Transforming How
Texas Government
Serves Texans

Data Ethics in Data Management

Data Management Advisory Committee

Office of the Chief Data Officer
Texas Department of Information Resources

July 2023

Table of Contents

Introduction....1
Tenets of Data Ethics....1
Tenet 1: Uphold applicable statutes, regulations, professional practices, and ethical standards. 1
Tenet 2: Respect the public, individuals, and communities. 2
Tenet 3: Respect privacy and confidentiality. 3
Tenet 4: Act with honesty, integrity, and humility. 3
Tenet 5: Hold oneself and others accountable. 4
Tenet 6: Promote transparency.....4
Tenet 7: Stay informed of developments in the fields of data management and data science....5
Establishing an Ethical Data Culture....6
Conclusion....8
References....8
Authors....8

Introduction

Ethics are principles of behavior based on ideas of right and wrong, and these principles often focus on concepts such as fairness, respect, integrity, transparency, and trust. Data ethics are the norms of behavior that promote appropriate judgments and accountability when creating or acquiring, managing, sharing, using, and disposing of data. Unethical handling of data can result in damage to the organization's reputation, financial loss due to penalties and fines, and loss of customers because it puts people whose data was exposed at risk. As such, data management professionals have an ethical responsibility to manage data in a way that reduces the risk that it may be inappropriately accessed, misused, misrepresented, or misunderstood.

While data may seem as if it is merely technical information, its use must be guided by ethical principles, or it will present a risk to an organization's success. All employees of Texas state agencies and institutions of higher education are accountable for ethical data handling and preserving privacy with an appropriate level of transparency.

The ethical use of data focuses on several core concepts:

Impact on people – Data represents characteristics of individuals and is used to make decisions that affect people's lives; therefore, it is imperative to manage its quality and reliability.

Potential for misuse – Misusing data can negatively affect people and organizations; therefore, it is imperative to prevent the misuse of data.

Economic value – Data has an economic value; therefore, it is imperative to determine how that value can be accessed and by whom.

Employees of Texas state agencies and institutions of higher education who use data in any way are responsible for having a foundational understanding of the basic tenets of data ethics that go beyond these core concepts so they may practice ethical behaviors when working with data. Data Management Officers, working with the organization's privacy, information security, and records management offices, may wish to consider developing a data ethics framework and policy for their organization that goes beyond basic mandated ethics training.

Tenets of Data Ethics

The Federal Data Strategy outlines seven tenets around the ethical use of data, and each tenet should be considered at every stage of the data lifecycle, from the creation or acquisition of data, to processing data, and using it to make business decisions, and even when disposing of data.

Tenet 1: Uphold applicable statutes, regulations, professional practices, and ethical standards.

Existing laws reflect and reinforce ethics. Therefore, information producers and consumers should adhere to all applicable laws and regulations, many of which are discussed in the Texas Data Literacy Program course on data privacy. Each organization's data governance program should establish its own standards and policies regarding the ethical handling of data specific to the organization.

Tenet 1 example: Imagine you're a data analyst working for a large Texas state agency. You have access to confidential information about the agency's clients and employees. One day, you're approached by a friend who runs a small business and is in desperate need of information about a potential client. Your friend offers to pay you for the information and now you're faced with a dilemma. While you could always use the extra money, you know that sharing confidential information would be a violation of your agency's policies and could even be illegal. In this scenario, choosing to uphold applicable statutes, regulations, professional practices, and ethical standards would mean declining your friend's offer and protecting the confidential information. This would demonstrate your commitment to following the laws and ethical principles as a data analyst, as well as protecting the rights of the agency's clients and employees.

Tenet 2: Respect the public, individuals, and communities.

Data-related activity can be lawful yet still be unethical. In other words, legal compliance does not always guarantee and enforce ethical behavior. Data users may collect, use, or share data legally, but when analyzing and making decisions based on that data, they must be aware of any potential negative consequences to people, the community, or the environment due to issues such as incomplete data, misinterpretation of results, or underlying biases and discrimination.

Data activities should have the overarching goal of benefiting the public good, and 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.

Tenet 2 example: Let's say you're a data scientist tasked with proposing improvements to the public transportation system in a big city. Excited about the opportunity to make a positive impact, you gather some data on bus and train routes and schedules, along with information on who uses the transport system and when. As you delve into the data, you realize that the bus routes serving low-income neighborhoods are less frequent and less reliable compared to those serving wealthier areas. Remembering data ethics, you meet with your supervisor, present your findings, and propose solutions to make the transportation system more equitable for everyone.

Here are some questions to consider as you work with data:

- ✓ Are you familiar with the laws, regulations, and ethical standards related to the data you use?
- ✓ Are there any underlying biases in the data collected that might have a negative impact or increase inequalities among certain populations or cultures?
- ✓ What are the potential benefits and drawbacks of using the data for a specific purpose?
- ✓ What would be the harm in excluding or not using certain data in your analysis?
- ✓ What are the environmental implications of the project and how might risks be mitigated?

Tenet 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, disclosure, 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, policies, and procedures.

Tenet 3 example: Imagine you work for an agency that manages personal health records for millions of patients and one day, a co-worker asks you to access the health records of a local celebrity to see what kind of treatments they have been receiving. You decline and remind your co-worker about the importance of keeping patient information confidential and only accessing it for legitimate and authorized purposes. Breaking this trust can have serious legal and ethical consequences. By upholding data ethics, you and your colleague ensure that the personal information of millions of patients remains protected.

Tenet 4: Act with honesty, integrity, and humility.

All data users are expected to exhibit honesty and integrity in their work with data, regardless of job title, role, or data responsibilities. Employees should not perform or condone unethical data behaviors. When sharing data and findings, employees should report information accurately and present any data limitations and known biases. All sources and methods used to obtain and analyze data, including any pre-processing, should be fully disclosed, and detailed explanations should be provided for any exclusions. Information producers and consumers are expected to exhibit humility when presenting data, be open to feedback, and when possible, invite discussion with the public.

Tenet 4 example: Imagine your organization manages a lot of sensitive customer information and one day your manager asks you to alter a customer activity report to make the organization look a little bit better. As a data professional, you know this goes against data ethics. You remind your manager it is the organization’s ethical responsibility to maintain the accuracy of the information and altering the report would be a breach of the customer's trust. Presenting the report as it is, even if it doesn't reflect the best results for the organization, is the right choice because the customers' trust is more valuable in the long run. You also remind your manager that being open to feedback provides opportunities for the organization to make improvements.

Here are some questions to consider as you work with data:

- ✓ Are you using data only for its explicitly defined and appropriate use?
- ✓ Could the data be combined with other datasets that could increase the risk of violating an individual’s privacy and confidentiality?
- ✓ Have you provided notes describing what data is included, what data is excluded, and why?
- ✓ Have you presented information in a way that could be misunderstood or is inaccurate?

Tenet 5: Hold oneself and others accountable.

Accountability requires that anyone working with or using data at any point in the data lifecycle, be aware of, and responsible to, any and all stakeholders. Stakeholders may include data providers, such as research subjects or persons receiving state-provided services, or data consumers, such as internal or external customers requesting data and information. Accountability includes:

- the responsible handling of classified and controlled information,
- upholding data use agreements made with data providers and consumers,
- 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.

Tenet 5 example: Let's say one of your colleagues approaches you and asks for some confidential customer data that you have access to, but you know that sharing this information without a legal data sharing agreement in place would be a violation of your organization's privacy policy of handling classified and controlled information. Your colleague says they need the data quickly for a very important project and they don't have time to wait for the legal documents to be created and approved. It's your responsibility to act with integrity and follow the rules, even if it means going against a colleague, so you decline the request. You explain why you can't provide the data and remind your colleague of their responsibility for demonstrating ethical behavior. Your colleague is frustrated at first but understands and respects your decision. By upholding the data ethics tenet of accountability, you protect your organization's reputation and the privacy of its customers, but you also demonstrate your own ethical character and commitment to doing the right thing.

Tenet 6: Promote transparency.

Individuals, organizations, and communities' benefit when the decision-making process is as transparent as possible to stakeholders. Transparency depends on clear communication of all aspects of data activities throughout the entire data lifecycle. Promoting transparency requires engaging stakeholders through easily accessible feedback channels and providing timely updates on the progress and outcomes of data use. Transparency also includes standard processes and documentation when correcting previously reported data that might contain errors, including providing explanations of what was inaccurate and how it was corrected.

Tenet 6 example: Imagine you're a data analyst and while analyzing some customer data for trends that could help improve the organization's provision of services, you realize that some of the information is missing. Instead of just ignoring it you decide to bring this to your boss's attention. You explain how the missing data could potentially impact the accuracy of your analysis. Your boss agrees and together, you come up with a plan to be transparent about the missing data in your final report. You include a note explaining that some data was missing and the potential impact it could have on the analysis, and on any business, decisions made based on the analysis. You also suggest ways to gather more complete data in the future. By promoting transparency, you're doing the right thing ethically. You are making sure the analysis is as accurate and trustworthy as possible, and acknowledging the limitations allows you to plan for more complete analyses in the future.

Here are some questions to consider as you work with data:

- ✓ Does a data sharing agreement or other legal document need to be in place?
- ✓ Could sharing confidential data with another organization have a negative impact on individuals or communities?
- ✓ Does the data need to be collected in the first place?
- ✓ Are the rights of individuals balanced against the need for transparency?
- ✓ Are you disposing of data in accordance with records retention schedules?

Tenet 7: Stay informed of developments in the fields of data management and data science.

Advanced technologies can provide great benefits to the public sector but should be deployed with a commitment to accountability and risk mitigation. Organizations are increasingly using interactive visualization tools to present data and tell their data stories, and visualizations of the data should be designed in such a way to minimize the risk of misinterpretation. In planning such projects or analyses, it is also critical to make sure the data feeding these visualization tools is secure, and analytic tools should be configured in such a way that unauthorized users will not be able to download sensitive, confidential, or protected data, such as personally identifiable information or personal health information, either intentionally or by accident.

New data innovations emerge every day with the potential to deliver insights more efficiently and effectively, so it is important for data users to remain informed of developments in the fields of data management and data science. Thorough documentation of reference and master data, metadata, and security classifications are essential for protecting sensitive, confidential, and regulated data from unauthorized or unethical use. The use of more advanced and emerging technologies requires deliberate awareness and oversight as they are often based on considerable amounts of data and employ complex algorithms which can increase opportunities for risk such as privacy and security breaches, or even misinterpretation of results based on incomplete or poor-quality data.

Further, certain privacy laws and regulations stipulate that an individual be notified on how their personal information will be used and additional consent should be obtained if their personal information will be used for purposes not described when the information is originally collected. Data users should work with the organization’s ethics, privacy, or legal office, and the information security office to ensure the data is handled responsibly and ethically in data science and other analytic practices.

Tenet 7 example: Let’s say you are the director of a client service delivery program within your organization, and you are excited about the possibilities offered by artificial intelligence and machine learning tools to inform and drive improvements in service delivery. You know that advanced analytics algorithms are based on vast amounts of data, so the quality of the data must be high to get accurate results. You also know it is essential to prioritize human-centered approaches to the development and use of these emergent technologies to reduce bias, increase transparency, and uphold other ethical tenets. You realize that the field of data management and data science is constantly evolving, and new techniques and tools are being developed all the time. To make sure you understand the appropriate and ethical use of these new techniques and tools, you make it a point to stay informed of the latest developments in the field by subscribing to industry newsletters and attending conferences and workshops to network with data professionals and learn about new trends and best practices in the field. By staying informed, you are enhancing your ability to make better business decisions using the most up-to-date technologies possible while also reducing bias, increasing transparency, and upholding other ethical tenets.

Here are some questions to consider as you work with data:

- ✓ What happens when AI and ML efforts are based on poor quality data, or if the quality isn't defined?
- ✓ What happens when analytic efforts are based upon biased or incomplete data?
- ✓ How might that impact business decisions based on AI and ML efforts?

Establishing an Ethical Data Culture

Establishing a culture of ethical data handling within the organization requires understanding existing practices, defining expected behaviors, codifying these in policies and a code of ethics, and providing training and oversight to enforce expected behaviors. The steps below describe some best practices to help Data Management Officers establish an ethical data culture in the organization.

1) Data Management Officers should be aware of the privacy and security laws and regulations pertinent to the data created, acquired, managed, and used by the organization, and work collaboratively with those offices to develop a data ethics framework.

2) Identify principles, practices, and risk factors within the organization that align with legal and regulatory compliance requirements. For example:

    a) Guiding Principle – People have a right to privacy with respect to information about their health. Therefore, the personal health data of patients should not be accessed except by people who are authorized to access it as part of administering the program or providing care for patients.

    b) Risk – If there is wide access to the personal health data of patients, then information about individuals could become public knowledge, thereby jeopardizing their right to privacy.

    c) Practice – Only personnel with authorized permission to access the data for administrative purposes, and only nurses and doctors, or others directly involved in patient care, are allowed to access the personal health data of patients.

  d) Control – There is an annual review of all users of the systems that contain personal health information of patients to ensure that only those people who need to have access do have access.

3) Review the current state of data handling practices to identify how well employees understand the ethical implications of existing practices in preserving the trust of customers, partners, and other stakeholders. Plan to address gaps in employees' understanding of data ethics when developing a training program on the ethical handling of data.

4) Adopt a socially responsible ethical risk model. Data professionals involved in business intelligence, analytics, and data science are often responsible for data that describes, who people are, what people do, where people live, and how people are treated. Data can be misused and counteract the principles underlying data ethics if not handled appropriately. A risk model can be used to determine whether to execute a particular project or not. It should also influence how to execute the project including how the data will be redacted, how the data will be limited to the purpose of the project, security of the data files, and review of the applicable privacy laws and regulations.

5) Create an ethical data handling strategy and roadmap that describe both ethical principles and expected behavior related to data, expressed in value statements and a code of ethical behavior. The components of a such a strategy may include:

  a) Value statements – Describe what the organization believes in. Examples include truth, fairness, or justice.

  b) Ethical data handling principles – Referencing the Federal Data Strategy's Tenets of Data Ethics, describe how an organization approaches challenges presented by data; for example, how to respect the right of individuals to privacy, or how to promote transparency and accountability.

  c) Compliance framework – Develop a framework that includes factors that drive organizational obligations to meet compliance requirements.

  d) Risk assessments – Define and outline on-going data ethics risk assessments to identify the likelihood and the implications of specific ethical issues arising within the organization.

  e) Training and communication – Develop training to review data ethics as it pertains to the data managed by the organization. All staff who work with data in any capacity should take the training and sign off that they are familiar with the code of ethics and understand the implications of unethical handling of the data.

  f) Auditing and monitoring – Monitor specific activities to ensure that they are being executed in compliance with ethical principles.

  g) Roadmap – Develop a roadmap that includes a timeline with activities such as execution of the training and communications plan, the identification and remediation of gaps in existing practices, risk mitigation, and monitoring plans.

Conclusion

As employees of Texas state agencies and institutions of higher education, it is our responsibility to be good stewards of our data and information. As more and more data are being collected on individuals, data ethics is an increasingly important issue in the public sector. Careful consideration of data ethics will go a long way to make sure we collect, handle, protect, and use data in an ethical manner.

This whitepaper analyzes seven core tenets that govern efficient, integrity-based handling methods and provides practical examples of their application, offering a comprehensive framework for ensuring ethical practices throughout organizational operations involving sensitive information.

The document identifies the importance of prioritizing moral obligations in response to potential fallout linked to privacy breaches or security risks. At all times, organizations must actively uphold transparency and ensure accurate representation, while consistently demonstrating accountability, especially when operating high-risk activities, such as prevalent AI-powered systems that dramatically increase the opportunities for bias among data users.

Organizations foster a culture of accountability when they prioritize adherence to the core principles of ethical data handling. The paper suggests adopting a compliance framework, conducting regular assessments of data handling practices, implementing an ethical risk model, and refining leadership strategies with potent strategy statements that direct everyone's decision-making process towards responsible use of information. By executing these measures, organizations build a robust foundation that promotes good stewardship.

Finally, the paper underlines the critical responsibility of stakeholders within the public sector to approach their duties surrounding sensitive information with the highest level of integrity. As technology and innovation continually shape our society, the need for adopting an ethical approach towards data handling has never been more significant.

If you have any questions about any of the ethical tenets presented here in this whitepaper, please consult your organization’s legal counsel.

For more information on data governance and data management principles, visit the Office of the Chief Data Officer website (dir.texas.gov/office-chief-data-officer).

References

Federal Data Strategy Data Ethics Framework
fds-data-ethics-framework.pdf

Authors

Balakrishnan Thiagarajan, Texas Health and Human Services Commission
Julie Leung, Texas Department of Housing and Community Affairs
Carrie Bradford, Texas Department of State Health Services
James Lown, Texas Lottery Commission
Monica Smoot, Texas Department of Information Resources