represented in public datasets and regularly update disclosure protection protocols to mitigate these risks to the greatest extent possible.
- Make clear the tradeoff between confidentiality and granularity of data in public data releases and, to the extent possible, provide tools to enable users to evaluate the impact data protection measures have on results obtained from the data.
- Establish protocols for notifying data providers and other relevant stakeholders if there is a breach that potentially impacts their privacy and provide mechanisms for accommodating victims of those breaches.
- Support and implement mechanisms that limit risks to privacy and confidentiality, such as disclosure limitations and controlled access to data.
- Comply with applicable legal authorities that govern the protection and use of sensitive data, establishing policies and procedures to prevent re-identification of sensitive data made public, maintain the minimum amount of sensitive data necessary, and adhere to data sharing and use agreements.

Recommendations for Federal Data Users:

- Appreciate the rights and responsibilities related to the collection, management, and use of sensitive data.
- Take training to ensure the appropriate handling of sensitive data during everyday activities.
- Take appropriate measures to comply with organizational policies and procedures to prevent re-identification of sensitive data made public, maintain the minimum amount of sensitive data necessary, and adhere to data sharing and use agreements throughout the data lifecycle.
- Use sensitive data only for authorized purposes and without violating protection assurances.
- Access data only with appropriate authorization or approval.

Legal Authorities:

- Defend Trade Secrets Act of 2016 (Pub. L. 114-153)
- E-Government Act (44 U.S.C. § 3501 (2002))
- Federal Information Security Modernization Act (FISMA) (44 U.S.C. § 3551-3558 (2014))
- Foundations for Evidence-Based Policymaking Act of 2018 (Pub. L. 115-435)
- Freedom of Information Act (FOIA) (5 U.S.C § 552 (1967))
- Health Insurance Portability and Accountability Act (HIPAA) (42 U.S.C. § 17932)
- Information Quality Act (IQA) of 2001 (Pub. L. 106-554)
- OMB Memorandum: Improving Implementation of the IQA (M-19-15)
- Federal Policy for Protection of Human Research Subjects ('Common Rule') (42 U.S.C. § 289 (1985))
- OMB Memorandum: Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002 (M-03-22)
- OMB Memorandum: Policy for Preparing for and Responding to a Breach in Personally Identifiable Information (M-17-12)
- OMB Circular: Managing Information as a Strategic Resource (A-130)
- OMB Memorandum: Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices (M-20-12)
- Presidential and Federal Records Act Amendments of 2014 (Pub. L. 113-187)
- Privacy Act (5 U.S.C. § 552a (1974))
- The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002 (Pub. L. 107-347)
- Title 13, United States Code (U.S. Census Bureau)

Additional Resources:

- NIST Privacy Framework
- Department of Homeland Security (DHS), the Fair Information Practice Principles: Framework for Privacy at the DHS

4. Act with Honesty, Integrity, and Humility

All federal leaders and data users are expected to exhibit honesty and integrity in their work with data regardless of job title, data responsibilities, and role in the organization. Federal leaders and data users should not perform or condone unethical data behaviors. When sharing data and findings, personnel should accurately report information and present data limitations, known biases, and methods of analysis that apply. They should also take care not to overgeneralize based on available data and recognize that no dataset can fully represent all facets of a person, community, or issue. Federal leaders and data users are expected to exhibit humility when presenting data, be open to feedback, and invite discussion with the public. In addition, federal data users should accurately and honestly represent their abilities when working with data.

Federal agencies should also support honesty and integrity by clearly defining processes for reporting data ethics concerns and violations, and federal leaders and staff should appropriately implement those processes. It is recommended that each agency develop and communicate policies and procedures to protect those reporting issues. Refer to Figure 6 for additional recommendations, legal authorities, and resources.

Figure 6: Recommendations and Resources for Tenet 4 – Act with Honesty, Integrity, and Humility

Recommendations for Federal Organization Leaders:

- Develop a culture of honesty and integrity within their organizations, setting the example of ethical data acquisition, management, and use for their colleagues to follow.
- Assign qualified personnel to conduct data activities.
- Implement policies and procedures for the appropriate use of data by agency employees.
- Understand and disclose any known limitations, defects, or biases.
- Provide mechanisms for employees to anonymously report ethical violations with data.
- Use techniques to limit bias throughout the data lifecycle.
- Support the objective interpretation of results during analysis.

Recommendations for Federal Data Users:

- Conduct data activities with honesty and integrity, and in accordance with organizational policies and procedures.
- Perform data responsibilities only for which one is qualified.
- Document the data collection and curation process to promote an understanding of the data used in analysis and reporting and to enable reproducibility of results.
- Document and communicate the data lineage to promote understanding of where the data came from, how the data was used, and who used it.
- Document and communicate any known data limitations, defects, or biases that could not be mitigated during data collection.
- Adhere to organizational policies and procedures to report when unethical behaviors are witnessed or suspected.
- Use established methods and protocols to limit bias during data collection wherever possible.
- Remain as objective as possible during analysis.

Legal Authorities:

- Executive Office of the President, Office of Science and Technology Policy (OSTP) Scientific Integrity Policy
- OMB Memorandum: Guidance for Providing and Using Administrative Data for Statistical Purposes (M-14-06)
- OMB Memorandum: Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices (M-20-12)

Additional Resources:

- Department of Homeland Security (DHS), Compliance, Computer Matching Programs
- Department of Education, National Center for Education Statistics (NCES), The Forum Guide to Data Ethics
- Federal Policy for Protection of Human Research Subjects ('Common Rule')
- Office of Government Ethics (OGE) – Where to Report Misconduct
- OMB Questions & Answers When Designing Surveys for Information Collections
- Principles of Artificial Intelligence Ethics for the Intelligence Community
- Artificial Intelligence Ethics Framework for the Intelligence Community

5. Hold Oneself and Others Accountable

Accountability requires that anyone acquiring, managing, or using data be aware of stakeholders and responsible to them, as appropriate. Remaining accountable includes the responsible handling of classified and controlled information, upholding data use agreements made with data providers, minimizing data collection, informing individuals and organizations of the potential uses of their data, and allowing for public access, amendment, and contestability to data and findings when consistent with privacy and national security concerns.

The list of data stakeholders below represents common groups to which anyone working with data might be responsible, though it is not exhaustive.

- Individuals and communities providing data as respondents or serving as research subjects
- Those directly impacted by data use, such as recipients of program services
- Members of the public who rely on data products
- Data consumers, including customers or clients requesting data, who may be internal (e.g., agency program manager) or external to the agency (e.g., Members of Congress or their staffs)

Refer to Figure 7 for additional recommendations, legal authorities, and resources.

Figure 7: Recommendations and Resources for Tenet 5 – Hold Oneself and Others Accountable

Recommendations for Federal Organization Leaders:

- Provide sufficient and appropriate data ethics training and skills analysis to personnel working with, interpreting, and communicating data findings.
- Assign accountability to specific individuals for ethical considerations through the data lifecycle.
- Establish procedures that allow for public access, amendment, and contestability to data and findings, where appropriate.
- Maintain data governance policies and practices over time, updating them when necessary.
- Ensure the ethical collection, handling, and use of classified and controlled organizational data.
- Ensure ethical principles are reflected in the terms and conditions of data sharing and use agreements, such as Memoranda of Understandings (MOU), Inter-Agency Agreements (IAA), and contracts.
- Document processes for data activities and decisions to enable accountability, auditing, and oversight.
- Consider providing centralized guidance of data ethics within agencies.

Recommendations for Federal Data Users:

- Consider stakeholders while conducting data activities and determine appropriate engagement, keeping their interests in mind and upholding the public trust.
- Take data ethics and skills training to improve one's ability to work with, interpret, and communicate data findings.
- Allow for public access, amendment, and contestability to data findings, where appropriate and in accordance with organizational policies.
- Uphold data governance policies and data ethics standards practices.
- Access, use, and share classified and controlled organizational data only for authorized purposes.
- Uphold data use agreements made with data providers.
- Document how data are collected, curated, and analyzed for accountability purposes.

Legal Authorities:

- Executive Order: Classified National Security Information (E.O. 13526)
- Executive Order: Controlled Unclassified Information (E.O. 13556)
- Foundations for Evidence-Based Policymaking Act of 2018 (Pub. L. 115-435)
- Information Quality Act (IQA) of 2001 (Pub. L. 106-554)
- OMB Memorandum: Improving Implementation of the IQA (M-19-15)
- OMB Circular: Managing Information as a Strategic Resource (A-130)
- OMB Memorandum: Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices (M-20-12)
- Paperwork Reduction Act (44 U.S.C. § 35 (1995))

Additional Resources:

- United States Census Bureau's Data Stewardship Executive Policy Committee
- NIST Privacy Framework
- United States Geological Survey's Data Sharing Agreements

6. Promote Transparency

Individuals, organizations, and communities benefit when the ethical decision-making process is as transparent as possible to stakeholders. Transparency depends on clear communication of all aspects of data activities and appropriate engagement with data stakeholders. Promoting transparency requires engaging stakeholders through easily accessible feedback channels and providing timely updates on the progress and outcomes of data use. Refer to Figure 8 for additional recommendations, legal authorities, and resources.

Figure 8: Recommendations and Resources for Tenet 6 – Promote Transparency

Recommendations for Federal Organization Leaders:

- Nurture a culture that supports appropriate, transparent reporting of the organization's data activities and products.
- Establish standards and provide training for data preparation, documentation, and presentation to promote accuracy and consistency, as well as improved understanding by stakeholders.
- Promote clear guidance that ensures data are made available for research equitably and objectively.
- Implement standards to document understandable descriptions of analytical methods and models to be shared with appropriate stakeholders.
- Establish procedures for making corrections to previous reporting that might contain errors, explaining what was inaccurate and corrected, if feasible.

Recommendations for Federal Data Users:

- Take training on data preparation and presentation.
- Follow standard data preparation and presentation methodologies.
- Make data available for research in an equitable and objective manner.
- Document data activities in ways that can be clearly communicated to stakeholders.
- Document metadata accurately, as necessary.
- Follow procedures to correct previously reported data that might contain errors, providing clear explanation of what was inaccurate and how it was corrected if feasible.

Legal Authorities:

- Foundations for Evidence-Based Policymaking Act of 2018 (Pub. L. 115-435)
- Information Quality Act (IQA) of 2001 (Pub. L. 106-554)
- OMB Memorandum: Improving Implementation of the IQA (M-19-15)
- OMB Memorandum: Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices (M-20-12)
- OMB Statistical Policy Directive: Release and Dissemination of Statistical Products Produced by Federal Statistical Agencies (4)
- Paperwork Reduction Act (44 U.S.C. § 35 (1995))
- Privacy Act (5 U.S.C. § 552a (1974))
- The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002 (Pub. L. 107-347)
- The Plain Writing Act of 2010 (Pub. L. 111-274)

Additional Resources:

- Principles of Intelligence Transparency for the Intelligence Community
- Commission on Evidence-based Policymaking (CEP) Final Report: The Promise of Evidence-Based Policymaking
- Environmental Protection Agency's (EPA) Strengthening Transparency in Regulatory Science

7. Stay Informed of Developments in the Fields of Data Management and Data Science

Advanced technologies provide great benefit to the public sector, but should be deployed with a commitment to accountability and risk mitigation. Emerging systems, technologies, and techniques require additional awareness and oversight as they can present new—and sometimes hard to detect—opportunities for bias. Below are examples of why emerging systems, technologies, and techniques require additional oversight.

- Responsibility for emerging systems, technologies, and techniques may fall to those who do not have the requisite training or experience to perform the work.
- Advanced technologies and computational methods can lead to algorithms and automations that make probabilistic recommendations and decisions on behalf of programs (e.g., acceptance of application for benefit program) with minimal human oversight.
- Automations and probabilistic recommendations may have differential impacts that can only be identified through deep analysis and review (e.g., facial recognition technology that benefits one group while having serious negative consequences for another).

Since new data innovations emerge every day, federal leaders and data users should keep abreast of innovations and how to use those methods in an ethical manner. Examples of advanced technologies, analytics, and computational methods that should be monitored and assessed are below.

- Artificial Intelligence (AI)
- Machine Learning (ML)
- Neural Networks (NNs)
- Robotic Process Automation (RPA)
- Internet of Things (IoT)
- Blockchain

Refer to Figures 9 and 10 for additional recommendations, legal authorities, and resources.

Figure 9: Recommendations for those using AI, ML, Advanced Analytics, and Computational Technologies

For Those Using AI, ML, Advanced Analytics, and Computing Technologies:

The Framework does not offer in-depth guidance on advanced technologies and computational methods. The Framework is generally applicable to those methods, but also references other federal initiatives that more thoroughly address the ethical considerations associated with the deployment of such methods. The Framework encourages those who use AI, ML, advanced analytics, and computing technologies to:

- Design, develop, and use those methods while respecting the law and acting with integrity.
- Provide appropriate transparency regarding methods, including models and underlying data, applications, and uses.
- Understand and disclose any known limitations, defects, or biases, including those of the underlying data or models.
- Develop and employ mechanisms to identify responsibilities and provide accountability.
- Institute rigorous protocols to evaluate outputs for bias and implement mitigation strategies as needed.
- Prioritize human-centered development and use.
- Leverage mechanisms to communicate the design, development, and use process to non-technical stakeholders.
- Engage other scientific and technology communities to leverage best practices.
- Seek ethical guidance around specific emerging techniques.

Figure 10: Recommendations and Resources for Tenet 7 – Stay Informed of Developments in the Fields of Data Management and Data Science

Recommendations for Federal Organization Leaders:

- Develop a diverse workforce to support the policies, oversight, and governance structure for any large-scale system that learns from data to limit bias and to best consider societal and cultural consequences.
- Provide or support training in data science and any new systems, technologies, or techniques if required for job function.
- Employ advanced methods in ways that fully comply with applicable legal authorities, policies, and procedures that protect privacy, civil rights, and civil liberties.
- Monitor advanced methods and how they might impact their organization's data activities.
- Hold employees responsible for staying abreast of advanced methods and how they might be used for data activities in their areas of work.
- Establish protocols explicitly designed to identify and mitigate bias, as well as assign accountability, when designing, developing, and deploying advanced methods.
- Ensure human involvement in development and use of advanced methods.

Recommendations for Federal Data Users:

- Take adequate training or develop required knowledge in data science and any systems, techniques, and technologies before applying such expertise in the field.
- Deploy advanced methods in ways that fully comply with applicable legal authorities and organization policies and procedures.
- Keep abreast of advanced methods and how they might impact data activities in their areas of work.
- Promote accountability and properly mitigate risks when designing, developing, and deploying advanced methods.
- Involve human judgement when advanced methods might interfere with privacy, civil rights, or civil liberties.

Legal Authorities:

- Executive Order on Maintaining American Leadership in Artificial Intelligence (AI)

Additional Resources:

- Department of Defense (DoD) Joint Artificial Intelligence Center (JAIC)
- Federal RPA Community of Practice (CoP)
- General Services Administration (GSA) AI Center of Excellence (CoE)
- NIST Cybersecurity Framework
- Principles of Artificial Intelligence Ethics for The Intelligence Community
- Artificial Intelligence Ethics Framework for the Intelligence Community

Data Ethics Tenets in Action

Use Cases

The Use Cases herein provide examples of ethical considerations with data encountered while doing federal work. Each Use Case is an open-ended scenario designed to help readers contemplate the issue and ethical considerations at each phase of the data lifecycle.

The questions included are examples of the important considerations for federal leaders and data users before, during, and after data initiatives. They are not meant to be all inclusive. In addition, it is recognized that ethical use of data is supported by the ethical behavior of organizations. The Framework's Use Cases aim to complement existing cultural norms and training initiatives that reinforce ethical behavior across the government.

1. Use Case: Artificial Intelligence & Bias

Artificial Intelligence & Bias	
Organization Type:	Government Benefit Agency
Primary Objective:	Improve administration of government benefits
Secondary Objective:	Leverage administrative data and advanced technologies (i.e., AI) to improve performance
Scenario:	An agency that oversees administration of benefits collects large amounts of applicant data on a daily basis. To streamline the application process, the agency engaged an outside party to create an automated tool that makes decisions on applicant eligibility for the benefits program. The tool relies on models that gather data from different parts of the organization, including applicant employment and financial records, and analyzes the chances of applicant success within the program. In operation for over two years, the tool helps eliminate numerous manual processes, identifies potential fraud, and better deploys limited resources. During this time, the tool has made thousands of decisions on applicant eligibility for benefits, impacting countless lives. Judy, who works for the agency's outreach department, has received an increasing volume of complaints from applicants in recent months. Applicants have consistently stated they are inappropriately screened and removed from the application process. Judy brings this issue to her management, which requests an internal review. The internal review finds that data sharing agreements required for the original, underlying data to be destroyed upon the tool's deployment on the agency's customer-facing website. As a result, the agency is unable to reproduce, assess, or scrutinize aspects of the tool and its supporting decision models.
General Use Case Questions:	• Was the agency justified in using applicant data to improve its own processes via automated decision models? *(Reference Tenet 2 – Respect the Public, Individuals, and Communities; Tenet 3 – Respect Privacy and Confidentiality; Tenet 5 – Hold Oneself and Others Accountable; Tenet 7 – Stay Informed of Developments in the Fields of Data Management and Data Science)*

	• Could altered data sharing agreements help enable reproducibility of the impactful decision models? *(Reference Tenet 5 – Hold Oneself and Others Accountable; Tenet 7 – Stay Informed of Developments in the Fields of Data Management and Data Science)* • What ethical considerations should have arisen during the design, development, and deployment of the automated tool? *(Reference Tenet 2 – Respect the Public, Individuals, and Communities; Tenet 3 – Respect Privacy and Confidentiality; Tenet 5 – Hold Oneself and Others Accountable; Tenet 7 – Stay Informed of Developments in the Fields of Data Management and Data Science)* • Does the tool support human judgement by agency personnel or does it make decisions – without a human-in-the-loop – on behalf of the program? *(Reference Tenet 7 – Stay Informed of Developments in the Fields of Data Management and Data Science)*
Data Lifecycle Questions	**Creation or Collection (Acquisition)** • Are there any limitations on the data the agency is allowed to collect for program administration purposes? *(Consider legal authorities, if this data has already been collected, and if it has been collected, by who)* • Should members of the public know how data collected during the application process is ultimately used? *(Consider the data subjects and the uses – and potential uses – of their data that have been communicated)* • Should all data being collected serve a distinct purpose? *(Consider why this data is being collected and if it needs to be collected in the first place)* • How could bias in the collection process be mitigated? *(Consider how bias might enter the data collection process and proper representation of data subjects)* **Processing** • Are there issues with leveraging data from other parts of the organization (e.g., employment and financial records) for model development? *(Consider the data subjects, the uses of their information that have been communicated, and any access or use limitations)* • Should there be oversight measures in place to ensure the correct records are linked across the organization? *(Consider how a lack of oversight could lead to mistakes in data linkage activities and negatively affect the impacted data subjects)* • Should processing activities and effects on data quality be documented? *(Consider what needs to be documented – and how – to promote accountability to data stakeholders)* **Dissemination** • What should the leaders and data users have considered when arranging the data sharing agreement with the outside party? *(Consider reproducibility, the data subjects, and organizational accountability)* **Use** • Is the automated tool allowing for the appropriate level of human judgement involved in decisions produced? *(Consider automation that helps flag certain items for human judgement, versus automation that makes decisions on behalf of programs)* • Should usage of data produced by the automated tool be monitored? *(Consider human-centered development and use)*

	• Should the probabilistic recommendations from the automated tool be reviewed and/or validated? *(Consider human-centered development and use)* • Should applicants be aware that their personal data are being used to drive automated decisions on benefits, and should they be allowed to contest those decisions? *(Consider what uses of their data have been communicated to data subjects, and appropriate measures for access, amendment, and contestability)* • How often is the tool assessed to consider data drift, model decay, and the tool's recommendations? *(Consider appropriate oversight of the tool and how often its outputs should be evaluated, along with mitigation strategies)* **Storage and Disposition** • Should the underlying data and code supporting the automated tool have been stored for accountability purposes? *(Consider reproducibility, the data subjects, accountability, and transparency)* • Should the agency have considered documenting analytical methods and results in this scenario? *(Consider what needs to be documented to promote accountability and transparency)* • Should the agency have procedures in place to dispose of the data after a certain timeframe? *(Consider records management risks and compliance)*

2. Use Case: Dissemination & Impacts

Dissemination & Impacts	
Organization Type:	Government Inspection Agency
Primary Objective:	Enforce the law and minimum standards of the Animal Welfare Act
Secondary Objective:	Publish reports to increase the public's understanding of adherence to the Act's standards
Scenario:	The Animal Welfare Act (AWA) regulates the treatment and care for certain animals bred for commercial sale, used in research, transported commercially, or exhibited to the public. The government inspection agency enforces this law, setting minimum standards as the baseline by which they assess a facility's care for animals. Each year, the agency collects, manages, and analyzes AWA records from animal facilities (e.g., animal shelters, retailers) to support its mission and perform reporting. In addition, the agency posts information and investigations and inspections results to a public website. The reports give the public a picture of the compliance of all entities licensed or registered under the AWA. Due to a pending lawsuit and privacy concerns, the government inspection agency removed certain reports and AWA records from its website that identified animal abuse complaints (i.e., complaints not convictions). Users, activists, and the public were upset, as posting the AWA records not only seemed like the right thing to do, but the agency has also taken measures to redact some information previously posted. The agency communications department is in a difficult situation—there is a demand for public transparency, but legal matters and a transition in agency leadership leave the team with conflicting views on how to proceed.
General Use Case Questions:	• Is the agency justified in removing certain reports and AWA records from its website, instead of trying to redact certain private information? *(Reference Tenet 1 – Uphold Applicable Statues, Regulations, Professional Practices, and Ethical Standards; Tenet 3 – Respect Privacy and Confidentiality; Tenet 6 – Promote Transparency)* • Should the agency have disclosure limitation methodologies in place to prevent this type of situation? *(Reference Tenet 5 – Hold Oneself and Others Accountable)* • How could the agency weigh consumers' right to know potential animal abuse violations and also protect privacy of those not convicted? *(Reference Tenet 1 – Uphold Applicable Statues, Regulations, Professional Practices, and Ethical Standards; Tenet 2 – Respect the Public, Individuals, and Communities; Tenet 3 – Respect Privacy and Confidentiality; Tenet 4 – Act with Honesty, Integrity, and Humility; Tenet 6 – Promote Transparency)* • Should the agency have policies and procedures in place to provide guidance in this situation, despite the transition in agency leadership? *(Reference Tenet 1 – Uphold Applicable Statutes, Regulations, Professional Practices, and Ethical Standards; Tenet 3 – Respect Privacy and Confidentiality; Tenet 5 – Hold Oneself and Others Accountable)*

Data Lifecycle Questions	**Creation or Collection (Acquisition)** • Does the agency have the authority to collect and release this information? *(Consider legal authorities, if this data has already been collected, and if it has been collected, by who)* • What compliance documentation already exists, and does any documentation need to be updated? *(Consider privacy, security, records management, and relevant legal authorities)* • How will data quality be checked and be documented? *(Consider how you know what you know, the data's lineage and provenance, and what activities with the data need to be documented)* • Who needs to be involved in the data collection? *(Consider coordination that needs to occur – internally and externally – to ensure accountability to data subjects)* • Have the entities (e.g., animal shelters, retailers) been informed of the data to be collected and how it will be used? *(Consider the data subjects and the uses – and potential uses – of their data that have been communicated)* **Processing** • Once acquired, what refinement does the data require before release? For example, what private or confidential information should be removed? *(Consider privacy and confidentiality protections, as well as what preparation must occur before data is shared publicly)* • What considerations need to be weighed in protecting privacy while analyzing the data? *(Consider appropriate access and use during analysis, and what identifying information should be left out to support objectivity)* • What additional information around data collection, maintenance, and use needs to be shared to promote public transparency? *(Consider the documentation of data collection, prep, and analysis activities that will support findings and public transparency)* **Dissemination** • What data need to be provided to the public? *(Consider what data products and activities need to be shared, along with the appropriate level of detail)* • How can data be shared with external stakeholders? *(Consider public feedback channels and appropriate posting locations for public use data)* • Are there open feedback channels for stakeholders to share insights or to report incorrect information? *(Consider the organization's feedback mechanisms and coordination with external consumers)* • Are the data presented in a consumable way? *(Consider presenting data in ways that are consumable for different stakeholder groups)* **Use** • Can data consumers use the data in a nefarious or harmful way? *(Consider ways the data can be used once made public)* • What types of decisions or uses could data consumers make with the data? *(Consider ways the data can be used once made public)*

	• Could the data be joined with other data to identify personal identities or confidential information? *(Consider other public datasets, the potential for re-identification, and the associated risks)* **Storage and Disposition** • Which records schedule applies to this data? *(Consider records management risks and compliance)* • How long should this data be maintained? *(Consider records management risks and compliance)* • Does the data storage mechanism document any changes in data collection or curation that could impact the long view of the data? *(Consider what needs to be documented to support future accountability and transparency)*

End Notes

[1] Federal Data Strategy 2020 Action Plan, December 2019, available at https://strategy.data.gov/action-plan/

[2] An article by David J. Hand on the ethical, social, and policy challenges associated with the rise of "big data," available at https://imstat.org/2018/12/14/hand-writing-right-legitimate-and-proper-the-new-world-of-data-ethics/

[3] Definition of data by Merriam Webster, available at https://www.merriam-webster.com/dictionary/data

[4] Evolution of Evidence Building in the United States, The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking, Pages 12-16, available at https://www.cep.gov/report/cep-final-report.pdf

[5] Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5544, available at https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

[6] Description of the Mosaic Effect, available at https://aspe.hhs.gov/report/minimizing-disclosure-risk-hhs-open-data-initiatives/c-mosaic-effect