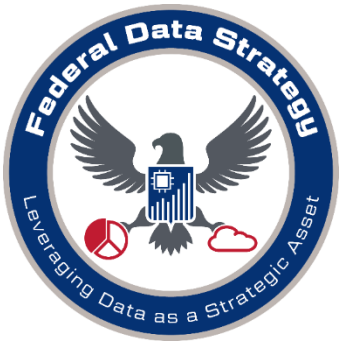
PRESIDENT'S MANAGEMENT AGENDA



Federal Data Strategy

Curated Data Skills Catalog

November 2020

Table of Contents

OVERVIEW AND PURPOSE **3**

DEVELOPMENT PROCESS **3**

FEDERAL DATA LIFECYCLE AND ASSOCIATED SKILLS **4**

FEDERAL DATA LIFECYCLE 4

ROLES AND SKILLS 7

Overarching Concept and Skills *7*

Data Roles and Skills *8*

Other Roles and Skills *12*

FEDERAL LEARNING OPPORTUNITIES **13**

RESOURCES AVAILABLE TO THE PUBLIC (INCLUDING ALL FEDERAL EMPLOYEES) **14**

CENSUS ACADEMY 14

US GEOLOGICAL SURVEY DATA MANAGEMENT TRAININGS 15

RESOURCES AVAILABLE TO SELECT FEDERAL EMPLOYEES **16**

CHIEF INFORMATION OFFICER (CIO) COUNCIL DATA SCIENCE TRAINING PROGRAM 16

EVALUATION AND EVIDENCE TRAINING SERIES 17

THE HHS DATA SCIENCE COLAB 18

THE TREASURY EXECUTIVE INSTITUTE (TEI) 19

APPENDIX A **20**

DEVELOPMENT OF THE FEDERAL DATA LIFECYCLE 20

APPENDIX B **21**

CROSSWALK OF FDS PRACTICES AND FEDERAL DATA LIFECYCLE ROLES 21

APPENDIX C **26**

CHIEF INFORMATION OFFICER (CIO) COUNCIL DATA SCIENCE TRAINING PROGRAM CURRICULUM 26

APPENDIX D **32**

SELECT FY2021 EVALUATION AND EVIDENCE TRAINING SERIES COURSES 32

APPENDIX E **33**

THE TREASURY EXECUTIVE INSTITUTE (TEI) MEMBERS FY20 33

APPENDIX F **35**

THE TREASURY EXECUTIVE INSTITUTE (TEI) SELECT FY20 DATA COURSES 35

Overview and Purpose

Improving staff data skills will allow agencies to better harness the power of data. To help Agencies address this need, Action 13 of the Federal Data Strategy (FDS) 2020 Action Plan called for the development of a curated data skills catalog (Catalog). This catalog offers a common vocabulary of data roles and skills in the Federal data ecosystem. It highlights examples of governmental learning opportunities to help agencies develop staff data competencies within the Federal Data Lifecycle. The Catalog can also help agencies develop training programs by highlighting existing government resources and curricula.

Increasing agency staff data capabilities supports the implementation of both the Evidence Act and related guidance [1-3] as well as the Executive Order on Maintaining American Leadership in Artificial Intelligence,[4] both of which require an assessment of staff capacity for various data-related functions.

The Catalog consists of two parts:

- **Federal Data Lifecycle and Associated Skills**, providing a common vocabulary of the data roles and skills in the Federal data ecosystem.

- **Federal Learning Opportunities**, detailing select existing Federal data training and courses.

Development Process

This Catalog is intended to help agencies develop competencies for managing data as a strategic asset and making data-driven decisions. The 2020 Action Plan specifically tasked the General Services Administration (GSA) with leading Action 13. A focus group of data experts from across 9 Federal agencies was convened to determine the significant roles in the data ecosystem within the Federal government, list skills affiliated with these key functions, and create the Catalog.[5]

The Catalog was released in multiple stages to allow useful content to be available as soon as it was ready rather than in a single, delayed release. The contents of the Catalog have initially been available to the Chief Data Officers (CDOs) of each agency as well as other relevant Federal communities as it was being built. This allowed Federal stakeholder input to be incorporated early in its design. Non-Federal input was also obtained and included prior to its final release. Demand for data skills identified in other parts of the FDS, including Action 4, have also informed the priorities for how the Catalog was built. The Catalog will be transferred to the CDO Council after December 31, 2020.

---

[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." Please refer to OMB M-19-23 and OMB Circular A-11 for more information on this requirement.

[4] https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/

[5] The focus group convened to create the Federal Data Lifecycle is described in Appendix A.

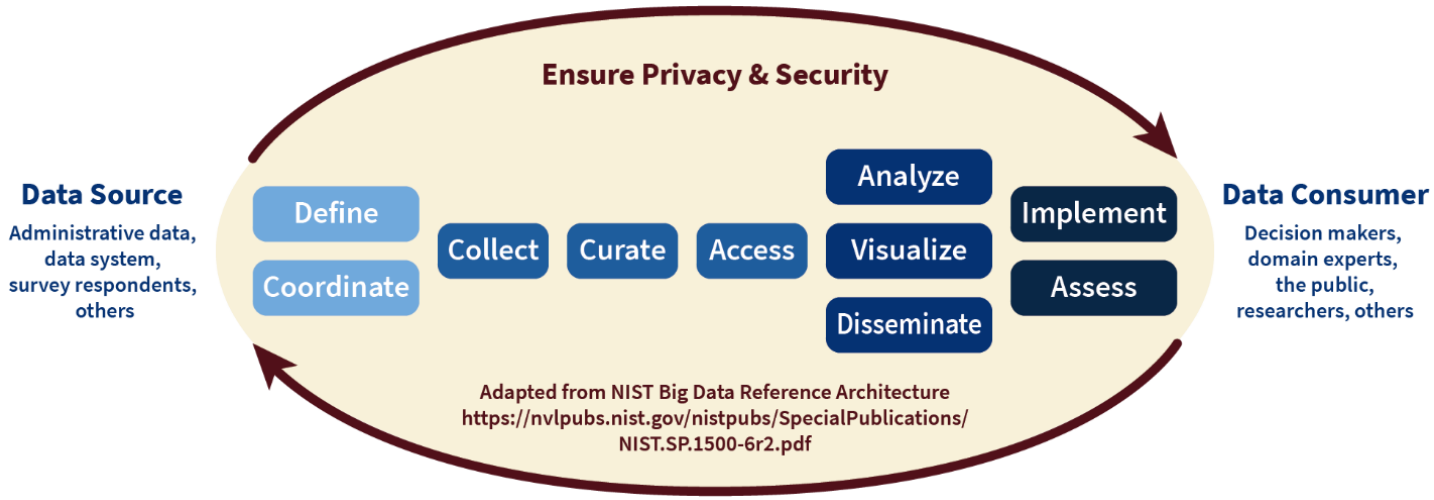
Federal Data Lifecycle and Associated Skills

Federal Data Lifecycle

This Lifecycle offers a common vocabulary for agencies to understand and evaluate data roles and skills. The Federal Data Lifecycle adapts the NIST Big Data Reference Architecture[6] and links subsequent data roles with the established FDS practices. The framework below lays out a variety of roles for those involved in data and provides the conventional responsibilities for each role. These include data-focused roles[7] and others in the organization such as leadership and domain experts. This framework can be helpful with assessing staff data skills and creating training programs to bridge any gaps.

The required data skills, whether technical skills associated with analysis, 'softer' skills associated with communication and coalition building, or otherwise are given for each of the roles beginning on page 7. FDS practices are related to many of these data roles, and a crosswalk of FDS Practices and the Federal Data Lifecycle provides further details of this connection in Appendix B.

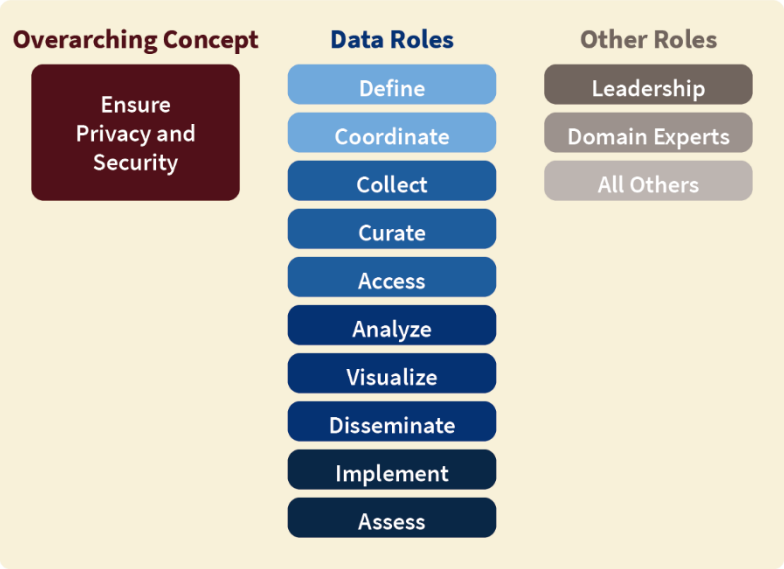
Figure 1: Federal Data Lifecycle



[6] 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.

[7] Agency personnel in data-focused roles as well as staff that access data 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 Federal 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 roles



Overarching Concept

- **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

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, validate, 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 to address problems with data collection/measurement, processing, and analyses

Other Roles

- **Leadership:** Cultivate and support a culture of enterprise-wide demand and use of data to maximize outcomes
- **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
- **All Others:** Each staff member values the use of data in their day-to-day work

Select Statutory authority and policy guidance related to data

- The Federal Records Act of 1950
- Privacy Act of 1974
- Paperwork Reduction Act (PRA) of 1995
- Information Quality Act of 2001
- Federal Information Security Modernization Act of 2014 (FISMA) Public Law 113–283 (2014)
- Geospatial Data Act of 2018
- Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) – Public Law 115–435 (2019)
    - TITLE I—FEDERAL EVIDENCE–BUILDING ACTIVITIES
    - TITLE II—OPEN GOVERNMENT DATA ACT
    - TITLE III—CONFIDENTIAL INFORMATION PROTECTION AND STATISTICAL EFFICIENCY
- OMB Memorandum M–19–15 Improving Implementation of the Information Quality Act issued on April 24, 2019. https://www.whitehouse.gov/wp-content/uploads/2019/04/M-19-15.pdf
- OMB Memorandum M–19–21 Transition of Electronic Records issued on June 28, 2019. https://www.whitehouse.gov/wp-content/uploads/2019/08/M-19-21-new-2.pdf
- OMB Memorandum M–19–23 Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance issued on July 10, 2019. https://www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf

Roles and Skills

Overarching Concept and Skills

Overarching Concept	Skills
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 data-focused roles as well as staff that access data need to have a basic understanding of their agency's cybersecurity and privacy policies that govern data as well as data ethics principles, and need to work closely with their agency experts in cybersecurity, privacy, and ethics throughout the Federal Data Lifecycle

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	• Communication • Knowledge of requirements for information collection review approvals including data requirements and business rules • Performance metrics • Planning • Problem solving • Strategic planning • Strategic thinking • Teamwork *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	• Business acumen • Coalition building • Communication • Federal enterprise architecture framework (FEAF) Data modelling • Horizon scanning • Knowledge of data archiving and preservation standards • Knowledge of data standards • Knowledge of novel data sources and collection methods • Knowledge of requirements for information collection review approvals • Knowledge of the budget process • Planning • Problem solving • Relationship building • Strategic thinking *Domain Expert input is important at this step*

Data Role	Skills
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 • Communication • Developing MOUs and Interagency Agreements • Frame development, questionnaire/instrument development • Interpersonal • Knowledge of Federal Data Collection Authorities • Knowledge of requirements for information collection review approvals • Real-time or near real-time collection methods • Reviewing existing data for potential new uses • Sampling design
Curate: Organize, refine, validate, 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	• Big data technologies • Data storage and preservation • Data warehouse/architecture design, development, construction, and maintenance • Develop scalable extract, transform, and load (ETL) processes • Information quality control and quality assurance methods and techniques • Knowledge of operating systems • Metadata skills including knowledge of the enterprise metadata standards and ability to produce unambiguous, useful metadata • Programming tools • Statistics, data checking, and internal controls on data reporting and quality, consistency, and logic edits • Techniques to understand and clean raw and unstructured data (data wrangling) • Understand and incorporate stakeholder needs and requirements *Domain Expert input is important at this step*

Data Role	Skills
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	• Business acumen • Communication • Data modeling and transformation • Data sharing policies • Database authentication methods • Disclosure risk limitation • File formats • Fundamental computer forensics • Intellectual property rights • Metadata repository development and maintenance • Problem solving • Tiered access to data methods • Understand stakeholders needs and requirements
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	• Data conditioning and data mining • Generate dashboards, charts, and graphs • Geospatial analysis • Machine learning algorithms, artificial intelligence, natural language processing and rapid process automation • Math and statistics • Predictive modeling and projections • Research design • Strategic communication • Trend or pattern analysis • Understand metadata • Specialized expertise in the application of data, such as for statistical production or conducting program evaluation • Knowledge and understanding on applying the appropriate and relevant analysis methods for the questions posed *Domain Expert input is important at this step*
Visualize: Present data insights for consumption by leaders and stakeholders ***Practices*** 6. Convey insights from data	• Communication • Data storytelling • Data visualization (including 508 compliance) • Programming or tools to create static or interactive reports, graphics, and dashboards • Understand stakeholder needs and requirements • User experience design • Web design and HTML • Sophisticated understanding of what data does and does not infer *Domain Expert input is important at this step*

Data Role	Skills
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	• Accessibility standards (508 compliance) • Communication • Companion materials creation • Data formats and API technology • Intellectual property rights • Negotiation • Relationship building • Understand stakeholder needs and requirements
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	• Coalition building and advocate for data use • Communication • Data management • Problem solving • Quality management • Scientific standards, theories, measurement, testing, and evaluation procedures • Strategic and innovative thinking

Other Roles and Skills

Other Roles	Skills
Leadership Cultivate and support a culture of enterprise-wide harmonization of data-related activities to maximize outcomes	• Communication • Data science process • Data security • Data sets, data priorities, and resources • Data use • Data-based decision making • Quantitative and qualitative data types • Relationship building • Strategic planning • Strategic thinking
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	• Communication • Data collection and use • Data Lifecycle • Knowledge of the possible data uses, context, and limitations • Mission-related subject matter expertise including policies and regulations • Operations and procedures • Relationship building • Strategic thinking • Understand stakeholder needs and requirements
All Others Each staff member values the use of data in their day-to-day work	• Basic terminology • Data risks or limitations • Data use and applications • Reading charts and graphs

Federal Learning Opportunities

The final section of this Catalog showcases representative Federal data training and courses and is not meant to be exhaustive. Federal agencies have access to a variety of learning providers and offerings (including those in the private sector) to increase data skills within their workforce. The following list highlights some of these available federal resources.

The Federal learning opportunities are divided into two categories. The first category includes programs that are accessible to all Federal employees as well as the public while the latter consists of training that is available only to select Federal employees. For each resource, the website, intended audience, and the data role(s) covered are described. Moreover, a brief description of each offering is provided.

This initial Catalog incorporates information about the significant roles in the data ecosystem into its organization and specifies skills and responsibilities necessary for these key roles. Using this common vocabulary, agencies can more easily align their data training needs to existing learning providers, programs, courses, and certifications. Additionally, this Catalog can provide agencies with possible government resources and curricula to assist with increasing their data skills capacity.

Resources available to the Public (including all Federal Employees)

Census Academy

Website: https://www.census.gov/data/academy.html

Audience: Federal employees interested in using Census data

Data Roles Covered: Varies by offering

The Census Academy is a publicly available training hub that is hosted by the U.S. Census Bureau. This free training site has been divided by its features and by topic as seen in the image below; features include the following:

- **Data Gems:** a series of "how-to" videos available for data users who are looking for an easy and quick way to enhance their knowledge of Census data.
- **Webinars:** a series of recorded and upcoming webinars on data topics related to the U.S. Census Bureau.
- **Courses:** online courses on data skills needed to work with Census data by topic, skill level, and audience/roles.
- **Resources:** topical one-pagers about the programs and surveys offered by the Census Bureau for survey respondents and users of Census data, as well as videos and visualizations.

You can request a data training by emailing census.askdata@census.gov or subscribe to the Census Academy's weekly updates by emailing census.academy@census.gov.

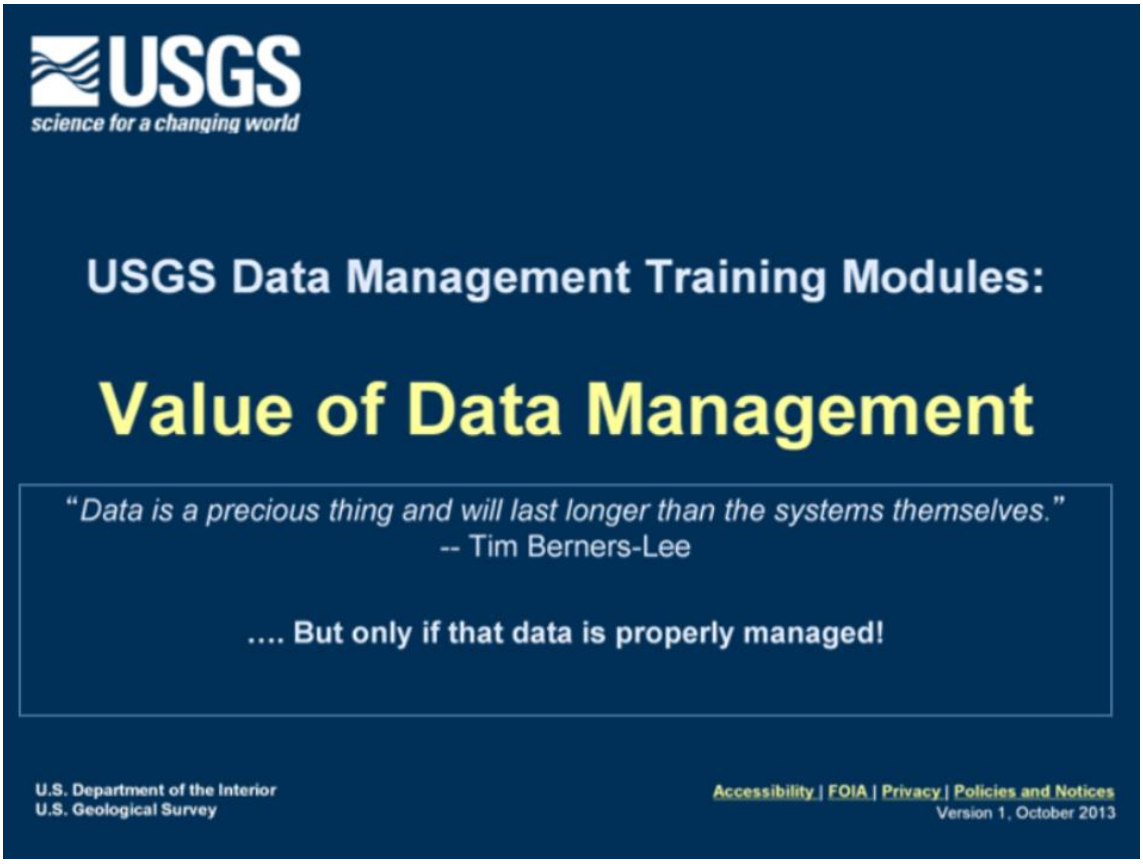
US Geological Survey Data Management Trainings

Website: https://www.usgs.gov/products/data-and-tools/data-management/training

Audience: Researchers, Data stewards, Data managers, and scientists inside and outside USGS

Data Roles Covered: Curate, Access, Disseminate, Implement & Assess

The USGS developed training modules about data management. The first section of trainings has a broad audience and covers science data management concepts and best practices. These trainings are geared towards researchers, data stewards, and the public. Some of the topics include an introduction to data management concepts, how to develop a data management plan for a research project, and a course on metadata for research projects. The second section of trainings covers tools and requirements for sharing USGS data and is geared towards USGS scientists and data managers. These training modules are available as video on-demand and include slides and supplementary materials.



Resources available to select Federal Employees

Chief Information Officer (CIO) Council Data Science Training Program

Website: https://www.cio.gov/programs-and-events/data-science-training-program/

Audience: Nominated Employees in CFO Act Agencies

Data Roles Covered: Curate, Analyze, and Visualize

The Data Science Training Program assists agencies in filling a critical skills gap by training current Federal employees in emerging data science skills. In support of the Federal Data Strategy and the President's Management Agenda, the Data Science Training Program fills a critical competencies skills gap for Federal agencies by training the Federal workforce in core, data science skills. This program trains 60 Federal employees. Participants are required to attend eight hours of training every week, and the training slots are available to the 24 CFO Act agencies.

The training program lasts six months and occurs in two phases. First, the participants take a series of online courses delivered through LinkedIn Learning in the fundamentals of data science. This curriculum covers basic data science tools, including Python, R, Tableau, and PowerBI. It also introduces fundamental data science concepts, including data visualization, machine learning, and data querying. The second phase incorporates an online capstone project. The curriculum can be found in Appendix C.

Evaluation and Evidence Training Series

Website: Training opportunities are available at: https://community.max.gov/x/n3l5fw (MAX login required) and https://oes.gsa.gov/events/

Audience: Limited only to Federal, Executive-Branch Staff; Geared towards Evidence Act Officials (Evaluation Officers, Statistical Officials, and Chief Data Officers); Agency staff and leaders who work on evidence and evaluation activities (e.g., evaluation, performance, regulatory, data, or statistical staff); and Program staff who may work with evaluation staff or use evaluation products

Data Roles Covered: Define, Coordinate, Collect, Analyze, Visualize, Disseminate, Implement and Assess

The Office of Management and Budget's Evidence Team and the General Services Administration's Office of Evaluation Sciences are continuing their partnership to offer the Evaluation and Evidence Training Series in FY 2021. More than 1,000 attendees participated in their FY 2020 workshops, and they will continue to provide these monthly trainings for Evidence Act Officials, their staff and agency partners, and other interested Federal staff in the coming fiscal year. The trainings cover topics related to the Evaluation process including an introductory session about Evaluation. Appendix D lists the upcoming course schedule.

The HHS Data Science CoLab

Website: https://www.hhs.gov/cto/initiatives/data-science-colab/index.html

Audience: Department of Health and Human Services employees

Data Roles Covered: Analyze and Visualize

The Data Science CoLab aims to build a community of HHS employees as skilled data scientists who continuously learn from experts and from each other. The learning cohorts bring in experts from the data science field to lead an eight-week intensive educational program. Participants will gain hands-on data science skills to address real problems in the HHS. CoLab members will engage in a cycle of receiving high quality education, solving real problems, and helping others solve theirs. Two virtual modules are available, and applicants may apply for one or both modules.

1. Introduction to R and Visualization includes an introduction to data science.

2. Machine Learning and Text Mining in R includes advanced data science topics as well as the development and refinement of a capstone analysis project.

The Treasury Executive Institute (TEI)

Website: https://home.tei.treasury.gov/about

Audience: SES, GS-15, and GS-14 employees from across over 40 different federal agencies (Appendix E)

Data Roles Covered: Define, Analyze, Visualize, Disseminate, Implement and Assess

In its offerings the Treasury Executive Institute (TEI) provides unique data-related courses every year. In FY20, TEI offered data literacy courses including a bootcamp designed for executives and leaders who need foundational knowledge and skills to create and lead a data-informed organization. Bootcamp topics included Artificial Intelligence, Machine Learning, and Business Intelligence & Analytics. A more detailed curriculum of past data courses is in Appendix F.

Appendix A

Development of the Federal Data Lifecycle

The GSA convened a focus group consisting of leaders from the Federal CDO, Statistical Officer and Chief Human Capital Officer communities to determine the significant roles in the data ecosystem within the Federal government and list skills and responsibilities associated with these key functions. The Federal Data Lifecycle was generated by adapting the NIST Big Data Reference Architecture which is a vendor-neutral, technology- and infrastructure-agnostic conceptual model. The Federal Data Lifecycle and corresponding skills framework form the backbone for the Catalog.

Focus Group Members	**Role & Organization**
Anne Levine	Deputy CDO, Federal Communications Commission
Brian Moyer	Director of National Center for Health Statistics & HHS Statistical Official
Chris Haffer	CDO, Equal Employment Opportunity Commission
Keith Krut	HR Division, National Aeronautics and Space Administration
Lance Harris	Statistical Official, United States Office of Personnel Management
Lucas Hitt	U.S. Bureau of Economic Analysis
Megan Dreher	Performance Team, Office of Management and Budget
Megan McNeely	Communications Directorate, U.S. Census Bureau
Sharon Boivin	Chief of Staff, Office of the CDO, U.S. Department of Education

Appendix B

Crosswalk of FDS Practices and Federal Data Lifecycle Roles

	Overarching Concept	Data Lifecycle								
	Ensure Privacy & Security	Define	Coordinate	Collect	Curate	Access	Analyze	Visualize	Disseminate	Implement & Assess
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.			✓							

	Overarching Concept	Data Lifecycle								
	Ensure Privacy & Security	Define	Coordinate	Collect	Curate	Access	Analyze	Visualize	Disseminate	Implement & Assess
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	✓									
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.			✓		✓					

	Overarching Concept	Data Lifecycle							
	Ensure Privacy & Security	Define	Coordinate	Collect	Curate	Access	Analyze	Visualize	Disseminate	Implement & Assess
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.			✓							
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			✓		✓					
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.				✓		✓				

	Overarching Concept	*Data Lifecycle*								
	Ensure Privacy & Security	***Define***	***Coordinate***	***Collect***	***Curate***	***Access***	***Analyze***	***Visualize***	***Disseminate***	***Implement & Assess***
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.		✓								
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.						✓			✓	

	Overarching Concept	Data Lifecycle								
	Ensure Privacy & Security	Define	Coordinate	Collect	Curate	Access	Analyze	Visualize	Disseminate	Implement & Assess
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.							✓			

Appendix C

Chief Information Officer (CIO) Council Data Science Training Program Curriculum

Topic/Courses	Learning Objectives
Data Science Foundations: Fundamentals -What is data science? -Applications of data science -Common data science tools -Foundational data concepts	-Assess the skills required for a career in data science -Evaluate different sources of data, including metrics and APIs -Explore data through graphs and statistics -Discover how data scientists use programming languages such as R, Python, and SQL -Assess the role of mathematics, such as algebra, in data science -Assess the role of applied statistics, such as confidence intervals, in data science -Assess the role of machine learning, such as artificial neural networks, in data science -Define the components of effective data visualization
Data Fluency: Exploring and Describing Data -How to use common data science tools -Use cases for data science -Developing data products -Best practices in data science	-The ROI of data fluency -Data ethics -Preparing data -Assessing the quality of data -Visualizing data with bar, pie, and line charts -Describing variability with the variance and standard deviation -Describing associations with correlations
Learning Data Visualization: -Fundamentals of visualization -Fundamentals of storytelling -Designing a presentation -How to present and stay on point -Improving the value of your time	-Channeling your audience -Understanding your data -Determining the information hierarchy -Sketching and wire-framing your ideas -Defining your narrative -Using typography, color, contrast, and shape to convey meaning -Making your visualization interactive
Design Thinking: -Understanding the Process -Data Intelligence	-Agile, lean, and design thinking -Preparing to sell design thinking to your organization -Finding the real problem -Ideation -Prototyping -Correcting course -Offshoring and outsourcing

Topic/Courses	Learning Objectives
	-Getting past organizational inertia and silos -Tracking your success -Design intelligence best practices -Avoiding data paralysis -Applying data intelligence to design -Partners in design intelligence -Visualizing data intelligence
Artificial Intelligence Foundations: Machine Learning	-Describe how to work with data -Apply machine learning principles -Distinguish different types of machine learning -Identify problems that use machine learning -Create decision trees -Explain how to select the best algorithm
Statistics Foundations: 2	-Apply the Central Limit Theorem to find the average of sample means -Analyze random samples during hypothesis testing -Assess individual situations to determine whether a one-tailed or two-tailed test is necessary -Define acceptance sampling
Learning Data Analytics	-Define data analysis and data analyst -List roles in data analysis -Explain data fields and types -Define syntax -Explain how to interpret existing data -Describe data best practices -Repurpose data -Create a data dictionary -Compare and contrast linking versus embedding charts and data -Build pivot charts with slicers
Python Essential Training -Introduction to Python	-Python anatomy -Types and values -Conditionals and operators -Building loops -Defining functions -Python data structures: lists, tuples, sets, and more -Creating classes -Handling exceptions -Working with strings -File input/output (I/O) -Creating modules

Topic/Courses	Learning Objectives
	-Integrating a database with Python db-api
Power BI Essential Training	-Working with cloud-based and on-premises data sources -Creating reports with data visualizations -Modifying existing reports -Creating and managing data dashboards - Asking questions with Power BI Q&A -Enabling Microsoft Cortana for Power BI -Sharing Power BI datasets, dashboards, reports, and workbooks -Using Power BI Mobile and Power BI Desktop
Data Science Foundations: Data Mining	-Data mining using R, Python, Orange, and RapidMiner -Data reduction -Data clustering -Anomaly detection -Association analysis -Regression analysis -Sequence mining -Text mining
Python: Data Analysis	-Writing and running Python in iPython -Using Python lists and dictionaries -Creating NumPy arrays -Indexing and slicing in NumPy -Downloading and parsing data files into NumPy and Pandas -Using multilevel series in Pandas -Aggregating data in Pandas
Learning Logistic Regression in R and Excel	-Recognizing the problems with ordinary regression on a binary outcome -Quantifying errors in forecasts -Managing different slopes -Forecasting odds instead of probabilities -Limiting probabilities on the upside and downside -Working with exponents and bases -Predicting the logit -Working with original data and coefficients -Establishing the Log Likelihood -Interpreting -2LL or deviance -Establishing a data frame with XLGetRange -Using the R functions mlogit or and glm -Understanding long versus wide shapes in data sets
Designing a Data Visualization	-Working with the data

Topic/Courses	Learning Objectives
	-Sketching and wireframing your design -Roughing out the visual design components -Manually creating the design in Illustrator -Using Illustrator scripting to improve accuracy, speed, and repeatability -Designing callout boxes, legends, labels, and more
Applied Machine Learning: Foundations	-What is machine learning (ML)? -ML vs. deep learning vs. AI -Handling common challenges in ML -Plotting continuous features -Continuous and categorical data cleaning -Measuring success -Overfitting and underfitting -Tuning hyperparameters -Evaluating a model
Applied Machine Learning: Algorithms	-Models vs. algorithms -Cleaning continuous and categorical variables -Tuning hyperparameters -Pros and cons of logistic regression -Fitting a support vector machines model -When to consider using a multilayer perceptron model -Using the random forest algorithm -Fitting a basic boosting model
ArcGIS Pro 2.4 Essential Training	-Choosing a template -Creating an ArcGIS project -Connecting to data -Rendering text files as points -Adding ECW and DEM data -Working with different layer types -Converting 2D maps into 3D scenes -Creating feature classes in a geodatabase -Exporting data -Creating heatmaps -Publishing to ArcGIS Online -Sharing layers -Creating project packages
Tableau Essential Training	-Managing data sources and visualizations -Managing Tableau worksheets and workbooks -Displaying source metadata -Creating custom calculations and fields -Analyzing data using statistical tools

Topic/Courses	Learning Objectives
	-Sorting and filtering Tableau data -Defining groups and sets -Creating and pivoting crosstabs -Formatting Tableau visualizations -Creating basic charts -Annotating and formatting charts -Mapping geographic data -Creating dashboards -Methods for defining an URL-based action
Data Visualization: Best Practices	-What charts and graphs work best for different types of data -Putting data into visual and textual context to ensure it is accurate -Visualizing data that doesn't lend itself to imagery -Adding visual appeal without sacrificing accuracy -Using the Adobe Illustrator graphing tools -Avoiding common data visualization mistakes
Creating Interactive Presentations with Shiny and R	-Explain the benefits associated with RStudio. -List the options for creating HTML R Markdown presentations. -Use RPubs to publish and share presentations through Twitter, Google+, and Facebook. -Prepare and share an interactive application with Shiny. -Name the three different layout options available with the Shiny app. -Compare different Shinyapps.io accounts to determine which fits your needs.
Database Foundations: Creating and Manipulating Data	-Storing dates, times, and text -Converting data types -Creating tables -Writing T-SQL commands -Selecting records with queries -Combining and sorting data -Creating views -Creating stored procedures and functions -Inserting and updating data in a table -Deleting records and tables
SQL: Data Reporting and Analysis	-Retrieving data with SELECT statements -Filtering and sorting your results -Transforming results with built-in SQL functions -Grouping SQL results -Merging data from multiple tables -Using variables, functions, and procedures
Designing a Presentation	-Outlining your topic -Creating a storyboard

Topic/Courses	Learning Objectives
	-Developing your design ideas with moodboards -Exploring slide typography -Using colors, words, and images effectively -Using charts and graphs -Effectively using animations and transitions -Beyond the presentation

Appendix D

Select FY2021 Evaluation and Evidence Training Series Courses

Please refer to the site with more information and registration links:
https://community.max.gov/x/n3l5fw (MAX login required)

Evaluation 101
Introduction to evaluation as a research method: what it is, what questions it can answer, and how it can help agencies better understand their programs, policies, and operations.

Understanding Null Results
Discuss and dispel misconceptions about evaluations that show no evidence of impact or provide evidence of no impact, and highlight uses for null results.

Evidence and Strategic Planning
Concrete strategies and approaches that agencies can use to maximize the benefit of evidence use in strategic planning.

Introduction to Quasi-Experimental Design
Introduction to impact evaluation designs, other than randomized controlled trials, to understand program impacts.

Using Evidence to Inform Agency Priorities
Highlight ways that agencies can use evidence to inform agency priorities, both mission and operational.

Introduction to Formative and Process Evaluations
Introduction to formative and process evaluations: common tools and methodologies used in these types of evaluations.

Appendix E

The Treasury Executive Institute (TEI) Members FY20

Department of Commerce — International Trade Administration
Department of Commerce — National Oceanic and Atmospheric Administration — National Environmental Satellite, Data, and Information Service
Department of Commerce — National Oceanic and Atmospheric Administration — National Ocean Service
Department of Commerce — National Oceanic and Atmospheric Administration —SO

Department of Defense — U.S. Air Force

Department of Education

Department of Energy

Department of Health and Human Services – Agency for Healthcare Research and Quality

Department of Homeland Security — Headquarters
Department of Homeland Security — Federal Emergency Management Agency
Department of Homeland Security — Immigration and Customs Enforcement
Department of Homeland Security — United States Citizenship and Immigration Service
Department of Homeland Security — United States Secret Service

Department of Housing and Urban Development

Department of Justice – Bureau of *Alcohol, Tobacco, Firearms* and Explosives
Department of Justice – Drug Enforcement Administration
Department of Justice – Executive Office for Immigration Review
Department of Justice – US Marshal Service

Department of Labor – Bureau of Labor Statistics
Department of Labor – Office of Inspector General

Department of Transportation – Federal Transit Administration

Department of Treasury – Bureau of Fiscal Service
Department of Treasury – Financial Crimes Enforcement Network
Department of Treasury – Internal Revenue Service
Department of Treasury – Internal Revenue Service Counsel
Department of Treasury – Office of the Comptroller of the Currency
Department of Treasury – Office of Inspector General
Department of Treasury – Tax and Trade Bureau
Department of Treasury – Treasury Inspector General for Tax Administration
Department of Treasury – United States Mint

Independent Agencies

Environmental Protection Agency

Export–Import Bank of the United States – Office of Inspector General

Federal Housing Finance Agency

General Services Administration

Millennium Challenge Corporation

National Archives and Records Administration

National Transportation Safety Board

Office of Government Ethics

Office of Management and Budget

Small Business Administration

Appendix F

The Treasury Executive Institute (TEI) Select FY20 Data Courses

Data Science and Why it Matters to Leaders

The Data Society provides data analytics programs and data services that transform the way government staff learns data science and has been recognized by FedScoop and Forbes for the innovative way that they develop and deliver data science trainings to unleash the workforce's potential to solve their organization's toughest problems and prepare for the future. In this interactive workshop, leaders will learn: what data science is and why it's important; data storytelling: how to use data to present and communicate results clearly to stakeholders in order to encourage change; and how to select the appropriate data visualization for a scenario.

Data Analytics for Government Leaders

This session will equip government leaders with data literacy skills. Understanding advanced data analytics and how to champion analytics in an organization will be covered. Learn how to create conditions for success and apply data analytics in new areas at your agency. Further, learn how to ensure appropriate data analytics safeguards for privacy, inaccuracy, and bias are implemented.

Data Literacy Bootcamp: Session 4-Communicating Data

This series is designed for executives and leaders who need foundational knowledge and skills to create and lead a data-informed organization. This series uses zoom. You do not have to attend all sessions in the series; the sessions do build upon each other.

This session will cover:
- Converting data to information
- Data visualization tips and techniques
- Storytelling with data
- Leading a data-informed culture

Data Literacy Bootcamp: Session 5-Advanced Data Design

This series is designed for executives and leaders who need foundational knowledge and skills to create and lead a data-informed organization. This series uses zoom. You do not have to attend all sessions in the series; the sessions do build upon each other.

This session will cover:
- How to design a business experiments
- Fundamentals of A/B testing

Data Literacy Bootcamp: Session 6-Advanced Data Analysis

This series is designed for executives and leaders who need foundational knowledge and skills to create and lead a data-informed organization. This series uses zoom. You do not have to attend all sessions in the series; the sessions do build upon each other.

This session will cover:
• Useful statistics—testing for correlation and significance
• Probability testing—understanding the basics behind statistics and predictive analytics
• Hypothesis testing—when data is used for researching alternatives
• Linear regression—identifying the critical drivers

Data Literacy Bootcamp: Session 7-Data Trends for Executives—Leading in the Digital Age

Learn what you need to know about Artificial Intelligence, Machine Learning, and Business Intelligence & Analytics. This session will also explore how you can foster data literacy throughout your organization to build a data-informed organization.

This is the last session in the series designed for executives and leaders who need foundational knowledge and skills to create and lead a data-informed organization. You did not have to attend the previous sessions in the series; however, the sessions do build upon each other. This series uses zoom.