

APDU Guiding Principles for Public Data

The Association of Public Data Users (APDU) affirms that accessible, high-quality, and responsibly governed public data are essential for informed decision-making in the public interest. These guiding principles articulate the values and expectations that should shape the stewardship, dissemination, and use of public data in the United States. They are intended to guide both data providers and users toward shared goals of transparency, trust, and societal benefit.

Our guiding principle is that public data should be public. Public data are of the people, by the people, and for the people.

That guiding principle is bolstered by six supporting principles, each of which is detailed below.

PUBLIC DATA USERS DEMAND:

1

Public Data Are Public and Accessible

WHAT THIS MEANS:

As noted in the guiding principle: public data should exist in the public domain and for the public good. Data collected by public entities should—in conformance with data governance standards that include appropriate privacy protections and access modalities—be made available to the public.

HOW WE EXPECT TO SEE IT PLAY OUT:

- Data providers shall build tiered access pathways to help the public discover what data exist and where they exist. At a minimum, data should be catalogued in open data portals like data.gov or researchdatagov.org.

- Data providers should offer multiple ways to access data, such as through APIs, downloadable data files, table-generators, prepared tables, reports, and (where appropriate) restricted access arrangements to meet the needs of users of all abilities. All data and metadata should be provided in machine-readable, open formats.
- Data, metadata, and related documentation and technical reports should align with open data / open science standards, such as the Sunlight Foundation's [Ten Principles For Opening Up Government Information](#) and [DCAT-US](#) (Project Open Data Metadata Schema).

2

Privacy and Confidentiality Are Protected

WHAT THIS MEANS:

Public data must be transparently governed, respecting the privacy and confidentiality of respondents (such as individuals, households, businesses, and organizations) as data are collected, stored, accessed, analyzed, and released. For the purposes of this document, privacy and confidentiality are defined as follows:

- **PRIVACY:** Assurance that the confidentiality of, and access to, certain information about an entity is protected.¹
- **CONFIDENTIALITY:** Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.²

Strong privacy and confidentiality protection requires continuous monitoring and refinement to keep pace with growing threats and evolving ethical standards for data use.

HOW WE EXPECT TO SEE IT PLAY OUT:

- Strong data governance shall be practiced throughout the data lifecycle, including the following:
 - Respect the data provider—data are used only for the benefit of the community, without harming the individual. Data collection decisions should include a robust discussion and understanding of potential harms.
 - People must be able to inspect and correct their data.
 - Statistical data may not be used for non-statistical purposes.
 - Data stewards shall identify what’s collected, why, and what consent mechanisms exist.

- Data stewards shall apply modern disclosure avoidance techniques (such as differential privacy, noise injection, or data suppression) to protect confidential information.
- Access tiers are clearly defined to balance privacy protection with data utility. For example, some data may be made available in public-use files, while other data may be restricted-use and subject to additional restrictions.
- Privacy policies are transparent and updated to address new data types, linkage methods, and analytical technologies.
- Data users follow ethical guidelines and comply with access agreements when using sensitive datasets.

In general, if the privacy-utility-accessibility tradeoff is unknown or unclear, privacy should be the priority.

3

Every Step of the Data Lifecycle Supports Ethical Data

WHAT THIS MEANS:

Ethical data shall demonstrate respect for all people—including consideration of those who might not be represented among decisionmakers. The [Federal Data Strategy - Data Ethics Framework](#) states that data activities “have the overarching goal of benefiting the public good.” Specifically, the framework defines data ethics as “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.”

In short: data are used for good, not evil.

1 National Institute of Standards and Technology. <https://csrc.nist.gov/glossary/term/privacy>

2 National Institute of Standards and Technology. <https://csrc.nist.gov/glossary/term/confidentiality>

HOW WE EXPECT TO SEE IT PLAY OUT:

Under ethical public data product development:

- Data stewards shall collect, tabulate, and disseminate data in accordance with stated goals. Data collection goals should be stated publicly.
- Data collectors shall uphold applicable statutes, regulations, professional practices, and ethical standards.
- Data stewards should demonstrate transparent intentionality to represent and respect communities, and data stewards are intentionally transparent about decisionmaking process and collection methods (see Section 4: Transparent, below).
- Data collections should respect data sovereignty—data about you is part of you, and therefore individuals should be able to own their data (up to, and including, the right to be represented, the right to access your data, and the right to be forgotten).
- Data stewards should demonstrate respect for the public, individuals, and communities at all stages of data collection, use, and dissemination, particularly when data may affect visibility, resources, or outcomes.
- Data stewards shall obtain informed consent when appropriate, with clear communication about how data will be collected, used, shared, and retained. Professional conduct is grounded in honesty, integrity, and humility, including openness about uncertainty, limitations, and potential harms, consistent with established ethical guidance such as the American Association of Public Opinion Research's (AAPOR) [Transparency Initiative](#).
- Data stewards should maintain accountability by holding themselves and others responsible for ethical data practices and for addressing harms or errors when they occur.
- Data stewards should engage with developments in data management, data science, and ethical standards that support responsible and current practice.
- Data stewards shall protect privacy and confidentiality through appropriate safeguards, minimizing risk of disclosure while preserving analytical value (see Section 2: Privacy and Confidentiality, above).

- Data stewards shall communicate known limitations, uncertainties, and caveats related to the data clearly, to support responsible use and interpretation (see supporting principle 4, below).

4

Public Data Are Transparent

WHAT THIS MEANS:

Transparency ensures that data systems, methodologies, and dissemination processes are open and understandable—with materials that are accessible to a wide range of constituents (including data users, affected populations, technical experts, and policymakers). This includes, but is not limited to, clear documentation of how data are collected, processed, and shared, as well as open communication about limitations, updates, and governance. Proposed changes to data collections are well-researched, and are announced publicly and with adequate opportunity for public review and comment. Additionally, dataset suspensions, schedule changes, and terminations are treated with a similar level of public transparency. Decisions are made with consideration of both research findings and public comment.

HOW WE EXPECT TO SEE IT PLAY OUT:

- Agencies publish metadata that conform to the [FAIR \(Findable, Accessible, Interoperable, Reusable\)](#) principles, including information about data provenance, collection instruments, and processing methods.
- Methodological notes, revisions, and error corrections are made public in a timely manner.
- Stakeholders have clear avenues for asking questions, providing feedback, and understanding changes to data products.
- Changes to data collections—whether new, revisions, or terminations—are presented for public review and comment, and public feedback is incorporated into decisions on any changes.

- Data should be published on a consistent and predictable schedule. Data release schedules are published and adhered to, which allows the public to track system performance.
- Governance structures are published and adhered to, fostering accountability at all stages of the data lifecycle (i.e., from collection to termination).

5

Public Data Are Fit-for-Purpose

WHAT THIS MEANS:

Data are timely, accurate, reliable, and fit-for-purpose. Although priorities for quality may vary based on use (such as timeliness being more important for data used during a natural disaster, completeness may be more important for executing statewide programs). Regardless of which dimension(s) are prioritized, data quality depends on rigorous collection methods, documentation, and continuous evaluation against established standards. Data providers should strive to identify fitness-for-purpose and communicate how well their data perform on each of the characteristics listed below.

Even when data are fit-for-purpose, there should be a continuous effort towards improvement.

HOW WE EXPECT TO SEE IT PLAY OUT:

Fitness for use reflects whether data are suitable for their intended purpose and is determined by a combination of characteristics that together determine their trustworthiness and fit-for-purpose. These dimensions span technical standards, representational adequacy, and stewardship practices rather than any single attribute:

- Real-world phenomena are reflected with minimal error or bias, supporting accurate measurement and interpretation.

- Data include all information necessary for their intended use—with missing data documented and addressed through efforts to reduce non-response or through transparent, methodologically sound imputation when required.
- Data are harmonized across time, sources, and systems to enable meaningful comparisons.
- Documented definitions, rules, formats, and accepted methodological standards are consistently followed, ensuring validity. Release timelines align with analytical and decision-making needs, supporting timely use of the data.
- Redundant or duplicate records are minimized, and linkage processes are designed to preserve uniqueness and overall data integrity.
- Safeguards protect against unauthorized alteration, and any alterations are made based on methodologically rigorous and stated rules, and in alignment with ethical principles (described in Section 3, above), ensuring trustworthiness over time.
- Appropriate levels of disaggregation are possible, enabling analyses by demographic, geographic, economic, industry, or other domains while protecting confidentiality and supporting community needs.
- All data are complete and no critical information is missing. Survey responses may be missing due to item non-response, but no data are excised after the fact. Efforts are made to reduce survey and item non-response or to fill in missing information using alternative sources or well-documented imputation methods when a complete count is required. Data that are imputed should be flagged as such.
- The communities or conditions of interest are reflected accurately, without systematic over- or under-representation of any group(s). Equitable and inclusive data practices are supported through meaningful opportunities for public input into how data are collected, governed, and used, including mechanisms such as advisory committees. Public data stewards should continuously monitor these dimensions, publish quality assessments, and involve users in identifying quality concerns and priorities.

Public Data Systems Are Sustainable

WHAT THIS MEANS:

Public data systems should produce data that are high-quality (i.e., accurate, reliable, and fit-for-purpose). Data should be maintained, accessed, and useful over the long term without degradation or loss. Sustainable data depend on rigorous collection methods, documentation, and continuous oversight/evaluation against established standards. Functioning at a high level requires resources—including adequate budgets, expert staff, access to cutting edge tools and technologies, and oversight from external experts.

HOW WE EXPECT TO SEE IT PLAY OUT:

■ Resources:

- Adequate and stable funding: Budgets are sufficient to support ongoing data collection, maintenance, data review, and user support. Funding—whether federal, state, or local—should be through regular budget allocations and not dependent on short-term grants.
- Expert staff: Organizations should employ qualified and experienced personnel in the applicable field. Staffing levels should be able to handle data collection, processing, documentation, and user assistance, as well as all data governance required to meet privacy, transparency, and quality needs.
- Modern tools and technology: Organizations should regularly invest in software, hardware, and analytical tools to enable data collection, processing, review, and dissemination.

■ Infrastructure

- IT and Data Storage: Data should be stored in a way that keeps the data safe from unauthorized access and in a way that is efficient and sustainable.
- Preservation: Systems and policies should include long-term preservation of data and allow historical data to be accessible.

- **User access:** Data are stored in reliable platforms that allow users to access data without interruptions or delay.

- **Monitoring:** Regular assessments should track programs over time to identify sustainability risk (such as ageing infrastructure, staff departures, funding gaps), and trigger timely interventions.

- **External Oversight:** Boards or committees composed of external experts should have the access and authority to evaluate whether programs have adequate resources and organizational support to maintain operations in the present and into the future.

One-time data collections may, on occasion, be appropriate (such as during a disaster or health emergency). However, even in one-time collections, data stewards should use the most rigorous data collection methods and practices available at the time of collection.



Related Resources

When compiling this framework, the APDU Board reviewed and synthesized existing literature including the following:

- [Confidential Information Protection and Statistical Efficiency Act \(CIPSEA\)](#)
- [Ethical Guidelines for Statistical Practice](#)
- [Federal Committee on Statistical Methodology \(FCSM\) data quality principles](#)
- [Federal Data Strategy](#)
- [Federal Data Strategy Ethics Framework](#)
- [Findability, Accessibility, Interoperability, and Reuse \(FAIR\) Principles](#)
- [SMARTT Framework](#)
- [Open Data Principles](#)
- [Principles and Practices for a Federal Statistical Agency, Eighth Edition \(2025\)](#)
- [Seven Key Metrics of Data Quality](#)
- [Transparency Initiative](#) (American Association of Public Opinion Research)
- [Trust Regulation](#)
- [United States Code](#), including sections 5, 13, 20, 26, and 44
- [When the Numbers Play Politics, We All Pay](#)