

To Whom It May Concern:

Thank you for this opportunity to provide written comments on the Federal Data Strategy. The Data Coalition was founded in 2012 to advocate on behalf of the private sector and public interest for the transformation of government information into standardized, open, and machine-readable data. Based in Washington D.C. our trade-association represents over forty-five technology and data analytic companies as well as public sector focused consulting and accounting firms. We empower these companies to help make our government more transparent, efficient, and effective.

Below we provide comments on both the Federal Data Strategy draft principles as well as provide six use cases. Please do not hesitate to reach out to our policy director Christian Hoehner with any questions or required clarification. Christian can be reached directly at (434) 466-6053 or christian.hoehner@datacoalition.org.

FEDERAL DATA STRATEGY - PRINCIPLES

OVERALL COMMENT:

Most of the Federal Government's data is produced as documents: regulations, guidelines, policies, reports, forms and filings, letters and much more. This information, authored in word processing programs and often stored in paper or PDF, is practically inaccessible to the public and even to other parts of the government.

To a large extent, the primary expense of big data projects comes from extracting information from different sources, transforming those data sets into the same format, and then loading them into new systems to be analyzed. For example, when one agency shares documents with another, or attempts to work collaboratively, much time is wasted on “low value” work – reformatting, rekeying, reprinting and more – rather than “high value” analysis and policy development. And when the information is produced (e.g. in response to FOIA requests) the response is often costly, wasteful and still incomplete. If federal data sets were standardized and made available in machine-readable formats, such expensive one-off projects would become less frequent.

The Federal Data Strategy, at its core, should recognize the benefits of open and machine-readable formats and data standardization projects.

Machine-readable formats impose uniform and semantic structure on data. This structure mitigates uncertainties and speeds development of any application or tool to extract meaning from this data. Structure also enables straight-through processing, allowing information to flow automatically from producers to original and secondary users. Due to its consistency, machine-readable data can be read and utilized by a wide variety of companies, organizations, and citizens--including hobbyist and academic researchers who can leverage the myriad of software tools now available. Bulk data is also simpler to access and use than an Application Program Interface (API), as it does not require the user to make thousands or millions of individual API calls to assemble a complete data asset for outside user or analysis.

Data standardization identifies common data elements shared across agencies and mission areas, adopt and enforce such standards, and ultimately make core operational and programmatic data available in open formats for internal efficiency and outside transparency.

COMMENT ON STEWARDSHIP PRINCIPLES:

As the Federal Data Strategy and corresponding data governance projects are refined, a key aspect of any newly refined or established processes should be the establishment of clear requirements and lines of communication for data owners to consult with end users. Such consultation should take place during the standards development phase and at regular intervals as standards and governance, access, and dissemination policies are continually updated. Specifically, feedback loops should be incorporated to enable end users to flag data quality issues and recommend structural improvements on an ongoing basis.

In regards to responsible and ethical use of public sector data, that Administration should consider how metadata associated with machine-readable data formats can more readily incorporate data protection and security considerations, as well as other parameters around end use.

COMMENT ON QUALITY PRINCIPLES:

A common barrier to the business challenges facing agency programs and management workflow is the issue of poor data; both ‘operational’ (i.e., mission support; mission agnostic data that represent the resources, decisions, transactions, outputs, and outcomes of work) and ‘programmatic’ (e.g., mission or material data; program specific data often representing persons, places, and things) must be of high quality in order to be useful to end users. This means it must be accurate, consistent, and controlled.

A lot of work and external stakeholder engagement has gone into the refinement of existing open data federal initiatives that are fully documented in this 2017 CIO Council “State of Federal Information Technology” report (see: section C “Open Data and Open Government” - <https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/1151/2017/05/CIO-Council-State-of-Federal-IT-Report-January-2017-1.pdf>).

Specifically, the government’s Project Open Data, an outgrowth of the 2013 Executive Order “Making Open and Machine Readable the New Default for Government Information” has documented a number of Open Data Principles (see: CIO.gov Open Data Principles - <https://project-open-data.cio.gov/principles/>) which should be expanded and enforced as a part of the Federal Data Strategy.

These principles included the following seven fundamental principles: a presumption in favor of openness (**Public**), availability in machine-readable and open formats (**Accessible**), full descriptions with robust metadata and appropriate documentation (**Described**), availability under the appropriate open licenses that do not restrict reuse while documenting any necessary intellectual property constraints (**Reusable**), publishing of primary or source level data with the finest level of granularity (**Complete**), prompt availability so as to preserve the value and enable downstream use (**Timely**), and appropriate management to enable reliable end use (**Managed Post-Release**).

To be clear, these open data principles benefit internal use (inter- and intra-agency data sharing and governance) just as much as external purposes (transparency, oversight, and private sector re-use).

Beyond publishing, federal data governance often requires some form of reporting and collections. To assure that source level collected and curated information is of high quality from the start, the Administration should work to centralize reporting and information collection activities to the maximum extent possible. This enables more streamlined and efficient coordination with the authoritative sources of the government's data sources. Furthermore, it is at the point of data ingestion into systems that established standards and thresholds for quality and data validity can be more reliably and comprehensively enforced. Where applicable, data that does not meet established thresholds should be rejected and fixed.

If the Administration embraces these core principles and works to build out accurate, consistent, and controlled data in priority areas, we can start to fix the structural conditions and help federal leaders champion tangible reforms.

COMMENT ON CONTINUOUS IMPROVEMENT PRINCIPLES:

Whatever the delivery mechanism of public sector data (e.g., APIs or bulk data dumps) a crucial consideration in the purveyance of this data should be the aim to make data open and usable from the very outset. In other words, it is much easier to design a system in which data is open and usable from the very beginning, rather than retrofitting or altering an existing closed data system to make it public. Designing a system that is equally open from day one prevents the formation of certain privileged user-provider relationships, and enables developers the maximum flexibility to create their own unique products or build custom platforms around the underlying data. The best approach is to make data available for bulk data download in open formats to enable equal access for all users.

FEDERAL DATA STRATEGY - USE CASES

The Administration has set out on an ambitious reform agenda to modernize and enhance how government agencies operate. Using data-driven processes that integrate government financial, performance, and regulatory data, the Administration can help lay the foundation for a more effective and accountable government.

One of the key drivers in establishing data-driven decision-making is standards setting. The DATA Act, for instance, has already resulted in tremendous work to tie together federal financial data by applying consistent data standards. By leveraging such efforts, the Administration can establish benchmarks and parameters to help meet management reform goals.

The Data Coalition's industry members hope to serve as a continual resource to the Administration to help set benchmarks for reporting and governance processes and pursue opportunities to modernize these requirements through a more data-centric, customer focused approach. As the Administration establishes standards and process for government-wide reforms, we are ready to partner with agencies as they utilize rich datasets to set benchmarks and drive decision making.

As the Administration makes plans to improve management practices we hope leadership will consider leveraging existing standards efforts, such as the US GAAP Taxonomy, the DATA Act Information Model Schema (DAIMS), and the National Information Exchange Model (NIEM), which have each begun the important work of using government data to drive oversight, reform, and accountability.

USE CASE #1: FIX PUBLIC COMPANY FILINGS (ACCESS, USE, AND AUGMENTATION)

Since 2009, the Securities and Exchange Commission (SEC) has been requiring public companies to report their financial statements in a standardized, non-proprietary machine-readable format called eXtensible Business Reporting Language (XBRL). Maintaining the standard is the independent Financial Accounting Standards Board (FASB), which has developed and continually worked to update a taxonomy, or list of data tags, reflecting the accounting concepts of U.S. Generally Accepted Accounting Principles (US GAAP). (For details, see:

<https://www.fasb.org/cs/ContentServer?c=Page&cid=1176169699514&d=&pagename=FASB%2FPage%2FSectionPage>) (see also: <https://www.workiva.com/blog/your-guide-2018-us-gaap-taxonomy-update>).

However, the implementation of this taxonomy has not fully delivered the promised benefits of comparability, transparency, and market efficiency. First, the US GAAP Taxonomy is too complicated to allow easy comparability and has thousands more electronic tags than are needed to accurately reflect corporate financial statements. Second, the SEC permits companies to create custom-built extension tags in far too many circumstances. The major information data aggregators, such as Bloomberg, Thomson Reuters, and Morningstar, still do not ingest the SEC's XBRL corporate financial data without significant edits. Smaller, creative intermediaries are making progress, but the reality is that the U.S. capital markets do not efficiently absorb XBRL data. As a result, U.S. public companies are not experiencing lower costs of capital from filing XBRL reports. Until the SEC fixes the US GAAP Taxonomy and addresses the overuse of extensions, its XBRL open-data reporting regime for financial statements will not deliver the promised benefits.

The Data Foundation's "Open Data for Financial Reporting: Costs, Benefits, and Future" (see: <http://www.datafoundation.org/xbrl-report-2017>) provides the history of the US GAAP Taxonomy, outlines the benefits to the SEC's disclosure modernization reform efforts, and provides actionable recommendations for improvement.

A recent report by member company Donnelley Financial Solutions, "A New Approach to Data Quality: How the SEC Can Prepare for a RegTech, SupTech and AI Future" (see: <https://www.datacoalition.org/wp-content/uploads/2018/07/DFS-Summary-June27Final.pdf>), outlines additional steps to address taxonomy improvements, while also outlining the future benefits to the regulators, including the applications of emerging tech for supervisory technology workflow (i.e., SupTech; the technology regulators use internally to conduct their enforcement, supervisory, and research activities).

ACTION:

The Administration should work with the SEC and the FASB to assure that the US GAAP Taxonomy:

First, enforces FASB codifications as the true reference for all elements in the taxonomy, thereby eliminating unnecessary tags, reducing the complexity of the taxonomy. For example, incorrect references should be replaced with missing references. The filer is in the better position to know the relevant codifications that shape their filing, and can help make self-tagged data more accurate than service-provider tagged data. Additionally, accessing elements through the codifications will make it easier to find the standard elements to use, thereby eliminating unnecessary extension elements. The SEC should make sure filers follow the recommendations of the FASB's Efficiency and Effectiveness preparers report.

Second, minimizes the creation of extension data elements, thereby improving comparability and data quality.

INTERNAL BENEFITS:

By reforming the US GAAP Taxonomy, the SEC will gain the benefits of ingesting more accurate and comparable corporate information to conduct anti-fraud investigations and identify systemic market risk.

EXTERNAL BENEFITS:

Investors and data aggregators will benefit from more comparable market information. Filers will benefit from a taxonomy that more closely mirrors the requirements of US GAAP and their internal accounting and audit practices.

USE CASE #2: DOCUMENTS TO DATA IN MANAGEMENT MEMORANDUM (DECISION-MAKING AND ACCOUNTABILITY)

Cross Agency Priority Goal 6, “Shifting From Low-Value to High-Value Work,” seeks to establish “regular processes to assess the burden [of OMB’s management guidance] on agencies and to rescind or modify requirements over time.” That is, the Administration should recognize the central role of documents in government data and encourage the application of modern “data-first” technologies for authoring documents and begin the process of defining a data standard and strategy specifically for guidance documents.

ACTION:

The Administration should publish federal management guidance in integrated, machine-readable data formats instead of documents. The Data Coalition’s work to pursue open data for laws and mandates provides a use case for exactly the same transformation, starting with Congressional laws, bills, and amendments. The United States Legislative Markup (USLM) provides an excellent framework for such a standard, and is already used for bills, laws, and applications within the United States Code and the Code of Federal Regulations.

INTERNAL BENEFITS:

By treating federal policy documents as data, and establishing standards for the drafting and interoperable dissemination/issuance of such documents, management and the agencies can more readily understand how policies integrate with existing policies and therefore work to comply more readily. Furthermore, executive leadership will have the ability to smartly and efficiently draft and issue policies while being able to measure compliance and successful realization of intended effects.

EXTERNAL BENEFITS:

The public and Congress will be able to more readily and fully understand the specific directives currently guiding and constraining agency programs and leadership. This will help guide and inform public policy conversations while further building trust in government.

USE CASE #3: ENTITY IDENTIFICATION WORKING GROUP (ENTERPRISE DATA GOVERNANCE)

The federal government currently uses a wide variety (see: <https://www.datacoalition.org/a-jungle-of-entity-identifiers-and-what-were-going-to-do-about-it/%5C>) of codes to identify companies, nonprofits, and other non-federal entities. Because the same entity is identified differently with each agency it reports to, entity matching is a time-consuming and expensive threshold challenge for nearly every federal data analytics project.

In 2010, to address the expense and opacity of incompatible entity identification systems, the U.S. Treasury Department and foreign financial regulators started a global nonprofit organization to create and maintain the Legal Entity Identifier (LEI) (see: <https://www.financialresearch.gov/from-the-director/2017/02/02/breaking-through-barriers-impeding-financial-data-standards/>), a versatile and non-proprietary code that can be adopted within any regulatory regime. Several federal agencies are now considering adopting the LEI alongside existing non-interoperable identification codes.

ACTION:

The Administration should establish a policy statement endorsing the LEI as the default identification code for legal entities. The White House OMB Office of the CIO should direct all agencies to report on the disparate identification codes they currently use and the feasibility of a full transition to the LEI. The federal government should adopt an open and common universal identification code to be integrated into agency systems. These actions should be directed by a universal entity identification working group

INTERNAL BENEFITS:

The establishment of a universal entity identification working group will speed up current federal agency initiatives seeking to standardize entity identification and encourage the formation of new ones. Ultimately, the adoption of a universal, non-proprietary entity identifier will enable agencies to aggregate, compare, and match data sets critical to their regulatory or programmatic missions.

EXTERNAL BENEFITS:

This will facilitate, where appropriate, the publication of bulk data sets which would otherwise be inhibited by the proprietary nature of currently-used entity identification systems.

USE CASE #4: MISSION SUPPORT OR OPERATIONAL DATA STANDARDS COORDINATION (DECISION-MAKING AND ACCOUNTABILITY)

The U.S. Treasury and the Office of Management and Budget have spent over four years working with agencies and outside stakeholders to establish and integrate the DAIMS into agency systems. With over 400 unique data elements the DAIMS represents the most comprehensive and widely adopted data standard for federal operations in the U.S. government. The DAIMS links budget, accounting, procurement, and financial assistance datasets that were previously segmented across agency systems and databases. OMB should rely on the DAIMS's open documentation architecture which allows for ready expansion and linkage to other administrative datasets. Furthermore, in his Senate confirmation

hearings OMB Director Mick Mulvaney suggested that future budget requests could be published in a machine-readable and DATA Act compatible format.

ACTION:

The Administration should utilize the DAIMS for modernizing the annual budget process, agency financial reporting, and agency performance reporting:

Annual Proposed Budget - The Administration should centrally publish the Administration's annual budget request including the corresponding standardized agency Congressional Justifications (CJs) in a bulk downloadable, machine-readable, and DAIMS-compatible format.

Congressionally Approved Budget - The Administration should publish the Congressionally produced federal budget in a manner that integrates with the DAIMS to enable comparisons with current agency finances.

Agency Financial and Performance Reporting - Machine-readable data, in a common schema (e.g., integrated with the DAIMS), should be part of OMB's Annual Performance and Annual Financial Report which agencies are required to submit. Data-driven reporting provides an opportunity to improve efficiency while creating a more sustainable, meaningful, and repeatable process. Creating the critical connection between finance and performance information in a compelling, engaging, and informative way provides the opportunity for agencies, constituents, OMB, Congress, and others to derive greater value and insight from these reports. OMB would need to replace the document-based reporting methods with a consistent data schema that enables data-centric agency budget and financial reporting. OMB and agencies should also tie their goals to broader program areas.

INTERNAL BENEFITS:

Annual Proposed Budget - Will enable the Administration to easily use data to compare and justify the proposed budget in support of strategic goals.

Congressionally Approved Budget - Will help agencies monitor current spending against the enacted budget, thus enabling visibility into funding distribution and timely planning.

Agency Financial and Performance Reporting - Will help agencies adopt efficient data-driven strategies, diminish duplication of efforts, increase intra- & inter-agency information sharing, and reduce the resources required for information collection.

EXTERNAL BENEFITS:

Annual Proposed Budget - Will enable Congress and the public to more efficiently ingest linked data comprising the annual federal budget, parse proposed changes to programs, quickly compare budget plans against current fiscal year spending activity, and perform external budgetary analysis utilizing accurate data linkages. Will enrich the national conversation around the government's budget by assuring more accurate interpretations and enable more informed discussions about the performance of particular government programs.

Congressionally Approved Budget - Will help users more efficiently interpret the enacted budget in the context of specific agency program funding and award opportunities that matter to their own organizations.

Agency Financial and Performance Reporting - Benefits include transparency and accountability from publishing public sector data which fosters innovation across the academic, research, and technical developer community. By putting digitally integrated performance and budget information in the hands of the public, agencies will have the opportunity to articulate what's working and what's not, and develop necessary buy-in for improvement plans.

USE CASE #5: MISSION OR PROGRAMMATIC DATA STANDARDS COORDINATION (ENTERPRISE DATA GOVERNANCE; DECISION-MAKING AND ACCOUNTABILITY; ACCESS, USE, AND AUGMENTATION)

The federal government has very few data formats for sharing mission or programmatic related data across multiple agencies. One prominent exception is the Department of Homeland Security lead National Information Exchange Model (NIEM).

To build a common approach to multi-agency data sharing, the Departments of Homeland Security, Justice, and Health and Human Services created the National Information Exchange Model (NIEM) (see: <https://www.niem.gov/>), which publishes a data dictionary of common fields (the "NIEM Core") and helps agencies create formats using those fields. (In the summer of 2017, version 4.0 of the NIEM Core will incorporate the Legal Entity Identifier.) NIEM is now used extensively to build formats for multi-agency counter-terrorism and law enforcement projects, but has tremendous potential for use beyond these mission areas.

ACTION:

The Administration should consider endorsing NIEM as the government-wide default for data standardization and publication projects. The Administration should explore moving the NIEM management team (three-five full-time equivalents) into a government-wide leadership office such as OMB. In coordination with the data standardization work of GSA's US Data Federation (an outgrowth of the Data.gov effort) and Project Open Data, NIEM stands poised to foster a base of standardized material data to inform the natural harmonization of common mission data within agency environments.

INTERNAL BENEFITS:

Adopting a NIEM-first policy for federal mission or programmatic related data sharing projects will ensure that agencies at least consider reusing the common data fields of the NIEM Core, rather than building their own, when they initiate data exchanges.

EXTERNAL BENEFITS:

The broader use of the NIEM Core will enable serendipitous reuse of federal open data at a far greater scale.

USE CASE #6: ESTABLISH A STANDARD BUSINESS REPORTING TASK FORCE TO STANDARDIZE REGULATORY COMPLIANCE (ENTERPRISE DATA GOVERNANCE; ACCESS, USE, AND AUGMENTATION)

U.S. regulatory agencies employ document-based forms to collect information from the private sector. However, the concept of Standard Business Reporting (SBR) demonstrates that regulatory agencies can reduce the compliance burden on the private sector by replacing forms with standardized data, governed by common data standards across multiple regimes. Australia and the Netherlands have implemented SBR, with their regulatory agencies agreeing on a common data structure of common fields and formats for the information they collect. As a result, Australian and Dutch business software can automatically compile and submit regulatory information to multiple regulators at once, saving compliance costs.

In April 2018, the Commerce Department's Trade Finance Advisory Council (TFAC) recommended to the Secretary of Commerce that the federal government should pilot and implement a standardized reporting regime similar to the SBR concept (see: https://www.datacoalition.org/wp-content/uploads/2018/07/TFAC_Data-Standardization_Final.pdf). The TFAC's recommendations provide an opportunity to move forward.

ACTION:

The Administration should convene a task force representing all major regulatory agencies to create a road map for standardizing the data fields and formats that they use to collect information from the private sector, modeled on best practices learned from foreign success cases like the Australian Standard Business Reporting project, to reduce compliance costs. The task force should evaluate ways to deploy new data reporting technologies, particularly distributed ledger/blockchain, throughout U.S. regulatory reporting. The Administration should consider a role for the Department of Commerce in the housing and leadership of the task force as recommended by the TFAC.

INTERNAL BENEFITS:

Although the full implementation of a U.S. SBR program would require a multi-year effort, the creation of an exploratory task force would help scope out the required policy barriers and necessary investments, and encourage regulatory agencies to voluntarily begin regulatory data standard setting work where applicable.

EXTERNAL BENEFITS:

The adoption of SBR in the U.S. has the potential for enormous regulatory compliance cost savings to the private sector. For instance, Deloitte has estimated that Australia's SBR program will have accrued \$5 billion AUD in compliance cost savings to the private sector by the end of 2018 (see: <https://comms.dfsco.com/LP=1923>).