GOVERNMENT OF CANADA

DATA COMPETENCY FRAMEWORK

GC DATA COMMUNITY

IN PARTNERSHIP WITH:
ENTERPRISE DATA COMMUNITY OF PRACTICE - PEOPLE AND CULTURE WORKING GROUP
CANADA SCHOOL OF PUBLIC SERVICE – DIGITAL ACADEMY

Vanessa Vermette
Vice President
Innovation & Skills Development Branch
Canada School of Public Service

Christopher Valiquet
Director
GC Data Community
Canada School of Public Service

Foreword

Data is ubiquitous in our everyday lives and is at the core of a digitally enabled, citizen-centred public service. Data competencies can no longer be considered as relevant to only a small group of specialists. Instead, developing proficiency in these competencies is necessary across professions and functional areas in the fast-evolving, data-informed world we live in.

Having a data literate workforce is at the core of modernization efforts. This Data Competency Framework is meant to support conversations and aims to advance data literacy by creating a shared understanding and language about data competencies for all federal public servants. This shared understanding can provide the foundation for data literacy initiatives, such as the development of learning paths and assessments, data talent management initiatives, and ultimately allow us to better harness and grow the value of data knowledge and expertise as an enterprise asset.

Recognizing that departments and agencies have different operating contexts and business lines, this Framework is meant to be a guide that can be tailored to specific organizational needs.

We would like to express our gratitude to the organizations and colleagues across the Government of Canada who helped with the development of this Framework.

We hope it will support our collective efforts to strengthen data literacy and capabilities across the public service for a data-empowered workforce.

Table of Contents

Background 4
Acknowledgements 4
Introduction 5
The Framework at a Glance 7
Proficiency Levels 7
1. Data Concepts and Culture 9
1.1 Data, Digital and Organizational Awareness 9
1.2 Data Ethics and Privacy 11
1.3 Evidence-Informed Decision Making 13
2. Data Governance, Collection and Stewardship 14
2.1 Data Governance, Stewardship and Standards 14
2.2 Data Collection 16
2.3 Data Quality, Value and Trust 17
2.4 Access, Security and Interoperability 19
3. Analytics and Evaluation 21
3.1 Asking Questions and Problem Framing 21
3.2 Data Analytics and Science 22
3.3 Storytelling and Visualization 25
3.4 Evaluating Outcomes 27
4. Data Systems and Architecture 29
4.1 Enterprise Data Architecture 29
4.2 Data Systems 31
Annex A:
Definitions of Key Terms 33
Annex B:
Additional Resources 36

Background

Acknowledgments

The GC Data Community would like to acknowledge the important contributions of our many partners:

- CSPS Digital Academy
- People and Culture Working Group
- Statistics Canada—Strategic Analysis, Publications and Training Division
- University of Dalhousie
- GC Enterprise Data Community of Practice
- Treasury Board Secretariat—Office of the Chief Information Officer
- Privy Council Office—Chief Data Office
- Data Science Network
- Information Management community
- Employment and Social Development Canada—Chief Data Office

Special recognition goes to Employment Social Development Canada (ESDC) for operationalizing the Framework as a business use case and for their horizontal open sharing to promote collective synergies among the data community and beyond.

Introduction

A data literate federal public service is key for modernization and delivering value to those we serve. Building data literacy requires common definitions and understanding of the competencies needed for public servants to work together in treating data as an enterprise asset. The Government of Canada Data Competency Framework provides a model to guide data literacy efforts, which involve all federal public servants working with and using data.

The development of this framework has been a community-driven effort that began after the 2018 publication of the Data Strategy Roadmap for the Federal Public Service. Led by the Digital Academy at the Canada School of Public Service (CSPS) and Statistics Canada, in collaboration with Dalhousie University, the first step developed a foundation using literature review on data literacy. In 2021, the GC Data Community and the Enterprise Data Community of Practice People and Culture Working Group conducted consultations with partners spanning the data, digital, information, and HR communities to update the framework to reflect the diverse organizational contexts and needs of the Government of Canada. In 2022, the GC Data Community further mapped the competences to align with proficiency levels, building on the work done by the Chief Data Office at Employment and Social Development Canada.

Feedback is welcome and will be used to inform future iterations of this Framework. Questions or comments can be sent to the GC Data Community at CSPS (gcdc-cdgc@csps-efpc.gc.ca).

How to Use the Framework

The Framework was developed to establish a common understanding of foundational, intermediate, and advanced data competencies required across the federal government. It catalogues the suite of knowledge and skills across the data life cycle required to enable effective and rigorous, evidenced-informed decision-making.

The Framework is intended for three main uses:

1. Supporting **employees and leaders at all levels** in better understanding the data landscape, and building data literacy across the public service.

2. Supporting **government departments and agencies** understand the breadth of data competencies required throughout an organization for effective data-informed decision-making, which includes leveraging data as an enterprise asset:
    - Managing and using data effectively as a strategic asset

- Ensuring good governance according to ethical and enterprise standards
- Designing and building sustainable and secure infrastructure for the collection, access, use, management, interoperability and preservation of data
- Supporting a data literate workforce
- Placing the trust, needs, and expectations of those served by the Government of Canada's first when it comes to policy, program and service delivery

3. Guiding departments and agencies in conducting **data literacy assessments** to identify skills availability and gaps.

It is important to note that because the framework has not been mapped to specific classifications or job descriptions, it is not intended to be used as a formal HR tool or to conduct performance evaluations. That said, it can be used to guide and inform conversations and the development of learning pathways around data skills needs.

The Framework contains **four categories** that each plays a pivotal role in the data life cycle:

1. **Data Concepts and Culture**
2. **Data Governance, Collection, and Stewardship**
3. **Analytics and Evaluation**
4. **Data Systems and Architecture**

The Framework at a Glance

1. **Data Concepts & Culture**
   - 1.1. Data, Digital and Organizational Awareness
   - 1.2. Data Ethics and Privacy
   - 1.3. Evidence-Informed Decision Making

2. **Data Governance, Collection & Stewardship**
   - 2.1. Data Governance, Stewardship and Standards
   - 2.2. Data Collection
   - 2.3. Data Quality, Value and Trust
   - 2.4. Access, Security and Interoperability

3. **Analytics & Evaluation**
   - 3.1. Asking Questions and Problem Framing
   - 3.2. Data Analytics and Science
   - 3.3. Storytelling and Visualization
   - 3.4. Evaluating Outcomes

4. **Data Systems & Architecture**
   - 4.1. Enterprise Data Architecture
   - 4.2. Data Systems

These categories include a total of **13 competencies**, each broken down into indicators across foundational, intermediate, and advanced levels. A glossary of terms is included in Annex A.

Proficiency Levels

Level 1—Foundational
Defining the core level of understanding and awareness

- Finds, enters, reads, understands data, charts, and graphs.
- Awareness of key data legislation, policies, directives, and standards
- Understands simple data a
- nd AI terms and concepts.
- Understands the steps to working with data.

- Completes simple tasks (e.g., performs Excel calculations, creates charts, produces reports, etc.)

Level 2—Intermediate
Putting theory into practice

- Fully grasps foundational data terms and concepts.
- Knows how to access data that has the best fit for purpose, knows how to use them, and is able to identify gaps.
- Completes various data tasks successfully (e.g., uses different statistical methods and analytical techniques, cleans and processes data, etc.)
- Can access and handles data from different sources confidently.
- Analyzes and draws conclusions from quantitative or qualitative data.
- Takes into account the source of the data, and understands that why and how they were collected can influence their quality and usefulness.

Level 3—Advanced
Advanced applications and enabling others

- Has deeper understanding of data concepts and applications, and demonstrates ability to access and translate data into actionable knowledge.
- Combines qualitative and quantitative data to grasp the complexity of an issue or problem from different angles and at micro and macro levels.
- Identifies, analyses and solves complex problems with data.
- Finds creative ways to employ data to solve problems or inform policy.
- Leads by example, helps and teaches others.

Context is critical in identifying the necessary competencies and proficiency levels needed. While some competencies are needed at the foundational level at a minimum for all employees, such as data awareness, some are only applicable to particular roles, for example data systems. Additionally the indicators are meant to provide a range of skills for a given competency. As such, the appropriate combinations of required competencies, skills, and proficiency levels will depend on specific needs of a role, operation, activity, or project. There are many situations where the full range of skills for a given competency is not required; for example, a typical survey research project will not require the AI-related skills listed in the Data Analytics and Science competency.

1. Data Concepts and Culture

1.1 Data, Digital and Organizational Awareness

Knowledge of how data are shaping government today and how to use it effectively; includes the skills that support a data-driven organization.

Foundational

1.1.1 Aware of key data and digital terms, standards, policies, documents, and communities, including

- Data terms
- Federal and departmental Chief Data Offices' role
- Federal and departmental Data Strategies
- Data Strategy Roadmap
- Statistics Act
- Policy on Service and Digital
- Canada's Digital Ambition and Digital Standards

1.1.2 Understands what data are, the data life cycle (Plan, Collect, Process, Use/Share) and the many types of data that exist.

1.1.3 Understand the concept of information and terms related to data value, information and analysis.

1.1.4. Understands the value of data as a strategic asset and the importance of data literacy to your organization in supporting decision-making, research, learning and development, service delivery and measuring results.

1.1.5 Understands the roles, responsibilities, and accountability around data.

1.1.6 Knowledge of organizational data roles, policies, standards, processes and their intent.

Intermediate

1.1.7 Maintains awareness of broad data trends and their potential impact on the Government of Canada.

1.1.8 Uses the services, resources and support available through Centres of expertise (Chief Data Office, Data Standards/Governance Division, etc.) and other trusted sources.

1.1.9 Understands complex data, statistical, and analytical concepts.

1.1.10 Knowledge of the data that exist within the organization and how they are being used to inform or support decision making.

1.1.11 Works in a multidisciplinary partnership with core partners such as Information Management, Digital, Information Technology, privacy and security, legal, subject matter experts and end users.

Advanced

1.1.12 Enables colleagues to use key data terms when speaking about and working with data.

1.1.13 Leads by example and contributes to building and enabling a data literate workforce and a data-driven culture by encouraging the use of data within home organization and across government.

1.1.14 Identifies and removes roadblocks so others can work at their most efficient capacity around data (e.g., ensuring access to training, infrastructure and software).

1.1.15 Defines and implements strategies to advance the talent, systems and infrastructure needed to improve organizational data literacy.

1.2 Data Ethics and Privacy

Understanding ethical considerations relating to the collection, use, interpretation and sharing of data.

Foundational

1.2.1 Is familiar with the meaning of data ethics, governance, consent, bias and discrimination, inclusiveness, fairness, accountability.

1.2.2 Understands and adheres to key ethics, privacy, legal, and security principles and standards, including but not restricted to:

- Privacy Act
- Statistics Act
- Policy on Government Security
- Policy on Privacy Protection
- Privacy Impact Assessments
- Levels of Security
- Gender-Based Analysis+
- Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans
- Model Policy on Scientific Integrity
- Directive on Automated Decision-Making
- Disaggregated Data

Intermediate

1.2.3 Knows how to protect and share confidential data.

1.2.4 Understands the ethical implications of using AI.

1.2.5 Identifies indicators of bias and ensures policy, programs, or services do not reinforce unintended biases.

1.2.6 Ensure data are used in alignment with their intended purpose by consulting with data owners or stewards.

1.2.7 Applies processes and procedures to ensure ethical approaches to research and data throughout the data life cycle.

1.2.8 Identifies ethical issues, privacy/security implications and barriers to accessibility.

Advanced

1.2.9 Understands advanced concepts such as necessity, proportionality, sensitivity, and explainability.

1.2.10 Assesses data for bias, representation, accuracy, and validity. Identifies and implements steps to resolve issues if needed.

1.2.11 Tests ideas and potential solutions with a wide range of diverse data to challenge assumptions and ensure solutions that are inclusive by design.

1.2.12 Designs and implements policies, processes and procedures to ensure ethical approaches to research and data use.

1.3 Evidence-Informed Decision Making

Evidence-informed decision-making is the process of distilling and disseminating the best available evidence from research, practice, and experience. It is the use of evidence to inform and improve policies, programs, operations and service delivery to Canadians.

Foundational

1.3.1 Prioritizes the use of knowledge and information gathered through data rather than simple anecdotal evidence.

1.3.2 Assists in answering and resolving business data questions.

1.3.3. Consults with appropriate authorities (e.g., subject matter experts, community leaders) to identify what is considered high-quality evidence in a given context.

Intermediate

1.3.4 Locates data to inform decision-making and supports assessment of their suitability. Documents when and where required data are not captured, collected, or are missing.

1.3.5 Uses data to understand users' needs and to design and develop products, programs and services that meet those needs (see Government of Canada Digital Standards: Playbook for more details).

1.3.6 Uses data and analytics to weigh the merit and impact of solutions or decisions prior to implementation.

Advanced

1.3.7 Communicates the complexities and nuances of analytical findings to non-technical audiences to promote understanding and the business value added.

1.3.8 Elaborates and implements policies or initiatives to develop the talent, systems, or infrastructure that allows for the use of and timely access to data for decision-making.

1.3.9 Evaluates data for suitability to inform decision-making by integrating ethical, methodological, and subject-matter or context-specific perspectives.

2. Data Governance, Collection and Stewardship

2.1 Data Governance, Stewardship and Standards

Managing, implementing and adhering to policies, procedures and standards that support the availability, usability, integrity, security and accessibility of data within the organization and across government.

Foundational

2.1.1 Is familiar with data governance and data sovereignty, including First Nations Data Sovereignty**,** standards, directive, processes, including accessibility standards.

2.1.2 Can identify and proactively flags data issues or conflicts pertaining to data governance, stewardship.

Intermediate

2.1.3 Applies key policies, procedures and standards for the collection, access, and management of data.

2.1.4 Collaborates and negotiates to:

- ensure common understanding of data manage access to data identify data privacy, security, or accessibility implications and barriers

Advanced

2.1.5 Ensures sound management and oversight of data while limiting data duplication and respecting the data's intended purpose.

2.1.6 Selects methods, processes or tools that minimize the duplication of data, and utilizes data where it resides whenever possible.

2.1.7 Monitors data issues and changes to ensure data is actively managed, accessible and enhanced to maintain trusted, usable data in the enterprise.

2.1.8 Assesses enterprise and organizational priorities for the operational and strategic use of data.

2.1.9 Assesses data requirements for projects, programs, departmental reporting and policies to ensure that data are managed according to value and use.

2.1.10 Assesses requirements for the retention and disposition of data.

2.1.11 Adapts or implements data governance within the operational business context.

2.1.12 Develops rules, governance principles and guidelines prior to the collection, creation, use, or sharing of data.

2.1.13 Develops policies, governance or standards that balance and ensure security, interoperability and access to data.

2.2 Data Collection

Utilizing methods, processes, tools, platforms and software to collect data.

Foundational

2.2.1 Understands the role of data collection and aware of common data collection methods and tools.

2.2.2 Identifies and uses existing data prior to collecting new data.

2.2.3 Adheres to policies, legislation, processes, and standards when collecting data.

- Privacy Act
- Statistics Act
- Departmental- or agency-specific acts on data collection
- Policy on Government Security
- Policy on Privacy Protection
- Privacy Impact Assessments
- Levels of Security
- Gender-Based Analysis+
- Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans
- Model Policy on Scientific Integrity

Intermediate

2.2.4 Extensive knowledge of data collection methods and tools, and how to apply them to ensure quality, consistency, and timeliness.

2.2.5. Supports with mitigation or resolution of data collection issues (e.g., sampling problems, missing data, errors and inconsistencies, etc.).

Advanced

2.2.6 Designs and implements data collection methods that directly align with a stated question(s).

2.2.7 Designs and implements automated data collection methods when appropriate.

2.2.8 Assesses or validates data collection methods and tools.

2.3 Data Quality, Value and Trust

Data trust means having confidence in their quality. Data cleaning, processing and transforming data ensures access to accurate, reliable, and high value data and information. Building solid partnerships fosters a shared responsibility towards data integrity.

Foundational

2.3.1 Is familiar with the data quality framework and the different dimensions through which one can evaluate quality (ex: metadata definitions, standards, interpretability, coherence, relevance, accessibility, timeliness).

2.3.2 Understands the negative impact of poorly managed data on organizational operations and decision-making.

Intermediate

2.3.3 Reviews data to ensure validity, accuracy, and completeness.

2.3.4 Aligns with data standards to ensure and improve data quality.

2.3.5 Engages directly with stakeholders and recognized authorities of data to build relationships that enhance trust within an organization, with stakeholders and with Canadians.

2.3.6 Identifies and analyzes outliers and anomalies within data and uses problem-solving approaches (e.g., Root Cause Analysis) to improve data quality.

Advanced

2.3.7 Ensures and assesses data for quality and interoperability to maximize value and trust.

2.3.8 Uses or develops master data management (MDM) and reference data management (RDM) solutions to ensure consistency and integrity in data and their use.

2.3.9 Raises awareness about data holdings for integration and re-use.

2.3.10 Develops or coordinates additional data entry training when entered data quality or validity is a concern.

2.3.11 Maintains data lineage and audit trailing to ensure the appropriate fit-for-purpose use of data.

2.3.12 Builds multi-disciplinary partnerships across the organization to foster a communal responsibility towards data governance, standards and stewardship, and to enhance trust throughout the organization.

2.3.13 Evaluate fitness for use of a dataset (i.e., sometime no data is better than the wrong data). Documents and shares rationale transparently.

2.4 Access, Security and Interoperability

Ensuring the access, security and accessibility of enterprise-wide and organizational data assets.

Foundational

2.4.1 Awareness with policies, directives and regulations relating to data security and storage.

2.4.2 Understands the costs and risks of data loss, and takes action to minimize data and information loss in their work.

2.4.3 Understands the risks associated with duplicate data and takes action to minimize duplication of data and information in their work.

2.4.4 Ensures the privacy and security of data using appropriate levels of permission and access.

Intermediate

2.4.5 Supports the use of self-serve data and data access by design.

2.4.6 Awareness of data holdings, open data and accessible data that are available for (re) use and integration.

2.4.7 Identifies and prioritizes the release of high-value datasets to the Open Government Portal for re-use and interoperability.

2.4.8 Applies approved security-enhancing tools to minimize risk when working with data, such as de-identification, anonymization, masking, or other similar methods.

2.4.9 Publishes data in machine-readable formats, to the extent possible.

Advanced

2.4.10 Understands a variety of conversion methods in support of interoperability.

2.4.11 Develops provenance and mappings to support the preservation of data.

2.4.12 Uses, creates or manages metadata, database indexing, or information analysis and synthesis to improve information access, retrieval and use.

2.4.13 Strives to identify and remove barriers to data sharing and release, and champions open data and transparency in government by sharing and working in the open by default.

3. Analytics and Evaluation

3.1 Asking Questions and Problem Framing

Understanding when data can be used to inform or to support, as well as the process of interpreting data into identifiable problems and research questions.

Foundational

3.1.1 Knowledge of when and how data can be used to inform a decision or not.

3.1.2 Uses critical thinking and asks questions to define the data needed, how it will be collected and how it will be used.

3.1.3 Considers the importance and use of data at the outset and not as an afterthought.

Intermediate

3.1.4 Frames questions around business and user needs that can be answered with the support of data.

3.1.5 Identifies pre-existing data that can be used to support or inform questions or decision-making.

3.1.6 Explores data and uses relevant frameworks to identify issues, support research or generate questions relating to practical situations.

Advanced

3.1.7 Works in multidisciplinary teams as an expert to inform a variety of areas using data-driven approaches.

3.1.8 Applies understanding of organizational and user needs, mandates, and directions to generate business requirements associated with data and to improve problem framing at the outset of new projects, policies, programs, service delivery and more.

3.2 Data Analytics and Science

Recognizing patterns and trends within data to identify relationships and generate insight. Accessing, manipulating, querying and analyzing data using a variety of methods, tools, and processes.

Foundational

3.2.1 Familiar with basic functionality of commonly used software and can perform simple calculations, create charts and tables.

3.2.2 Can analyze data to answer simple business questions.

3.2.3 Considers the use of open source and freely available tools as compared to commercially provided.

3.2.4 Aware of the Directive on Automated Decision-Making and Algorithmic Impact Assessments.

Intermediate

3.2.5 Uses basic analytical methods and related tools, qualitative or quantitative, to generate insights.

3.2.6 Takes stated questions and develops an analysis plan including assessing relevant data and methods.

3.2.7 Understands various statistical methods, analyses, and related tools and when their use is appropriate.

3.2.8 Disseminates data and analytical findings openly to support other government endeavours as per the Digital Standards.

3.2.9 Recognizes and explores patterns, relationships and trends within data and across data sources to generate insights using a variety of methods.

3.2.10 Sets clear goals prior to analytical endeavours to ensure value added analyses and consults with all stakeholders to ensure data collection and analytical plans align with this value.

3.2.11 Conducts analysis using relevant data and methods for consumption across a wide range of audiences.

3.2.12 Evaluates results of analyses and compares with other findings.

3.2.13 Shares data and analytical findings openly to support other teams' work.

3.2.14 Can access and manipulates data from different sources, for example using flat files or Structured Query Language (SQL) queries.

3.2.15 Applies common analytical methods and tools to data, qualitative or quantitative, to generate insights.

3.2.16 Applies with the Directive on Automated Decision-Making, including the completion of an Algorithmic Impact Assessment, to identify, evaluate and mitigate risks associated with deploying an automated decision-making system.

Advanced

3.2.17 Uses a wide range of advanced analytical methods and their related tools, qualitative or quantitative, to generate insights.

3.2.18 Links together data from multiple sources to increase their usefulness and value.

3.2.19 Conducts exploratory analysis to support problem framing, identify areas of risk and more.

3.2.20 Builds and validates statistical models from data.

3.2.21 Uses appropriate, accurate, valid and efficient modelling solutions across a variety of complex data types.

3.2.22 Utilizes advanced tools and techniques to perform data exploration such as data mining, web scraping or machine learning.

3.2.23 Assesses and selects tools based on their benefits, use cases, ease of access and inclusivity, and considers open source options when available.

3.2.24 Considers digital and platform design as well as organization-wide and government enterprise-wide considerations when selecting tools.

3.2.25 Assesses data for their fit for use, quality, accuracy, relevancy and potential bias.

3.2.26 Ability to apply automation for the collection of data.

For more details on data science competencies, please refer to the Competency Profiles developed by the Data Science Network for the Federal Public Service.

3.3 Storytelling and Visualization

Translating data into an accessible format to help others to see and understand trends, outliers, and patterns in data.

Foundational

3.3.1 Presents information with accessible visuals, presentations, or stories to:

- help others understand a subject matter
- inform a discussion report on progress
- support decision-making or problem solving

3.3.2 Builds in accessibility and inclusivity in visuals and in content following best practices, such as the Government of Canada Digital Standards: Playbook.

3.3.3 Creates tables and graphical representations of data that are accurate and informative.

3.3.4 Includes correct and relevant references, labels and citations.

Intermediate

3.3.5 Displays holistic information, telling complete stories rather than presenting selective or incomplete evidence.

3.3.6 Ensures data presentations link directly to the original questions or line of thinking.

Advanced

3.3.7 Assesses audience needs, familiarity with data and understanding of subject matter.

3.3.8 Evaluates storytelling and visualizations for accuracy and misrepresentation.

3.3.9 Communicate best data visualization practises and tools among data teams and others to avoid common mistakes and make data visualizations more effective.

3.3.10 Provides drill down capability in reports and summary information to allow further investigation.

3.3.11 Develops dashboards, infographics, and interactive visualizations using different software, including common business intelligence tools (e.g., PowerBI, Tableau) and/or specialized libraries (e.g., D3.js, seaborn, plotly).

3.4 Evaluating Outcomes

Critically assesses the efficacy of policies, programs, services, and decisions using research, evaluation, and analytical techniques.

Foundational

3.4.1 Understands the principles around evaluation and how data are used, interpreted, and applied as part of program or project monitoring, improvements or to demonstrate results.

3.4.2 Aware of basic analytical methods and techniques to measure and track the implementation or performance of a project or program.

Intermediate

3.4.3 Uses data for the purpose of measuring outcomes relating to policy, program implementation, legislation, regulation, decision-making and more.

3.4.4 Engages regularly with users, collecting feedback and data throughout the implementation of ideas or solutions to take into account how well they are working/performing, and utilizes the feedback to adjust course as needed (see Government of Canada Digital Standards: Playbook for more details).

3.4.5 Cross-compares results with other research and findings.

3.4.6 Retains and maintains original data and information utilized in decision-making to analyze outcomes compared to intent.

3.4.7 Collects follow-up data to assess the efficacy of decisions or solutions.

3.4.8 Identifies key takeaways from charts, tables and graphs to integrate with other information into future decision-making processes.

Advanced

3.4.9 Analyses data to understand the impact that identified solutions may have on diverse groups of people, policies, programs, services, processes, and more.

3.4.10 Evaluates and measures outcomes of decision-making using data.

3.4.11 Uses end-user feedback to establish evolving requirements to adjust policies, programs, services, and more to meet users' needs.

3.4.12 Obtains stakeholder agreement on criteria for assessment prior to implementation.

3.4.13 Implements changes to policies, programs, service delivery and other government functions based on ongoing feedback and user research.

3.4.14 Monitors the outcomes of automated decisions to safeguard against unintentional outcomes.

4. Data Systems and Architecture

4.1 Enterprise Data Architecture

Understanding principles of data systems and designing solutions to maximize the usability, integrity, security, scaleability, and accessibility of data and information.

Foundational

4.1.1 Understands data management tools, including data catalogues, data lakes and data warehouses.

4.1.2 Familiar with the Government of Canada Enterprise Architecture Review Board and its role, mandates, and function within government.

4.1.3 Understands the role the Enterprise Data Lake plays in the research and exploration of data, including the use of artificial intelligence (AI).

Intermediate

4.1.4 Considers architectural needs across business, IT and data when developing enterprise solutions.

4.1.5 Considers full system life cycle in the development or deployment of organizational or enterprise architecture.

4.1.6 Assesses system requirements (such as storage, retention and schedule) to ensure that organizational or enterprise architecture is designed for usability, interoperability and to provide value to users.

Advanced

4.1.7 Develops and maintains monitoring of AI Systems, which includes but is not limited to health, bias, and data drift.

4.1.8 Designs, implements, or monitors systems that allow for access at source, using appropriate RDM and MDM, reducing the duplication of information.

4.1.9 Anticipates and identifies problem points in systems and solution development.

4.1.10 Prototypes and demonstrates solutions for clients their environments to enable further development.

4.1.11 Builds automation software to operate systems needed for data storage, data management, data science codes, distributed training, model repository, feature repository, continuous delivery, model serving, and monitoring.

4.1.12 Operates production AI systems and makes sure they are available, scalable, and performing effectively.

4.1.13 Builds and shares the technical expertise necessary to analyze and recommend enterprise-grade solutions for operationalizing AI or advanced analytical models.

4.1.14 Sets the required architecture and deployment processes for AI, from data ingestion to production and maintenance.

4.1.16 Provides technical advice to management and other data scientists as it pertains to the operationalization of models.

For more details on data science competencies, please refer to the Competency Profiles developed by the Data Science Network for the Federal Public Service.

4.2 Data Systems

Utilizing tools, software, platforms and processes to collect, organize, store, manage and protect data.

Foundational

4.2.1 Aware of basic approaches, tools, and techniques to ensure sound data management.

Intermediate

4.2.2 Ensures that system designs aligns with organizational needs.

4.2.3 Ensures security of data and information through appropriate levels of permissions and access.

4.2.4 Develops, implements, tests and troubleshoots system workflows.

4.2.5 Develops end-to-end data pipelines based on an in-depth understanding of the available tools and data management ecosystem, data life cycle, and business problems to ensure analytics solutions are efficient, predictable and sustainable.

4.2.6 Is comfortable:

- using best coding practices to generate reproducible, verifiable work
- extracting, transforming and loading data with appropriate tools and practices
- using data analysis tools and techniques, including open source

4.2.7 Considers all organizational and enterprise-wide interdependencies that require processes, standards or the preservation and storage of data to maximize organizational value.

4.2.8 Extracts data from data sources utilizing appropriate tools and practices.

4.2.9 Transforms data using a variety of tools.

4.2.10 Knowledge of extract, transfer, load (ETL) functions and their appropriate use.

Advanced

4.2.11 Assesses, builds or implements data structures such as data lakes, data marts, relational databases, or in-memory databases.

4.2.12 Designs or implements data systems and processes that utilize data at source rather than replicate.

4.2.13 Designs, implements, or assesses data collection and transformation tools, methods and processes.

4.2.14 Assesses the value of cloud storage vs on premise storage.

4.2.15 Uses Cloud and Software as a Service (SaaS), such as Enterprise Resource Planning (ERP), Service Management applications, or Distributed Architecture.

4.2.16 Maintains and supports data availability and longevity through backup systems and appropriate data recovery plans.

4.2.17 Develops and recommends performance indicators to measure the efficiency, efficacy and impact of Information Management solutions.

4.2.18 Determines problems with stakeholders, understands where data science can add value in supporting strategic and operational decision-making to create impact, and designing data science solutions and metrics to these problems.

Annex A: Definition of Key Terms

Artificial Intelligence
Information technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours, or solving problems.

Algorithm
A set of instructions used to solve a problem or perform a computation.

Analytics
The discovery, interpretation and communication of meaningful patterns in data. Concerned with turning data into useful information for making better decisions. Analytics often relies on the application of statistics, algorithms, and software to research potential trends, analyze the effects of decisions, or evaluate the performance of a policy, program, or service.

Dashboard
An information management tool which is used to visually track and represent key performance indicators, metrics and key data points.

Data
Data are facts, figures, observations, or recordings that can take the form of image, sound, text or physical measurements (ex: distance, weight, wave lengths).

Database
A large and organized collection of data. It allows data to be easily accessed, manipulated and updated.

Dataset
A collection of data. It often corresponds to the content of a single database or data file.

Data cleaning
The process of preparing data for analysis by removing, correcting, and documenting data that are incomplete, incorrect, duplicated or improperly formatted.

Data drift
Variations in the production data from the data that was used to test and validate the model before deploying it in production. Data drift leads to model performance degradation.

Data lake
A storage system or repository that holds large amounts of raw data.

Data literacy
The ability to understand, create and communicate data as meaningful information.

Data mart
A subset of a data warehouse focused on a particular line of business, department, or subject area.

Data science
A multidisciplinary field that combines the scientific method, computer programing, statistics, and business to extract meaningful insights from data and inform decision-making. Although data science shares commonalities with statistics, these two terms should not be used interchangeably as they differ in their processes and applications.

Data warehouse
A system that aggregates data from multiple sources into a single, central, consistent data store to support data mining, artificial intelligence, and machine learning.

Enterprise data lake
An enterprise-wide data lake for information storage and sharing.

Extract, transfer, load
A data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system.

Information
Data converted into a meaningful and useful context. Knowledge captured in any format, such as facts, events, things, processes, or ideas, that can be structured or unstructured, including concepts that within a certain context have particular meaning.

Interoperability
The ability of different systems, devices, applications or products to connect and communicate in a coordinated way, without effort from the end user.

Master data management
The technologies, tools, and processes to ensure master data is coordinated across the enterprise.

Model
A mathematical representation of a real-world process. A predictive model forecasts a future outcome based on past behaviours.

Open data
Structured data that is machine-readable, freely shared, used and built on without restrictions.

Personal information
Recorded information about an identifiable individual other than business contact information. It includes any information that can be linked back to or identify an individual through reference or association.

Qualitative data
Data describing the attributes or properties that an object possesses. They represent subjective variables that are sometimes not easily measured (e.g. taste, eye colour).

Quantitative data
Data expressing a certain quantity, amount or range. Usually there are measurement units associated with the data (e.g., meters for heights, kilograms for weight or Celsius for temperature).

Raw data
A collection of numbers or characters before it has been cleaned or processed.

Reference data management
An approach to maintain accurate, consistent, and standard references to data across the enterprise.

Root cause analysis
A range of approaches, techniques, and tools to systematically identify the causal mechanisms that underlie the potential roots of a problem.

Statistics
In the context of data analysis, statistics refer to a type of information obtained through mathematical operations on data. Statistics also refers to the science of developing and studying methods for collecting, analyzing, interpreting, and presenting data.

Structured data
Data that has been organized into fixed fields within a database or file so that it can be readily searchable and easily analyzed; for example, data an excel spreadsheet.

Structured language query
A programing language used to manage relational databases and conduct operations on their data**.**

Unstructured data
Data that do not have a pre-defined data model or that are not organized in a pre-defined manner. They are thus not stored in a database in fixed fields. Examples include the text included in a report.

Annex B: Additional Resources

GC Data Community

Enterprise Data Community of Practice

Data Ecosystem Resource Directory

Research Data Canada

Library and Archives Canada Glossary for Information

CODATA Research Data Management Terminology

Consortia Advancing Standards in Research Administration Information Research Data Management Glossary