Together, Advancing Data-Driven Government

Guiding Principles to advance culture which data:

• **Drive Decisions** – data is integrated into policymaking and other processes to inform direction of mission.

• **Fosters Collaboration** – data is a mechanism within agencies and across the government to breakdown silos and bring diverse thoughts and perspectives across mission priorities.

• **Informs and Engages the Public** – data is a government resource made accessible to civic and tech communities for spurring innovation in other sectors.

• **Increases Capabilities** – data is an enterprise asset which is responsibly managed, accessible, and leveraged to improve efficiency and effective.

• **Facilitates Product and Service Development** – data is used to facilitate the delivery of products and services to internal and/or external users.
Data-Driven Culture

Mission
Responsibly unleash the power of data to benefit all Americans

Challenge
To meet our nation’s greatest economic and social challenges, the federal government must leverage their data as a mechanism for engaging with other agencies and the public.

Agencies
Maximizing the value of data and ensuring it is responsibly accessed in the most meaningful ways.
- Mission imperatives – Identify data needs against mission priorities (i.e., maturity model)
- Build data capacity – Adopt data science best practices to drive innovative culture (i.e., CDO role, data resource management, and the Interagency Open Data Working Group)
- Peer-to-peer learning – Interagency Open Data Working Group listserv & biweekly meetings
- Public-Private Collaborations – Federal agencies partnering formally and informally with data leaders in private, non-profit, and academic sectors or other governments

EOP
Empower agencies with the people, processes and practices necessary to build their data capacity.
- CIO Council – Principal interagency forum for improving practices in the design, modernization, use, sharing, and performance of information resources; includes Innovation Committee
- Open Data CAP Goal – Federal government-wide objectives, strategies and major metrics to increase data interoperability, openness, and discoverability
- Privacy Council and the OMB Office of Information and Regulatory Affairs
- Data Cabinet – Community of Practice to advance adoption of data science best practices
- NSTC Data Science Interagency Working Group – Senior data leaders prioritize the critical science and technology areas of investment to build capacity
The Data Cabinet

PURPOSE
Serve as the principal community of practice for data professionals from across federal government to share proven practices among agencies and critical insights into what’s working “on the front lines”, the Data Cabinet plays an essential role in advancing data-driven management, maximizing the use of the data, and enhancing problem-solving throughout government.

OBJECTIVES
Bring together authorities, case studies, and best practices to scale data talent capacity in the federal government through recruitment, training, and hiring best practices. As well as guide Agencies in adoption of data policies and tools necessary to ensure data is responsibly gathered, processed, leveraged, made discoverable accessible and reusable in a timely fashion.

• Information sharing and collaboration on issues of mutual importance.
• Pursue solutions at the agency level, including pilot programs and ‘proof of concept’ experiments that apply data science techniques in innovative and scalable ways.
• Inform the development and execution of federal data standards and policies.
• Identifying and scaling promising techniques, technologies, and solutions by facilitating adoption within multiple agencies.
### Building Federal Data Capacity — Data Maturity Model

<table>
<thead>
<tr>
<th>Analytics Capability</th>
<th>Data Culture</th>
<th>Data Elements</th>
<th>Data Personnel</th>
<th>Systems/Technology</th>
<th>Data Governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Capability</td>
<td>Low Capability</td>
<td>Low Capability</td>
<td>Low Capability</td>
<td>Low Capability</td>
<td>Low Capability</td>
</tr>
<tr>
<td>High Capability</td>
<td>High Capability</td>
<td>High Capability</td>
<td>High Capability</td>
<td>High Capability</td>
<td>High Capability</td>
</tr>
</tbody>
</table>

#### Analytics Capability

**Low Capability**
- Siloed manual reporting on aggregate-level data
- Data typically used in a reactive manner, viewed as necessary for compliance
- Only aggregate-level data is collected and reported
- Data responsibilities are secondary for most personnel
- Proprietary siloed systems maintained within the agency
- Bureau-level collaboration, data ownership and stewardship

**High Capability**
- Automated reporting with some advanced analytics
- Data used proactively to drive decision making; data regarded as an asset by leadership and program managers
- Agency-wide data models, dictionaries and taxonomies are used to standardize data collection and transactional-level data is captured
- Established roles and development paths for data professionals
- Open source technologies and/or agency-wide and cross-functional use of systems

#### Data Culture

**Low Capability**
- Agencies participate in inter-agency communities that build data-focused culture government-wide, employees of all levels regard data as an important tool

**High Capability**
- Cross-functional data models, data standards, dictionaries, and taxonomies used within and across agencies.

#### Data Elements

**Low Capability**
- Only aggregate-level data is collected and reported

**High Capability**
- Agency-wide data models, dictionaries and taxonomies are used to standardize data collection and transactional-level data is captured

#### Data Personnel

**Low Capability**
- Data responsibilities are secondary for most personnel

**High Capability**
- Data professionals are partnered with subject matter experts in multi-disciplinary teams to solve agency mission and operational challenges

#### Systems/Technology

**Low Capability**
- Proprietary siloed systems maintained within the agency

**High Capability**
- Open source technologies and/or agency-wide and cross-functional use of systems

- Multi-agency or cross-functional use of centralized systems
Data Science as a Discipline

Data needs the same dedicated focus that technology has enjoyed if we truly want it to achieve its full potential. Now is the time to embrace the next generation in the data journey that must be focused on how to best manage that abundance of data and put it to work for the business.

-- Cathryne Clay Doss, First Appointed Chief Data Officer (CDO) at Capital One, quoted in Aiken and Gorman (2013).

Building Government Career Roadmap

- What are the core skills for data professionals?
- How to develop core Data Science skills?

Defining Data Science Community

- CDO Role & Responsibilities
- Data Scientist roles and responsibilities
- Other Data roles/job series—Data analysts, Data stewards
- Certification to Professionalize

Hiring and Recruiting

- Hiring Playbook
- Organizational structure
- Onboarding (building data community within agency)
Data Cabinet – Data Value Proposition

The value of data lies in the Agencies ability to use it. Raw data adds value as it moves up the pyramid to become the cornerstone of high-level, high value decisions by leadership.

Valuable Data comes from three pillars:
- Data Management leads to high quality data.
- Data Governance ensure the entire agency understand the value and handling of data
- Data In Action is the Value derived from the Agencies efforts
Data Resource Management Playbook

Objective:

• Build an initial repository and a shared place for further development of sustainable, collaborative data governance; repeatable data management processes, and best-practices for releasing data power across the Federal space.

<table>
<thead>
<tr>
<th>Roles</th>
<th>DRM Aspects</th>
<th>Questions</th>
<th>Resources</th>
<th>Impacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive</td>
<td>Data Governance</td>
<td>What I need to know</td>
<td>• Existing resources</td>
<td></td>
</tr>
<tr>
<td>Data Steward</td>
<td>Data Management</td>
<td>What I need to start</td>
<td>• DRM tools</td>
<td></td>
</tr>
<tr>
<td>Data Professional</td>
<td>Data-in-Action</td>
<td>What I need to do</td>
<td>• Data products/reports</td>
<td></td>
</tr>
<tr>
<td>Implementer</td>
<td></td>
<td>What I can expect</td>
<td>• Repeatable processes</td>
<td></td>
</tr>
</tbody>
</table>

(48-cell matrix)
Exemplars
D2D supports GSA across the full data lifecycle

**Area**
- Any software\(^1\) used to create an application\(^2\) intended to collect and restore data

**D2D Involvement**
- Interested in applying backend storage in new application development in terms of:
  - What data the application is storing
  - How data may be used for reporting

**Level of Involvement**
- Low
- Medium
- High

**GSA-specific examples for context**
- Alfresco, Salesforce, Oracle
- CHRIS, Pegasis, GREX, FRPP, GSA Advantage
- Financial reports, human capital reports, construction project status, fleet reports
- Labor projections, construction project status on cost information

**Entering data into an application**
- Includes tapping into a single database to create an executive level report used to make specific functional decisions\(^3\)

**Report data for specific functional reasons**
- Multiple data sets combined for cross-functional reporting in a logical data warehouse\(^4\)

**Data Transformation**
- Process by which transactional data transformed to warehouse data

**Access data from origin to be combined**
- Data displayed in a consumable, analytical way by individuals with knowledge of both technical & business line aspects

**Data Visualization & Analytics**
- Provide a centralized location where business line information can be stored/shared
- Coordinate a training plan, at various levels, on analytics tools
- Present data across common terminologies, visualizations, data usage, and governance

---

\(^1\) Alfresco, Salesforce, Oracle
\(^2\) CHRIS, Pegasis, GREX, FRPP, GSA Advantage
\(^3\) Financial reports, human capital reports, construction project status, fleet reports
\(^4\) Labor projections, construction project status on cost information
Multiple offices provide support for D2D efforts, including training and change management

<table>
<thead>
<tr>
<th>Data migration</th>
<th>CIO</th>
<th>CDO</th>
<th>Data Team</th>
<th>Security Team</th>
<th>HSSOs</th>
<th>Data Stewards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Support</td>
<td>Guidance and Approval</td>
<td>Execute dataset migrations</td>
<td>Security Assessment and Approval</td>
<td>Executive Support</td>
<td>Contribute requirements, define limitations, underlying issues</td>
<td></td>
</tr>
<tr>
<td>Executive Support</td>
<td>Guidance and Approval</td>
<td>Post finalized dashboards to D2D portal</td>
<td>Contribute security reviews</td>
<td></td>
<td>Prioritize, provide high-level requirements</td>
<td></td>
</tr>
<tr>
<td>Executive Support</td>
<td>Guidance and Approval</td>
<td>Test for quality; flag issues</td>
<td>Contribute security validation</td>
<td>Executive Support</td>
<td>Implement improvement plan</td>
<td></td>
</tr>
<tr>
<td>Guidance and Approval</td>
<td>Formulate options/guidance</td>
<td>Surface options, pros/cons</td>
<td>Contribute security assessments</td>
<td></td>
<td>Contribute requirements; user feedback</td>
<td></td>
</tr>
<tr>
<td>Guidance and Approval</td>
<td>Executive Support</td>
<td>Guidance and Approval</td>
<td>Contribute security perspective</td>
<td>Contribute requirements; user feedback</td>
<td>Contribute SSO perspectives</td>
<td></td>
</tr>
</tbody>
</table>

- **Final senior approval of work products**
- **Responsibility for on-going execution**
- **Senior-level consultative support**
- **Working level inputs, reviews, consultations**
D2D recognizes potential process-related risks and has set plans to offset and mitigate these problems

<table>
<thead>
<tr>
<th>Risk type</th>
<th>Risk description</th>
<th>Potential mitigations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schedule</td>
<td>• Project schedule slips due to increase in scope or other unforeseen events</td>
<td>• Weekly CDO status meetings, regular management meetings, close monitoring of task orders and deliverables</td>
</tr>
<tr>
<td>Costs &amp; Funding</td>
<td>• Insufficient funding to complete the project. Uncontrollable out-year costs for operation and maintenance (O&amp;M).</td>
<td>• Look for offsets; submission of budget requests; look for funding partners; request life-cycle funding for post-development O&amp;M; re-compete O&amp;M contract for out-years, if necessary.</td>
</tr>
<tr>
<td>Technology</td>
<td>• Requirements not sufficiently clear or misunderstood by the developers. • Technological advances may render D2D solutions as obsolete. • Understanding of manual vs. automated efforts, risk of accepting all data at the same level of quality</td>
<td>• Component-based architectural design allows for the addition and subtraction of components over time • Scheduled evaluations and market research to ensure we can deliver and manage the necessary technical capabilities through D2D</td>
</tr>
<tr>
<td>Ongoing Support</td>
<td>• Lack of resources to adequately support the new system and its users once it’s in production - a great behind-the-scenes platform that no one knows how to use</td>
<td>• GSA has assembled an integrated project team that possesses skills and experience with IT project management and planning that will assess the requirements for ongoing support and include that in our project plan.</td>
</tr>
<tr>
<td>Organizational Change Management</td>
<td>• Disruption to the business due to implementation of new technologies • Hesitance to release data, inaction despite data suggesting otherwise, ability to position as “normal work”, not as an additional workstream</td>
<td>• Regular engagement and communication with users to understand their requirements and demonstrate future capabilities within D2D. Targeted, comprehensive training programs to avoid any business disruptions. • Maintaining legacy tools until D2D meets demand</td>
</tr>
</tbody>
</table>
GSA is given the opportunity to fully incorporate D2D through active change management efforts by the team.

<table>
<thead>
<tr>
<th>Summary</th>
<th>Feature capabilities</th>
<th>Understand needs/current processes</th>
<th>Propose enhancements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agency engagement to level-set on key facts around D2D [e.g. costs, time to develop, dashboards]</td>
<td>Clear expression of D2D capabilities and demonstrations across GSA offices [e.g. contracting data sets, real estate trends visualizations]</td>
<td>Commitment to determine priority goals and needs in overall timeline of D2D progress</td>
<td>Opportunity to leverage &amp; build on existing agency information to make more insightful &amp; data-driven</td>
</tr>
</tbody>
</table>

**Examples**
- D2D is expensive
- D2D takes a lot of time to develop a dashboard
- Viewing D2D requires Tableau license
- Received feedback to incorporate crime, traffic, sustainable use into buildings dashboard
- Ability to do so versus necessary steps

**Questions for SteerCo**
- What are those misconceptions?
- How can we best convey all of D2D’s abilities?
- What are the target offices?
- What are two or three processes that are open and available for similar developments?

**Examples**
- D2D as a data:
  - Repository
  - Visualization tool
  - Collection/access point
- CXO SLAs:
  - Always existed with GSA
  - Review sessions yielded need for more granularity: now drill down data points/dashboards
The long process of data-driven decision making

How do you help the affected individuals understand and communicate the rationale?

Question/Hypothesis Identification
Why do you want the information? What will you do in the face of different results? How do you ensure sufficient buy-in? Is a governance process needed?

Is this a question that requires quantitative data to answer? Do you need to collect more data, or do you have enough to start?

Data Discovery & Collection
Does your workforce have the skills sets to analyze the data? How will you handle change management?

What kind of governance structure will you set up for data access? How will you secure it?

Analysis

Who are the decision maker(s) and are they comfortable with data?

Data Security & Storage

Implement

Who are the decision maker(s) and are they comfortable with data?