

October 21, 2024

# Standardizing Federal Data: Categorizing Approaches

Congress has at times enacted laws that specifically require federal agencies to use *data standards*, which establish rules to enhance the usability of data. This In Focus provides an overview of three topics related to data standards in the federal context: (1) their role in federal data management, (2) the terminology surrounding data standards and some of the resulting implications for Congress, and (3) categories for the data standards Congress has required in statute. Further discussion of federal data standards is available in CRS Report R48053, *Federal Data Management: Issues and Challenges in the Use of Data Standards*.

## Federal Data Governance and Data Management

The Government Accountability Office (GAO) has identified data standards as a key practice for governing and managing data. GAO has described *data governance* as a framework for ensuring that an agency's data are transparent, accessible, and of sufficient quality to support its mission; improve the efficiency and effectiveness of its operations; and provide useful information to the public. Data governance includes the authorities, roles, responsibilities, organizational structures, policies, processes, standards, and resources for the definition, stewardship, production, security, and use of data. As such, data governance is concerned with how to manage data and is a precursor to *data management*, which is concerned with implementation of those decisions.

**Information Resources Management.** Data governance generally operates within a broader framework for information resources management (IRM), which is directed by the Paperwork Reduction Act (PRA; codified at 44 U.S.C. §§3501-3521). IRM is the process of managing information and related resources (e.g., information technology) to accomplish agency missions and to improve agency performance (44 U.S.C. §3502(7)). The Office of Management and Budget (OMB) is tasked with developing, coordinating, and overseeing IRM policies among executive branch agencies (44 U.S.C. §3504(a)(1)(A)). When a law requires data standards for some federal purpose, OMB often plays a role in forming such standards.

Agencies are also expected to manage information to meet certain objectives, such as improving the use of information within and outside of the agency (44 U.S.C. §3506(b)). Under the PRA, program officials are responsible for defining the program's information needs and developing the strategies, systems, and capabilities to meet those needs in consultation with the agency's chief information officer and chief financial officer (44 U.S.C. §3506(a)(4)). Congress has also established more specific data governance and management responsibilities for individual agencies or programs that operate outside of the PRA.

**Chief Data Officers.** Within agencies, chief data officers (CDOs) are responsible for data management (44 U.S.C. §3520). A 2017 House committee report suggested that CDOs would improve data interoperability in the executive branch and the transparency of federal data by centralizing data management. Among other activities, an agency CDO may work with stakeholders in the agency to demonstrate how data analytics can address challenges and priorities, including the role of data standards in these types of projects; initiate the development of data standards to educate stakeholders about the value of data management, data architecture, and data-driven decisionmaking; and facilitate a common language for data among data stewards—those that have day-to-day data management and data analysis roles.

## Defining Data Standards for Federal Purposes

OMB, the General Services Administration, and National Archives and Records Administration (NARA) jointly maintain online resources for federal data management. They have characterized the universe of data standards as “large, varied, and complex” and indicated there is no single, simple definition to adequately convey their purposes for all the ways agencies may use them to manage and use data. For example, data standards can dictate data definitions, data types, data formats, and data structures and relationships. Data standards include metadata standards such as those required by NARA for permanent electronic records transferred to it.

The adequacy of the term *data standards* is not always straightforward. For example, the Bureau of the Fiscal Service (BFS) maintains the standards for federal financial spending data pursuant to requirements in the Digital Accountability and Transparency Act of 2014 (DATA Act; P.L. 113-101). While the law uses the term *data standards*, BFS initially named its implementation of the requirements the “DATA Act Information Model Schema” (DAIMS). In 2023, BFS said it “rebranded” DAIMS as the “Governmentwide Spending Data Model” due to new legislation and policies that went beyond the DATA Act. Thus, three terms were used interchangeably: *data standards*, *information model schema*, and *data model*.

Similarly, the Financial Data Transparency Act of 2022 (P.L. 117-263; 136 Stat. 3421) required several financial regulatory agencies to promulgate joint standards for certain data reported by financial entities to these agencies. In a *Federal Register* notice of the proposed data standards, the agencies noted that the area of data standards is “rich with well-established practices and also rapidly evolving” and discussed interpreting the meaning of certain words used in the law. Specifically, the act indicates that the data standards should, to the extent practicable, “enable high

quality data through schemas, with accompanying metadata documented in machine-readable taxonomy or ontology models, which clearly define the semantic meaning of the data” (12 U.S.C. §5334(c)(1)(B)(ii)). The agencies said that these words are used in various and sometimes conflicting ways within the field of data science. For example, they noted that *taxonomy* sometimes refers only to a description of the semantic meaning of a data asset, that *ontology model* may also refer to this description of semantic meaning, and that *taxonomy* can at other times refer to a description that goes beyond semantic meaning alone to include data syntax and hierarchical structure. Similarly, they noted that *schema* can sometimes refer to only a description of data syntax, while it can at other times include a description of syntax, semantic meaning, and structure. As such, a lack of consensus in practice may pose a challenge for the implementation of data standards required by a law. Given some of the challenges with terminology, lawmakers are faced with making decisions about how to specify the data standards that might be necessary to achieve their policy goals—or whether to make such specifications at all.

### Categorizing Federal Data Standards

GAO has identified general categories of data standards used by federal agencies: (1) those that are specific to a program, (2) those that are specific to an agency, and (3) those that are government-wide. Government-wide data standards attempt to consistently specify requirements for data across multiple agencies. In contrast, each agency or program may manage the same kinds of data (e.g., place of performance) but use different data standards (e.g., formats for state name) that are unique to the agency or to the program (e.g., California, Calif., CA, Ca., or 06). These dissimilarities make it difficult to process data from different programs and agencies.

**Program Data Standards.** At times, Congress has directed a federal program or a specific program activity to use data standards. Sometimes, a law requires agencies to establish data standards for federal programs using the rulemaking process. In its final rule on the unemployment data standards required by Title 42, Section 1111, of the *U.S. Code*, the Department of Labor (DOL) noted a relationship between data standards and states’ information systems and that implementation of the standards would require substantial changes to many state systems. DOL also noted that the data standards in this case had implications for *collections of information* (44 U.S.C. §3502(3)), possibly imposing new *burdens* (44 U.S.C. §3502(2)) and thus potentially adding to the costs of collecting the underlying data. DOL also claimed that agencies need flexibility to determine what data standards will produce the best results and the ability to balance issues such as state capacity and costs, which may be constrained when required to implement data standards as regulations.

In 2018, the Administration for Children and Families (ACF) sought public comment on the statutory requirements for certain programs under Title VI of the Social Security Act to designate data standards, as a handful of laws over several Congresses had established such

requirements on a program-by-program basis. ACF stated that the benefits of state agencies sharing data for state-administered federally funded programs are well understood and that data standards make data sharing easier, which may increase program effectiveness and be cost effective in the long run. However, implementation of data standards could also introduce time and cost considerations. ACF said it would seek to balance the benefits of standardization with the burden of implementation.

**Agency Data Standards.** Congress has at least once directed an agency to use data standards agency-wide: The Department of Homeland Security (DHS) Data Framework Act of 2018 (P.L. 115-331) required the development of a framework for integrating existing DHS datasets and systems, codifying preexisting efforts to promote the exchange of information within the agency. DHS was directed to promulgate data standards and to instruct its components to make data available in a machine-readable format (6 U.S.C. §126(b)(3)).

Agencies may use data standards for certain agencywide operations or functions. For example, the Department of State identified a need for an agency-wide approach to data standards because “current approaches are bespoke to specific data products and are not applied uniformly nor broadly understood.” In 2019, OMB issued a memorandum describing a federal data strategy that was intended to enable agencies and the government more broadly to use and manage federal data. The strategy called for agencies to use data standards, echoing previous OMB guidance.

**Government-Wide Data Standards.** Congress has enacted laws that require several agencies to use the same data standards for certain activities. In most cases, implementation of these government-wide data standards is an ongoing process. One example is the DATA Act discussed above. Another is the Geospatial Data Act of 2018 (codified at 43 U.S.C. §§2801-2811) that requires a federal committee to establish standards for geospatial data that are to be adopted by any executive department (as defined by 5 U.S.C. §101 but excluding the Department of Defense) that collects, produces, acquires, maintains, distributes, uses, or preserves geospatial data directly or through a relationship with another organization. These standards are intended to support national geospatial data infrastructure. A third example is the Grant Reporting Efficiency and Agreements Transparency Act (P.L. 116-103), which requires standards for managing data related to federal grants, potentially reducing duplicative reporting and assisting in aggregating and comparing grant data from different grantmaking agencies. In some cases, Congress has established a role for agency inspectors general and GAO to report on the implementation of government-wide data standards.

---

Natalie R. Ortiz, Analyst in Government Organization and Management

IFI2787

## Disclaimer

This document was prepared by the Congressional Research Service (CRS). CRS serves as nonpartisan shared staff to congressional committees and Members of Congress. It operates solely at the behest of and under the direction of Congress. Information in a CRS Report should not be relied upon for purposes other than public understanding of information that has been provided by CRS to Members of Congress in connection with CRS's institutional role. CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.