

MP 190492

July 4, 2019



Response of The MITRE Corporation to the Request for Comments on the Cross-Agency Priority Goal: Leveraging Data as a Strategic Asset: Phase 3

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

This page intentionally left blank.

Contents

Introduction	1
Input on Federal Data Strategy Action Plan	3
Input on Request 1: Identify additional actions needed to implement the Federal Data Strategy that are not included in this draft Action Plan and explain why.	3
Input on Item 2: Identify additional actions that would align with or complement ongoing federal data initiatives or the implementation of new legislation, such as the Foundations for Evidence-based Policy Making Act of 2018, and explain why.	4
Input on Item 3: Identify any actions in this draft Action Plan that should be considered for omission and explain why.	6
Input on Item 4: For each action, provide any edits and additional detail to ensure that they accurately and effectively describe needed activities, responsible entities, metrics for assessing progress, and timelines for completion.	6
Input on Item 5: For each action, provide information about the implementation resources necessary to ensure success of these action steps.	7
Action Edits	7
Action 1: Create an OMB Data Council	7
Action 2: Develop a Curated Data Science Training and Credentialing Catalog	8
Action 3: Develop a Data Ethics Framework	9
Action 4: Develop a Data Protection Toolkit	9
Action 5: Develop a Repository of Federal Data Strategy Resources and Tools	9
Action 6: Pilot a One-stop Standard Research Application	10
Action 7: Pilot an Automated Inventory Tool for Data.gov	11
Action 8: Pilot Standard Data Catalogs for Data.gov	11
Action 9: Improve Data Resources for AI Research and Development	11
Action 10: Improve Financial Management Data Standards	12
Action 11: Improve Geospatial Data Standards	12
Action 12: Constitute a Diverse Data Governance Body	12
Action 13: Assess Data and Related Infrastructure Maturity	13
Action 14: Identify Opportunities to Increase Staff Data Skills	13
Action 15: Identify Data Needs to Answer Key Agency Questions	14

MITRE Comments on Federal Data Strategy Action Plan

Action 16: Identify Priority Datasets for Agency Open Data Plans 14

Relationships Between Actions 14

Dependency and Timeline for Actions 20

Additional Thoughts 22

Introduction

The MITRE Corporation is a not-for-profit company that works across government to tackle difficult problems that challenge the safety, stability, security, and well-being of our nation through its operation of multiple federally funded research and development centers (FFRDCs) as well as public-private partnerships. With a unique vantage point working across federal, state, and local governments, as well as industry and academia, MITRE works in the public interest to discover new possibilities, create unexpected opportunities, and lead by pioneering together for public good to bring innovative ideas into existence in areas such as artificial intelligence, intuitive data science, quantum information science, health informatics, policy and economic expertise, trustworthy autonomy, cyber threat sharing, and cyber resilience.

MITRE has direct experience assisting federal agencies leverage government and private-sector data to meet critical mission needs. 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. 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.

As requested in the Request for Comments (RFC), MITRE has reviewed the draft action plan and the questions specifically posed in the RFC. We generally feel that the draft actions are proper, though some require fine-tuning. We also feel that the measurements need to be enhanced so that they better support their goals and the dependencies across the actions need to be better understood and properly staged. Please let us know how we may be of future assistance.

Input on Federal Data Strategy Action Plan

The following response reflects MITRE’s input requested toward the 2019–2020 Federal Data Strategy Action Plan.

[Input on Request 1](#): Identify additional actions needed to implement the Federal Data Strategy that are not included in this draft Action Plan and explain why.

We suggest adding Action “Define Baseline Metadata for the Federal Government” (newly suggested Action 17) as an inter-agency Action, and the Office of Management and Budget (OMB) as responsible. We suggest adding this Action for several reasons, including:

MITRE Comments on Federal Data Strategy Action Plan

- Baseline metadata such as Dublin Core¹ has been defined and widely adopted nationally and internationally. This information includes title, description, author/owner, date created, and date last updated, among other things.
- To effectively search for and share information, both structured and unstructured, having the same baseline set of metadata is key.
- The Foundations for Evidence-Based Policymaking Act of 2018 mandates that agencies publish data catalogs for key datasets (as part of the federal data catalog) by 7/14/2019.² Required metadata, in addition some Dublin Core metadata, includes restrictions, access method(s), responsible agency(s), and location of the dataset.
- Several other Actions in the Federal Data Strategy Action Plan either call specifically for metadata or benefit from close coordination in defining metadata. These include Actions 5, 8, 11, 12, 13, and 14. Please refer to Section “Relationship Between Actions” for details on suggested coordination.

We also suggest the following for this Action:

- **Responsible:** Office of Management and Budget
- **Measurement:** Inventory of baseline metadata – broadly applicable – across agencies
- **Timeline:** Completed within 6 months but dependent on “Action 12: Constitute a Diverse Data Governance Body.” See Section “Dependency and Timeline for Actions” for details.
- **Implementation Resources:** We suggest leveraging existing metadata standards, including:
 - Dublin Core³
 - National Information Exchange Model⁴
 - Treasury/Fiscal Service Data Registry⁵
 - National Institute of Standards and Technology (NIST) schema for evaluating federal attributes⁶
 - [ISO Geographic Metadata](#)

Input on Item 2: Identify additional actions that would align with or complement ongoing federal data initiatives or the implementation of new legislation, such as the Foundations for Evidence-based Policy Making Act of 2018, and explain why.

We suggest newly added Action “Develop Roadmap for the Government to Become Fully Digital.” Such an Action will align closely with a number of existing policies, guidelines, and mandates that aim to

¹ See <http://dublincore.org/>, accessed June 2019

² See <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>; section “3511. Data inventory and Federal data catalogue,” accessed June 2019

³ See <http://dublincore.org/>, accessed June 2019

⁴ See <https://www.niem.gov/>, accessed June 2019

⁵ See <https://www.transparency.treasury.gov/dataset/data-registry/registry#meta-data>, accessed June 2019

⁶ See <https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8112.pdf>, accessed June 2019

MITRE Comments on Federal Data Strategy Action Plan

move the federal government away from working with paper and toward digitization. A few examples include:

- The National Archives and Records Administration (NARA) mandates of 2016, 2019, and 2022⁷ are mandating moving the government away from managing paper and toward more digital management by requiring that email and permanent records be managed digitally as records and by mandating only the acceptance of digital content after 2022.
- The Foundations for Evidence-Based Policymaking Act of 2018⁸ requires agencies to make data machine readable and publish a data catalog, among other mandates.
- The DATA Act⁹—Government-Wide Financial Data Standards of 2014 requires the Department of the Treasury and the White House Office of Management and Budget to transform U.S. federal spending from disconnected documents into open, standardized data, and to publish that data online. It requires government-wide financial data standards for any federal funds and mandates common data elements for financial and payment information reported.
- The Open Data Policy—Managing Information as an Asset¹⁰ Executive Order of 2013 mandates that federal agency information resources be accessible, discoverable, and usable. It states that agencies collect and create information in a way that supports downstream processing and dissemination. In addition, this order cites a number of related guidelines and mandates, all significantly addressing how data and information must be managed. Thus, managing data is a de facto part of federal strategic plans.
- Electronic health records (EHR) interoperability¹¹ is the ability of different information systems and applications to communicate, exchange data, and use the information that has been exchanged. EHR interoperability relies upon an exchange schema and exchange standards to permit data to be shared across clinicians, labs, hospitals, pharmacies, and patient regardless of the application or application vendor.

The federal government spends an enormous amount of time, effort, and money managing paper. This includes storing large volumes of paper in warehouses, pulling boxes from storage for virtually every FOIA request, manually searching for and finding relevant paper content, verifying that paper and digital

⁷ NARA's 2016 mandate states that by 12/31/2016 agencies must manage email as records. NARA's 2019 mandate states that by 12/31/2019, agencies must manage permanent eRecords effectively. For details, see <https://www.archives.gov/files/records-mgmt/m-12-18.pdf> and "2018-2022 NARA Strategic Plan," Section 3.2 <https://www.archives.gov/files/about/plans-reports/strategic-plan/2018/strategic-plan-2018-2022.pdf>, which states that NARA will no longer accept analog data after 12/31/2022.

⁸ See <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>, accessed June 2019

⁹ See "Digital Accountability and Transparency Act of 2014" or the "DATA Act," <https://www.congress.gov/113/bills/s994/BILLS-113s994es.xml>, accessed June 2019

¹⁰ See Presidential Memorandum "Open Data Policy – Managing Data as an Asset," May 9, 2013, <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>, accessed June 2019

¹¹ See "What Is EHR and Why Is It Important?" <https://www.healthit.gov/faq/what-ehr-interoperability-and-why-it-important>, accessed July 2019

MITRE Comments on Federal Data Strategy Action Plan

(manually transcribed database) content match, and often mailing and managing content. These costs can be mitigated by managing content digitally.

We suggest that this Action include a list of barriers, such as policies requiring paper and possible mitigations. For example, some agencies, like the IRS (Internal Revenue Service), must accept paper by law. Alternatively, the government could provide kiosks (for example, at post offices) that would allow anyone to file tax forms or benefit claims digitally.

We also suggest the following for this Action:

- **Responsible:** Office of Management and Budget
- **Measurement:**
 - Inventory of laws and regulation that require paper
 - High-level roadmap for digitization
- **Timeline:** Completed within 6 months but dependent on “Action 12: Constitute a Diverse Data Governance Body.” See Section “Dependency and Timeline of Actions” for details.

[Input on Item 3:](#) Identify any actions in this draft Action Plan that should be considered for omission and explain why.

We suggest that “Action 7: Pilot an Automated Inventory Tool for Data.gov” and “Action 8: Pilot Standard Data Catalogs for Data.gov” be consolidated into a single Action “Pilot a Data Catalog for Data.gov Using an Automated Inventory Tool.” Following are reasons for consolidating Actions 7 and 8:

- These two actions are intricately linked. Developing a data catalog is made substantially easier by deploying an automated tool that inventories both data and metadata.
- Conversely, “...populating needed information on agency data inventories...,” as Action 7 requests, is significantly aided by a data catalog.

Additionally, we suggest that there is a timeline dependency between several actions. This causes Actions 7 and 8 (suggested combined), 13, and newly suggested Actions 17 and 18 to shift right in the timeline. Due to these shifts, the combined Actions 7 and 8, which are directly dependent on prior year-one Actions that complete at the end of year one, should be postponed to year two. For details, please refer to Section “Dependency and Timeline of Actions.”

[Input on Item 4:](#) For each action, provide any edits and additional detail to ensure that they accurately and effectively describe needed activities, responsible entities, metrics for assessing progress, and timelines for completion.

- Please refer to Section “Action Edits.”

[Input on Item 5:](#) For each action, provide information about the implementation resources necessary to ensure success of these action steps.

- Please refer to Section “Action Edits.”

Action Edits

This section addresses suggested edits and implementation resources as requested by Items 4 and 5 of the Federal Data Strategy Action Plan.

Action 1: Create an OMB Data Council

We suggest the following edits to Action 1:

- Creation of an OMB Data Council should be done in close coordination with other data management bodies. For example:
 - Coordinate/integrate with the Advisory Committee on Data for Evidence Building and Chief Data Officer Council mandated as part of the Foundations for Evidence-Based Policymaking Act of 2018.¹²
 - Coordinate/integrate with initiatives such as the General Services Administration’s (GSA’s) Technology Transformation Services.¹³
 - Coordinate with the IC CDO Council, as outlined in the “Intelligence Community Information Environment (IC IE) Data Strategy.”¹⁴
- Additionally, we suggest that the creation of an OMB Data Council be coordinated with other Federal Data Strategy Actions, including Actions 5, 7, 10, 12, 13, and 16 and newly suggested Action 18. Please see Section “Relationship Between Actions” for details on suggested coordination.
- We also suggest that the OMB Data Council develop and publish data quality framework and implementation guidance for authoritative data sources and open government datasets. We note:
 - The current OMB Office of Information and Regulatory Affairs guidance provides for quality of data that is disseminated, particularly financial and statistical data, but falls short on quality used by artificial intelligence (AI) and machine learning, and other data-reliant projects.
 - According to Forbes¹⁵ and other sources, analysts spend about 80% of their time wrangling the data (e.g., filling in missing values, correcting flaws, addressing anomalies). Wrangling the data typically does not include determining and correcting root cause, and continuous monitoring for data quality.

¹² See <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>, accessed June 2019

¹³ See <https://www.gsa.gov/about-us/organization/federal-acquisition-service/technology-transformation-services>, accessed, June 2019

¹⁴ See https://www.dni.gov/files/documents/CIO/Data-Strategy_2017-2021_Final.pdf, accessed June 2019

¹⁵ See <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#5e62dbf6f637>, accessed June 2019

MITRE Comments on Federal Data Strategy Action Plan

- Guidance leading to robust and consistent data quality practices across agencies will reduce data preparation time, particularly for AI and machine learning projects.

We also suggest the following additional metrics for this Action:

- **Measurement**
 - Document a list of other data management bodies with which the creation of the OMB Data Council has been coordinated/integrated.
 - Document an inventory of data-related policy, directives, mandates, and guidelines from the past 10 years. Highlight overlaps and conflicts. Draft suggested prioritization and streamlining.
 - This will be a significant help to accomplish providing “...guidance on government-wide data standards and improvements required by statute, such as the Digital Accountability and Transparency Act, the Foundations for Evidence-Based Policymaking Act (hereinafter ‘Evidence Act’), and the Geospatial Data Act.”¹⁶
 - Include the following in “Consistent Governance Approach Developed”:
 - Develop a common portal for agencies to document their key open datasets; this requires coordination with “Action 16: Identify Priority Datasets for Agency Open Data Plans.”
 - Develop guidelines on prioritization of actions and required coordination between them.
 - Define and publish a governance process for vetting/curating the best/suggested practices from those ideas that should be promoted.

Action 2: Develop a Curated Data Science Training and Credentialing Catalog

- We strongly suggest that the government broaden this Action to “Data Management” vs. only “Data Science.”
- We suggest that GSA:
 - Organize the catalog by level of proficiency, ranging from basic data management skills to machine learning, etc.
 - Outline key formal certifications.
- We suggest closely coordinating with Actions 14 and 17. Please refer to Section “Relationship Between Actions” for details on suggested coordination.

We also suggest the following additional metrics and implementation resources for this Action:

- **Measurement**
 - Develop a roadmap for leveraging/consolidating/integrating available inventory of offerings; includes leveraging academic partnerships and certifications.
 - Document the expected learning need by key job categories.
- **Implementation Resource**
 - We suggest compiling a consolidated, government-wide, quick-reference handbook of terms and definitions similar to Treasury’s Fast Book of symbols and titles.¹⁷

¹⁶ As stated in “Action 1: Create an OMB Data Council” in the DRAFT 2019-2020 Federal Data Strategy Action Plan

¹⁷ See https://fiscal.treasury.gov/files/fast-book/Fastbook_March_2019.pdf, accessed June 2019

MITRE Comments on Federal Data Strategy Action Plan

Action 3: Develop a Data Ethics Framework

- We suggest that the ethics framework should explicitly include building a synergistic approach between “Machine Ethics” and “Data Ethics” as a foundation against discriminatory outcomes in artificial intelligence.
- We suggest that the framework address ethical overlaps and conflicts (e.g., attorney code of ethics vs. agency-specific ethics).

We also suggest the following implementation resources:

- **Implementation Resources**
 - Each ethics use case should be presumed unique. Requiring a second set of eyes is a useful component of a data ethics framework.
 - Ambiguity of the “right answer” is inherent in diverse data environments. Thus, “one size does not fit all” and flexibility is essential.

Action 4: Develop a Data Protection Toolkit

- No edits.

Action 5: Develop a Repository of Federal Data Strategy Resources and Tools

- We suggest the government begin with a common industry framework (such as DAMA’s Guide to the Data Management Body of Knowledge¹⁸). If the government decides to publish its own framework, it is our recommendation that it draws on multiple industry frameworks. The new framework can include:
 - Examples of tool kits, case studies, and best practices that would be stored in the repository
 - A set of use cases for when to utilize this content to help to illustrate the use of the repository
- We suggest close coordination with several Federal Data Strategy Actions, including:
 - “Action 1: Create an OMB Data Council” to, for example, include a discussion about how new resources/ideas can be shared/loaded to the repository by any agency
 - “Action 12: Constitute a Diverse Data Governance Body” to incorporate the role of a governance body as part of vetting/curating content
- There is a likelihood of duplication of tools across Actions. We suggest:
 - Incorporating tools of other Actions – e.g., Action 7 (automated inventory tool) and Action 8 (data catalog tool) in the repository
 - On a case-by-case basis, linking to or from other relevant repositories; alternatively making this repository a one-stop shop for users, with links to other, relevant agency repositories

We also suggest the following additional metrics and implementation resources for this Action:

- **Measurement**
 - The proposed metric, “Number of agencies that access and use the tool within 3/6/9/12 months,” does not adequately measure value, because it ignores impact. Agencies could

¹⁸ <https://dama.org/content/dmbok>, accessed July 2019

routinely check the repository and find nothing of value. We suggest survey mechanisms that allow more in-depth analysis and feedback. Another approach from social media that may prove valuable is for users to rate (1-5 stars) and review a given tool.

- The Federal Data Strategy may also consider a metric that measures the extent to which this repository links and/or interfaces with relevant repositories, with an aim to reduce duplication and create a “one-stop shop” repository. An example measure might be “Number of other sites containing similar content as this repository” (the lower the better).
- **Implementation Resources**
 - Evaluate and select an appropriate repository capability capture and management tool (e.g., a knowledge repository).
 - Evaluate and select an appropriate hosting architecture/platform and software to be managed as an ongoing concern.
 - Deploy the repository incrementally and ensure value-added functionality in each release. To achieve this, requirements for the repository should be vetted and prioritized with users.

Action 6: Pilot a One-stop Standard Research Application

- We suggest that this Action adopt agile software development methodology that enables developers to start with a subset of agencies and add harmonization in workflows incrementally.
- We also suggest that this Action coordinate closely with data governance bodies at both the agency and Federal Data Strategy levels (Actions 12 and 1, respectively).

We also suggest the following additional metrics and implementation resources for this Action:

- **Measurement**
 - Implement a scale that measures improvement in user experience based on post-application surveys of participants who have used the old process.
 - Similarly, implement a scale that allows users to assess the quality of available datasets.
 - Survey the time required for users to complete new workflow vs. the legacy application process.
- **Implementation Resources**
 - MITRE developed the “Blue Ridge Architecture,”¹⁹ which provides an architectural model that will support a diversity of research projects and allow for their coordination across multiple research domains, including:
 - Varied partitioning methods for the analytic environment based on the expected needs of the research threads and projects
 - A method for orchestrating across multiple cloud-computing environments
 - Accommodating the varied types of data and workflows that researchers require across potentially multiple data sources
 - Providing common storage for results that will be visualized and used to improve or suggest ultimate decisions
 - Create a data dictionary, use cases, and user stories.
 - Create data models and application process workflow models.

¹⁹ See <https://www.mitre.org/sites/default/files/publications/pr-15-1254-architectural-model-mitre-research-blueridge.pdf>, accessed June 2019

MITRE Comments on Federal Data Strategy Action Plan

- Design a platform (e.g., web application, middleware, and common data services) to accommodate each agency/data provider.

Action 7: Pilot an Automated Inventory Tool for Data.gov

- As detailed in Section “Input on Item 3...,” we suggest that this Action be consolidated with Action 8, as piloting a standard data catalog and inventory tool for Data.gov are intimately linked.
- Further, we suggest the following close coordination:
 - Action 7 mandates “...an automated tool that leverages agency Information Collection Review (ICR) processes and documentation...” For any tool to effectively “leverage” automated processes, consistent metadata is required, as newly suggested Action 17 states.
 - Piloting an automated inventory tool for Data.gov also requires coordination among agencies that submit those datasets. We suggest this as a well-suited task for the newly created OMB Data Council (Action 1).

Action 8: Pilot Standard Data Catalogs for Data.gov

- As detailed in Section “Input on Item 3...,” we suggest that this Action be consolidated with Action 7, as piloting a standard data catalog and inventory tool for Data.gov are intimately linked.
- We suggest coordinating this Action closely with work being done on digital.gov.²⁰
- We also suggest close coordination with:
 - “Action 9: Improve Data Resources for AI Research and Development” because improving data resources for AI research and development (Action 9) is significantly aided by a standard data catalog.
 - “Action 16: Identify Priority Datasets for Agency Open Data Plans,” which asks GSA to “...identify then use their initial list of highest priority datasets as the focus for enhancing their data inventories and catalogs...”
 - Newly suggested “Action 17: Define Baseline Standard Metadata for Government” to identify key metadata to be applied to every Data.gov (etc.) dataset.
- We suggest that this Action extend beyond those datasets that may or should be accessible via Data.gov. For example, we suggest coordination with “Action 6: Pilot a One-stop Standard Research Application.”

We also suggest the following additional metrics and implementation resources for this Action:

- **Measurement**
 - Remove “Number of additional datasets on or available through Data.gov.” Consider the following metrics instead:
 - Number of metadata sources identified/leveraged/harmonized
 - Number of datasets that leverage (complete, partial, no) key metadata
 - Measure usage in terms of rating or downloads

Action 9: Improve Data Resources for AI Research and Development

- We suggest this Action coordinate closely with other Federal Data Strategy Actions 3, 8, 12, and 16, as detailed in Section “Relationships Between Actions.”

²⁰ See <https://digital.gov/about/>, accessed June 2019

MITRE Comments on Federal Data Strategy Action Plan

- We suggest implementing a model store to house packaged AI. Such a model store can house standardized AI and machine learning (ML) training data and ML models.

We also suggest the following additional metrics and implementation resources for this Action:

- **Measurement**
 - Consider creating a scoring/points framework for agency AI adoption.
- **Implementation Resources**
 - Consider inventorying agency current and near future AI needs.
 - Consider developing use cases to expose the type and volume of data that will enrich the analytic pipeline.
 - Consider leveraging NIST or similar AI lexicons to foster a collaborative approach across federal/commercial industry and development community and then publishing a data dictionary for each dataset that addresses metadata, including pedigree and provenance of data.

Action 10: Improve Financial Management Data Standards

- This Action calls for, among other things, “...standardiz[ing] grant reporting data and improv[ing] data collection...” and “...improv[ing] Federal IT Spending Transparency...” We suggest these requirements can most effectively be met through closely working with the newly created OMB Data Council (Action 1).
- We suggest that OMB ensure the standards leads for financial management, grants management, and acquisition established under PMA CAP Goal 5 also work to have their standards incorporated into government-wide data collection and reporting systems such as USASpending.gov.

Action 11: Improve Geospatial Data Standards

- No edits.

Action 12: Constitute a Diverse Data Governance Body

- We suggest that this key Action closely coordinate with Actions 1, 9, 13, 14, 17, and 18, as detailed in Section ““Relationships Between Actions.”
- We see a direct dependency of this Action on “Action 1 Create an OMB Data Council.” We suggest that Action 12 alignment to agency priorities can’t be fulfilled until Action 1 is completed. We therefore suggest a shift of Action 12 to begin in month four, as detailed in Section “Dependency and Timeline for Actions.”

We also suggest the following updates for this Action:

- **Measurement**
 - Augment “Established diverse and empowered governance board” to “Established diverse and empowered governance board with roles and responsibilities outlined.”
 - Consider replacing “Number of actions taken by body each quarter” with “Percentage of (target) organizations actively participating.”
 - We suggest that the following may be valuable metrics in the future:
 - Percentage of data for which stewards are identified and published
 - Percentage of data for which candidate or approved authoritative data sources are identified and published

MITRE Comments on Federal Data Strategy Action Plan

- **Timeline** – consider increasing the timeline to 1 year and including the following:
 - Executive Charter including:
 - Roles and responsibilities
 - Timeline to execute the vision (draft principles and policies, triage/escalation approach, etc.)
 - Identification of data stewards – matrixed data governance (DG) members from across the organization
 - Recommended data governance operations staff
 - Employee goals/objectives to support DG (representing their organization but also the enterprise as a whole)
 - Management-level Charter including:
 - Roles and responsibilities
 - Identified business units represented
 - Agreement on DG member/steward time commitments
- **Implementation Resources**
 - Identifying Data Governance Office staff with governance/data management experience up front
 - Identifying data-literate governance members with ability to authoritatively speak for the business unit they represent
 - Committed senior management body to execute the vision (draft principles and policies, identify stewards and authoritative data stores, triage issues, etc.)

Action 13: Assess Data and Related Infrastructure Maturity

- We suggest that this Action should be carried out in close coordination with numerous other Actions, including 1, 12, 13, 14, 15, 17, and 18, as detailed in Section “Relationships Between Actions.”
- We see a direct dependency of this Action on “Action 12: Constitute a Diverse Data Governance Body.” We suggest that the Action 13 maturity assessment can’t be fulfilled until Action 12 is completed. We therefore suggest a shift of Action 13 to begin after Action 12 (which is itself dependent on Action 1) in month seven, as detailed in Section “Dependency and Timeline for Actions.”

We also suggest the following updates for this Action:

- **Measurement**
 - Assessment documented and available, including:
 - Initial identified focus areas, roadmaps, and timeframes for improved maturity in key low-scoring areas
- **Implementation Resources**
 - If a commercial maturity model is used, consulting assistance may be needed to execute it well.

Action 14: Identify Opportunities to Increase Staff Data Skills

- To execute this Action effectively, we suggest close collaboration with Actions 2, 5, 12, 13, 15, and 17. Please refer to Section “Relationships Between Actions” for details.

Action 15: Identify Data Needs to Answer Key Agency Questions

- Action 15 calls for agencies to “...consider what data are currently available; any issues around data quality or coverage; and if data are not available, how they might be collected or acquired.” We suggest that understanding which key questions to ask requires an understanding of an agency’s data management maturity (Action 13) as well as the skill to know the right questions to ask (Action 14).

We also suggest the following updates to metrics and implementation resources for this Action:

- **Measurement**
 - Update metric to “Agency has engaged stakeholders and consulted with key leaders such as CDO to begin identifying the data needed and document existing challenges (e.g., existing legislation, regulations, or policies) that limit or prevent access to or collection of the data; new technology/business processes that would be needed; funding amount for acquisition/subscription to the data to answer key agency questions.”
- **Implementation Resources**
 - Have OMB compile responses across agencies, identify and prioritize common/shared data needs, and develop an action plan to address the most common, high-priority data needs.

Action 16: Identify Priority Datasets for Agency Open Data Plans

- We suggest that this Action should be carried out in close coordination with numerous other Actions, including 1, 8, 9, and 18, as detailed in Section “Relationships Between Actions.”

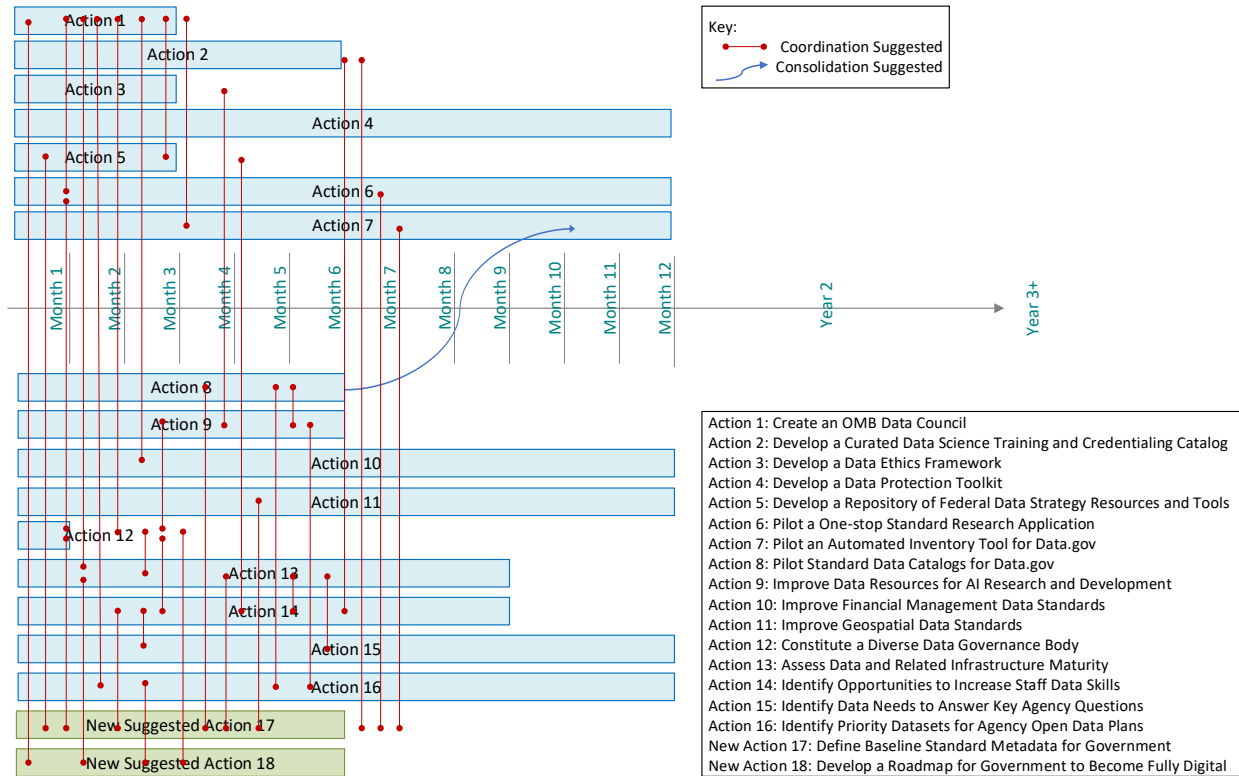
We also suggest the following updates to metrics and implementation resources for this Action:

- **Measurement**
 - Augment metric “Priority agency datasets identified with stakeholder input” to:
 - “Agency has engaged stakeholders and consulted with key leaders, such as Chief Data Officers and key healthcare industry leaders and organizations, to identify priority datasets needed to answer key agency questions.”
 - We are not clear on what is meant by “Agency has categorized data needs for key questions by data type” and suggest rewording this metric.
 - For “Priority agency datasets identified with stakeholder input”:
 - Rather than a simple “Y/N,” document the list of key datasets identified, including metadata and provenance.
 - Also document key issues with making these datasets open.

Relationships Between Actions

For optimal execution of the year-one actions, we suggest coordinating a number of key Actions. We also suggest consolidating Actions 7 and 8. The diagram below provides an overview of this coordination. The table that follows the diagram provides additional detail.

MITRE Comments on Federal Data Strategy Action Plan



Action	Coordination/Consolidation
<p>Action 1: Create an OMB Data Council</p>	<ul style="list-style-type: none"> • Action 1 requires: <ul style="list-style-type: none"> ○ Coordinat[ing] information policy development and implementation activities and provid[ing] guidance on government-wide data standards and improvements ○ Provid[ing] a coordinated voice (or response) and common guidance regarding the implementation of the Federal Data Strategy ○ Ensur[ing] relevant participants are engaged in data governance ○ Provid[ing] a way to address issues that cross agencies ○ Help[ing] inform government-wide management and budget priorities for data management <p>As such, we suggest tightly coordinating on several Actions, including Actions 5, 7, 10, 12, 13, and 16, and newly suggested Action 18. To denote a few examples:</p> <ul style="list-style-type: none"> • “Develop[ing] a Repository of Federal Data Strategy Resources and Tools” (Action 5) requires data governance at the inter-agency level to ensure consistency. • “Pilot[ing] an Automated Inventory Tool for Data.gov” (Action 7) requires a consistent approach to applying metadata across Data.gov datasets, so that needed data can be easily found.

MITRE Comments on Federal Data Strategy Action Plan

Action	Coordination/Consolidation
	<ul style="list-style-type: none"> • “Constitut[ing] a Diverse Data Governance Body” (Action 12) requires guidance from interagency bodies, such as the OMB Data Council, on priorities and data governance framework. • “Defin[ing] Baseline Standard Metadata for Government” (newly suggested Action 17), requires metadata consistency across agencies, for which the OMB Data Council is ideally suited to guide implementation. • We also suggest coordinating this Action with newly suggested “Action 18: Develop a Roadmap for Government to Become Fully Digital.”
Action 2: Develop a Curated Data Science Training and Credentialing Catalog	<ul style="list-style-type: none"> • Action 14 requires agencies to identify “critical data skills, [assess] current staff capacity, and [develop] an initial plan to reskill their workforce or hire to meet their data demands.” To do so, agencies will benefit by being able to access “a curated catalog of federal and non-federal training offerings in data science,” as Action 2 states. • We also suggest close coordination between Action 2 and newly suggested “Action 17: Define Baseline Standard Metadata for Government,” as metadata and education on metadata are fundamental to the Federal Data Strategy.
Action 5: Develop a Repository of Federal Data Strategy Resources and Tools	<ul style="list-style-type: none"> • Action 5 requires “...create[ing] a government-wide repository of tools and resources to assist agencies with implementing the Federal Data Strategy.” We suggest that developing such a repository be done in close collaboration with the OMB Data Council (Action 1). We also suggest that one of those tools be a set of standard baseline metadata, consistently applied across the government, as newly suggested Action 17 states.
Action 6: Pilot a One-stop Standard Research Application	<ul style="list-style-type: none"> • Action 6 calls for “...a one-stop standard application for accessing federal data assets for research and other evidence building purposes.” We agree. For researchers to effectively navigate this site, we suggest they coordinate with the OMB Data Council (Action 1) and have access to a standard set of baseline metadata, as newly suggested Action 17 states.
Action 7: Pilot an Automated Inventory Tool for Data.gov	<ul style="list-style-type: none"> • Action 7 mandates “...an automated tool that leverages agency Information Collection Review (ICR) processes and documentation...” For any tool to effectively “leverage” automated processes, consistent metadata is required, as newly suggested Action 17 states. • Piloting an automated inventory tool for Data.gov also requires coordination among agencies that submit those datasets. We suggest this as a well-suited task for the newly created OMB Data Council (Action 1). • We also suggest that this Action be consolidated with Action 8, as piloting a standard data catalog and inventory tool for Data.gov are intimately linked.
Action 8: Pilot Standard Data Catalogs for Data.gov	<ul style="list-style-type: none"> • Action 8 calls for “...an improved standard Federal Data Catalog kit pilot for metadata management and data hosting capabilities, in support of legally required federal data catalog requirements.” We suggest that Action 8 coordinate closely with Actions 9, 16, and 17. For example: <ul style="list-style-type: none"> ○ Improving data resources for AI research and development

MITRE Comments on Federal Data Strategy Action Plan

Action	Coordination/Consolidation
	<p>(Action 9) is significantly aided by a standard data catalog.</p> <ul style="list-style-type: none"> ○ Identifying priority datasets for open data plans (Action 16) is closely linked to piloting a standard Data.gov data catalog. ○ Having government-wide baseline metadata (newly suggested Action 17) significantly contributes to an “improved data catalog kit for metadata management.” <ul style="list-style-type: none"> ● We also suggest that this Action be consolidated with Action 7, as piloting a standard data catalog and inventory tool for Data.gov are intimately linked.
<p>Action 9: Improve Data Resources for AI Research and Development</p>	<ul style="list-style-type: none"> ● This Action mandates, among other things, “increase[d] access ... while protecting safety, security, privacy, and confidentiality based on OMB guidance....” This requires close coordination with “Action 12: Constitute a Diverse Data Governance Body.” ● Similarly, the above mandate requires close coordination with/adherence to “Action 3: Develop a Data Ethics Framework.” ● To accommodate “...increase[d] access and use by the greater non-Federal AI research community...” we also suggest closely coordinating with “Action 8: Pilot Standard Data Catalogs for Data.gov,” and “Action 16: Identify Priority Datasets for Agency Open Data Plans.”
<p>Action 10: Improve Financial Management Data Standards</p>	<ul style="list-style-type: none"> ● This Action calls for, among other things, “...standardiz[ing] grant reporting data and improv[ing] data collection...” and “...improv[ing] Federal IT Spending Transparency....” We suggest these requirements can most effectively be met through closely working with the newly created OMB Data Council (Action 1).
<p>Action 11: Improve Geospatial Data Standard</p>	<ul style="list-style-type: none"> ● Action 11 calls for the Federal Geospatial Data Committee to “establish content standards for metadata.” We suggest that this Action coordinate closely with newly suggested “Action 17: Define Baseline Standard Metadata for Government.”
<p>Action 12: Constitute a Diverse Data Governance Body</p>	<ul style="list-style-type: none"> ● This key Action requires an “appropriately inclusive and empowered data governance body.” As such, it is most closely linked to “Action 1: Create an OMB Data Council.” ● Furthermore, we see this key Action as closely linked to Actions 9, 13, 14, 17, and 18. For example: <ul style="list-style-type: none"> ○ Coordinating the “safety, security, privacy, and confidentiality” requirements stated by Action 9 ○ Facilitating the execution of the maturity assessment as required by Action 13 ○ Staffing the Data Governance Body with the right people, and thus closely aligning with “identifying opportunities to increase staff data skills,” as required by Action 14 ○ Coordinating with the OMB Data Council to facilitate the creation of a baseline government metadata set, as noted by newly suggested Action 17 ○ Facilitating the government’s move away from paper-oriented

MITRE Comments on Federal Data Strategy Action Plan

Action	Coordination/Consolidation
	toward digital data and information, as noted by newly suggested Action 18
Action 13: Assess Data and Related Infrastructure Maturity	<ul style="list-style-type: none"> • We agree that this Action is fundamental and key for the Federal Data Strategy. We suggest that it should be carried out in close coordination with numerous other Actions, including 1, 12, 13, 14, 15, 17, and 18. For example: <ul style="list-style-type: none"> ○ We suggest that effectively executing a maturity assessment requires close coordination with the OMB Data Council (Action 1) as well as with the Agency Data Governance Body (Action 12). ○ Effectively assessing data maturity is closely linked to identifying and leveraging the right staff skills (see Action 14) and aligning people with the maturity assessment. ○ Action 15 requires agencies to “consider what data [have] issues around data quality...,” so any maturity assessment can be considered a related effort. ○ Action 17 suggests standardized, baseline metadata across the government. We suggest that assessing the maturity of such metadata is a key component of Action 13. ○ Across the federal government, mandates such as the NARA mandates of 2016, 2019, and 2022 are moving the government away from managing paper and toward more digital management. This is the essence of newly suggested “Action 18: Develop a Roadmap for Government to Become Fully Digital.” Assessing each agency’s maturity in digital vs. paper data management is closely aligned with Action 13.
Action 14: Identify Opportunities to Increase Staff Data Skills	<ul style="list-style-type: none"> • This Action requires agencies to “...identify critical data skills required to support high quality analysis and evaluation, data management, and privacy protection.” In particular, this requires close coordination with “Action 2: Develop a Curated Data Science Training and Credentialing Catalog.” Additionally, to execute this Action effectively we suggest close collaboration with Actions 2, 5, 12, 13, 15, and 17. For example: <ul style="list-style-type: none"> ○ The better the staff data skills... <ul style="list-style-type: none"> ▪ ...the better the resulting government-wide repository of tools and resources for the Federal Data Strategy (Action 5) ▪ ...the better agency Data Governance Bodies will be (Action 12) ▪ ...the better future data management maturity assessments will be (Action 13) ▪ ...the better the identification and quality assessments of available datasets will be (Action 15) ▪ ...the better the definition and management of a baseline federal metadata set will be (newly suggested Action 17)

MITRE Comments on Federal Data Strategy Action Plan

Action	Coordination/Consolidation
Action 15: Identify Data Needs to Answer Key Agency Questions	<ul style="list-style-type: none"> Action 15 calls agencies to “...consider what data are currently available; any issues around data quality or coverage; and if data are not available, how they might be collected or acquired.” We suggest that understanding which key questions to ask requires an understanding of an agency’s data management maturity (Action 13) as well as the skill to know the right questions to ask (Action 14).
Action 16: Identify Priority Datasets for Agency Open Data Plans	<ul style="list-style-type: none"> This action calls on agencies to “...identify then use their initial list of highest priority datasets as the focus for enhancing their data inventories and catalogs and approaches to secure data access and sharing.” We suggest that to do so, agencies will want to closely align this effort with piloting the Data.gov pilot catalog (Action 8) as well as increased access and use of federal data and data models (Action 9). We also suggest close coordination with the OMB Data Council (Action 1) and with identifying paper-based data that should be digitized and be part of open data (Action 18).
New Suggested Action 17: Define Baseline Standard Metadata for Government	<ul style="list-style-type: none"> Newly suggested Action 17 calls for agencies to develop baseline metadata that can be applied broadly to both structured and unstructured datasets across the government. It should inventory existing metadata standards as input for the baseline federal metadata. As such, it should coordinate closely with Actions 5, 8, 11, 12, 13, and 14. For example: <ul style="list-style-type: none"> Standardized baseline metadata across the government will be a key tool in a “government-wide repository of tools and resources” (Action 5). Piloting standard data catalogs for Data.gov (Action 8) relies significantly on associated catalog metadata. Action 11 explicitly calls for establishment of “content standards for metadata” for geospatial data. Helping to establish baseline metadata should be a key activity of a “Diverse Data Governance Body” (Action 12). Assessing metadata maturity is integral to a data management maturity assessment (Action 13). Coordinating closely with Action 14 ensures leveraging the right skills to define the baseline metadata.
New Suggested Action 18: Develop a Roadmap for Government to Become Fully Digital	<ul style="list-style-type: none"> Newly suggested Action 18 calls for OMB to craft a roadmap for the federal government to become fully digital. Doing so will mitigate the cost of data management by systematically decreasing dealing with paper and should include developing a list of barriers, such as laws, policies, and mandates, that require agencies to accept and/or work with paper. We suggest, that to build such a roadmap effectively, this Action coordinate closely with Actions 1, 12, 13, and 16. For example: <ul style="list-style-type: none"> Working with the OMB Data Council (Action 1) ensures a consistent roadmap across the government.

MITRE Comments on Federal Data Strategy Action Plan

Action	Coordination/Consolidation
	<ul style="list-style-type: none">○ Working with agencies' Data Governance Body (Action 12) further enhances collaboration with the OMB Data Council.○ Taking into account agencies' data management maturity (Action 13) informs the roadmap.○ Identifying paper-based data that should be digitized and be part of open data (Action 16) further enhances federal open data.

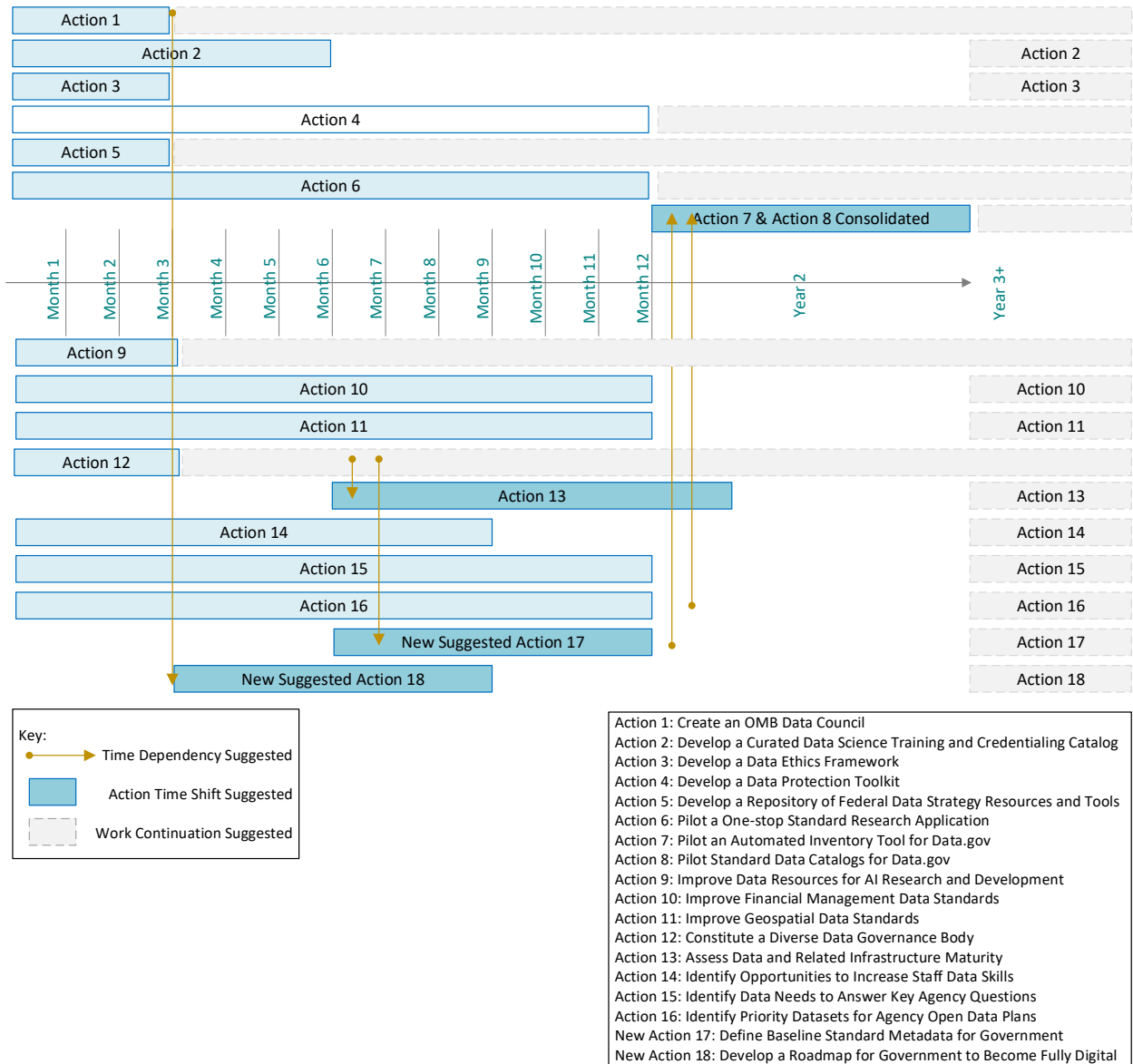
Dependency and Timeline for Actions

A key component of coordination between Actions is timing, as some Actions are directly dependent on others. The diagram below provides one approach to how the timing coordination between Actions might be executed.

A second key component of a data strategy is that it is continuous. We suggest that some Actions warrant uninterrupted continuation, while others require regular but intermittent attention. For example, "Action 1: Create an OMB Data Council" cannot end once the Council is created; rather the Council must execute data governance. Conversely, the catalog in "Action 2: Develop a Curated Data Science Training and Credentialing Catalog," once created, can be updated periodically. To this end, the diagram below also highlights how we suggest Actions should continue over time.

The table that follows the diagram provides additional detail.

MITRE Comments on Federal Data Strategy Action Plan



Action	Timeline Dependency
<p>Actions 7 and 8: Pilot a Data Catalog for Data.gov Using an Automated Inventory Tool</p>	<ul style="list-style-type: none"> We suggest that leveraging an automated tool to inventory Data.gov and piloting the associated data catalog is directly dependent on “Action 16: Identify Priority Datasets for Agency Open Data Plans,” since those priority datasets essentially constitute Data.gov and foster any related cleanup. Additionally, we suggest that these combined Actions directly depend on defining baseline metadata (Action 17), since a usable Data.gov critically depends on consistent metadata. We therefore suggest that the combined Actions 17 and 18 begin in year two, after the completion of Actions 16 and 17.

MITRE Comments on Federal Data Strategy Action Plan

Action	Timeline Dependency
Action 13: Assess Data and Related Infrastructure Maturity	<ul style="list-style-type: none"> We suggest that effectively executing a maturity assessment is directly dependent on the agency Data Governance Body (Action 12), since this is the body that will execute the assessment. Therefore, we suggest shifting Action 13 to the right to begin after the completion of Action 12 in month seven.
New Suggested Action 17: Define Baseline Standard Metadata for Government	<ul style="list-style-type: none"> We suggest that the agency Data Governance Body (Action 12) is also the skilled body that defines baseline metadata for its agency. Due to this direct dependency, we suggest that Action 17 be shifted to the right, and begin in month seven after the Data Governance Body is established.
New Suggested Action 18: Develop a Roadmap for Government to Become Fully Digital	<ul style="list-style-type: none"> Newly suggested Action 18 calls for OMB to craft a roadmap for the federal government to become fully digital. We suggest that the OMB Data Council is the right body to build such a roadmap effectively. This dependency shifts Action 18 to the right, beginning in month four, after the establishment of the OMB Data Council.

Additional Thoughts

The following additional ideas could help the government:

- While some Actions are funded, all Actions require ongoing work, even if intermittently. Having sufficient budget and skilled staff to carry these Actions forward is key to success. We feel that the sum total of these action items may be difficult for some agencies to achieve within the timeframes laid out when considering each agency’s current workload and competing priorities. It would be beneficial to lay out an Action Strategy that:
 - Starts with the agencies with the most capacity for additional workload and the ability to execute
 - Takes a continuing, long-term initiative into account and funds and staffs that initiative