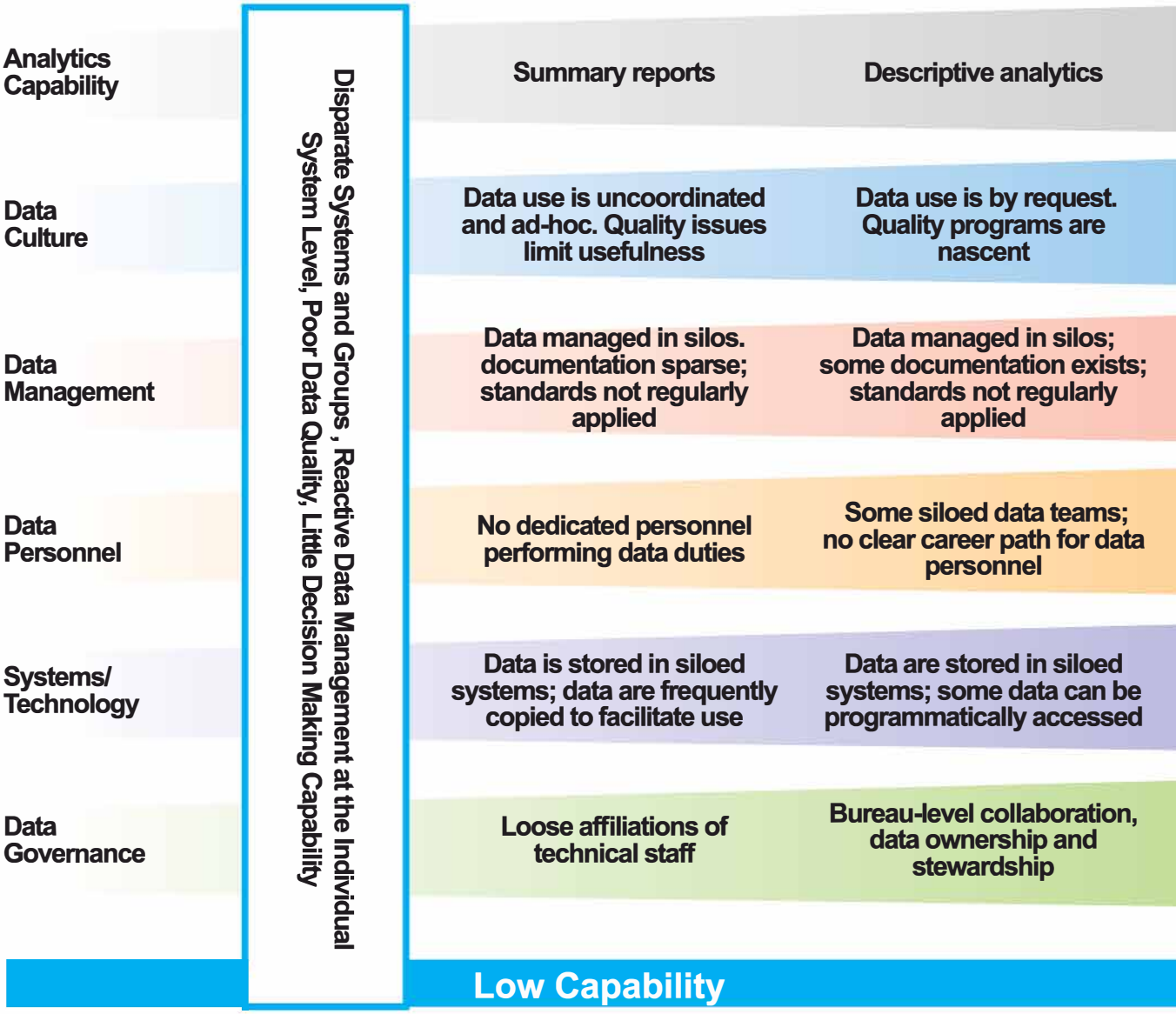
Data Cabinet

The Federal Government Data Maturity Model



Federal Government Data Maturity Model:

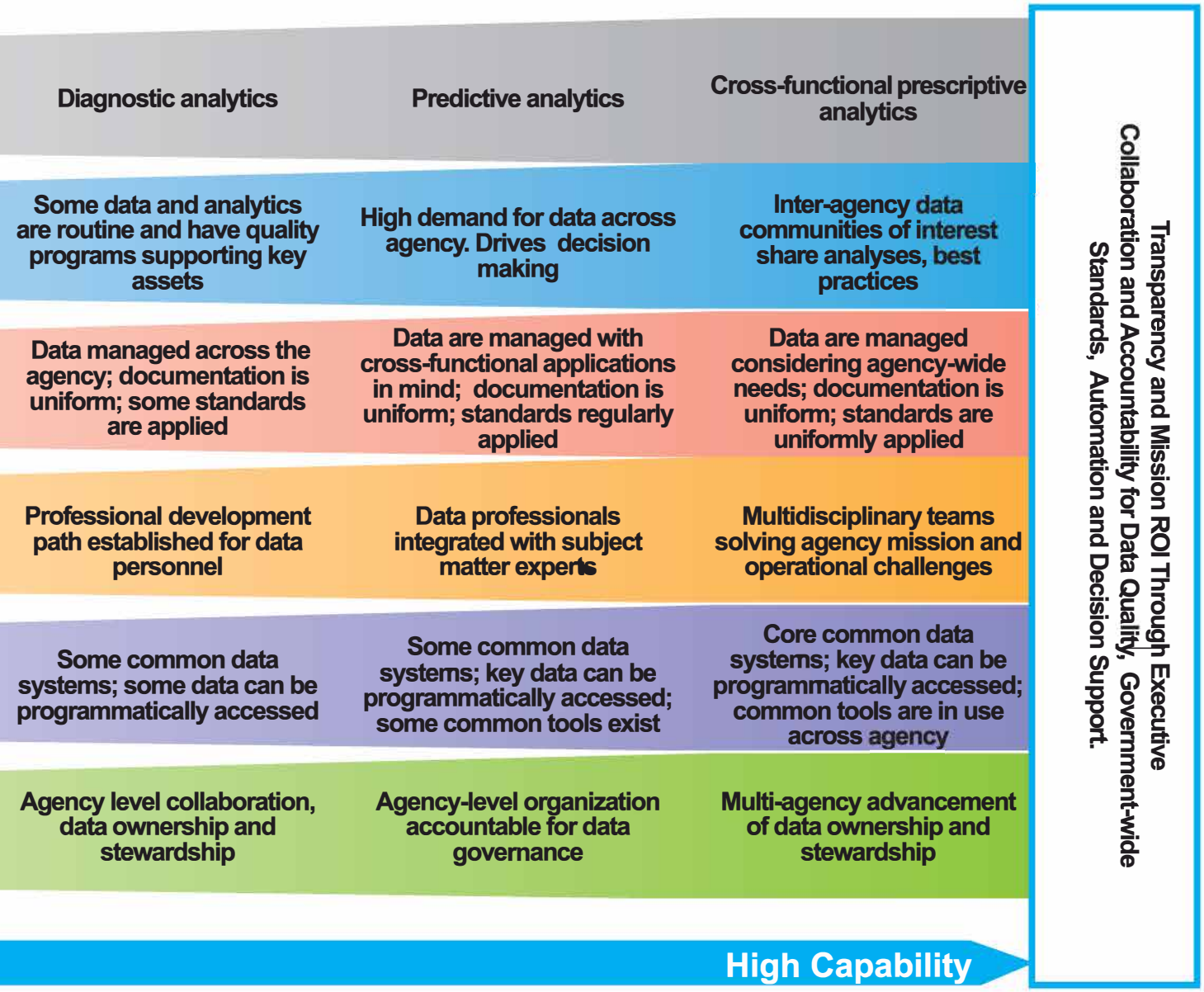
The following document details the six lanes of the Federal Government Data Maturity Model, including each of the five milestones within the lanes. The six lanes are: Analytics Capability, Data Culture, Data Management, Data Personnel, Data Systems and Technology, and Data Governance. The power of the model is in its simplicity, therefore, this document provides more context and detail around the lanes and milestones so that concepts are well defined, allowing for a common language and understanding to be established among practitioners.

Purpose

The purpose of this model is threefold.

First, this model helps agencies with a high-level assessment of current capabilities and supporting processes. The framework allows agencies to easily understand their "current state" and helps users conceptualize where they want their organization to be in the long term. It then provides some practical steps for getting there.

Second, this model helps with strategic communication between agency data professionals and agency leadership. It can also serve to communicate to the broader agency about the strategic direction of data improvement initiatives.



Finally, the model provides a common language and framework to help promulgate common solutions and best practices across federal agencies toward advancing data-driven decision making. This document contains helpful guidance within each of the lanes that address numerous organizational dynamics when creating organizational change.

How to Use this Model

It is important to note that the purpose of this model is not to provide a rigid or prescriptive "one size fits all" approach to improving data capabilities within federal agencies. Nor does it maintain that all agencies need to reach or complete all milestones within all lanes to achieve optimal data capabilities. It simply provides a framework of organized ideas and suggestions to help agencies consider what works best for them as they carve out a path to success.

Outcome Measure

The top lane of the model, "Analytics Capability" should be considered an outcome measure for capability. This lane provides a continuum of demonstrated capability from the simplest summary reporting to the most complex prescriptive analytics requiring a vast amount of data and supporting processes. The other five lanes all help to support and enable greater capability, and are essential to achieving lasting change across an organization.

Analytics Capability

As governance, technology, personnel, and culture become more established and overall capacity increases, the capability of agencies to collect, process, manage, analyze and visualize data will also increase. As data quality increases and more data sources come online, the capability to do more sophisticated analyses will begin to shape the way managers and leaders make decisions. The progression of capability will proceed from the project, program, bureau, agency and inter-agency levels.

- **Summary reports.** Little or no analysis. Basic reporting capability, aggregated data, lack of reliable data management or storage capability to provide comparable data from year to year
- **Descriptive analytics.** Descriptive analytics help agencies compare data and metrics across years. Enough transactional or line-item data exists to create metrics or compare data points from year to year. Descriptive analytics provides hindsight and answers "what happened?"
- **Diagnostic analytics.** Diagnostic analytics help agencies better understand mission and operations functions in real time. Agencies have built out their data collection and analytics capabilities to conduct more sophisticated analyses that can identify causes of poor performance or operational problems in near real-time. Diagnostic analytics provides oversight and answers "what is happening?" and "why did it happen?"
- **Predictive analytics.** Predictive analytics help agencies forecast the impact of current policy and decision making. Agencies use algorithms and complex data models to provide scenario-based, agency-level, decision making support and predict the impact of decisions and policy. Predictive analytics answers "what will happen given our trajectory?"
- **Cross-functional prescriptive analytics.** Prescriptive analytics are used to optimize and influence outcomes. Sophisticated models using multiple data sources are employed. Prescriptive analysis answers "how can we make it happen" or "how can we optimize what is already happening?"

Data Culture

Improving data capability is not only about enhanced technology and standards. Agencies must help create a shift in mentality across all employees at all levels of the organization. Data must be viewed as a strategic asset that is owned and managed by data professionals as well as program managers and executives. Agencies must understand and educate employees on the value of data, the importance of data quality, and how to transform the way they do business.

- **Data use is ad-hoc and not coordinated. Quality issues limit usefulness.** Pockets of the organization use data, but most data are poor quality and therefore not trusted to use in decision making. There is a lack of coordination and leadership around improving data capabilities. Data is more "maintained" than managed and data quality is considered the system owner or CIO's job. Little or no program management or executive accountability for data quality
- **Data use is by-request. Quality programs are nascent.** Program managers request and use data where the quality is acceptable and provides some value. Some program managers and executives undertake initiatives to collect improved data that can enhance decision making
- **Some data and analytics are routine and have quality programs supporting key assets.** The organization has some high performing programs or bureaus that transform quality data into useful information that supports decision making for managers and executives. There is a defined data quality strategy in pockets of the organization where key data assets exist. User feedback is collected and incorporated into programs that produce data-related products and services. Leadership and management begin to drive data quality improvements throughout the organization. Data quality requirements are being realized in less mature parts of the organization
- **High demand for data across agency. Data drives decision making.** Data are managed as a strategic asset across the organization. Data quality becomes an agency priority and is monitored and measured. Full value chain, customer feedback loop and accountability for data and data quality are realized at all levels and all bureaus. Data are in high demand by agency leaders and data are being used to drive both strategic and operational decision making. This milestone may also include but is not limited to the following:
  - An agency level data quality strategy is followed across the organization and is accompanied by corresponding policies, processes and guidelines
  - Data quality is included in the roles and responsibilities for data owners and data stewards
  - Training programs for data competencies exist for non-technical, program management and executive level staff
  - A defined process is established for determining costs and benefits of data quality initiatives
  - Data quality policies and processes are followed across the data lifecycle and corresponding processes are mandated in the system development lifecycle
  - Data quality metrics are employed to analyze proposed changes to the data quality strategy
  - Data quality issues are prioritized and evaluated using quantitative methods*
- **Inter-agency data communities of interest share analyses, best practices.** Data owners, data stewards, analysts, program managers, and system owners establish data communities to help build out a pervasive data culture, educate all employees and improve capability throughout the organization and with other agencies. This milestone may also include but is not limited to the following:
  - Stakeholder and customer reports of data quality issues are collected systematically
  - Data quality program milestones are regularly reviewed by agency executives and continuous improvements are implemented

** Where applicable, the Data Culture lane contains milestones included in the CMMI Institute's Data Management Maturity Model*

- The agency shares best practices and successful approaches to improving data quality with government peers and industry*

Data Management

Core data management activities such as metadata management, development of business glossaries, data development, and data standards are key to facilitating the documentation and use of an agency's data. Agencies must develop and execute the necessary policies and procedures to manage the full lifecycle needs of an agency in an effective manner.

- **Data managed in silos; documentation sparse; standards not regularly applied.** Data is aggregated and collected manually for use by certain programs or bureaus. Data cannot be easily shared or analyzed. Business glossaries and metadata documentation are not used or do not exist. Standards are not applied or used
- **Data managed in silos; some documentation exists; standards not regularly applied.** Transactional and line item data is used for specific business needs. Line item data collection is established within bureaus or by programs but not shared within the agency or across government. Business terms are defined for a particular purpose and logical data models are created with reference to defined and approved business terms. Metadata documentation is developed, stored, and accessible for certain programs or bureaus
- **Data managed across the agency; documentation is uniform; some standards are applied.** Common intra-agency data categories and data element definitions are defined within agency functions. This milestone may also include but is not limited to the following:
  - A process is established, documented, and followed to define, manage, use, and maintain the business glossary
  - Standard business terms are readily available and promulgated to relevant stakeholders
  - Each business term added to the business glossary has a unique name and definition.
  - New development, data integration, and data consolidation efforts apply standard business terms as part of the data requirements definitions process.
  - Metadata management process is established and followed
  - Metadata documentation captures data interdependencies
  - Metadata are used to perform impact analysis on potential data changes
  - Metadata categories, properties and standards are established and followed*
- **Data are managed with cross-functional applications in mind; documentation is uniform; standards regularly applied.** Cross-functional data models, dictionaries and taxonomies are defined, approved, and documented. Where possible, common inter-agency data categories and data element definitions are defined for common mission and mission support functions, (i.e. A-11, Treasury Financial Manual, GAO accounting standards for Financial Management). This milestone may also include but is not limited to the following:
  - The organization uses the approved business glossary in the development of shared repositories, data transfer standards, semantic models, etc.
  - Organization-wide data governance for compliance with the business glossary process is implemented and followed.
  - Impact assessments are conducted and governance approval is obtained prior to implementing changes to business terms.
  - Compliance monitoring processes are used to verify correct use of terms, highlight exceptions, and ensure they are addressed
  - The business glossary is integrated into the organization's metadata repository with appropriate access permissions.
  - A metadata management strategy for the organization is established and managed and maintained by data governance with input from stakeholders
  - Data governance approves changes and additions to metadata
  - Measure and metrics are used to evaluate the accuracy and adoption of metadata*
- **Data are managed considering agency-wide needs; documentation is uniform; standards are uniformly applied.** Fully standardized mission support and asset management data across government to conduct full range of performance benchmarking and analysis. Mission data standardized by domain and shared where common programs and mission objectives dictate. This may also include but is not limited to the following:
  - The business glossary is integrated into the organization's metadata repository with appropriate access permissions
  - The business glossary uses standard industry business terms and definitions as appropriate
  - Organization has developed and integrated metadata model deployed across all platforms
  - The metadata repository include exchange data representation standards used by the organization*

** Where applicable, the Data Management lane contains milestones included in the CMMI Institute's Data Management Maturity Model*

Data Personnel

Personnel are perhaps the most important part of building data capacity within an organization. An agency's data personnel may exist within any job series (HR, Program Analyst, IT) or they may have "data" in their job title (Data Scientist, Data Analyst).To advance in any other lane, an agency must understand what data expertise is needed and how those data skills are integrated into agency operations. In addition, advancing an agency's data capacity requires a professional development career path which advances those data skills regardless of job title.

- **No dedicated personnel performing data duties.** Data analysis typically falls on many personnel as an additional duty or secondary job. People are in data roles with little training or experience in the technology or methods
- **Some siloed data teams; no clear career path for data personnel.** There are some data professionals in scattered, homogenous teams throughout the organization. Data teams are not integrated into operations. The career path for data professionals is not well defined. Data personnel spend most of their time on data maintenance, data cleaning and producing reports
- **Professional development path established for data personnel.** Agencies establish professional development training and/or a career path for data professionals. Highly trained data professionals are hired into organizations and use their skills to help solve operational and mission challenges using complex analytics within their business line or bureau
- **Data professionals integrated with subject matter experts.** Agency designates a Chief Data Officer (CDO) and establishes a CDO Office with authorities, budget and clear reporting lines. The CDO, among other things, is accountable for ensuring data professionals are integrated throughout agency to tackle complex intra-agency challenges and deliver actionable reports to leadership
- **Multidisciplinary teams solving agency mission and operational challenges.** Fully integrated and multidisciplinary teams are using data as a strategic asset across functions to drive decision making at all levels of organization

Systems/Technology

Systems and technology support optimal curation and use of data and are critical for improving government-wide data management. Furthermore, successful implementation of technology enables effective collaboration and cultural adoption within agencies and across government. While systems and technology are not the responsibility of the Chief Data Officer, a common approach to the Federal Enterprise Architecture (FEA) should be considered by the Chief Data Officer working in coordination with the agency CIO and Chief Enterprise Architect.

- **Data are stored in siloed systems; data are frequently copied to facilitate use.** Database systems typically work independently of each other and commonly have repetitive information that does not stay in sync. Reporting is primarily used for status checks and rarely used for analysis of a business\mission function. Data efforts focus on the "what" and "when"
- **Data are stored in siloed systems; some data can be programmatically accessed.** Bureau/Sub-component organizations use data for decision making on regular basis. Reporting efforts in the bureau focus on "what" and "when", regardless of data quality or causes of poor data quality
- **Some common data systems; some data can be programmatically accessed.** Data is used as an asset for some common systems at the agency and bureau levels. (i.e. mission systems leveraging data from mission support systems such as Human Resources, IT, Financial, etc.)
- **Some common data systems; key data can be programmatically accessed; some common tools exist.** Agencies are leveraging shared services as established by FEA goals, where applicable, and increasing mission ROI by becoming more efficient with data management and data analysis
- **Core common data systems; key data can be programmatically accessed; common tools are in use across agency.** Where relevant and mutually beneficial, agencies leverage shared platforms, such as the Govcloud, to further efficiency while using government-wide data standards, open data protocols, and APIs

Data Governance

Establishing a governing structure for addressing both agency and government-wide data quality issues helps establish and enforce policies, common processes, and enhance communication, coordination, and cooperation amongst data leaders and experts. Effective governance provides the framework in which stakeholders can collaborate and align intra-agency and government-wide objectives that move all agencies in a shared, common direction. Improving data quality and analytics capacity requires a high volume of quality data where solutions go well beyond any one agency or organization.

- **Loose affiliations of technical staff.** Little coordination or collaboration within the bureau or agency. Data is maintained by technical staff at the system level
- **Bureau-level collaboration, data ownership and stewardship.** Program managers and system owners take accountability for data sets based on projects. Program staff work with technical staff on incremental policy and data management improvements based on specific business needs or project assignments
- **Agency-level collaboration, data ownership and stewardship.** Agency leadership takes an active role in agency-wide data sharing policy, data management, and improvement initiatives. This may also include but is not limited to the following:
    - A defined and documented Data Governance structure is in place
    - The organization follows the principles of data lifecycle management

- Governance roles and responsibilities are established by data subject areas
- Data subject area representatives participate in data governance, and follow data lifecycle management processes
- A review and evaluation process is established for data governance for continuous improvement*

- **Agency-level organization responsible for data governance.** Agency establishment of an executive level office, accountable for the establishment and oversight of data policy, data governance, data management, data integrity and agency-wide data capacity and improvement initiatives. This may also include but is not limited to the following:
  - The governance structure includes executive level representatives and representatives from all business units which are suppliers or consumers of high profile data subject areas
  - Standard data governance policies and processes are followed.
  - Classroom, mentoring, or on the job training in data governance processes is required for new governance members and other stakeholders
  - Metrics are established and statistical analysis is used to evaluate the effectiveness of governance activities and to make changes or adjustments*

- **Multi-agency advancement of data ownership and stewardship.** Drive the continued advancement of the federal sector through cross-agency communities such as the Federal Chief Data Officers (CDO) Council by sharing best practices, promoting smart government-wide data policy, promoting standardized data quality metrics, and leading government-wide improvement and standardization efforts. Additional examples may include:
  - Representatives of the agency governance structure also participate in communities of interest and government-wide councils and committees that promulgate cross agency standards and solutions
  - External governance structures and industry case studies are evaluated for best practices and lessons learned, providing ideas for improvement*

** Where applicable, the the Data Governance lane contains milestones included in the CMMI Institute's Data Management Maturity Model*

References

Open Data Maturity Model, The Open Data Institute

Data Management Maturity Model, CMMI Institute

IBM Data Governance Council Maturity Model, IBM

Analytics Excellence Roadmap, Harvard Ash Center

Data Management Body of Knowledge, Data Management Association

Data Maturity Framework, University of Chicago Center for Data Science and Public Policy

Analytics Maturity Model/Business Intelligence Maturity Model, The Data Warehousing Institute

Analytics Maturity Model, INFORMS

For more information please contact Ryan Swann (ryan.swann@gsa.gov) or Todd Coleman (todd.coleman@gsa.gov)

https://www.ntis.gov/TheDataCabinet