PRESIDENT'S MANAGEMENT AGENDA

Federal Data Strategy

Improving Agency Data Skills Playbook

May 2020

Contents

Improving Agency Data Skills Playbook

Improving staff data skills will allow agencies to better harness the power of data. To help agencies address this need, this playbook provides advice and tools to help agencies implement Action 4 of the 2020 Federal Data Strategy Action Plan (hereafter "2020 Action Plan"), "Identify Opportunities to Increase Staff Data Skills." Executing the activities contained in this document will help agencies progress in fulfilling requirements of the Foundations for Evidence-Based Policymaking Act of 2018 in addition to related guidance[1-3].

This playbook's structure mirrors and expands upon the four major steps of a data skills capacity assessment:

1. identify critical data skills needed for the agency,
2. assess the current staff capacity for those data skills,
3. perform a data skills gap analysis to prioritize the agency's needs, and
4. identify and execute approaches to meet those needs.

The "who, what, and how" of each step is broken out below, followed by practical activities agencies will find helpful in executing the respective step. These activities are not strict, prescriptive instructions that require exact execution. Rather, agencies should prioritize and adapt the individual approaches in ways that best apply to their own organizations. While the scope for this assessment is intended to be at the agency level, some agencies may find it more helpful to do a more granular analysis at the operating or business unit level, particularly when skills and mission differ across an agency. If an agency takes such a distributed approach, a centralized body should analyze whether gaps can be filled by another part of the organization (e.g., the analytics office can provide advanced statistical analysis to other units).
As each agency is part of a larger Federal Government community, agencies should look to one another for inspiration and lessons learned as they identify and meet their data skills needs. Coordination and knowledge sharing will be critical to the success and longevity of efforts to develop data skills. Some councils that will be important for this include:

- Chief Data Officers Council
- Interagency Council on Statistical Policy
- Human Capital Planning and Resources councils

Agencies can also leverage centralized data governance bodies for knowledge sharing within the organization.

[1] Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5529.
[2] OMB M-19-23, Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance, July 10, 2019, found at: whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf.
[3] The Evidence Act also requires a capacity assessment, which in part determines "the extent to which evaluation and research capacity is present within the agency to include personnel" and "the extent to which the agency has the capacity to assist agency staff and program offices to develop the capacity to use evaluation research and analysis approaches and data in the day-to-day operations." This can help guide agencies in fulfilling Step 2 – "Assessing current staff capacity for needed data skills" and Step 3 – "Perform a data skills gap analysis" of this Action.

Figure 1: Process to identify opportunities to increase staff data skills

Step 1 – Identify any critical data skills needed for the agency

What? The agency formulates a clear idea of how data skills play a role in the execution of its mission(s) and supporting leadership priorities and its functions at the present and continuing into the future.

Who? The agency's Chief Data Officer, Statistical Official, Evaluation Officer, Chief Human Capital Officer, Chief Learning Officer, and Data Governance Body.

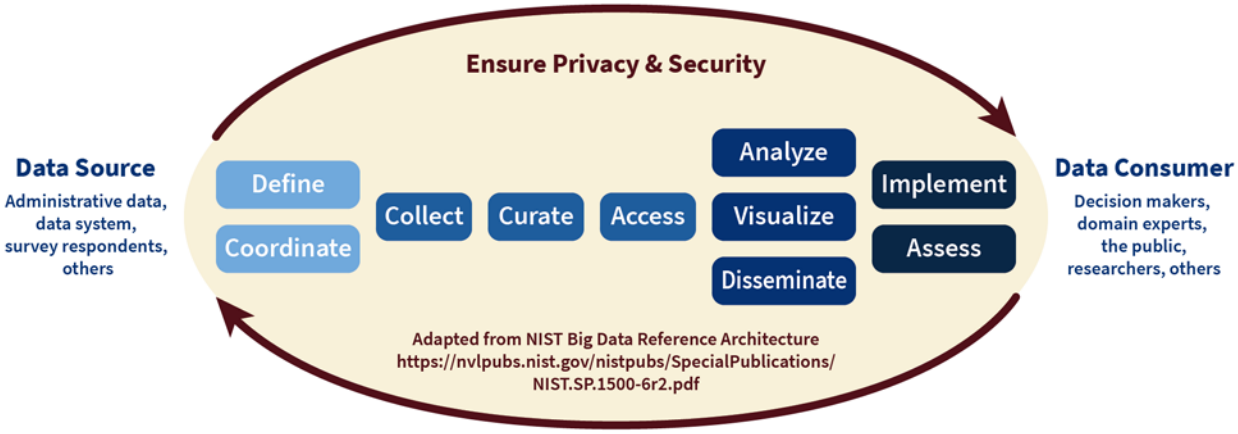
How? Identify agency leadership priorities and mission imperatives around data[4] and then convene agency staff to outline the ways data flows into and within the agency and is used to support decision-making or otherwise create value towards these priorities. It is important to take a holistic, integrated view for assessing data skills – both horizontally across operating units, and vertically within the agency hierarchy. An agency-level assessment is required; however, this assessment can be performed at the level that makes sense for the agency. If a distributed approach is used, a centralized group should review the needs or gaps from each of the operating units of the agency and compile. A common vocabulary for the federal data lifecycle was generated to support both this playbook and also for use in Action 13 of the Federal Data Strategy 2020 Action Plan, "Develop a Curated Data Skills Catalog," by adapting the NIST Big Data Reference Architecture[5] and linking subsequent data roles with Federal Data Strategy practices. The framework below lays out a variety of roles for those involved in the federal data lifecycle (Appendix A). These include data-focused roles[6] (given on the next page) and others in the organization such as leadership and domain experts. The required data skills are given for each of the roles in Appendix B.

[4] The learning agenda development process will be a helpful resource for identifying these priorities. See M-19-23, Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance, July 10, 2019, found at: whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf.

[5] U.S. Department of Commerce. National Institute of Standards and Technology (NIST). NIST Big Data Program. Retrieved from nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-6r2.pdf.

[6] Agency personnel in all data roles need to have a basic understanding of their agency's cybersecurity and privacy policies that govern data and need to work closely with their agency experts in cybersecurity and privacy throughout the data lifecycle More information on the Federal Information Security Modernization Act of 2014 can be found at congress.gov/113/plaws/publ283/PLAW-113publ283.pdf

Figure 2: Federal data lifecycle



Data roles

- **Define:** Identify agency and stakeholder needs for data of sufficient quality for intended uses
- **Coordinate:** Assess the ability of data resources and infrastructure to meet agency and stakeholder needs
- **Collect:** Organize, plan, and execute data collections and acquisitions to meet agency and stakeholder needs
- **Curate:** Organize, refine, and maintain agency data resources with sufficient quality to meet agency and stakeholder needs
- **Access:** Identify and develop multiple data access methods for agency staff and stakeholders
- **Analyze:** Optimize the ability of staff and stakeholders to use agency data to generate insights
- **Visualize:** Present data insights for consumption by leaders and stakeholders
- **Disseminate:** Provide multiple avenues for release of data and insights
- **Implement & Assess:** Maximize the use of data for decision-making, accountability, and the public good and continuously improving the data process

Other roles

- **Leadership:** Cultivate and support a culture of enterprise-wide demand and use of data to maximize outcomes
- **All Others:** Each staff member values the use of data in their day-to-day work
- **Domain Expert:** Understands the context around the data, the needs of all involved stakeholders, and often takes on many roles throughout the process in an advisory or lead capacity to inform collection, data systems, data dictionaries, data set design, and analysis
- **Privacy and Security*:*** Ensure that agencies are consistently adopting and using the most up-to-date methods to protect data and comply with all applicable laws and regulations

In addition, the activities and resulting conclusions from this Step may help inform the agency's efforts to address other Actions from the Federal Data Strategy 2020 Action Plan, including Action 1 "Identify Data Needs to Answer Priority Agency Questions," Action 2 "Constitute a Diverse Data Governance Body," Action 3 "Assess Data and Related Infrastructure Maturity," and Action 5 "Identify Priority Datasets for Agency Open Data Plans."

The following checklist and questions further break down how an agency might approach Step 1.

Checklist

- ☐ Identify agency leadership priorities and mission imperatives around data.
- ☐ Outline, using a shared vocabulary (e.g., the federal data lifecycle presented herein), the data needed and the processes or activities agency staff perform to transform raw data into useful knowledge to address leadership priorities.
- ☐ Discuss and document the data skills needed to execute these processes or activities. The skills should be documented regardless of whether agency personnel currently possess them. If a distributed approach is taken, compile the results at an agency level and review with the agency data governance body.

Example

Agency X is responsible for ensuring facilities are maintained and safely able to support partner agencies' missions. The table shown on page 6 represents Agency X's notes from a meeting convened to think through Step 1 of this Action. (Data roles are not necessarily aligned with a person's job role as one person may perform more than one data role.)

Questions for further thought

- How do an agency's or operating unit's public and private peers assess their own data skills needs?
- How well do the agency's position descriptions match up with the data skills uncovered in the above exercises?

Role	Critical processes in our context	Related data skills (technical and non-technical)
Define	Identify questions for data analysis to evaluate strategic and operational outcomes for Agency X	Understand the strategic and operational objectives of Agency X; facilitation; communication
Coordinate	Identify data sets internal and external to Agency X required to answer key questions	Be familiar with enterprise data architecture sources and systems used by Agency X; problem solving; understand how to assess data fitness for use
Collect	Gather information about facilities from site inspection report systems and purchase data from external sources for benchmarking and work with CFO office for cost data	Work with and query the SQL databases where these reports are stored; pull data from a NoSQL database; knowledge of data acquisitions; facilitate data use agreements internally
Curate	Combine, organize, refine, and maintain Agency X data resources about facilities and cost	Handle raw and unstructured data; knowledge of programming tools; ability to develop scalable ETL packages; knowledge of data warehouse and big data technologies; ability to understand stakeholder needs and requirements
Access	Store sensitive facility and cost data in a secure way that ensures analysts have the proper access	Understand the laws, policies, and supportive processes, e.g., Privacy Act[7], Information Quality Act[8,9], disclosure avoidance techniques
Analyze	Model real rate of facility degradation in comparison with theoretical design life and use data analysis techniques to explore data collected to answer other priority questions	Import data into statistical software; model facility/component condition, and prepare results for visualization using trend or pattern analysis and predictive modeling
Visualize	Design and communicate data analysis that tells a story and is relevant to data user needs to depict the past, current, and expected future health of facilities and equipment	Use robust data storytelling techniques to communicate results with the team members; programming or tools to create static or interactive reports, graphics, and dashboards; ability to gather/understand audience requirements
Disseminate	Share data about facilities with the operating unit-level facilities office and create an open data set for public release	Develop and implement communication products; ability to develop, maintain, and utilize relationships with key stakeholders; familiarity with different data formats and API technology
Implement and Assess	Communicate the implications of performing maintenance activities in the current vs. future fiscal years	Present insight from data to a range of audiences including senior leadership; strategic and innovative thinking; process improvement

[7] Privacy Act of 1974.
[8] Information Quality Act of 2001.
[9] OMB Memorandum M-19-15 Improving Implementation of the Information Quality Act issued April 24, 2019.

Step 2 – Assess current staff capacity for needed data skills

What? Having identified the processes and the associated data skills that the agency performs to turn raw data into actionable knowledge, the agency needs to determine which of those skills staff and managers currently possess. Without a clear understanding of how the current workforce's data skills match up with needed data skills, the agency cannot determine how to invest in filling gaps.

The agency should view the skills assessment in terms of a network or ecosystem of staff and skills. For example, increasing staff capability and capacity for predictive analytics alone will not result in effective use of predictive analysis products. Managers and leaders, program staff, and mission support staff also need specific applicable data skills in order for these advanced data products to be effectively used.

Who? The agency's Chief Data Officer, Statistical Official, Evaluation Officer, Chief Human Capital Officer, Chief Learning Officer, and Data Governance Body.

How? Develop and implement a process to determine alignment of staff data skills to critical agency data skills identified in Step 1.

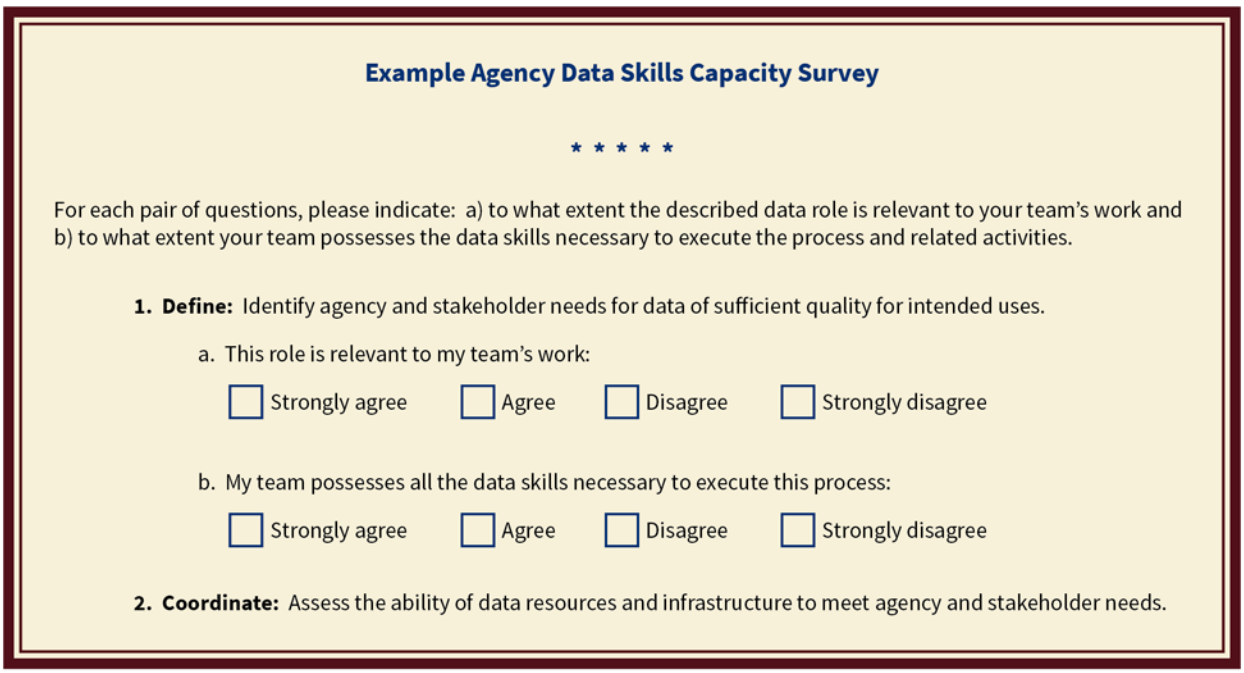
Checklist

- ☐ Decide how best to assess staff capacity to perform the data skills identified in Step 1 across the agency. Possible options include analyzing agency position descriptions and personnel data, querying supervisors or staff through surveys or focus groups, and measuring knowledge through pre- and post-training tests. Often a combination of methods will be needed to fully assess capacity. Creating survey and test questions and focus group protocols are examples of activities that require using appropriate data skills to properly develop.
- ☐ Perform the assessment and review the results.

Example

Agency X took a two-pronged approach to Step 2 – Assess the current staff capacity for needed data skills. Using notes from its Step 1 discussions for inspiration, the agency Evaluation Officer, Chief Data Officer and Statistical Official created a survey with 10 questions capturing a) to what extent program managers consider a given data process is relevant to their work, and b) to what extent, if relevant, they believe their staff have the requisite skills to execute that process. In addition to this widely-distributed survey, the team coordinated with the Chief Human Capital Officer and Learning Officer, and conducted several hour-long focus groups to get a deeper understanding of current data skills capacity. To round out the analysis, they worked with the Chief Human Capital Officer to analyze personnel data to determine potential staff turnover using retirement eligibility in key data roles. This approach was feasible for Agency X; yet, a different method may be necessary for other agencies.

Figure 3: Example survey



Example Agency Data Skills Capacity Survey

* * * * *

For each pair of questions, please indicate: a) to what extent the described data role is relevant to your team's work and b) to what extent your team possesses the data skills necessary to execute the process and related activities.

1. **Define:** Identify agency and stakeholder needs for data of sufficient quality for intended uses.

   a. This role is relevant to my team's work:

   ☐ Strongly agree ☐ Agree ☐ Disagree ☐ Strongly disagree

   b. My team possesses all the data skills necessary to execute this process:

   ☐ Strongly agree ☐ Agree ☐ Disagree ☐ Strongly disagree

2. **Coordinate:** Assess the ability of data resources and infrastructure to meet agency and stakeholder needs.

Questions for further thought

- How will critical data skills be measured individually and in aggregate?
- Who in the agency has critical data skills? What roles are they in right now and in which operating units? What is their workload?
- Where are those people working? Centrally within a data office? Embedded into programs and mission support functions?
- What barriers exist that might prevent effective skill development or application of those skills in routine work activities?

Step 3 – Perform a data skills gap analysis

What? After determining the required and available data skills, the agency should identify and prioritize any unmet needs. It is important to take a holistic, integrated view for assessing data skills – both horizontally across offices and teams, and vertically within the agency hierarchy. There are, after all, many actors who must work collaboratively to make good and ethical use of data associated with analysis activities.

Who? The agency's Chief Data Officer, Statistical Official, Evaluation Officer, Chief Human Capital Officer, Chief Learning Officer, and Data Governance Body.

How? Perform a gap analysis between the critical agency skills identified in Step 1 – Identify any critical data skills needed for the agency, and the available staff skills determined by Step 2 – Assess the current staff capacity for needed data skills.

Once the gap analysis is complete, the results should inform how the agency will develop the plan to provide and supplement data skills training and related resources to their staff and managers. The agency should also consider how to retain, recruit, or otherwise obtain talent to meet their data skill needs, for example, referencing recent OPM guidance[10] relevant to data science skills. For the results of the gap analysis to be actionable, a crosswalk should be developed to clearly show existing data skills versus data skills needed, assessing not only where there is unmet need for skills, but also where skills gaps exist.

Checklist

- ☐ Discuss and prioritize the gaps that exist between the current capacity at the agency-level (i.e., Step 2), and the data skills and associated resources the agency requires (i.e., Step 1).
- ☐ Consider how those gaps will change as the workforce turns over and technology evolves across the agency.

Example

Agency X returned to the framework and the documentation of what activities the agency did in relation to each process. With the survey results and focus group notes in hand, Agency X prioritized the following critical gaps:

1) ***Curate:***
    a) Agency X staff are well-versed in working with structured data using SQL databases, but partner organizations who provide data to Agency X are increasingly collecting unstructured data and storing them in NoSQL technologies. Agency X has very few employees comfortable with this technology.

[10] U.S. Office of Personnel Management. (2019). *Data Scientist Titling Guidance*. Retrieved from chcoc.gov/content/data-scientist-titling-guidance.

   b) The current data engineer staff have a high modal age. Employees who know how to work with the agency's legacy database architecture are starting to retire, taking valuable know-how with them.
2) ***Analyze:*** Agency leadership is interested in taking advantage of machine learning approaches. Agency X does not have many employees skilled in machine learning approaches or related coding languages such as Python.
3) ***Visualize:*** Some of the Agency's stakeholders have begun asking for interactive visualizations and the agency's CIO has recently approved and procured software to support this function. However, Agency X does not have many employees with experience using software that creates interactive reports or actually designing interactive visualizations.
4) ***Leadership:*** Leaders in Agency X don't feel confident in their understanding of how advanced analytics can be used or the power and limitations of predictive modeling techniques. This prevents some leaders from making decisions based on analyses.

Questions for further thought

- Are data skills gaps concentrated in specific parts of the agency, either organizationally or geographically?
- Are data skills gaps or strengths concentrated in one part of the workforce (e.g., among permanent employees, term employees, contract workers)?

Step 4 – Identify and execute ways to meet those needs

What? After identifying gaps, the agency should develop and execute strategies to close those gaps.

Who? The agency's Chief Data Officer, Statistical Official, Evaluation Officer, Chief Human Capital Officer, Chief Learning Officer, and Data Governance Body.

How? Determine the approaches to addressing agency data skills deficiencies. Agencies can address gaps by retaining talent, recruiting new employees, and building upon existing skill sets. The Curated Data Skills Catalog from Action 13 of the Federal Data Strategy 2020 Action Plan should aid the agency in meeting those needs by providing information on existing learning providers, programs, courses, and opportunities for employees to practice and apply new skills.

Based on OPM's guidance on Competency Modeling and Gap Analysis,[11] the agency identifies core and technical competency models necessary for mission achievement.

Chief Human Capital Officers and Chief Learning Officers play a big role in this phase and should work collaboratively to execute ways to meet those needs as resources permit. Using the data gathered from the data skills gap analysis, the Chief Human Capital Officer and Chief Learning Officer should establish a comprehensive and sustainable strategic approach to recruit and develop existing talent within organizations, and bridge identified skills gaps, working across human resources disciplines, just as data practitioners are being asked to do[12].

Some of the strategies to close the gaps include, but are not limited to:

- Use the agency's data governance body to analyze whether gaps can be filled by another part of the agency (e.g., the analytics office can provide analysis support to other units).
- Review the Curated Data Skills Catalog from Action 13 of the Federal Data Strategy 2020 Action Plan to identify learning providers, programs, and courses that can provide the necessary skills to bridge any gaps.
- Coordinate and share lessons learned in addressing skills gaps through government-wide councils such as the Chief Data Officers Council.
- Convene subject matter expert meetings to discuss topics ranging from developing workforce planning, upskilling, and reskilling to improving recruitment processes, job analysis documentation, and other human resources initiatives.
- Bring the voice of the employee to the table in data skill discussions and define data skills that all staff should have versus those appropriate for more advanced staff. Then, customize the data skills training program based upon the level of competency needed by different groups of employees.
- Identify successful change management strategies for managers to help them better understand how data can enhance the work of the organization.

[11] U.S. Office of Personnel Management. Workforce & Succession Planning: Competencies and Gaps. Retrieved from opm.gov/services-for-agencies/workforce-succession-planning/competencies-and-gaps/.

[12] Elkeles, T., & Phillips, J.J. (2017). *Chief Talent Officer: The Evolving Role of the Chief Learning Officer* (2nd ed.). New York, NY: Routledge.

- Work with professional societies and other external partners to train existing staff and recruit new talent.
- Enhance communication channels between the Chief Human Capital Officer and Chief Learning Officer, as well as those between managers and employees, to clearly define data skills needed for better workflows, processes, and overall effectiveness.
- Adapt or design, evaluate, and scale successful data skill improvement programs and initiatives. Equip employees with the tools and training they need to embrace these new data skills, including new technologies.
- Use OPM guidance[13] to help leaders understand how to retain, recruit, or otherwise obtain talent to meet their data skill needs.

Checklist

- ☐ Discuss and list potential gap-closing approaches that the agency is interested in pursuing.
- ☐ Set outcome milestones and timelines for achieving them.
- ☐ As the agency executes selected gap-closing approaches, evaluate success against milestones.

Example

This table shows how Agency X is pursuing three overall strategies for addressing four identified gaps, which it plans to start rolling out in FY20 Q3. For the intersection of each strategy and gap, Agency X laid out milestones.

[13] U.S. Office of Personnel Management. (2019). *Data Scientist Titling Guidance*. Retrieved from chcoc.gov/content/data-scientist-titling-guidance.

Figure 4. Example gap-closing strategy

		Gap-closing Strategy		
		Hiring Incentives	**Training**	**Phased Retirement**
Gap	1. Need more skill with NoSQL	**X** Hire two new engineers with NoSQL experience by FY21 Q1		
	2. Must retain some capacity to work with legacy data architecture			**X** Communicate availability of phased retirement to relevant individuals by FY20 Q4
	3. Want more skill with machine learning	**X** Hire two new analysts with Python experience by FY21 Q1	**X** Enroll 10+ staff in week-long machine learning trainings by FY20 Q4	
	4. Want more capacity to design interactive dashboards		**X** Enroll 5+ staff in 2-day data visualization course by FY21 Q1	
	5. Need more data understanding for leadership		**X** Enroll leadership into data fluency course by FY20 Q3	

Questions for further thought

- Which gaps are temporary, and which are chronic? Could the temporary gaps be addressed by hiring private companies, supplemental staff, or hiring retirees as temporary workers?
- What are the constraints (e.g., budgeting, time) for training and resources?
- If one strategy is failing, how might resources be shifted to a more successful strategy? At what point would that decision be made (i.e., when are the exit gates)?

Metrics and Related Practices and Actions

Metrics

Beyond the metrics presented in this playbook, agencies should develop performance metrics, assign responsibility, audit practices, collect implementation and outcome data, and document and learn from results. Much of this information will be useful for any future data infrastructure (Action 3) or data skills (Action 4) assessments and sharing with other agencies. Agencies should share processes, metrics, and results with OMB and other agencies where possible.

Related strategy practices and actions

Action 4 of the Federal Data Strategy 2020 Action Plan is designed to guide agencies in addressing their data skill needs. This should not be taken as an isolated action, as many of the other Federal Data Strategy actions and principles inform agency-specific implementation of this action. Agencies are encouraged to use this Action as a way to build an organizational culture that values data, to be as inclusive as possible with their workforce, and to make data skills broadly relevant within their organization on a day-to-day basis.

The Federal Data Strategy principles serve as motivational guidelines in the areas of ethical governance, conscious design, and a learning culture. They underlie the comprehensive data strategy that encompasses federal and federally-sponsored program, statistical, and mission-support data. These principles include concepts reflected in existing frameworks, such as those for the protection of personal information, for the management of information as an asset, for federal statistical agencies, and for federal evidence building. These principles inform the Federal Data Strategy's practices and actions.

Increasing the data skills of the federal workforce requires putting these principles into practice. Federal Data Strategy practices are involved in many of the data roles, whether technical skills associated with analysis, 'softer' skills associated with communication and coalition building, or otherwise. Refer to the Crosswalk of Federal Data Strategy Practices and the Federal Data Lifecycle in Appendix C for details.

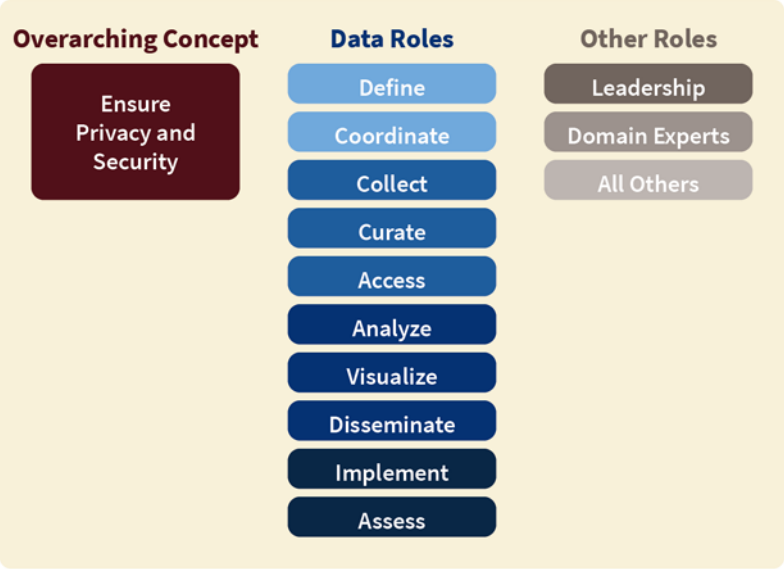
Additionally, other Federal Data Strategy actions may require training or otherwise be related to an agency's data skills, priority data analytics lifecycles, and evidence-building functions. For example:

- **Action 1** (*Identify Data Needs to Answer Priority Agency Questions*) and **Action 5** (*Identify Priority Datasets for Agency Open Data Plans*) can help identify priority data analytics lifecycles.
- **Action 2** *(Constitute a Diverse Data Governance Body)* creates a data governance body that will play a central role in **Action 4** Step 1.
- **Action 3** *(Assess Data and Related Infrastructure Maturity)* overlaps significantly with the current **Action 4** Step 2, which can be seen as a subcomponent of **Action 3**.
- **Action 8** *(Improve Data and Model Resources for AI Research and Development)* would benefit from having internal knowledge of Artificial Intelligence to assist non-federal uses of federal datasets.
- **Action 13** *(Develop a Curated Data Skills Catalog)* will provide a source for enhancing workforce data skills.

Agencies are encouraged to examine the other practices and actions, relate them to data skills, and determine what data skills will be required, both in the short and long term.

Appendix A

Federal Data Lifecycle Roles



Appendix B

Roles and Skills

Overarching Concept	Description
Ensure Privacy and Security: Ensure that agencies are consistently adopting and using the most up-to-date methods to protect data and comply with all applicable laws and regulations ***Practices*** 11. Prioritize Data Governance 12. Govern Data to Protect Confidentiality and Privacy 13. Protect Data Integrity	Agency personnel in all data roles need to have a basic understanding of their agency's cybersecurity and privacy policies that govern data, and need to work closely with their agency experts in cybersecurity and privacy throughout the data life cycle

Data roles and skills

Data Role	Skills
Define: Identify agency and stakeholder needs for data of sufficient quality for intended uses ***Practices*** 1. Identify Data Needs to Answer Key Agency Questions 2. Assess and Balance the Needs of Stakeholders 28. Align Quality with Intended Use	• Strategic planning • Performance metrics • Strategic thinking • Problem solving • Planning • Communication • Teamwork • Knowledge of requirements for information collection review approvals *Domain Expert input is important at this step*
Coordinate: Assess the ability of data resources and infrastructure to meet agency and stakeholder needs ***Practices*** 1. Identify Data Needs to Answer Key Agency Questions 8. Monitor and Address Public Perceptions 9. Connect Data Functions Across Agencies 10. Provide Resources Explicitly to Leverage Data Assets 15. Assess Maturity 16. Inventory Data Assets 17. Recognize the Value of Data Assets 18. Manage with a Long View 19. Maintain Data Documentation 20. Leverage Data Standards 21. Align Agreements with Data Management Requirements 36. Leverage Partnerships 37. Leverage Buying Power 38. Leverage Collaborative Computing Platforms	• Horizon scanning • Strategic thinking • Problem solving • Planning • Communication • Knowledge of the budget process • Business acumen • Knowledge of data standards • Federal enterprise architecture framework (FEAF) Data modelling • Relationship building • Coalition building • Knowledge of requirements for information collection review approvals *Domain Expert input is important at this step*
Collect: Organize, plan, and execute data collections and acquisitions to meet agency and stakeholder needs ***Practices*** 23. Allow Amendment 25. Coordinate Federal Data Assets 26. Share Data Between State, Local, and Tribal Governments and Federal Agencies 29. Design Data for Use and Re-Use 30. Communicate Planned and Potential Uses of Data	• Basic computer literacy • Collecting new source data such as through web scraping • Frame development, questionnaire/instrument development • Sampling design • Real-time or near real-time collection methods • Developing MOUs and Interagency Agreements • Communication • Interpersonal • Knowledge of requirements for information collection review approvals • Reviewing existing data for potential new uses

Data Role	Skills
Curate: Organize, refine, and maintain agency data resources with sufficient quality to meet agency and stakeholder needs ***Practices*** 16. Inventory Data Assets 19. Maintain Data Documentation 20. Leverage Data Standards 21. Align Agreements with Data Management Requirements 24. Enhance Data Preservation 32. Harness Safe Data Linkage 33. Promote Wide Access	• Techniques to understand and clean raw and unstructured data • Knowledge of operating systems • Develop scalable extract, transform, and load (ETL) processes • Big data technologies • Data warehouse/architecture design, development, construction, and maintenance • Programming tools • Statistics, data checking, and internal controls on data reporting and quality, consistency, and logic edits • Information quality control and quality assurance methods and techniques • Understand and incorporate stakeholder needs and requirements *Domain Expert input is important at this step*
Access: Identify and develop multiple data access methods for agency staff and stakeholders ***Practices*** 5. Prepare to Share 22. Identify Opportunities to Overcome Resource Obstacles 25. Coordinate Federal Data Assets 26. Share Data Between State, Local, and Tribal Governments and Federal Agencies 31. Explicitly Communicate Allowable Use 33. Promote Wide Access 34. Diversify Data Access Methods 35. Review Data Releases for Disclosure Risk	• Disclosure risk limitation • File formats • Business acumen • Communication • Data sharing policies • Database authentication methods • Tiered access to data methods • Understand stakeholders needs and requirements • Data modeling and transformation
Analyze: Optimize the ability of staff and stakeholders to use agency data to generate insights ***Practices*** 27. Increase Capacity for Data Management and Analysis 31. Explicitly Communicate Allowable Use 38. Leverage Collaborative Computing Platforms 39. Support Federal Stakeholders 40. Support Non-Federal Stakeholders	• Math and statistics • Trend or pattern analysis • Predictive modeling and projections • Geospatial analysis • Machine learning algorithms, artificial intelligence, natural language processing and rapid process automation • Data conditioning and data mining • Communication • Generate dashboards, charts, and graphs *Domain Expert input is important at this step*

Data Role	Skills
Visualize: Present data insights for consumption by leaders and stakeholders ***Practices*** 6. Convey insights from data	• Programming or tools to create static or interactive reports, graphics, and dashboards • Web design and HTML • Data visualization (including 508 compliance) • Understand stakeholder needs and requirements • User experience design • Communication • Data storytelling
Disseminate*:* Provide multiple avenues for release of data and insights ***Practices*** 6. Convey Insights from Data 7. Use Data to Increase Accountability 14. Convey Data Authenticity 30. Communicate Planned and Potential Uses of Data 31. Explicitly Communicate Allowable Use 33. Promote Wide Access 34. Diversify Data Access Methods 35. Review Data Releases for Disclosure Risk	• Communication • Negotiation • Understand stakeholder needs and requirements • Relationship building • Data formats and API technology • Companion materials creation • Accessibility standards (508 compliance)
Implement & Assess: Maximize the use of data for decision-making, accountability, and the public good, and continuously improving the data process ***Practices*** 2. Assess and Balance the Needs of Stakeholders 3. Champion Data Use 4. Use Data to Guide Decision-Making 7. Use data to Increase Accountability	• Strategic and innovative thinking • Problem solving • Data management • Scientific standards, theories, measurement, testing, and evaluation procedures • Communication • Coalition building and advocate for data use

Other roles and skills

Other Roles	Skills
Leadership Cultivate and support a culture of enterprise-wide harmonization of data-related activities to maximize outcomes	• Data use • Quantitative and qualitative data types • Communication • Relationship building • Data science process • Data sets, data priorities, and resources • Data-based decision making
All Others Each staff member values the use of data in their day-to-day work	• Data use and applications • Basic terminology • Data risks or limitations • Reading charts and graphs

Other Roles	Skills
Domain Expert • Understands the context around the data, the needs of all involved stakeholders, and often takes on many roles throughout the process in an advisory or lead capacity to inform collection, data systems, data dictionaries, data set design, and analysis • Prioritizes work to improve operational efficiency, mitigate bias and maximize effective outcomes • Provides a variety of insights to maximize data work (use), including knowledge of customers/beneficiaries, full understanding of processes and requirements, connections with a larger domain ecosystem of experts and peers, and more • Fully understands the landscape of the data ecosystem and the needs of all involved stakeholders • Acquires a complete understanding of all data assets and data-related business processes and their respective responsiveness to internal and external stakeholder needs/requirements	• Mission-related subject matter expertise • Operations and procedures • Data lifecycle • Strategic thinking • Data collection and use • Understand stakeholder needs and requirements • Relationship building • Communication • Knowledge of the possible data uses and limitations

Appendix C

Crosswalk of Federal Data Strategy Practices and the Federal Data Lifecycle

	Overarching Concept	*Data Lifecycle*								
	Ensure Privacy & Security	***Define***	***Coordinate***	***Collect***	***Curate***	***Access***	***Analyze***	***Visualize***	***Disseminate***	***Implement & Assess***
	Ensure that agencies are consistently adopting and using the most up-to-date methods to protect data and comply with all applicable laws and regulations	*Identify agency and stakeholder needs for data of sufficient quality for intended uses*	*Assess the ability of data resources and infrastructure to meet agency and stakeholder needs*	*Organize, plan, and execute data collections and acquisitions to meet agency and stakeholder needs*	*Organize, refine, and maintain agency data resources with sufficient quality to meet agency and stakeholder needs*	*Identify and develop multiple data access methods for agency staff and stakeholders*	*Optimize the ability of staff and stakeholders to use agency data to generate insights*	*Present data insights for consumption by leaders and stakeholders*	*Provide multiple avenues for release of data and insights*	*Maximize the use of data for decision-making, accountability, and the public good and continuously improving the data process*
1. Identify Data Needs to Answer Key Agency Questions: Use the learning agenda process to identify and prioritize the agency's key questions and the data needed to answer them.		✓	✓							
2. Assess and Balance the Needs of Stakeholders: Identify and engage stakeholders throughout the data lifecycle to identify stakeholder needs and to incorporate stakeholder feedback into government priorities to maximize entrepreneurship, innovation, scientific discovery, economic growth, and the public good.		✓								✓
3. Champion Data Use: Leaders set an example, incorporating data in decision-making and targeting resources to maximize the value of data for decision-making, accountability, and the public good.										✓
4. Use Data to Guide Decision-Making: Effectively, routinely, transparently, and appropriately use data in policy, planning, and operations to guide decision-making; share the data and analyses behind those decisions.										✓
5. Prepare to Share: Assess and proactively address the procedural, regulatory, legal, and cultural barriers to sharing data within and across federal agencies, as well as with external partners.						✓				
6. Convey Insights from Data: Use a range of communication tools and techniques to effectively present insights from data to a broad set of audiences.								✓	✓	
7. Use Data to Increase Accountability: Align operational and regulatory data inputs with performance measures and other outputs to help the public to understand the results of federal investments and to support informed decision-making and rule-making.									✓	✓
8. Monitor and Address Public Perceptions: Regularly assess and address public confidence in the value, accuracy, objectivity, and privacy protection of federal data to make strategic improvements, advance agency missions, and improve public messages about planned and potential uses of federal data.			✓							
9. Connect Data Functions Across Agencies: Establish communities of practice for common agency data functions (e.g. data management, access, analytics, informatics, and user support) to promote efficiency, collaboration, and coordination.			✓							
10. Provide Resources Explicitly to Leverage Data Assets: Ensure that sufficient human and financial resources are available to support data driven agency decision-making, accountability and the ability to spur commercialization, innovation, and public use.			✓							
11. Prioritize Data Governance: Ensure there are sufficient authorities, roles, organizational structures, policies, and resources in place to transparently support the management, maintenance, and use of strategic data assets.	✓									
12. Govern Data to Protect Confidentiality and Privacy: Ensure there are sufficient authorities, roles, organizational structures, policies, and resources in place to provide appropriate access to confidential data and to maintain public trust and safeguard privacy.	✓									
13. Protect Data Integrity: Emphasize state-of-the-art data security as part of Information Technology security practices for every system that is refreshed, architected, or replaced to address current and emerging threats; foster innovation and leverage new technologies to maintain protection.	✓									
14. Convey Data Authenticity: Disseminate data sets such that their authenticity is discoverable and verifiable by users throughout the information lifecycle, consistent with open data practices, and encourage appropriate attribution from users.									✓	
15. Assess Maturity: Evaluate the maturity of all aspects of agency data capabilities to inform priorities for strategic resource investment.			✓							
16. Inventory Data Assets: Maintain an inventory of data assets with sufficient completeness, quality, and metadata to facilitate discovery and collaboration in support of answering key agency questions and meeting stakeholder needs.			✓		✓					
17. Recognize the Value of Data Assets: Assign value to data assets based on maturity, key agency questions, stakeholder feedback, and applicable law and regulation to appropriately prioritize and document resource decisions.			✓							

	Overarching Concept	Data Lifecycle								
	Ensure Privacy & Security	Define	Coordinate	Collect	Curate	Access	Analyze	Visualize	Disseminate	Implement & Assess
	Ensure that agencies are consistently adopting and using the most up-to-date methods to protect data and comply with all applicable laws and regulations	Identify agency and stakeholder needs for data of sufficient quality for intended uses	Assess the ability of data resources and infrastructure to meet agency and stakeholder needs	Organize, plan, and execute data collections and acquisitions to meet agency and stakeholder needs	Organize, refine, and maintain agency data resources with sufficient quality to meet agency and stakeholder needs	Identify and develop multiple data access methods for agency staff and stakeholders	Optimize the ability of staff and stakeholders to use agency data to generate insights	Present data insights for consumption by leaders and stakeholders	Provide multiple avenues for release of data and insights	Maximize the use of data for decision-making, accountability, and the public good and continuously improving the data process
18. Manage with a Long View: Include data investments in annual capital planning processes and associated guidance to ensure appropriated funds are being used efficiently to leverage data as a strategic long-term asset.			✓							
19. Maintain Data Documentation: Store up-to-date and comprehensive data documentation in accessible repositories to facilitate use and document quality, utility, and provenance in support of informing key agency questions and meeting stakeholder needs.			✓		✓					
20. Leverage Data Standards: Adopt or adapt, create as needed, and implement data standards within relevant communities of interest to maximize data quality and facilitate use, access, sharing, and interoperability.			✓		✓					
21. Align Agreements with Data Management Requirements: Establish terms and conditions for contracts, grants, cooperative agreements, and other agreements that meet data management requirements for processing, storage, access, transmission, and disposition.			✓		✓					
22. Identify Opportunities to Overcome Resource Obstacles: Coordinate with stakeholders to identify mutually-acceptable cost recovery, shared service, or partnership opportunities to enable data access while conserving available resources to meet user needs.						✓				
23. Allow Amendment: Establish clear procedures to allow members of the public to access and amend federal data about themselves, as appropriate and in accordance with federal laws, regulations and policies, in order to safeguard privacy, reduce potential harm from inaccurate data, and promote transparency.				✓						
24. Enhance Data Preservation: Preserve federal data in accordance with applicable law, regulation, policy, approved schedules, and mission relevance.					✓					
25. Coordinate Federal Data Assets: Coordinate and share data assets across federal agencies to advance progress on shared and similar objectives, fulfill broader federal information needs, and reduce collection burden.				✓		✓				
26. Share Data Between State, Local, and Tribal Governments and Federal Agencies: Facilitate data sharing between state, local, and tribal governments and the Federal Government, where relevant and appropriate and with proper protections, particularly for programs that are federally funded and locally administered, to enable richer analyses for more informed decision-making.				✓		✓				
27. Increase Capacity for Data Management and Analysis: Educate and empower the federal workforce by investing in training, tools, communities, and other opportunities to expand capacity for critical data-related activities such as analysis and evaluation, data management, and privacy protection.							✓			
28. Align Quality with Intended Use: Data likely to inform important public policy or private sector decisions must be of appropriate utility, integrity, and objectivity.		✓								

	Overarching Concept	Data Lifecycle								
	Ensure Privacy & Security	Define	Coordinate	Collect	Curate	Access	Analyze	Visualize	Disseminate	Implement & Assess
	Ensure that agencies are consistently adopting and using the most up-to-date methods to protect data and comply with all applicable laws and regulations	*Identify agency and stakeholder needs for data of sufficient quality for intended uses*	*Assess the ability of data resources and infrastructure to meet agency and stakeholder needs*	*Organize, plan, and execute data collections and acquisitions to meet agency and stakeholder needs*	*Organize, refine, and maintain agency data resources with sufficient quality to meet agency and stakeholder needs*	*Identify and develop multiple data access methods for agency staff and stakeholders*	*Optimize the ability of staff and stakeholders to use agency data to generate insights*	*Present data insights for consumption by leaders and stakeholders*	*Provide multiple avenues for release of data and insights*	*Maximize the use of data for decision-making, accountability, and the public good and continuously improving the data process*
29. Design Data for Use and Re-Use: Design new data collections with the end uses and users in mind to ensure that data are necessary and of high enough quality to meet planned and future agency and stakeholder needs.				✓						
30. Communicate Planned and Potential Uses of Data: Review data collection procedures to update and improve how planned and future uses of data are communicated, promoting public trust through transparency.				✓					✓	
31. Explicitly Communicate Allowable Use: Regularly employ descriptive metadata that provides clarity about access and use restrictions for federal data, explicitly recognizes and safeguards applicable intellectual property rights, conveys attribution as needed, and optimizes potential value to stakeholders to maximize appropriate legal use.						✓	✓		✓	
32. Harness Safe Data Linkage: Test, review, and deploy data linkage and analysis tools that use secure and privacy-protective technologies to address key agency questions and meet stakeholder needs while protecting privacy.					✓					
33. Promote Wide Access: Promote equitable and appropriate access to data in open, machinereadable form and through multiple mechanisms, including through both federal and non-federal providers, to meet stakeholder needs while protecting privacy, confidentiality, and proprietary interests.					✓	✓			✓	
34. Diversify Data Access Methods: Invest in the creation and usability of multiple tiers of access to make data as accessible as possible while minimizing privacy risk and protecting confidentiality.						✓			✓	
35. Review Data Releases for Disclosure Risk: Review federal data releases to the public to assess and minimize the risk of re-identification, consistent with applicable laws and policies, and publish reviews to promote transparency and public trust.						✓			✓	
36. Leverage Partnerships: Create and sustain partnerships that facilitate innovation with commercial, academic, and other partners to advance agency mission and maximize economic opportunities, intellectual value, and the public good.			✓							
37. Leverage Buying Power: Monitor needs and systematically leverage buying power for privatesector data assets, services, and infrastructure to promote efficiency and reduce federal costs.			✓							
38. Leverage Collaborative Computing Platforms: Periodically review and optimize the use of modern collaborative computing platforms to minimize costs, improve performance, and increase use.			✓				✓			
39. Support Federal Stakeholders: Engage with relevant agencies to share expert knowledge of data assets, promote wider use, improve usability and quality, and meet mission goals.							✓			
40. Support Non-Federal Stakeholders: Engage with industry, academic, and other non-federal users of data to share expert knowledge of data assets, promote wider use, improve usability and quality, and advance innovation and commercialization.							✓			