

Approved for Public Release. Distribution Unlimited

Case Number: PR_18-2319-2

©2018 The MITRE Corporation. All Rights Reserved.

MP180465

July 27, 2018



Response of The MITRE Corporation to the Commerce RFI regarding Leveraging Data as a Strategic Asset (CAP Goal #2)

For additional information about this response, please contact:

Duane Blackburn, S&T Policy Analyst
The MITRE Corporation
7596 Colshire Drive
McLean, VA 22102-7539

dblackburn@mitre.org
(434) 964-5023

The MITRE Corporation is a not-for-profit, non-commercial company chartered to operate in the public interest. Our focus is on solving problems for a safer world. For the past 60 years, we have advanced federal and national needs by operating multiple federally funded research and development centers. The breadth of our work gives us a unique insight across missions and across functional and technology domains. Our charter, and our systems thinking perspective, has made it possible for us to see challenges from a federal, state and local, academic and private sector point of view. An important aspect of our data-specific capabilities has been the ability to obtain sensitive data from the public and private sectors, and to leverage the combination in the advancement of national interests.

Introduction

Government priorities, such as national security, economic growth,¹ and government reform, require a renewed focus on a common asset: federal data. This includes not only better organization and sharing of data within the federal government, but also significant enhancements of current practices to enable the private sector to innovate using federal data. Organized and managed properly, federal data can become a significant national asset.

Federal data efforts received a boost from the Cross-Agency Priority (CAP) Goal of “Leveraging Data as a Strategic Asset” in the 2018 *President’s Management Agenda (PMA)*. The overall objective of the CAP Goal is to develop a four-part Federal Data Strategy, based upon best practices and principles. This paper answers a Department of Commerce-issued request for community input into those practices and principles, and provides examples of “use cases” for the government to study.

The MITRE Corporation is uniquely positioned to respond to the Request for Information (RFI) and assist with the Federal Data Strategy because of our unique vantage point as the operator of multiple federally funded research and development centers (FFRDCs). Per the Federal Acquisition Regulation, FFRDCs can have unique access to both sensitive government data and proprietary private sector data – and both the government and the public sector have regularly trusted MITRE to access and leverage their data. Thus, we have combined and leveraged a variety of data sources in support of research, analysis, and the development of new operational capabilities on important national issues. For example, MITRE has:

- Developed the FAA’s Aviation Safety Information Analysis and Sharing (ASIAS)² platform, which serves as a trusted broker to exchange information amongst the federal government and private-sector aviation community
- Helped develop the Structured Threat Information Expression/Trusted Automated Exchange of Indicator Information (STIX/TAXII) standard to support the automated exchange of cyber-threat data

In addition, MITRE has been involved in advising, or helping to create, numerous public-private partnerships for data sharing, analysis and use in other functional domains.

MITRE’s access to and use of disparate data sources has given us insight into data’s untapped potential, as well as the challenges associated with greater use of government data (alone and in combination with private-sector data). Our experiences show that high-quality data combined with best practices will increase the effectiveness of the federal government, enhance accountability, and promote transparency. Furthermore, we believe that successfully

¹ “Data Is Giving Rise to a New Economy,” *The Economist Magazine*, May 6, 2017: <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy> (accessed July 2018)

² https://www.faa.gov/news/fact_sheets/news_story.cfm?newsid=22454

meeting this CAP Goal will significantly help the government achieve additional CAP Goals within the PMA.

Questions Posed in the RFI

Best Practices Related to the Four Pillars of the Federal Data Strategy

The following responses reflect best practices and ideas for the four pillars of the Federal Data Strategy.

RFI Request 1: *Enterprise Data Governance. Briefly describe which best practices the Federal Government should consider as it sets priorities for managing government data as a strategic asset, including establishing data policies, specifying roles and responsibilities for data privacy, security, and confidentiality protection, and monitoring compliance with standards and policies throughout the information lifecycle.*

Data governance is a well-covered topic with proven best practices. Rather than restating these, MITRE focused on providing concepts and practices that are more unique to the federal domain:

- Target data governance as a governmentwide issue
- Develop a capability to assess the strategic value of federal data
- Leverage public-private partnerships

Each of these inputs is described in the following subsections.

Target data governance as a governmentwide issue

The Federal Data Strategy should acknowledge that data considerations exist in multiple tiers:

- a) Governmentwide
- b) Sector/Domain, which are typically interagency in nature
- c) Individual agencies
- d) Individual datasets

The preponderance of prior data efforts have focused at the individual datasets and agencies tiers, without the benefits of being driven by, and leveraging, strategies and taxonomies in the first two tiers. White House leadership in developing domain-wide data management plans has

proven³ to drive better governance within agencies and datasets, thus leading to significant enhancements of whole-of-government and individual agency missions. The Federal Data Strategy should establish such sector/domain activities as an expectation, as well as create similar efforts at the governmentwide level.

OMB should consider establishing a Chief Data Officer (CDO) role, which is often done in the private sector, to lead the development and implementation of the Federal Data Strategy. The CDO could then establish an ongoing governmentwide forum to create recommended baselines, data management and data quality guidance, best practices, and overarching principles for applications that individual agencies can leverage and benefit from. Doing so can remove duplicative or contrasting, incompatible approaches that have historically been created as individual agencies investigate data governance on their own.

This governmentwide forum should then establish (or absorb existing) subordinate interagency communities of practice (COP) around sectors or domains of priority concern, such as for geospatial, income, or health data. Statutory limitations most often come into play on sensitive topics, and often impact multiple agencies. COPs can help OMB with cross-agency governance, as well as assist agencies in making decisions about value and protection of data in a more holistic manner. The cross-agency forum could oversee each of the COPs, ensuring that their data catalogs and other data assets are shareable, and potentially linkable, with a common interface.

The forum should also establish cross-functional information risk teams, enabling greater visibility into government-wide impacts on data sharing issues of pending agency decisions. Agreement to proceed should be jointly conferred, in a manner similar to how OMB currently supports interagency review of an agency's rulemaking plans.

Develop a capability to assess the strategic value of federal data

The strategic value of federal data sets isn't uniform, and the government would be best served by devoting its resources to activities that yield the most return on investment (ROI) or that meet other priority needs of the federal enterprise. Unfortunately, there are no commonly-existing measures for assessing the strategic value of federal data, a gap that should be overcome.

Federal data should be viewed as an asset by, for example, quantifying the cost of originating and managing data, the risk-cost of specific inadequate data management and non-compliance,

³ For example, prior to 9/11 federal biometric datasets were kept isolated within agencies – occasionally using approaches specifically intended to make interagency sharing difficult. Subsequent OSTP- and NSC-led efforts created a new paradigm with common standards and mutual collaboration to ensure that biometric data on terrorists was collectively updated and promulgated to all government screening systems.

or the value of data to third parties.⁴ Related to this, the community should create a capability to calculate the costs (detrimental impacts) that occur when data is compromised or lost.⁵

Once established, insight of this determined “value” of available data can guide agencies (individually and collective) on how to best align the thrusts of their data governance programs to meet both their most pressing needs and national-level objectives.⁶

Leverage public-private partnerships

The federal government doesn’t have a monopoly on the need to manage and, optimally, leverage the data that it holds. Other governments and the private sector also deal with these issues, and the government should leverage its experiences and innovative ideas to the maximum extent possible. Public-private partnerships (PPP) “combine the skills and resources of both the public and private sectors through sharing of risks and responsibilities. This enables governments to benefit from the expertise of the private sector, and allows them to focus instead on policy, planning and regulation by delegating day-to-day operations.”⁷

Recognizing that data is one of Europe’s key economic assets, the European Commission has already established a Big Data PPP to connect governments, industry, researchers and academia to implement a Strategic Research and Innovation Agenda. “The PPP will help focus investment and work on datasets that are strategic for European industry and where Europe can take a business or technological lead (e.g. manufacturing or agriculture). The PPP works across sectors and across borders, enabling companies from different industrial sectors to collaborate and benefit.”⁸ Given this effort, the United States is already falling behind its economic competitors. Already-proposed PPPs focused on data efforts, such as an initial activity within the GEAR Center⁹, need to be started quickly and prioritized.

Beyond the development of formal PPP(s), the federal government should also encourage its data managers to regularly interact with their external peers in one-on-one conversations as well as in open fora. Newly-established COPs and cross-functional information risk teams should also consider establishing Federal Advisory Committees¹⁰ to support their efforts in a recurring, formal manner.

⁴ Fleckenstein, M. and Fellows, L., *Modern Data Strategy*, Springer, 2018 chapters 3 (“Valuing Data as an Asset”) and 4 (“Physical Asset Management vs. Data Management”).

⁵ DHS created a precedence for this in cybersecurity when it created the Section 9 list (in response to EO 13636) and identified assets that, if disabled, would cause catastrophic damage to the economy.

⁶ Fleckenstein, M. and Fellows, L., p.40, “*Business Strategy as a Driver for Data Strategy*”

⁷ <https://ppp.worldbank.org/public-private-partnership/node/335/> (accessed July 27, 2018)

⁸ http://europa.eu/rapid/press-release_MEMO-14-583_en.htm (accessed July 27, 2018)

⁹ The GEAR Center is a federal proposal within *Delivering Government Solutions in the 21st Century, Reform Plan and Reorganization Recommendations*, 2018, p. 112. Available online: <https://www.performance.gov/GovReform/Reform-and-Reorg-Plan-Final.pdf>

¹⁰ Federal Advisory Committee Act (FACA; 5 U.S.C. Appendix—Federal Advisory Committee Act; 86 stat. 770)

Additional data governance practices

- Leverage existing data management resources, which are often already performing, albeit informally, some data governance.¹¹ For example, employees with deep knowledge of the mission as well as related IT applications often take it upon themselves to ensure that only high-quality data is entered into the application. Such experts are prime candidates for data stewards in a more formal data governance framework.
- Ensure that data governance decisions are jointly made by business/operations and IT staff. Consider also developing a data governance operations (DGO)¹² office to ensure coordination across all levels of the data governance framework.
- Data governance should be structured in a hierarchical framework, with an executive layer, a management layer, a data steward layer, and a data governance operations function to support and coordinate across these layers.¹³
- Consider better aligning federal data governance and federal records management. They often require a common taxonomy, the same approach to managing information, and similar resources.
- Enable the public to be an active stakeholder in properly managing some of the data about themselves, in order to enhance its accuracy. Allow citizens and organizations to own and update information about themselves centrally, similar to existing methods that credit bureaus use.¹⁴ Allow each citizen and organization to determine how widely they wish to share most data. (For example, some citizens may opt to share birth dates broadly.)

RFI Request 2: *Access, Use, and Augmentation.* List a few best practices that the Federal Government should consider as it develops policies and practices to enable interested parties to effectively and efficiently access and use data assets by: (1) Making data available more quickly and in more useful formats; (2) maximizing the amount of non-sensitive data shared with the public; and (3) leveraging new technologies and best practices to increase access to sensitive or restricted data while protecting privacy, security, and confidentiality, and the interests of data providers.

(1) Make data available more quickly and in more useful formats

¹¹ Ladley, J., *Data Governance—How to Effectively Design, Deploy, and Sustain an Effective Data Governance Program*, Morgan Kaufmann, 2012, p. 52.

¹² Fleckenstein, M. and Fellows, L., *Modern Data Strategy*, Springer, 2018, pp. 71-72.

¹³ Ibid, p. 71.

¹⁴ For citizens and organizations that don't have ready access to a portal, the government can leverage kiosks at post offices and/or other federal sites that citizens can visit to make updates.

- Define a common taxonomy for key data assets. Leverage and update the Federal Enterprise Architecture Data Reference Model v3 to define a common ontology for high-value or highly shared datasets.
- Simplify data classification categories for federal data and records using a common taxonomy.
- Apply World Wide Web Consortium (W3C) semantic web technologies and standards to strengthen machine readability of federal data and better enable linking available federal data by researchers, scientists, and other data consumers.
- Encourage federal agencies and federal COPs to document where key datasets (e.g., case data, claim data, profile data) reside and identify possible linkages across the federal enterprise.
- Encourage federal agencies to identify authoritative data sources. Reduce the number of authoritative data sources for each dataset to as few as possible. Apply continuous monitoring methods to ensure high-quality data.
- Encourage federal agencies to apply standard metadata consistently to describe unstructured data, semi-structured data (e.g., free-form text, documents, images), structured data (e.g., tabular data) to better enable search, discovery, and analytics.
- Encourage agencies to effectively apply retention policies to identify current and key data. Delete or archive outdated data.

(2) Maximize the amount of non-sensitive data shared with the public

- Provide robust, centralized search interfaces internally and externally, leveraging the same taxonomy that drives data governance, so that data is more easily surfaced.
- Increase the use of application program interfaces (APIs) for public and organizational access to data rather than making the data available as a download. This will allow the government to gather metrics on data use.

(3) Leverage new technologies and best practices to increase access to sensitive or restricted data while protecting privacy, security, and confidentiality, and the interests of data providers

Technologies and other approaches used to make sensitive or restricted data available should have their capabilities and limitations thoroughly and independently evaluated prior to use. Sensitivity determinations should be based on case-specific assessments utilizing sufficiently expansive risk models, including the potential for external aggregation, rather than a cursory review of data elements. Innovative approaches to dealing with the potential erosion of privacy and civil liberties caused by increased data sharing will be essential and difficult. Additional research and Institutional Review Board (IRB)-like practices will likely be required.¹⁵

¹⁵ For example, the government could consider participating in or leveraging the work being done by the W3C Data Activity group on Scalable policy-aware linked data architecture for privacy, transparency, and compliance (SPECIAL) (<https://www.specialprivacy.eu/>)

Once privacy and security-related concerns are properly accounted for, novel ideas to increase access to sensitive or restricted data include:

- Investigate the use of blockchain technology to foster data security, exchange, and provenance. According to a World Economic Forum report, “Blockchain, or distributed ledger technology, could soon give rise to a new era of the internet even more disruptive and transformative than the current one. Blockchain's ability to generate unprecedented opportunities to create and trade value in society will lead to a generational shift in the internet's evolution, from an internet of information to a new generation internet of value.”¹⁶
- Consider a solution similar to Snapchat, where the sensitive or restricted data is available on a secured shared space and will ‘self-destruct’ or be removed after a certain time limit.
- Leverage secure cloud technology to foster anytime, anywhere access on multiple platforms. Storing data in, and syncing data to, the cloud also fosters data retention by the government, ensuring that data can be more readily discovered.
- Leverage artificial intelligence (AI) technologies to categorize content and redact sensitive data.
- Apply effective algorithms to aggregate and anonymize data to protect privacy. For example, the freely-available MITRE Identification Scrubber Tool¹⁷ automatically redacts personally identifiable information in free-form text.

RFI Request 3: *Decision-Making and Accountability.* Which best practices should the Federal Government consider to improve the use of data assets for decision-making and accountability? Specifically, list best practices for:

- **(A)** *Providing high quality and timely information to inform decision-making and learning;*
- **(B)** *facilitating external research on the effectiveness of government programs and policies which will inform future policymaking; and*
- **(C)** *fostering public accountability and transparency by providing accurate and timely spending information, performance metrics, and other administrative data.*

In the bulleted statements that follow, MITRE identifies several best practices for consideration. The letters at the end of each statement track to the three bullets above.

- Apply data quality improvement best practices to ensure that federal data is fit for use for its intended purposes. The quality of data plays an important role in analytics and

¹⁶ <https://www.weforum.org/whitepapers/realizing-the-potential-of-blockchain> (accessed July 24, 2018)

¹⁷ Available at <http://mist-deid.sourceforge.net/>

decision making. For example, preparing data so that it can be effectively analyzed can consume 50 to 80 percent of an analyst's time.¹⁸ Ensuring "fit for use data" up-front is less expensive and more timely and provides reliable data. (A)

- Consider providing guidance to federal agencies that focuses first on key foundational data management components, including data governance, data quality, data architecture, and metadata management, before focusing on more advanced data management concepts, including analytics and big data, as the more advanced data management concepts heavily rely on these foundations.¹⁹ (A)
- Require and audit the existence and quality of service-level-agreements (SLA) for data exchanged between agencies and with external organizations. Standardize and publish this metric. (A, C)
- Provide guidance to agencies that demonstrates the right analytic repository and data access for each specific mission need. Highlight an analytic data environment that reflects a "data highway" with multiple caches of increasing latency.²⁰ This environment can include data stores ranging from data lakes to data warehouses and store a variety of data. Some data, such as fraud data, is available in real time and raw form. Other data, like financial data, is highly cleansed and available daily or periodically. (B, C)
- Employ decision traceability to help combat algorithmic bias by revealing the reasoning steps leading to a given decision, and support redress requests from individuals negatively affected by decisions. (C)
- Link government agency spending data with agency strategic outcomes to inform performance, decision making, and policymaking. Currently, data available from different sources is managed inconsistently. As a result, the data does not link well. Cross-agency guidance may be needed to ensure common data management, common data quality practices, and the use of common standards to enable better linkages of data across multiple government sources. (C)
- The government could improve data quality and provenance using Explainable AI (XAI). "The current generation of AI systems offer tremendous benefits, but their effectiveness will be limited by the machine's inability to explain its decisions and actions to users. Explainable AI will be essential if users are to understand, appropriately trust, and effectively manage this incoming generation of artificially intelligent partners."²¹ (C)

RFI Request 4: *Commercialization, Innovation, and Public Use. Outline best practices that the Federal Government should consider to facilitate the use of Federal Government data interested parties to enhance the accessibility and usefulness of the data through commercial*

¹⁸ Lohr, S., "For Big Data Scientists, 'Janitor Work' Is Key Hurdle to Insight," *New York Times*, August 17, 2014.

¹⁹ See, for example Aiken, P., "Succeeding at Data Management—BigCo Attempts to Leverage Data," *Journal of Data and Information Quality (JDIQ)*, Vol. 7, Is. 1-2, May 2016, <http://dl.acm.org/citation.cfm?id=2893482>

²⁰ Kimball, R., "Newly Emerging Best Practices for Big Data," The Kimball Group, September 30, 2012.

²¹ See David Gunning, DARPA. https://www.darpa.mil/attachments/XAIIndustryDay_Final.pptx

ventures, or innovation, or for additional public uses. Of particular interest are examples of how the Federal Government can promote data use by the private sector and scientific and research communities, by state and local governments for public policy purposes, for education, and in enabling civic engagement. Please include up to four examples of:

- *How enabling external users to access and use government data for commercial or additional public purposes spurs innovative technological solutions and fills gaps in government capacity and knowledge; and*
- *how supporting the production and dissemination of comprehensive, accurate, and objective statistics on the state of the nation helps businesses and markets operate more efficiently.*

For federal data to have enhanced impacts, it needs to be both easily accessible and useful for external entities. Fortunately, the same holds true outside the federal space, where several ongoing **best practices** and **commercial offerings** could potentially help the government quickly. These include:

- Make government data available in on-line “sandbox” environments to research organizations. These isolated environments allow others to access a copy of the federal data for research purposes, but in a simulated environment with restrictions on what they can or cannot do.
- The computational intelligence community often leverages Wolfram Alpha,²² which could be similarly leveraged for government data. This approach uses a variety of open data sources to compute answers using algorithms and AI technology with a high accuracy rate, on a number of subjects (e.g., science, math, finance, health).
- The following W3C products could be leveraged:
 - An emerging W3C Data Catalog Vocabulary, which is designed to facilitate interoperability between data catalogs published on the web
 - Other ongoing activities highlighted in their study titled “W3C Study of Practices and Tolling for Web Data Standardization”²³
- Consider U.S. participation in the Open Data Institute²⁴ to gain ideas and exchange knowledge, and/or consider creating similar partnership(s) with U.S. states and municipalities to help ensure their access to federal data and to promote idea exchange.
- Consider setting up a fund to reward exploration and development of innovative ideas in the manner of Innovate UK.²⁵

²² <http://www.wolframalpha.com/about/> (accessed July 25, 2018)

²³ <http://www.w3.org/2017/12/odi-study/> (a July 25, 2018)

²⁴ <https://theodi.org/>

²⁵ <https://www.gov.uk/government/organisations/innovate-uk>

Examples of how enabling external users to access and use government data for commercial or additional public purposes spurs innovative technological solutions and fills gaps in government capacity and knowledge

1. The U.S. Department of Veterans Affairs Center for Innovation (VACI), the Dean Center for Tick Borne Illness at Spaulding Rehabilitation Network/Harvard Medical School Department of Physical Medicine and Rehabilitation, MIT Hacking Medicine, and the Open Medicine Institute are collaborating and sharing data to help solve challenges in the prevention and treatment of Lyme disease.
2. The U.S. Patent and Trademark Office (USPTO) open data portal has improved the discoverability, accessibility, and usability of public patent and trademark data. The Developer Hub established a shareable and “social” platform to showcase unique ways data can be combined with other data sets, such as economic and geographic data. The portal also includes APIs for innovators to use in further mining this data, helping to inform USPTO customers where to spend their limited research and development resources, and providing a much more detailed view of the competitive landscape than previously available.
3. The collaboration between the National Institutes of Health (NIH), the UK-based Wellcome Trust, the Howard Hughes Medical Institute (HHMI), and The Open Science Prize has successfully stimulated international collaborations around open data. Prototypes of new tools, products, and services that use open digital content are being developed to help solve pressing public health and biomedical research challenges.
4. Government research, related to leveraging its data, continues to foster economic growth in significant ways. A prime example is NOAA’s providing access to weather data, which has significantly lowered the economic and human costs of weather-related damage through forecasts, enabled the development of a multi-billion-dollar weather derivative financial industry dependent on seasonal data records, and catalyzed a growing million-dollar industry of tools and applications derived from NOAA’s real-time data.

Examples of how supporting the production and dissemination of comprehensive, accurate, and objective statistics on the state of the nation helps businesses and markets operate more efficiently.

1. Business data collected by the U.S. Census Bureau from various sectors of the economy, such as manufacturing, construction, retail trade, and health care, along with data collected from state and local governments and international trade, helps businesses compare their operations to industry norms, find new markets, and inform key decisions. Examples are available at <https://www.census.gov/programs-surveys/economic-census/guidance/data-uses.html>

2. Data from the Securities and Exchange Commission (SEC) has powered investment firms for decades, and it is now possible to combine SEC data with other data sources for faster, more accurate, and more usable analysis.
3. Statistics on a rolling five-year projection of the manpower needed in specific geographies and skill areas was developed by the Brazilian government²⁶ in collaboration with private companies, trade associations, and labor unions in order to identify the best training provider to co-develop a curriculum with selected companies to meet emerging manpower needs.
4. Economic Indicators published by the Department of Commerce's Economic & Statistics Administration are accessed and used by businesses every day to make decisions. Retailers, such as Target, use the government data on neighborhood-level demographics, such as population density, owner-occupancy, and household size, to determine the optimal mix of goods with which to stock its stores throughout the country.²⁷

Feedback on Principles for a Comprehensive Federal Data Strategy

RFI Request 5: *Principles. The interagency team on Leveraging Data as a Strategic Asset has written a draft set of principles for a comprehensive data strategy. Please review and provide feedback on their clarity, appropriateness, completeness, and potential duplications.*

MITRE reviewed the draft set of principles for a comprehensive Federal Data Strategy. Our feedback is provided in the following table, which reflects the RFI-specified data management principles as well as suggested updates. To highlight each principle's match in terms of clarity, appropriateness, completeness, and potential duplications we use the following definitions:

- Clarity: Clear, concise, and written in a manner that will resonate with the audience
- Appropriateness: Addresses the question: does this principle makes sense for a Federal Data Strategy?
- Completeness: Having all the necessary components to fully embody the principle
- Potential Duplications: Ensures that this principle does not duplicate another principle

²⁶ <https://www.mckinsey.com/industries/public-sector/our-insights/government-by-design-four-principles-for-a-better-public-sector>

²⁷ <http://www.esa.gov/sites/default/files/the-value-of-the-acs.pdf>

| RFI-Specified Principle | Clarity | Appropriate-ness | Completeness | Potential Duplications | Proposed Update | Further Comments |
|---|---------|------------------|--------------|------------------------|--|--|
| Stewardship (Title) | ✓ | ✓ | x | x | Stewardship, <u>Security, and Privacy</u> | Security and Privacy are critical components of data management and they are currently embedded in the Stewardship principle. They should be reflected in the title. Alternatively, a separate Security & Privacy principle can be stated. |
| 1. <i>Exercise Responsibility:</i> Practice effective data stewardship and governance by maintaining modern data security practices, protecting individual privacy, and maintaining promised confidentiality. | x | ✓ | x | x | <i>Establish Ownership and Accountability:</i> Practice effective data stewardship within an established governance framework so that data assets are understandable, trusted, accessible, and interoperable. | |
| *NEWLY INTRODUCED PRINCIPLE* (based on proposed update above) | | | | | <i>Security & Access Rights:</i> Guard data under modern security practices. Implement mindful policies protecting individual privacy, maintaining promised confidentiality, and ensuring appropriate access levels. | New principle based on the proposed update to the title of the section (Stewardship & Security) above. |
| 2. <i>Uphold Ethics:</i> Consider, monitor, and assess the implications of federal data practices for the public and provide sufficient checks and balances to protect and serve the public interest. | ✓ | ✓ | ✓ | ✓ | (No modifications suggested; however, item needs to be renumbered due to newly introduced principle above.) | |
| 3. <i>Promote Transparency:</i> Articulate purposes for acquiring, using, and disseminating data and comprehensively document processes and products to inform data users. | ✓ | ✓ | ✓ | ✓ | (No modifications suggested; however, item needs to be renumbered due to newly introduced principle above.) | |
| *NEWLY INTRODUCED PRINCIPLE* <i>Align Data Needs to Business Goals Continually</i> | | | | | Align data needs to business goals continually. Business goals should be revisited as they change, often annually. | |

| RFI-Specified Principle | Clarity | Appropriate-ness | Completeness | Potential Duplications | Proposed Update | Further Comments |
|--|---------|------------------|--------------|------------------------|--|--|
| | | | | | Coordinate and prioritize data needs. Revisit and update data management principles and their rationale and implications regularly. | |
| Quality (Title) | | | | | | |
| 4. <i>Integrate Intentionality</i> : Create, acquire, use, and disseminate data deliberately and thoughtfully, considering quality, consistency, privacy, value, reuse, and interoperability from the start. | x | ✓ | x | x | <i>Integrate Intentionality for Value</i> : Create, acquire, use, and disseminate data deliberately and thoughtfully, according to a defined strategy that considers quality, consistency, security, privacy, reuse, redundancy, multiple sources, and interoperability from the outset. | |
| 5. <i>Ensure Relevance</i> : Validate that data are high quality, useful, understandable, timely, and needed. | x | ✓ | x | x | <i>Ensure Relevance for Use</i> : Validate that data is of appropriate (fit for use) consistency, currency, precision, and completeness and that it is delivered in a timely manner for use. | |
| 6. <i>Create Value</i> : Coordinate and prioritize data needs and uses, harness data from multiple sources, and acquire new data only when necessary. | x | ✓ | x | ✓ | (Delete) | Recommend deleting this Principle. It has been merged into the proposed #4 and 'Align Data Needs to Business Goals Continually' above. |
| Continuous Improvement | | | | | | |
| 7. <i>Demonstrate Responsiveness</i> : Improve data sharing and access with ongoing input from users and other stakeholders. | ✓ | ✓ | ✓ | ✓ | (No modifications suggested) | |
| 8. <i>Prioritize Best Practices</i> : Model, assess, and continuously update best practices throughout the data lifecycle. | ✓ | ✓ | ✓ | ✓ | (No modifications suggested) | |

| RFI-Specified Principle | Clarity | Appropriate-ness | Completeness | Potential Duplications | Proposed Update | Further Comments |
|---|---------|------------------|--------------|------------------------|------------------------------|------------------|
| 9. <i>Invest in Learning:</i> Promote a culture of continuous and collaborative learning with data and about data. | ✓ | ✓ | ✓ | ✓ | (No modifications suggested) | |
| 10. <i>Practice Accountability:</i> Audit data practices, document and learn from results, and make changes as needed based on findings. | ✓ | ✓ | ✓ | ✓ | (No modifications suggested) | |

MITRE recommends that all data management principles include their rationale and implications.²⁸ MITRE recognizes that the set of principles will vary over time, based on changing business priorities. Mindful of this variance, MITRE understands how important it is that the government put initial data management principles forward for input and also revisit these principles periodically. It is in this context that we suggested above adding the new principle “Align Data Needs to Business Goals Continually.”

The government may wish to consider numerous additional data management principles. For example, “Treat Data as an Asset” would be in line with objectives expressly stated in this RFI. Other categories of principles might focus on training, standardization, or preservation. Conversely, the most impact will be gained by having the final set of principles as small as possible, consisting only of those most critical to mission success.

Feedback on Use Cases for a Comprehensive Federal Data Strategy

RFI Request 6: *Call for Use Cases.* For the purposes of the Federal Data Strategy, a “Use Case” is a data practice or method that leverages data to support an articulable Federal agency mission or public interest outcome. The Federal Data Strategy is seeking best practices, missed opportunities, common solutions, and game changers that can help inform the four strategy areas:

1. Enterprise Data Governance. What data governance and stewardship practices should the Federal Government be employing and why?
2. Use, Access, and Augmentation. What data interoperability techniques or coordination tactics would better serve agency missions and the public?

²⁸ For an example see Fleckenstein, M. and Fellows, L., *Modern Data Strategy*, Springer, 2018, Appendix C.

3. Decision-making and Accountability. How can the Federal Government better assist policy-makers with data?
4. Commercialization, Innovation, and Public Use. What data solutions could address a pervasive problem in government service delivery or the public sphere?

| Brief Description of Opportunity | Relevance and Importance <ul style="list-style-type: none"> • Is it a best practice, missed opportunity, common solution, or game changer? • How could it be used to help address the strategy area? | Recommended Next Step Towards Leveraging as a Use Case |
|---|---|--|
| <p>1. Enterprise Data Governance. What data governance and stewardship practices should the Federal Government be employing and why?</p> | | |
| <p>Investigate and leverage governance and stewardship practices from similar domains.</p> <p>Example: National Information Exchange Model (NIEM),²⁹ which was initially developed by the Departments of Justice and Homeland Security</p> | <ul style="list-style-type: none"> • Best Practice/Common Solution • NIEM is a use case example of data governance, stewardship, and information exchanges for all levels of government (federal, state, local, and tribal). | <p>Consider leveraging and expanding NIEM as a framework and common vocabulary for data sharing and information exchanges beyond its current mission/business domains.</p> |
| <p>Consider leveraging existing interagency governance bodies (directly or by learning from their approaches) to quickly initiate action and minimize the chances of duplicative efforts.</p> <p>Examples: The Federal Chief Information Officer (CIO) Council from the IT domain or the National Council of Information Sharing and Analysis Center (ISAC) in the cyber domain</p> | <ul style="list-style-type: none"> • Best Practice/Common Solution • The Federal CIO Council has several communities of practice in place focused on security, innovation, services, strategies and infrastructure, and workforce. • ISAC could serve as a model for establishing information exchanges across federal agencies. | <p>Investigate if the CIO Council structure and procedures would enable a more rapid development of a data governance forum.</p> <p>ISAC policies and practices can be studied while developing data-centric versions.</p> |

²⁹ <https://www.niem.gov/>

| Brief Description of Opportunity | Relevance and Importance <ul style="list-style-type: none"> Is it a best practice, missed opportunity, common solution, or game changer? How could it be used to help address the strategy area? | Recommended Next Step Towards Leveraging as a Use Case |
|---|---|--|
| <p>Find easy ways to spur collaboration with state governments.</p> <p>Example: The National Association for State Chief Information Officers (NASCIO) is an organization of state governments that serves as the common forum for collaboration and the exchange of ideas for state CIOs.</p> | <ul style="list-style-type: none"> Common Solution Federal and state governments regularly need to exchange data in a number of different areas. Finding ways to make the sharing (and oversight) easier would be beneficial. | <p>NASCIO could be leveraged to share enterprise data governance practices. Alternatively, a similar organization can be created specifically supporting Chief Data Officers.³⁰</p> |
| <p>Enhance the focus on the quality of federal data.</p> <p>Example: HUD’s Homeless Management Information System (HMIS) provides a data quality framework, tools, and standards to enable national, state, and local communities to monitor data quality on a continuous basis.³¹</p> | <ul style="list-style-type: none"> Best Practice HMIS may serve as a model to leverage for government data quality practices. Consistent data quality practices would naturally lead to better data to support evidence-based decision making. | <p>Identify other government best practices for improving the quality (i.e., condition) of data to ensure that it is fit for use for its intended purpose. Develop new guidance or strengthen existing guidance on data quality for high-value and highly shared datasets.</p> |

³⁰ Johns Hopkins University is already attempting to expand its local community of CDOs to include the U.S. state CDOs.

³¹ <https://www.hudexchange.info/news/homelessness-data-collection-reporting-updates-and-deadlines/>

| Brief Description of Opportunity | Relevance and Importance <ul style="list-style-type: none"> Is it a best practice, missed opportunity, common solution, or game changer? How could it be used to help address the strategy area? | Recommended Next Step Towards Leveraging as a Use Case |
|--|--|--|
| <p>2. Use, Access, and Augmentation. What data interoperability techniques or coordination tactics would better serve agency missions and the public?</p> | | |
| <p>ASIAS is a trusted public-private platform in which airline and FAA safety data is gathered, cleansed, integrated, anonymized, and analyzed. The data is then presented back to the airlines for their use.</p> | <ul style="list-style-type: none"> Game Changer The approach enables the sharing of important, actionable information while protecting sensitive data and the source of the data. | <p>MITRE (which originally developed ASIAS) has already ported the concept to other domains, such as cybersecurity. It could be further extended to other domains where community sharing of sensitive data is required.</p> |
| <p>3. Decision-making and Accountability. How can the Federal Government better assist policy-makers with data?</p> | | |
| <p>Develop methods for policymakers to more easily gather data from traditionally unstructured formats.</p> <p>Example: The non-profit Sunlight Foundation’s “Open Law Library” is creating a platform to make policies and legislation, which are often available only in .PDF format, machine readable and easily accessible.³²</p> | <ul style="list-style-type: none"> Game Changer Most governments publish their laws and policymaking documents in .PDF form, making it difficult to access the information contained within and use it at a later date. “The ability to share and access legal data generated by other jurisdictions trying to answer the same question provides valuable context and can greatly improve decision making.”³³ | <p>Explore the creation of a platform that makes policies and legislation machine readable.</p> |
| <p>Leverage insights gained from prior data investigations.</p> <p>Example: the Pew Charitable Trusts’ 2018 report, “How States Use Data to Inform Decisions.”³⁴</p> | <ul style="list-style-type: none"> Missed Opportunity The Pew report includes ideas and use cases that the federal government may be able to leverage and adopt. | <p>Explore the Pew report for ideas and use cases that could also be applicable in the federal domain.</p> |

³² <https://sunlightfoundation.com/2017/10/30/lawmaking-is-data-making/>

³³ Ibid.

³⁴ <http://www.pewtrusts.org/en/research-and-analysis/reports/2018/02/how-states-use-data-to-inform-decisions>

| Brief Description of Opportunity | Relevance and Importance <ul style="list-style-type: none"> • Is it a best practice, missed opportunity, common solution, or game changer? • How could it be used to help address the strategy area? | Recommended Next Step Towards Leveraging as a Use Case |
|--|--|--|
| 4. Commercialization, Innovation, and Public Use. What data solutions could address a pervasive problem in government service delivery or the public sphere? | | |
| <p>Many municipalities are also exploring open data activities. For example, the city of Syracuse, N.Y., is using and sharing municipal data via a portal that provides a central location for open data, maps, and visualization tools to help residents understand what is happening in their city.³⁵</p> | <ul style="list-style-type: none"> • Common Solution • The government should consider allowing citizen and organizations to own and update information about themselves, as the credit bureaus do. Such engagement of the public will significantly raise data management awareness and ultimately improve data quality. If the government considered itself to be such an information broker, it could reconcile mistakes in the data across multiple systems and agencies. | <p>Explore and adopt ideas from municipalities' open data initiatives.</p> |
| <p>While using data to support healthcare enhancements is common, opportunities exist to investigate potential new efforts for communities that tend to get less attention, such as those with disabilities³⁶ or need the assistance of caregivers.³⁷</p> | <ul style="list-style-type: none"> • Missed Opportunity • Navigating requirements and obtaining information from numerous sources is often an onerous process for those requiring assistance. Can agencies work together to develop interfaces to data that would help? | <p>Initiate public-private efforts to uncover data stores that could assist the public on these complex issues, and develop easy to use interfaces to access them.</p> |
| <p>Support Commerce's Low Earth Orbit (LEO) Situational Awareness and Space Traffic Control systems.</p> | <ul style="list-style-type: none"> • Game Changer • Data management and commercialization will be key to this effort's success. | <p>Investigate to determine if it could become a useful case study, where lessons learned through supporting this initiative could be applied elsewhere.</p> |
| <p>Improper Payments is a massive (\$141 billion annually) and data-heavy federal issue. It's also the focus of CAP Goal #9.</p> | <ul style="list-style-type: none"> • Game Changer • Data-centric efforts on this problem space will reap benefits for two CAP goals. | <p>Establish connections with managers of CAP Goal #9 to develop a mutually beneficial project.</p> |

³⁵ <http://data.syr.gov.net/pages/tntdata>

³⁶ For example: <http://www.who.int/news-room/fact-sheets/detail/disability-and-health>

³⁷ For example: <http://www.who.int/news-room/fact-sheets/detail/assistive-technology>

Feedback on Stakeholder Engagement for a Comprehensive Federal Data Strategy

RFI Request 7: *Stakeholder Engagement. What are the best mechanisms for engaging stakeholders in the development of the data strategy? What platforms and processes are both comprehensive and efficient for collecting stakeholder feedback on interim work products and input on next steps?*

There are many potential mechanisms for engaging the public and stakeholders in the development of a data strategy, each of which has benefits and disadvantages. No one approach by itself will be the optimal solution through all stages in the development of the data strategy. Rather, multiple approaches may be leveraged, chosen according to which would yield the best outcomes for the current step.

- RFIs have proven to be extremely beneficial when asking very specific questions, or when requesting review and feedback on draft materials. They are less valuable for open-ended inquiries, as the responses collected will be unorganized and conflicting, making federal analysis extremely complicated. RFIs are issued within the Federal Register, and thus predominantly attract entities that normally interact with the federal government. They thus tend to be more valuable for gathering experienced insights than for eliciting innovative concepts from those working beyond the federal government domain.
- Town Halls, which can be conducted in-person or online, enable the host organization to gather a wide array of information from other organizations (government, commercial, non-profit) and individuals relatively quickly. Interactive Town Halls also offer the potential to produce synergistic results, as one idea prompts another during the discussion and interaction. Hosts of Town Halls will find it helpful to be specific about the discussion topic beforehand, perhaps even providing questions for discussion in advance. Less structured Town Halls often run the risk of discussions running off point. Finally, Town Hall organizers should establish, in advance, clear expectations for the type of information being solicited and how the information will ultimately be analyzed and used.
- Conferences can be useful in bringing together an entire community to devote a couple of days to information exchanges amongst experts, rapid training of new members, and developing community-wide plans. Such conferences also require significant pre-planning and come with a cost-recovery risk.
- The interagency identity community recently piloted a structured “meetup” format to foster public-private conversations on a specific, priority topic as well as to promote community building. The meetup began with federal presentations to explain the topic and the government’s goals for the meetup, shifted to a series of short presentations from the private sector to provide some lessons learned or novel ideas, and concluded

with small-group brainstorming and plan development.³⁸ The approach proved useful for fostering public-private communication and developing multiple draft approaches on how to overcome issues on a specific topic. A drawback of this approach is that there are no built-in mechanisms to continue the momentum gained during the meetup.

- Stakeholder events can be used to gather inputs from specific domains. They help their hosts understand important nuances within the domains that are often lost in broader discussions. To be beneficial they must be specifically targeted, however, so for broader domains with multiple issues the government might need to schedule several of them.
- “Road shows” are beneficial in gathering inputs from specific geographic areas, which may have enhanced knowledge or interests in an issue. They also encourage inputs from outside the D.C. region, which can be extremely beneficial in generating innovative ideas. These add travel costs to federal programs and require significant promotion in the targeted geographic areas (many of which may not be used to the federal government soliciting their ideas).
- Online fora or conversations via social media can be useful for enabling individual citizens to provide inputs, but also suffer from a higher likelihood that the inputs received will be off-target. Great ideas can be obtained via these approaches, but it can take effort to find them within the large volume of responses.
- Examples of applications of commercial capabilities leveraged by government entities includes DHS’s ISAOs’ use of Slack and the U.K. government’s use of the SMART Stakeholder Engagement Platform.

Additional Thoughts

The following are additional ideas that could help the federal government meet the objectives within the President’s Management Agenda and leverage data to meet federal priorities.

- Enhance the data knowledge of the federal workforce.
- The federal government could work more closely with academia to educate future federal employee data management professionals. The government could, for example, collaborate with universities to create a workforce that understands data management and has the skills necessary to use data as a strategic asset, as prior programs had helped develop teachers and cyber professionals.
- The government can develop and make available data management educational modules and content for current federal employees.
- The government should update the 2200 job series for data management to better reflect current roles and responsibilities, such as data governance and data quality.

³⁸ A summary and lessons-learned from this meetup are available at <http://events.afcea.org/FedID18/CUSTOM/pdf/June2018MeetupSummary.pdf>. This community plans to replicate this approach for ten additional topics at its annual conference in late September.

- The government could establish a Center of Excellence for data management, helping agencies jumpstart data projects by discovering government best practices and shareable artifacts.
- Optimizing the use of data isn't an issue that only the United States faces or is attempting to address. For example, Estonia is working a digital society program for efficient, secure, and transparent delivery of government services.³⁹ The federal government could establish international workshops to uncover novel applications or identify consistent best practices quite useful.

³⁹ <https://e-estonia.com/>